

POLITECHNIKA KOSZALIŃSKA

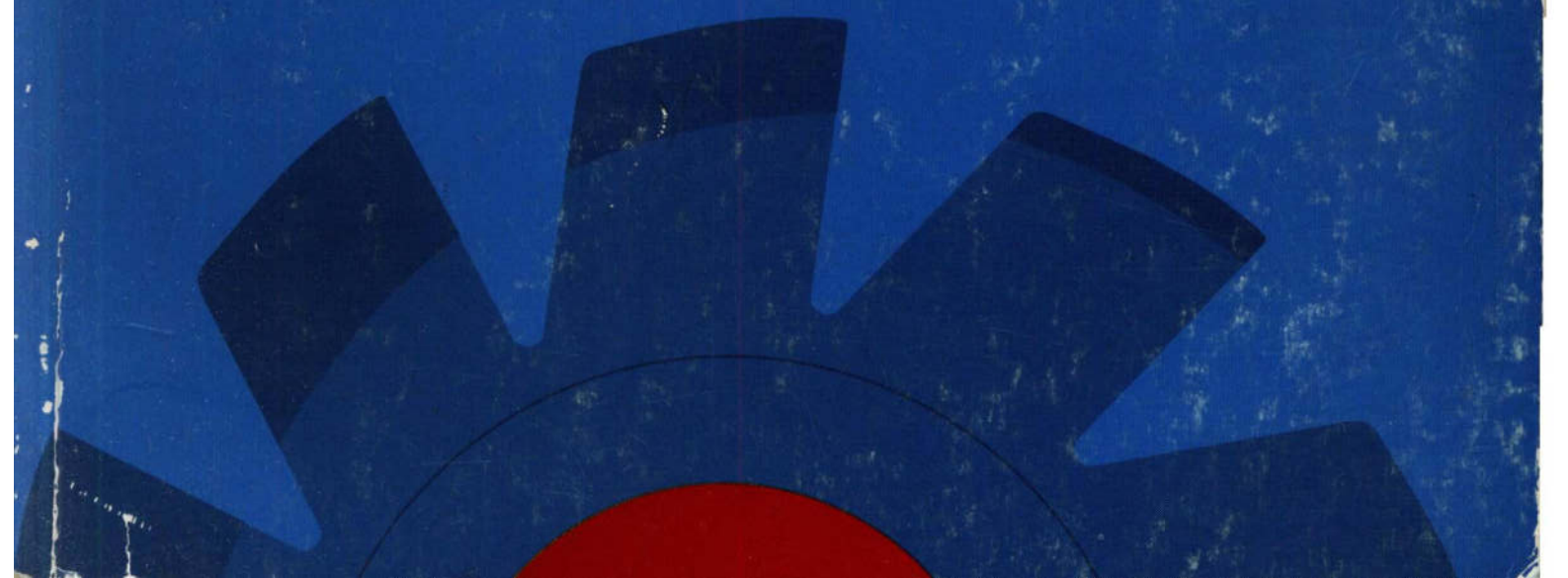
Z E S Z Y T Y
N A U K O W E
W Y D Z I A Ł U
M E C H A N I C Z N E G O
N R 27

C5661

0100

Poliptymalizacja
i Komputerowe Wspomaganie Projektowania

MIELNO 2000



Politechnika Koszalińska
Katedra Systemów Sterowania
Komitet Budowy Maszyn PAN
Sekcja Podstaw Konstrukcji

ZESZYTY NAUKOWE WYDZIAŁU MECHANICZNEGO

Prace XVIII Ogólnopolskiej Konferencji

POLIOPTYMALIZACJA
I KOMPUTEROWE WSPOMAGANIE PROJEKTOWANIA
MIELNO 2000

Koszalin–Mielno 14–17 czerwca 2000

ZESZYTY NAUKOWE WYDZIAŁU MECHANICZNEGO NR: 27

ISSN 1640 - 4572

Materiały recenzowane

Redaktor
Wojciech Tarnowski

Wydanie materiałów dofinansowane przez KBN

WYDAWNICTWO UCZELNIANE POLITECHNIKI KOSZALIŃSKIEJ
75-620 Koszalin, ul. Raławicka 15-17

Koszalin 2000, wyd. I, ark. wyd. 21,32, format B-5, nakład 150 egz.
Druk: INTRO-DRUK, Koszalin

KOMITET OPINIODAWCZY

przewodniczący

prof. dr hab. inż. Wojciech TARNOWSKI

Politechnika Koszalińska

członkowie

dr inż. Jerzy BALICKI	Akademia Marynarki Wojennej w Gdyni
dr hab. inż. Zbigniew BANASZAK	Politechnika Zielonogórska
prof. dr hab. inż. Michał BIAŁKO	Politechnika Koszalińska
prof. dr hab. inż. Bogdan BRANOWSKI	Politechnika Poznańska
dr hab. inż. Andrzej BUCHACZ	Politechnika Śląska
dr hab. inż. Tadeusz BURCZYŃSKI	Politechnika Śląska
prof. dr hab. inż. Edward CHLEBUS	Politechnika Wrocławska
dr hab. inż., prof. WSM Romuald CWILEWICZ	Wyższa Szkoła Morska w Gdyni
dr hab. inż. Lucjan DĄBROWSKI	Politechnika Warszawska
prof. dr hab. inż. Zbigniew DĄBROWSKI	Politechnika Warszawska
prof. nzw. dr hab. inż. Jarosław DIAKUN	Politechnika Koszalińska
dr hab. inż., prof. PW, Wiesław GRZESIKIEWICZ	Politechnika Warszawska
dr inż. Tadeusz JELENIEWSKI	Politechnika Wrocławska
prof. dr hab. inż. Stefan JENDO	IPPT PAN w Warszawie
prof. dr hab. inż. Wojciech KACALAK	Politechnika Koszalińska
dr hab. inż. Ignacy KALISZEWSKI	Krajowy Depozyt Papierów Wartościowych S.A.
dr hab. inż., prof. P.P., Zbigniew KIERZKOWSKI	Politechnika Poznańska
prof. dr hab. Ewa KONARZEWSKA-GUBAŁA	AE we Wrocławiu
prof. dr hab. inż. Józef Korbicz	Politechnika Zielonogórska
prof. dr hab. inż. Zbigniew KOWALSKI	Politechnika Gdańska
dr inż. Paweł KRASOWSKI	Wyższa Szkoła Morska
prof. dr hab. inż. Ludomir LAUDAŃSKI	Politechnika Rzeszowska
dr hab. inż. prof. PŁ, Józef MATUSZEK	Politechnika Łódzka, filia w Bielsku Białej
dr hab. inż. Tadeusz MIKULCZYŃSKI	Politechnika Wrocławska
prof. dr hab. inż. Krzysztof MARCHELEK	Politechnika Szczecińska
prof. dr hab. inż. Antoni NIEDERLIŃSKI	Politechnika Śląska
dr inż. Zenon OCIEPA	Politechnika Koszalińska

prof. zw. dr hab. inż. Andrzej OŁĘDZKI	Politechnika Warszawska
dr hab. inż. Marian OSTWALD	Politechnika Poznańska
prof. dr hab. inż. Andrzej OSYCZKA	Politechnika Krakowska
dr inż. Tadeusz PANKOWSKI	Politechnika Poznańska
prof. dr hab. inż. Witold POGORZELSKI	Politechnika Warszawska, OND w Płocku
doc. dr inż. Jerzy PUŁACZEWSKI	Politechnika Warszawska
prof. dr hab. inż. Ryszard ROHATYŃSKI	Politechnika Wrocławska
dr inż. Edward RÓJ	Instytut Nawozów Sztucznych w Puławach
prof. dr hab. inż. Eugeniusz RUSIŃSKI	Politechnika Wrocławska
prof. dr hab. inż. Adam SIELICKI	Politechnika Zielonogórska
dr hab. inż., prof. PG Jan SIKORA	Politechnika Gdańska
prof. dr hab. inż. Roman SŁOWIŃSKI	Politechnika Poznańska
dr hab. inż., prof. PK, Jerzy SMYCZEK	Politechnika Koszalińska
prof. dr hab. inż. Borys STORCH	Politechnika Koszalińska
prof. dr hab. inż. Czesław SZYMCZAK	Politechnika Gdańska
prof. dr hab. inż. Ryszard TADEUSIEWICZ	AGH w Krakowie
dr inż. Zbigniew TARAPATA	Wojskowa Akademia Techniczna
prof. dr hab. inż. Eugeniusz TOCZYŁOWSKI	Politechnika Warszawska
prof. dr hab. Wiesław TRACZYK	Politechnika Warszawska
dr inż. Ewa WACHOWICZ	Politechnika Koszalińska
prof. dr hab. inż. Zenobia WEISS	Politechnika Poznańska
dr hab. inż., prof. PŚL., Konrad WOJCIECHOWSKI	Politechnika Śląska
prof. dr hab. inż. Józef WOJNAROWSKI	Politechnika Śląska
prof. dr hab. inż. Jerzy WRÓBEL	Politechnika Warszawska
prof. dr hab. inż. Mirosław ZABOROWSKI	Politechnika Śląska

SPIS TREŚCI

Arendt R. <i>Zastosowanie systemu ekspertowego do generacji modeli symulacyjnych podsystemów energetycznych statków</i>	13
Balicki J. <i>Adaptacyjny algorytm ewolucyjny do optymalizacji przydziału zadań w systemie czasu rzeczywistego</i>	21
Bartkiewicz S. <i>Optymalny wybór i kierunki rozwoju zintegrowanego systemu uzbrojenia</i>	29
Bielecki M. <i>Komputerowe monitorowanie znoszenia jaj w hodowli doświadczalnej</i>	37
Borowiecki T., Saniuk S., Banaszak Z. <i>Zastosowanie technik programowania w logice z ograniczeniami do zarządzania przepływem wieloasortymentowej produkcji rytmicznej</i>	45
Fligiel M. <i>Kształtowanie konstrukcji metodą optymalnego rozkładu rodzaju materiału</i>	53
Grzesikiewicz W., Wakulicz A. <i>Dynamika układu sprężysto–ciernego</i>	61
Kaliński K., Kucharski T., Sawiak S. <i>Dopasowanie prędkości obrotowej wrzeciona w celu przeciwdziałania drganiom narzędzie–przedmiot</i>	69
Kania L. <i>Charakterystyki zastępcze wałeczków przy obliczaniu nośności łożysk wieńcowych metodą elementów skończonych</i>	77
Kasperska R., Ostwald M. <i>Optymalizacja wielokryterialna trójwarstwowej paneli walcowej przy obciążeniach złożonych</i>	85
Kiczkowiak T., Sobczuk M. <i>Rozwiązanie problemu typu <i>Picus viridis</i> w środowisku MatlaB[®] Simulink[®]</i>	93
Kołopieńczyk A. <i>Wirtualna organizacja prac projektowych przy wykorzystaniu dynamicznych dokumentów</i>	101
Kosmowski K.T., Kwiesielewicz M. <i>Metoda oszacowania ryzyka w systemach technicznych z wykorzystaniem analizy przedziałowej</i>	107
Kotarski W., Pyszny J. <i>Efektywne metody wyznaczenia brzegu Pareto i punktu Salukwadze</i>	115

Kowalski Z., Meler Kapcia M., Zieliński S.	123
<i>Metoda określania podobieństwa statków w procesie doboru układów automatyki systemów energetycznych</i>	
Kruczkiewicz Z.	131
<i>Metoda oceny równoważności modeli oprogramowania obiektowego</i>	
Kubik T.	139
<i>Automatyczna generalizacja linii w systemach typu GIS</i>	
Kula K., Morawski L., Tomera M.	147
<i>Sterowanie rozmyte statkiem wzdłuż zadanej trajektorii z zadaną prędkością</i>	
Kwiesielewicz M., Uden E.	155
<i>Szeregowanie czynników z wykorzystaniem metody porównywania parami w sensie AHP</i>	
Lasota A.	163
<i>Koncepcja tworzenia reguł heurystycznych dla projektowania ergonomicznego</i>	
Matuszek J., Kukła S.	167
<i>Wykorzystanie modelowania i symulacji do analizy odlewniczych systemów wytwarzania</i>	
Matuszek J. Plinta D.	175
<i>Koszty produkcji w modelowaniu i symulacji procesów produkcyjnych</i>	
Michalski R., Mikołajczak P.	183
<i>Sterowanie stanem w systemie utrzymania maszyn krytycznych przemysłu spożywczego</i>	
Montusiewicz J.	191
<i>Filtracja zbioru rozwiązań optymalnych w sensie Pareto metodą przedziałów nierozróżnialności</i>	
Osyczka A., Krenich S., Montusiewicz J.	199
<i>Filtracja rozwiązań Pareto optymalnych w optymalizacji wielokryterialnej z wykorzystaniem algorytmów genetycznych</i>	
Piasecki M., Rogaliński P.	207
<i>Dydaktyka eksperymentalna z wykorzystaniem programowalnych klocków Lego MindStorms™</i>	
Piasecki M., Rogaliński P.	215
<i>Komputerowa analiza działania robotów w zawodach RoboCup</i>	
Piątkowski T., Sempruch J.	223
<i>Dynamiczny model procesu zgarniania jako obiekt optymalizacji ładunków jednostkowych</i>	
Pławiak–Mowna A.	231
<i>Problem efektywności dostępu do informacji składowanych w relacyjnej bazie danych – analiza operacji zapytań</i>	

Popławski A. <i>Optymalizacja procesu dokształcania z wykorzystaniem technologii nauczania na odległość</i>	237
Popławski A. <i>Pewna metoda optymalizacji struktur danych przechowywanych w pamięci systemu komputerowego</i>	241
Puścian R. <i>Zastosowanie algorytmu genetycznego w zagadnieniu selekcji strategii postępowania</i>	247
Rój E. <i>Analiza wpływu aerodynamik/hydrauliki na efektywność pracy urządzeń przemysłowych</i>	253
Rudzki K. <i>Komputerowe wspomaganie doboru parametrów żeglugi statków żaglowo-motorowych</i>	261
Sielicki A. <i>Ewolucyjna metoda znajdowania rozwiązań dopuszczalnych</i>	269
Sienkiewicz Z., Domski J. <i>Odpowiedź dynamiczna ośrodka gruntowego w postaci półprzestrzeni sprężystej obciążonej impulsowo</i>	277
Sitek P., Wikarek J., Zaborowski M. <i>Model symulacyjny tłoczni lekkiej w fabryce samochodów osobowych</i>	285
Skolud B. <i>Komputerowe wspomaganie podejmowania decyzji o przyjmowaniu zleceń produkcyjnych</i>	293
Staniec I. <i>Dychotomiczna klasyfikacja kredytobiorców przy użyciu sztucznych sieci neuronowych</i>	301
Szwed C., Toczyłowski E. <i>Optymalizacja rozdziału zasobów lokalowych w warunkach elastycznego studiowania</i>	309
Tarnowski W. <i>Pomiar zmiennej wyjściowej w układach regulacji z obiektami ciągłymi</i>	317
Tytyk E. <i>Ergonomiczne wspomaganie procesów projektowania</i>	325
Walczak A., Toczyłowski E. <i>Wspomaganie decyzji planowania dobowego na hurtowym rynku energii elektrycznej</i>	333
Warczyński J., Janiszewski M. <i>Automatyzacja pasteryzacji i sterylizacji konserw w autoklawach poziomych</i>	341
Witkowska D., Staniec I. <i>Dychotomiczna klasyfikacja kredytobiorców za pomocą wybranych metod klasyfikacji</i>	349
Wojnarowski J., Zawiślak S. <i>Zastosowanie matroidów w wybranych zagadnieniach modelowania układów mechanicznych i optymalizacji</i>	357

- Wójcik R.** 365
Podjęcie grafowe do unikania blokad w systemach procesów współbieżnych
- Wójcik T., Kucharski T.** 373
Optymalizacja parametrów okrętowych układów napędowych w celu rozprzężenia drgań skrętnych i wzdłużnych
- Wróblewska K.** 381
Aplikacja wspomagająca dobór generatorów dla siłowni okrętowej
- Zajac W.** 387
Estymacja optymalnych parametrów algorytmu korekcji obrazu w oparciu o odebrany sygnał
- Ciebielski M.** 393
Wprowadzenie ograniczeń do funkcji błędu sieci neuronowej na przykładzie syntezy regulatora neuronowego według schematu Widrow'a–Nguyen'a

KOMUNIKATY

- Warczyński J., Purol R.** 401
Automatyzacja sterowania magazynem zbożowym zakładu młynarskiego
- Zaborowski M., Pusiewicz A.** 409
Dydaktyczny model symulacyjny procesu montażu samochodów

CONTENTS

Arendt R. <i>An application of an expert system for creation of simulation models of power ship subsystems</i>	13
Balicki J. <i>Adaptive evolutionary algorithm for task assignment optimization in real time systems</i>	21
Bartkiewicz S. <i>The optimum selecting and the developmental direction's of integrated weapon's system</i>	29
Bielecki M. <i>Eggs monitoring computer system for a research farm</i>	37
Borowiecki T., Saniuk S., Banaszak Z. <i>Application of constraint logic programming for repetitive production flow management</i>	45
Fligiel M. <i>Forming construction by method of optimal distribution of sort of material</i>	53
Grzesikiewicz W., Wakulicz A. <i>Dynamics of elastic system with dry friction</i>	61
Kaliński K., Kucharski T., Sawiak S. <i>Matching the spindle speed in order to counteract the tool-workpiece vibration</i>	69
Kania L. <i>Equivalent characteristics of rimmed bearing rollers for large diameter bearings capacity calculations with the use of FEM</i>	77
Kasperska R., Ostwald M. <i>Multicriteria optimization of sandwich cylindrical panels under combined loads</i>	85
Kiczkowiak T., Sobczuk M. <i>Solution of <i>Picass Vividis</i> type problem in Matlab[®] Simulink[®] environment</i>	93
Kołopieńczyk A. <i>Virtual organisation of design activities using dynamic documents</i>	101
Kosmowski K.T., Kwiesielewicz M. <i>Risk assessment method in technical systems using interval analysis</i>	107
Kotarski W., Pyszny J. <i>On effective methods of determining of Pareto boundary and Salukwadze point</i>	115
Kowalski Z., Meler Kapcia M., Zieliński S. <i>Method of the ships similarity determination during ship automation design process</i>	123
Kruczkiewicz Z. <i>The equivalence examination method of the object oriented software models</i>	131

Kubik T. <i>Automated line generalization in geographical information systems</i>	139
Kula K., Morawski L., Tomera M. <i>Ship track-keeping fuzzy control with assumed velocity</i>	147
Kwiesielewicz M., Uden E. <i>Ranking factors using pairwise comparison method in sense of AHP</i>	155
Lasota A. <i>The concept of creating heuristic rules for the ergonomic design</i>	163
Matuszek J., Kukla S. <i>Using of modelling and simulation in analysis of foundry manufacturing systems</i>	167
Matuszek J. Plinta D. <i>Production costs in modelling and simulation of production processes</i>	175
Michalski R., Mikołajczak P. <i>Controlling state in the maintenance system of critical machines in food industry</i>	183
Montusiewicz J. <i>Filtering of the set of optimal solutions in Pareto sense by means of the undifferentiation interval method</i>	191
Osyczka A., Krenich S., Montusiewicz J. <i>Genetic algorithm based method for multicriteria optimization with filtration Pareto set</i>	199
Piasecki M., Rogaliński P. <i>Learning by doing with programmable bricks LEGO MindStorms™</i>	207
Piasecki M., Rogaliński P. <i>Computer analysis of robot's behaviour in RoboCup competition</i>	207
Piątkowski T., Sempruch J. <i>Dynamic model of scraping process as an object of distribution process optimization of unitary load stream</i>	215
Pławiak-Mowna A. <i>RDB access effectivity problem analysis – request optimisation method</i>	231
Popławski A. <i>Optimisation of education process by using a distance learning technology</i>	237
Popławski A. <i>A method of optimisation of data structures stored in computer memory</i>	241
Puścian R. <i>Application of genetical algorithm to solve the strategy selection problem</i>	247
Rój E. <i>Analysis of the aerodynamic/hydraulic influence on the efficiency of industrial equipment performance</i>	253

Rudzki K. <i>Computer-aided selection of sailing parameters for sail-motor vessels</i>	261
Sielicki A. <i>Evolutionary method of searching for acceptable solutions</i>	269
Sienkiewicz Z., Domski J. <i>Dynamic response of an elastic half-space soil medium under impulsive loadings</i>	277
Sitek P., Wikarek J., Zaborowski M. <i>Simulation model of light press in a car factory</i>	285
Skolud B. <i>Computer aided decision making about production order validation</i>	293
Staniec I. <i>Dychotomus classification borrower with use artificial neural networks</i>	301
Szwed C., Toczyłowski E. <i>Optimization of space allocation in the course timetabling problem</i>	309
Tarnowski W. <i>Outcome variable measurement problem in industrial control systems with the space continuous objects</i>	317
Tytyk E. <i>Ergonomic aiding of design processes</i>	325
Walczak A., Toczyłowski E. <i>Unit - commitment decision support system on energy wholesale market</i>	333
Warczyński J., Janiszewski M. <i>On automating the pasteurization and sterilization processes in horizontal autoclaves</i>	341
Witkowska D., Staniec I. <i>Dichotomus classification of borrowers using selected methods</i>	349
Wojnarowski J., Zawiślak S. <i>An application of matroids to some optimization problems and to modelling of mechanical systems</i>	357
Wójcik R. <i>Graph based approach to deadlock avoidance in systems of concurrent processes</i>	365
Wójcik T., Kucharski T. <i>Optimum design of the ship propulsion system to uncouple torsional and axial vibration</i>	373
Wróblewska K. <i>Application for supporting ship electric energy balance</i>	381
Zajac W. <i>Definition of the optimal parameters of an algorithm for correction of images on the received signal</i>	387

Ciebielski M. 393
The constraints inserting into a neural network error function on the example of neural controller synthesis in the Widrow–Nguyen scheme

ORAL PRESENTATIONS

Warczyński J., Purol R. 401
On automatic control of a grain storehouse of a flour–mill workshop

Zaborowski M., Pusiewicz A. 409
Educational simulation model of the car assembly process

ZASTOSOWANIE SYSTEMU EKSPERTOWEGO DO GENERACJI MODELI SYMULACYJNYCH PODSYSTEMÓW ENERGETYCZNYCH STATKÓW

1 Wstęp

Zastosowanie systemu ekspertowego do projektowania układów automatyki statków poprzez wykorzystanie reguł wiedzy i udostępnienie dużego zbioru informacji zawartego w bazach danych i bazach wiedzy o zbudowanych statkach, stosowanych systemach, układach i elementach wspomaga proces projektowy [5]. Wykorzystanie modeli matematycznych elementów, układów i systemów statków może stanowić źródło głębokiej wiedzy w systemie ekspertowym. Zastosowanie badań symulacyjnych w fazie projektowej umożliwi: ocenę efektywności i poprawności przyjętych rozwiązań projektowych, dobór parametrów nastaw układów regulacji, ocenę statycznych i dynamicznych właściwości projektowanych systemów oraz opracowanie metod diagnostyki systemów. Implementacja badań symulacyjnych w systemie ekspertowym wymagała systematyzacji [4,6] i rozwiązania szeregu zagadnień między innymi:

- doboru struktur modeli elementów składowych systemu zapewniających bliską topologię modelu symulacyjnego względem schematów projektowych, co ułatwia opracowanie reguł wiedzy zestawiania modeli symulacyjnych systemu [2];
- określenia kryteriów wiarygodności modelu – bliskości charakterystyk modelu względem obiektu rzeczywistego [3];
- opracowania metod doboru parametrów modelu zapewniających wiarygodność modelowania określonych typów obiektów rzeczywistych [3];
- opracowania bibliotek modeli matematycznych elementów składowych [1];
- formalizacji kryteriów oceny systemów zawartych w przepisach towarzystw klasyfikacyjnych.

System ekspertowy w opracowanej aplikacji traktowany jest jako inteligentne środowisko programowe umożliwiające współpracę z różnymi programami zewnętrznymi. W trakcie sesji projektowej zestawiane są w trybie konwersacyjnym dane opisujące projektowany podsystem energetyczny, sprawdzana jest poprawność merytoryczna wykonanego projektu, tworzony jest model matematyczny zaprojektowanego podsystemu i przywoływana jest sesja badań symulacyjnych w środowisku Matlab/Simulink.

W referacie omówiono zagadnienia automatycznego generowania modeli symulacyjnych projektowanych podsystemów energetycznych statków w języku środowiska programowego Matlab/Simulink.

* Katedra Automatyki Wydział Elektrotechniki i Automatyki Politechniki Gdańskiej, Gdańsk,
Narutowicza 11/12, e-mail: rarendt@ely.pg.gda.pl

2 Aplikacja systemu ekspertowego wspomagająca proces projektowania podsystemów energetycznych statków

Opracowana aktualnie aplikacja w systemie ekspertowym Exsys Developer V.7.0, Multilogic Inc. zawiera opcję wprowadzania danych w trybie konwersacyjnym przez projektanta podsystemu energetycznego statku [6]. Program pyta kolejno o:

- liczbę poszczególnych typów ogólnych elementów składowych podsystemu;
- nazwy własne stosowanych elementów;
- kolejne elementy przetwarzające energię począwszy od elementu początkowego do napotkania elementu końcowego (opis struktury podsystemu).

Zebrałe dane zestawiane są w ramie struktur.frm i wyświetlane w trakcie sesji na ekranie w celu akceptacji poprawności zestawionych danych przez projektanta.

Do oceny poprawności zestawionych danych wykorzystano wiedzę zestawioną w postaci reguł. W regułach wiedzy oceniających poprawność zastosowanych elementów i przyjętej struktury projektowanych podsystemów energetycznych przyjęto szereg założeń np.:

- moment obrotowy jednego silnika nie powinien być przekazywany za pośrednictwem wału śrubowego do drugiego silnika;
- na wyjściu wału śrubowego nie powinno się stosować sprzęgła;
- nie powinno się łączyć dwóch sprzęgieł szeregowo;
- pomiędzy sprzęgłem, a śrubą powinien wystąpić element pośredniczący – wał śrubowy.

Reguły oceniające poprawność zestawianych danych sygnalizują osiągnięcie założonego celu, przy przyjętej pewności stwierdzenia (0/10 do 10/10). W zależności od wskazanego przez reguły celu i stopnia jego pewności reguły warunkują dalszą pracę programu. Możliwe są następujące przypadki:

1. zestawione dane są niekompletne, źle zestawione, bądź zaprojektowany podsystem ma niedopuszczalną strukturę;
2. zestawione dane wskazują na nietypową strukturę projektowanego podsystemu energetycznego;
3. wprowadzone dane są poprawne i kompletne.

W przypadku 1 projektant może zakończyć pracę programu lub rozpocząć od nowa zestawianie danych opisujących projektowany podsystem energetyczny. W przypadku 2 projektant może rozpocząć od nowa zestawianie danych opisujących projektowany podsystem, bądź przejść do sesji symulacyjnej. W przypadku 3 program bez ingerencji operatora przechodzi do sesji symulacyjnej.

3 Algorytmy zestawiania M - plików programu Simulink

Zestawione w ramie struktur.frm dane opisujące elementy i strukturę projektowanego podsystemu stanowią podstawę opisu modelu symulacyjnego. Proces zestawiania M - pliku programu symulacyjnego można podzielić na dwie podstawowe części:

- kolejne dołączanie M - plików z opisami modeli elementów składowych do M - pliku modelu symulacyjnego projektowanego podsystemu napędowego;
- utworzenie obrazu graficznego modelu projektowanego podsystemu.

Głównym źródłem informacji przy zestawianiu M - pliku programu symulacyjnego jest rama położenie.frm (tablica 1).

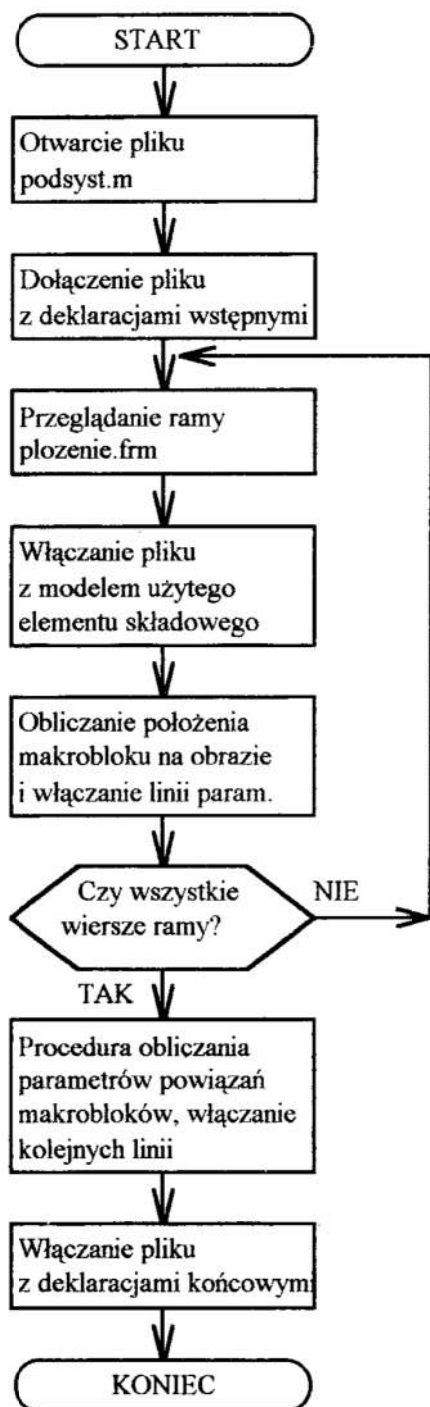
Tablica 1. Rama reprezentująca szkic obrazu graficznego przykładowego podsystemu energetycznego

LP	ELEMENT	WIERSZ	KOLUMNA
1	SILNIK1	1	1
2	SILNIK2	0	0
3	SILNIK3	0	0
4	SILNIK4	0	0
5	WALSRUB1	1	2
6	WALSRUB2	0	0
7	WALSRUB3	0	0
8	WALSRUB4	0	0
9	SPRZEGLO1	0	0
10	SPRZEGLO2	0	0
11	SPRZEGLO3	0	0
12	SPRZEGLO4	0	0
13	SRUBANAST1	1	3
14	SRUBANAST2	0	0
15	SRUBANAST3	0	0
16	SRUBANAST4	0	0
17	SRUBASTAL1	0	0
18	SRUBASTAL2	0	0
19	SRUBASTAL3	0	0
20	SRUBASTAL4	0	0
21	PRZEKLADNIASUM1	0	0
22	PRZEKLADNIASUM2	0	0
23	PRZEKLADNIASUM3	0	0
24	PRZEKLADNIAROZ1	0	0
25	PRZEKLADNIAROZ2	0	0
26	PRADNICAWAL1	0	0
27	PRADNICAWAL2	0	0

Tworzony obraz graficzny modelu podsystemu energetycznego podzielony jest na zdyskretyzowane pola, którym przyporządkowane są indeksy wiersza i kolumny (m,n). Dla elementu początkowego SILNIK1 przyporządkowane jest pole (1,1), dalsze powiązane elementy z silnikiem opisywane są polami (1,2)...(1,n), aż do napotkania elementu końcowego. Wystąpienie w wierszu przekładni rozdzielającej (2 wyjścia), powoduje przyporządkowanie elementom powiązanim z drugim wyjściem pól z indeksem (2,n). Kolejne silniki umieszczane są w polach (m+1, 1), przy czym proces przyporządkowywania indeksów pól kolejnym elementom jest powtarzany. Źródłem informacji o kolejnych powiązaniach elementów jest rama struktur.frm. Procedury zestawiania ramy położenie.frm nie wykorzystują reguł wiedzy.

W celu określenia algorytmów tworzenia M - pliku programu symulacyjnego z modelem podsystemu dokonano analizy formalnego opisu M - plików oraz zawartych informacji pod kątem semantycznym i lingwistycznym. Interesujący dla stosowanych procedur jest następujący podział M - pliku:

- nazwa pliku, deklaracje i formalne opisy wstępne;
- zasadnicza część pliku opisująca strukturę modelu symulacyjnego;



Rys. 1. Algorytm zestawiania M - pliku modelu symulacyjnego podsystemu

Fig. 1. Algorithm of M - file creation with a simulation model

- linia deklarująca parametry położenia makrobloku na obrazie graficznym;
- deklaracje i formalne zamknięcie pliku.

Zgodnie z powyższym zredagowano następujące pliki:

- wpodzesp.m – plik zawierający nazwę podsyst.m, deklaracje i opisy wstępne;
- wsiln1.m, wsiln2.m, wsiln3.m, wsiln4.m – pliki zawierające modele silników;
- wwal1.m, wwal2.m, wwal3.m, wwal4.m – pliki zawierające modele wałów śrubowych;
- wsprz1.m, wsprz2.m, wsprz3.m, wsprz4.m – pliki zawierające modele sprzęgła ciernego i podatnego;
- wsrst1.m, wsrst2.m, wsrst3.m, wsrst4.m – pliki zawierające modele śrub o skoku stałym i nastawnym;
- wpsum1.m, wpsum2.m, wpsum3.m – pliki zawierające modele przekładni sumujących momenty obrotowe;
- wproz1.m, wproz2.m – pliki zawierające modele przekładni rozdzielających momenty obrotowe;
- wprwal1.m, wprwal2.m – pliki zawierające modele prądnic wałowych;
- wkpodzesp.m – plik z deklaracjami i formalnym zamknięciem M - pliku.

Procedura zestawiania M - pliku z opisem modelu symulacyjnego projektowanego podsystemu energetycznego otwiera plik o nazwie podsyst.m (kasuje zawartość istniejącego) i wstawia komentarz „%Plik opisujący model symulacyjny podsystemu dla Matlaba utworzony przez Exsys”. Następnie włączana jest do pliku podsyst.m nazwa, deklaracje i formalne opisy wstępne poprzez wpisanie zawartości pliku wpodzesp.m. Przeglądane są kolejne wiersze ramy plozenie.frm i przy napotkaniu niezerowych indeksów

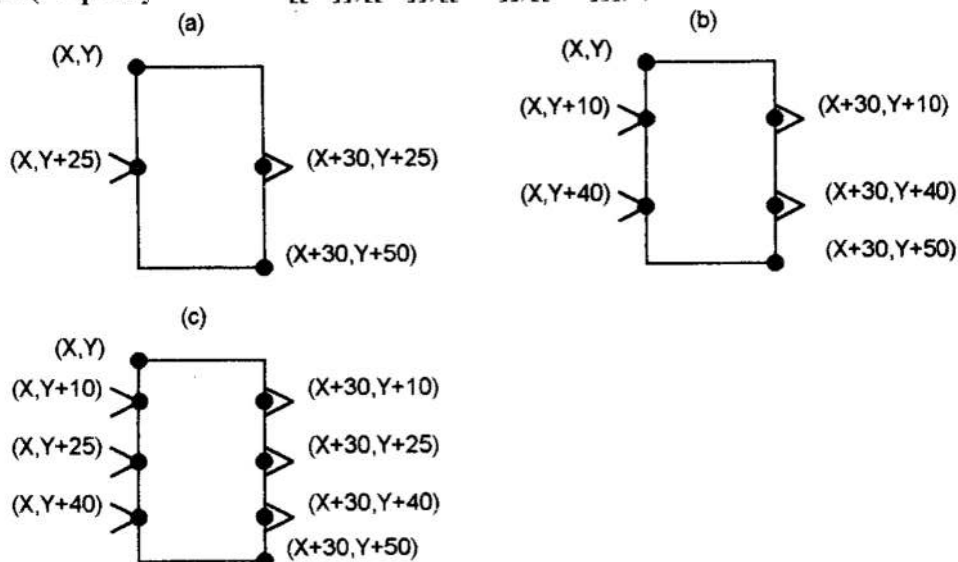
kolumn i wierszy włączane są kolejne pliki z modelami elementów składowych. Po każdym włączeniu pliku uruchamiana jest procedura obliczająca położenie makrobloku na obrazie graficznym. Dołączony plik uzupełniany jest linią deklarującą parametry

położenia. Po przeglądnięciu całej ramy położenie.frm uruchamiana jest procedura obliczania i dołączania do pliku podsyst.m kolejnych linii określających powiązania makrobloków. Procedura zestawiania pliku podsyst.m kończy się po dołączeniu pliku z deklaracjami i formalnym zamknięciem ciągu instrukcji (wkpodzes.m). W procedurach włączania kolejnych plików biorą udział reguły wiedzy. Algorytm zestawiania pliku podsyst.m z opisem modelu symulacyjnego projektowanego podsystemu energetycznego przedstawia rysunek 1.

4 Algorytmy obliczania parametrów położenia makrobloków na obrazie graficznym

Dla każdego makrobloku przyporządkowane jest pole graficzne 100x120 pikseli. Rozmiar i położenie makrobloku (obrazu graficznego zwiniętego modelu) definiowane jest przez podanie współrzędnych X, Y lewego górnego rogu i X1, Y1 dolnego prawego rogu. Na podstawie ramy położenie.frm obliczane są kolejno współrzędne [X], [Y], [X1], [Y1] lewego górnego i dolnego rogu makrobloku. Obliczane wartości są dopisywane do pliku w postaci linii przez instrukcję:

BB-WR(-Bpodsyst.m -A' [[X]],[[Y]],[[X1]],[[Y1]])').



Rys. 2. Obrazy graficzne makrobloków i względne współrzędne położenia ich wejść i wyjść

Fig. 2. Graphical images of macroblock and their relative placement of inputs and outputs

Należy zwrócić uwagę na fakt, że program Simulink ma obraz graficzny oparty o raster (5x5) pikseli, w związku z czym zadeklarowane położenie makrobloku może zostać skorygowane względem siatki. Korekcja położenia prowadzona jest w taki sposób, aby wejścia i wyjścia makrobloku leżały na liniach siatki.

W celu obliczenia parametrów powiązań makrobloków przeprowadzono analizę ich używanych typów, ze względu na liczbę wejść i wyjść (rys.2).

Następujące modele elementów składowych są przyporządkowane makroblokom:

- a) 1 wejście i 1 wyjście: wysokoprężny silnik spalinowy; śruba o skoku nastawnym; śruba o skoku stałym; prądnicza wałowa;

- b) 2 wejścia i 2 wyjścia: wał śrubowy; sprzęgło;
- c) 3 wejścia i 3 wyjścia: przekładnia sumująca momenty obrotowe; przekładnia rozdzielająca momenty obrotowe;

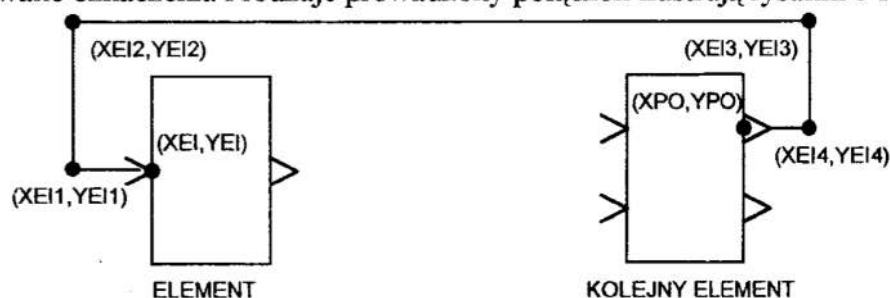
Procedura obliczania parametrów powiązań makrobloków przegląda kolejne wiersze ramy położenie.frm. Gdy napotkany element ma niezerowe indeksy wiersza i kolumny uruchamiane są reguły, które obliczają współrzędne:

- [XEI, YEI] – położenia wejścia makrobloku;
- [XEO, YEO] – położenia wyjścia makrobloku;
- [XEI1, YEI1] – położenia 1 punktu załamania linii prowadzonej od kolejnego elementu;
- [XPI4, YPI4] – położenia ostatniego punktu załamania linii prowadzonej do kolejnego elementu.

Określony jest również kolejny powiązany element, dla którego reguły określają współrzędne:

- [XPI, YPI] – położenia wyjścia makrobloku kolejnego elementu;
- [XPO, YPO] – położenia wejścia makrobloku kolejnego elementu;
- [XEI4, YEI4] – położenia pierwszego punktu załamania linii prowadzonej od kolejnego elementu;
- [XPI1, YPI1] – położenia ostatniego punktu załamania linii prowadzonej do kolejnego elementu.

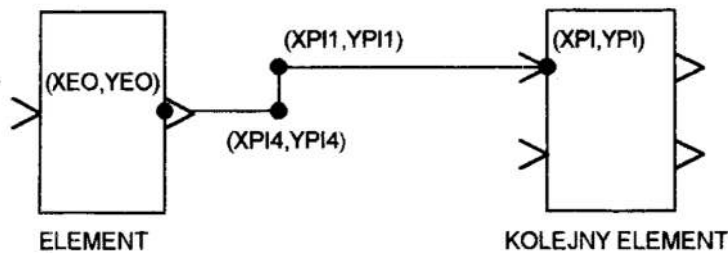
Stosowane oznaczenia i rodzaje prowadzonych połączeń ilustrują rysunki 3 i 4.



Rys. 3. Prowadzenie połączenia do wejścia elementu
Fig. 3. A draw of a connection to an input of element

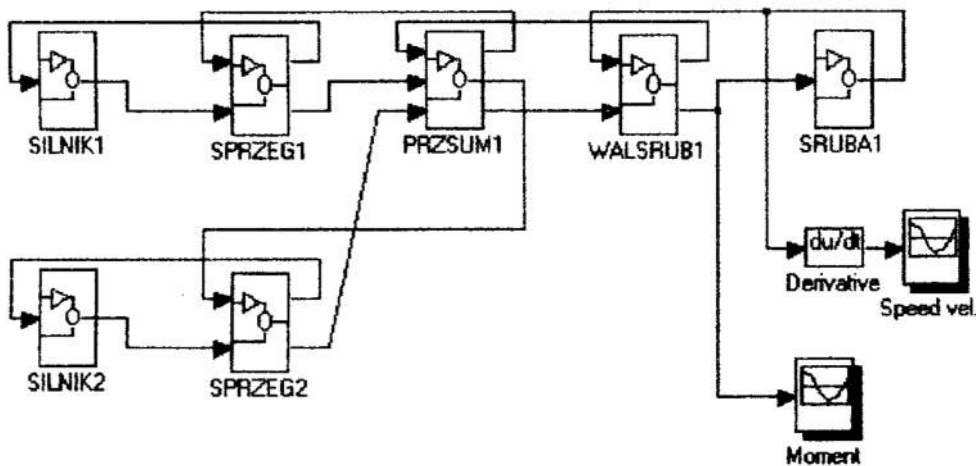
W zależności od numerów wierszy i kolumn oraz stosowanych elementów reguły obliczają parametry punktów pośrednich załamania linii (XEI2, YEI2), (XEI3, YEI3), (XPI1, YPI1), (XPI4, YPI4). W linii parametrów połączenia należy podawać kolejno współrzędne punktów począwszy od wyjścia elementu (wyjścia kolejnego elementu), a kończąc na wejściu kolejnego elementu (wejściu elementu). Kolejność odwrotna nie będzie przez program Simulink akceptowana, ponieważ połączenia odwzorowują kierunek przepływu sygnałów.

Po obliczeniu współrzędnych punktów połączeń uruchamiane są reguły R229 – R234 dołączające kolejne linie pliku podsyst.m z parametrami powiązań. Przedstawione na rysunkach 1.7 i 1.8 powiązania i omówione procedury obejmują makrobloki 0 1-2 wejściami i 1-2 wyjściami. W przypadku makrobloku o 3 wejściach i 3 wyjściach proces łączenia jest powtarzany dwukrotnie i wykorzystywane są zmienne sterujące procedury łączenia.



Rys. 4. Prowadzenie połączenia do wejścia kolejnego elementu
 Fig. 4. A draw of a connection to an input of next element.

Po zestawieniu pliku podsystem.m z opisem modelu symulacyjnego projektowanego podsystemu wywoływany jest program Matlab instrukcją: RUN(„C:\MATLAB\matlab-s.exe”). Po wywołaniu programu Matlab procedura Startup.m wywołuje automatycznie program Simulink z modelem podsystem.m, wyświetlając na ekranie obraz graficzny modelu zaprojektowanego podsystemu napędowego. Jako przykład zaprojektowano podsystem napędowy statku złożony z dwóch silników pracujących równoległe poprzez sprzęgła przekładnię sumującą momenty, wał śrubowy na śrubę o skoku stałym. Po wprowadzeniu danych system ekspertowy zestawia model symulacyjny (rys.5).



Rys. 5. Struktura modelu symulacyjnego zaprojektowanego podsystemu napędowego
 Fig. 5. The structure of a simulation mode of the designed power subsystem

5 Uwagi i wnioski

Opracowana aplikacja badań symulacyjnych umożliwia ich prowadzenie w sposób bardzo efektywny. Proces zestawiania danych opisujących projektowany podsystem energetyczny i automatyczne przejście do sesji badań trwa zaledwie kilka minut. Badania pozwalają na ocenę postawionych wymagań technicznych dotyczących procesów przejściowych w podsystemach pojawiających się przy: załączaniu, wyłączaniu, zmianie parametrów pracy podsystemu, jak również wystąpieniu czynników zakłócających jego działanie.

Prezentowana aplikacja badań symulacyjnych w systemie z bazą wiedzy realizowana jest w ramach Projektu Badawczego Komitetu Badań Naukowych.

Literatura

1. Arendt R.: *Modelowanie właściwości okrętowych zespołów napędu głównego statku z zastosowaniem programu SIMULINK*. XII Krajowa Konferencja Polioptymalizacja i Komputerowe Wspomaganie Projektowania Mielno'94', 22-25 czerwca 1994, Zeszyty Naukowe Wydziału Mechanicznego WSI Koszalin nr 18, s.7-18.
2. Arendt R.: *The choice of component elements for model structures which assure the topological compatibility of a simulation model with the configuration of a power unit*. Polioptymalizacja i Komputerowe Wspomaganie Projektowania MIELNO'98 Mielno, czerwiec 1998r.
3. Arendt R., Kwiesielewicz M.: *Weryfikacja modeli procesów technologicznych z użyciem algorytmów genetycznych*. II Konferencja "Automatyka w Rolnictwie i Przetwórstwie Spożywczym" Mielno wrzesień 1998r.
4. Arendt R., Kowalski Z.: *Simulation investigations in an expert system for ship automation aided design*. Proceedings of the 4th International Conference on Unconventional Electromechanical and electrical Systems "UEES'99", 21-24.06.1999, V2 pp. 541-546.
5. Lee D., Lee K. H.: *An approach to case-based system for conceptual ship design assistant*. Expert Systems with Applications, 1999, Vol. 16, 97-104.
6. Arendt R.: *An application of an expert system for aided design of power ship subsystems automation*. Referat zgłoszony na 7 International Symposium on Models in Automation and Robotics, 2000, Międzyzdroje, Poland.

Streszczenie

Zastosowanie systemu ekspertowego do projektowania układów automatyki statków poprzez wykorzystanie reguł wiedzy i informacji zawartych w bazach danych i bazach wiedzy wspomaga proces projektowy. Wykorzystanie badań symulacyjnych systemów statków może stanowić źródło głębokiej wiedzy w systemie ekspertowym. W referacie przedstawiono osiągnięte wyniki związane z aplikacją badań symulacyjnych w systemie ekspertowym dotyczące algorytmów tworzenia M - plików modeli symulacyjnych programu Simulink.

AN APPLICATION OF AN EXPERT SYSTEM FOR CREATION OF SIMULATION MODELS OF POWER SHIP SUBSYSTEMS

Summary

An application of an expert system to design ship automation systems with a use of knowledge rules and an information contained in data bases and knowledge bases aids the design process. A use of simulation investigations concerning ship sub-systems may be a source of a deep knowledge in the expert system. Results concerning an application of simulation investigation in the expert system, concerning creation of M - files with simulation model for Simulink program are presented in this paper.

Referat recenzowany

ADAPTACYJNY ALGORYTM EWOLUCYJNY DO OPTYMALIZACJI PRZYDZIAŁU ZADAŃ W SYSTEMIE CZASU RZECZYWISTEGO

1. Wprowadzenie

Dynamiczny rozwój technologii w zakresie sieci komputerowych owocuje nie tylko powszechnym dostępem do Internetu, ale także wspomaga rozwój teorii i zastosowań programowania rozproszonego. Wprowadzane standardy komunikacji między procesami CORBA czy DCOM umożliwiają komunikowanie się aplikacji napisanych w różnych językach programowania [4]. Skrócenie czasu realizacji złożonego programu w rozproszonym systemie komputerowym jest możliwe poprzez podzielenie programu rozproszonego na odrębne moduły oraz wykonanie tych modułów na różnych komputerach. Modułem może być m.in. aplet Javy [9] stosujący protokół komunikacyjny TCP/IP [4], segment programu napisanego w języku C++ wykorzystujący mechanizmy komunikacyjne systemu operacyjnego QNX [14], a nawet arkusz kalkulacyjny MS Excel przesyłający dane do innej aplikacji MS Office użytkowanej na innym komputerze lokalnej sieci komputerowej.

Odpowiednia alokacja modułów może wykorzystać specyficzne cechy i komponenty zestawów komputerowych takie, jak wydajność procesorów, wielkość pamięci operacyjnej, czas dostępu do rekordu na dysku, wydajność karty graficznej czy szybkość drukarki. Korzyści wynikające z przydziału modułów do odpowiednich komputerów mogą być zniwelowane jedynie przez powstające opóźnienia komunikacyjne. Tym niemniej w odniesieniu do racjonalnie zaprojektowanych przydziałów globalne skrócenie czasu przetwarzania danych może być znaczące.

W rozproszonych systemach komputerowych jako kryterium oceny jakości przydziałów segmentów programu przyjmuje się łączne obciążenie zadaniami całego systemu, obciążenie najbardziej eksploatowanego komputera, koszt użytkowania programu czy miarę niezawodności, np. prawdopodobieństwo poprawnego przebiegu programu. W systemach czasu rzeczywistego dąży się do skrócenia czasu reakcji systemu, co jest związane z minimalizacją obciążenia niewralgicznego komputera.

Minimalizacja obciążenia „wąskiego gardła” systemu może być również osiągnięta za pomocą właściwego doboru komputerów do pracy w sieci. Koszt komputera zależy w dużym stopniu od jego wydajności oraz od wielkości zasobów. Jednakże zbyt niskie obniżenie kosztu komputerów wchodzących w skład systemu rozproszonego może znacząco wydłużyć czasy wykonania poszczególnych modułów. Dodatkowo rozpatruje się dążenie do maksymalizacji wydajności systemu, gdyż pewne zadania mogą pojawiać się sporadycznie i system powinien dysponować wystarczającym zapasem mocy obliczeniowej, aby je zrealizować.

* Akademia Marynarki Wojennej, 81-919 Gdynia, ul. Śmidowicza, jbalicki@amw.gdynia.pl

W zaistniałej sytuacji konfliktowej sformułowano problem wyznaczania przydziałów modułów programów do komputerów w postaci zagadnienia optymalizacji trójkryterialnej, w którym poszukuje się reprezentacji rozwiązań optymalnych w sensie Pareto. Projektant rozproszonego systemu informatycznego może z wyznaczonego zbioru przydziałów modułów wybrać to rozwiązanie, które spełnia dodatkowe wymaganie lub zastosować podejście selekcji bazującej na wykorzystaniu punktów referencyjnych [1].

2. Ocena przydziału zadań w systemie

Modułem nazywamy sekwencję instrukcji programu wyodrębnioną z tego programu za pomocą ograniczników i mającą identyfikator, przez który możliwe jest odwołanie się tego modułu. Program podzielony na moduły jest elastyczny, gdyż niezbędne modyfikacje tego programu mogą być wykonane tylko w pewnym podzbiórze jego modułów. Ocenia się, że koszt modyfikacji istniejących programów stanowi ponad 50% kosztu ich wartości. Moduły pozwalają podnieść jakość i przejrzystość pisanych programów. Obniżają również ryzyko popełnienia błędu. Jeżeli programista dokładnie przetestuje moduł, to może wielokrotnie wykorzystywać raz napisany kod.

Zadaniem (procesem) nazywamy wykonanie modułu programu. Moduł programu jest obiektem pasywnym, podobnie jak zawartość pliku przechowywanego na dysku. Natomiast zadanie jest obiektem aktywnym ze stosem przechowującym dane tymczasowe (parametry procedur, adresy powrotne, zmienne tymczasowe), sekcją zmiennych globalnych oraz licznikiem rozkazów określającym następny rozkaz do wykonania. Każdy moduł może być wielokrotnie wywoływany podczas przebiegu programu, a to oznacza, że może mu odpowiadać wiele zadań.

Niech $\mathfrak{M} = \{m_1, \dots, m_v, \dots, m_V\}$ reprezentuje zbiór wykonywanych zadań w systemie. Przyjmuje się, że zadanie m_v może być wykonywane na różnych typach komputerów należących do zbioru dostępnych typów komputerów $\Pi = \{\pi_1, \dots, \pi_j, \dots, \pi_J\}$. Komputery mogą być usytuowane tylko w ustalonych węzłach należących do zbioru $W = \{w_1, \dots, w_i, \dots, w_I\}$. Pary węzłów są połączone kanałami umożliwiającymi realizację interakcji między modułami programu np. synchronizację przetwarzania modułów, komunikację pomiędzy modułami oraz rywalizację o zasoby. Przyjmuje się założenie, że w węzle może być usytuowany dokładnie jeden komputer wybrany ze zbioru możliwych typów komputerów.

Przydział zadań do węzłów systemu rozproszonego określa się za pomocą binarnego wektora, jak niżej:

$$x^m = [x_{11}^m, \dots, x_{1i}^m, \dots, x_{1I}^m, \dots, x_{v1}^m, \dots, x_{vI}^m]^T, \quad (1)$$

gdzie $x_{vi}^m = \begin{cases} 1 & \text{gdy } m_v \text{ przydziela się do } w_i, \\ 0 & \text{w przeciwnym razie,} \end{cases}$ dla $v = \overline{1, V}$, $i = \overline{1, I}$.

Przydział typów komputerów do węzłów w postaci binarnego wektora przydziału, który zapisuje się w następujący sposób:

$$x^\pi = [x_{11}^\pi, \dots, x_{1j}^\pi, \dots, x_{1I}^\pi, x_{21}^\pi, \dots, x_{ij}^\pi, \dots, x_{IJ}^\pi]^T, \quad (2)$$

gdzie $x_{ij}^\pi = \begin{cases} 1 & \text{gdy do } w_i \text{ przydzielono komputer typu } \pi_j, \\ 0 & \text{w przeciwnym razie,} \end{cases}$ dla $i = \overline{1, I}, j = \overline{1, J}$.

Łączny przydział zadań do węzłów oraz przydział komputerów do węzłów nazywamy przydziałem zadań do komputerów. Binarny wektor przydziału zadań do komputerów oznaczamy, jak niżej:

$$x = [x_{11}^m, \dots, x_{1i}^m, \dots, x_{1J}^m, \dots, x_{vi}^m, \dots, x_{vJ}^m, x_{11}^\pi, \dots, x_{1j}^\pi, \dots, x_{1J}^\pi, \dots, x_{ij}^\pi, \dots, x_{IJ}^\pi]^T. \quad (3)$$

Wprowadzcie podstawowe preferencje projektowe zwykle polegają na minimalizacji czasu przetwarzania danych, to ze względu na brak analitycznej zależności do wyznaczenia tego czasu w systemach czasu rzeczywistego przyjmuje się założenie o minimalizacji obciążenia najbardziej eksploatowanego komputera. Chu i Lan zastosowali to kryterium po raz pierwszy do optymalizacji rozmieszczenia 23 zadań na 3 komputerach w systemie obrony powietrznej DPAP [3].

Maksymalne obciążenie węzła dla zadanego przydziału modułów do komputerów x wyraża się następującą formułą:

$$Z_{\max}(x) = \max_{i \in \overline{1, J}} \left\{ \sum_{j=1}^J \sum_{v=1}^V t_{vj} x_{vi}^m x_{ij}^\pi + \sum_{v=1}^V \sum_{u=1}^V \sum_{\substack{i_2=1 \\ u \neq v, i_2 \neq i}}^I \tau_{vu} x_{vi}^m x_{ui_2}^m \right\}, \quad (4)$$

gdzie

t_{vj} – szacowany czas wykonania v -tego modułu na komputerze j -tego rodzaju,

τ_{vu} – szacowany czas komunikacji między v -tym modułem a u -tym modułem.

Drugim rozważanym kryterium jest koszt zakupionych komputerów, który dla przyjętych oznaczeń jest wyznaczany w poniższy sposób:

$$F_2(x) = \sum_{i=1}^I \sum_{j=1}^J \kappa_j x_{ij}^\pi, \quad (5)$$

gdzie κ_j – koszt komputera typu π_j [JM-jednostka monetarna].

Trzecim rozważanym kryterium jest sumaryczna wydajność przydzielonych komputerów, którą nazywamy mocą obliczeniową rozproszonego systemu komputerowego wyznaczaną w poniższy sposób:

$$\tilde{F}_2(x) = \sum_{i=1}^I \sum_{j=1}^J \mathcal{G}_j x_{ij}^\pi, \quad (6)$$

gdzie \mathcal{G}_j - wydajność komputera typu π_j [Flops].

Wydajność komputerów może być mierzona za pomocą różnych testów wydajności. Zestawienie wydajności najpopularniejszych komputerów wg wybranych benchmarków zamieszczono w [13].

3. Sformułowanie zadania optymalizacji wielokryterialnej

Rozważa się problem wyznaczania reprezentacji rozwiązań optymalnych w sensie Pareto sformułowany w postaci następującego zagadnienia.

Dla danych: $\mathbf{T} = [t_{vj}]_{V \times J}$, $\tau = [\tau_{vu}]_{V \times V}$, $\mathcal{G} = [\mathcal{G}_1, \dots, \mathcal{G}_j, \dots, \mathcal{G}_J]$,
 $\mathcal{K} = [\mathcal{K}_1, \dots, \mathcal{K}_j, \dots, \mathcal{K}_J]$ należy wyznaczyć reprezentację zbioru rozwiązań
niezdominowanych \mathcal{X}_N^{\leq} problemu

$$(\mathcal{X}, F, R_{\leq}), \quad (7)$$

gdzie

1. \mathcal{X} - zbiór rozwiązań dopuszczalnych

$$\mathcal{X} = \{x \in \mathcal{B}^{I(V+J)} \mid x = [x_{11}^m, \dots, x_{1i}^m, \dots, x_{1J}^m, \dots, x_{vi}^m, \dots, x_{VJ}^m, x_{11}^\pi, \dots, x_{1j}^\pi, \dots, x_{1J}^\pi, \dots, x_{ij}^\pi, \dots, x_{IJ}^\pi]^T$$

$$\left. \sum_{i=1}^I x_{vi}^m = 1 \text{ dla } v = \overline{1, V}; \sum_{j=1}^J x_{ij}^\pi = 1 \text{ dla } j = \overline{1, J} \right\}.$$
2. F - kryterium jakości, $F(x) = [Z_{\max}(x), F_2(x), -\tilde{F}_2(x)]^T \in \mathcal{R}^3$ dla $x \in \mathcal{X}$,
3. R_{\leq} - relacja Pareto w zbiorze ocen $\mathcal{V} \subset \mathcal{R}^3$.

Zadanie optymalizacji (7) charakteryzuje się skończonym zbiorem rozwiązań dopuszczalnych oraz nieliniową funkcją celu. Dla $V \geq 1$, $J \geq 1$ i $I \geq 1$ istnieje co najmniej jedno rozwiązanie dopuszczalne. Liczba przydziałów modułów do komputerów wynosi $I^V J^I$, rośnie wykładniczo wraz ze wzrostem liczby modułów, wykładniczo wraz ze wzrostem liczby węzłów oraz wielomianowo wraz ze wzrostem liczby typów komputerów.

Dla pewnego przydziału modułów do komputerów $x \in \mathcal{X}$ można wykazać, że istnieje dodatkowo $I!-1$ permutacji przydziałów modułów do komputerów o jednakowych wartościach czasu obciążenia najbardziej zajętego komputera $Z_{\max}(x)$.

Zbiór możliwych wariantów w problemie (7) dla reprezentacji binarnej przydziałów modułów do komputerów zawiera $2^{I(V+J)}$ alternatyw. Przyjmując reprezentację całkowitoliczbową przydziału modułów do komputerów, uzyskuje się rozwiązanie dopuszczalne, a zbiór rozwiązań dopuszczalnych zawiera $I^V J^I$ przydziałów.

W tej sytuacji można zastosować jedna z nowoczesnych metod heurystycznych takich, jak metody bazujące na sieciach neuronowych, tabu search, symulowane wyżarzanie, algorytm genetyczny [8] czy strategia ewolucyjna [7]. Wyniki o wysokiej jakości uzyskano za pomocą zmodyfikowanej wersji wielokryterialnego algorytmu ewolucyjnego [2].

4. Wielokryterialne algorytmy genetyczne

Fourman [5] i Schaffer [10] opracowali niezależnie, prawdopodobnie jako pierwsi, algorytm genetyczny GA do rozwiązywania problemów polioptymalizacji. Metoda

VEGA (ang. vector evaluated genetic algorithm) [10] wykorzystuje *procedurę niezależnej selekcji* osobników z różnych podzbiorów, którymi przypisano poszczególne kryteria cząstkowe. Wadą metody Schaffera jest *dyskryminowanie pośrednich rozwiązań* optymalnych w sensie Pareto, które są dobre ze względu na wszystkie kryteria, ale nie są najlepsze ze względu na którekolwiek kryterium skalarnie. Dlatego, metodę Schaffera można wykorzystać do wyznaczenia rozwiązań leksykograficznych lub rozwiązań leksykograficznych z parametrem [1].

Natomiast Fourman rozważał selekcję wykorzystującą turniej leksykograficzny polegający na porównaniu par rozwiązań z danej populacji i wybraniu rozwiązania dominującego w sensie relacji leksykograficznej do zbioru potencjalnych rodziców [5]. Prawdopodobieństwo selekcji wyznaczone jest dla najważniejszego celu. Losowanie danej pary odbywa się dwukrotnie zgodnie z regułą ruletki. Selekcja leksykograficzna dla priorytetów celów ustalonych *a priori* prowadzi do migracji rozwiązań w kierunku rozwiązań hierarchicznych.

Goldberg opisał *procedurę nadawania rang* rozwiązaniom optymalnym w sensie Pareto [6], która eliminuje dyskryminację pośrednich rozwiązań optymalnych w sensie Pareto. Procedura nadawania rang opiera się na założeniu, że w żadnym pokoleniu nie powinna istnieć tendencja skierowana przeciwko rozwiązaniom optymalnym w sensie Pareto. Zatem wszystkie osobniki optymalne w sensie Pareto powinny posiadać jednakowy potencjał reprodukcyjny. Jednym ze sposobów osiągnięcia jednakowego potencjału reprodukcyjnego dla wszystkich rozwiązań optymalnych w sensie Pareto jest specjalna procedura sortująca nadająca rangi wszystkim rozwiązaniom [12].

Procedura nadawania rang polega na iteracyjnym wyborze rozwiązań optymalnych w sensie Pareto z określonego podzbioru populacji oraz nadaniu im jednakowej rangi. W pierwszym kroku rozpatruje się całą populację, wyznacza się zbiór Pareto, a rozwiązania ze zbioru Pareto otrzymują rangę 1. W drugim kroku rozważa się pozostałą część populacji, z której wyznacza się rozwiązania optymalne w sensie Pareto i nadaje się im rangę 2. Proces ten jest kontynuowany aż do wyczerpania wszystkich rozwiązań w danej populacji.

Niech dla danej populacji zostanie wygenerowanych R podzbiorów rozwiązań o różnych rangach. Wartość funkcji przystosowania f oblicza się za pomocą poniższej formuły:

$$f(x) = -r(x) + R + 1, \quad (8)$$

gdzie $r(x)$ jest rangą nadaną rozwiązaniu x .

5. Wielokryterialny algorytm ewolucyjny ze zmianą wartości parametrów

Do rozwiązania zagadnienia optymalizacji wektorowej (7) proponuje się zastosować wielokryterialny algorytm ewolucyjny WAE+. Algorytm ten w nieco zmodyfikowanej wersji użyto także do wyznaczania reprezentacji rozwiązań efektywnych w zagadnieniu wyznaczania przydziałów modułów przy minimalizacji łącznego kosztu wykonania programu i maksymalizacji wydajności systemu [2]. Algorytm WAE+ wykorzystuje wiedzę specyficzną dla zagadnienia optymalizacji wektorowej, co zasadniczo różni go od wielokryterialnego algorytmu genetycznego z procedurą nadawania rang.

W algorytmie WAE+ stosuje się bardziej naturalną reprezentację rozwiązań, odchodząc od binarnego kodowania chromosomu, dopuszcza się zmianę prawdopodobieństwa krzyżowania i tempa mutacji, stosuje się zmodyfikowane operacje krzyżowania i mutacji tak, aby nie naruszyć spełnionych ograniczeń. Celem modyfikacji algorytmu jest przyspieszenie jego zbieżności oraz poprawienie jakości uzyskanych rozwiązań.

Przyjmuje się, że chromosom jest reprezentowany w postaci pary wektorów (X^m, X^n) o całkowitoliczbowych współrzędnych. Jeżeli zmienne decyzyjne są liczbami naturalnymi oraz $0 < X^m(v) \leq I, v = \overline{1, V}, 0 < X^n(i) \leq J, i = \overline{1, I}$, to każdy wektor (X^m, X^n) jest rozwiązaniem dopuszczalnym zagadnienia (7). Generowanie rozwiązań początkowych polega na przydziale każdego zadania do losowanego węzła oraz na alokacji komputera wylosowanego typu w każdym węźle. W wyniku przyjętej reprezentacji chromosomów oraz za pomocą zaproponowanego sposobu generowania populacji początkowej uzyskuje się rozwiązania dopuszczalne.

Krzyżowanie chromosomów polega na wylosowaniu liczby naturalnej z przedziału $[1; V+I]$, a następnie na przeprowadzeniu tej operacji z prawdopodobieństwem p_c , jak w krzyżowaniu standardowym par chromosomów algorytmu genetycznego. Mutacja polega na zmianie wartości każdej zmiennej decyzyjnej z prawdopodobieństwem p_m w ten sposób, że jeżeli wylosuje się współrzędną $X^m(v)$, to przypisuje jej się losową liczbę naturalną z przedziału $[1; I]$. Jeżeli jednak wylosuje się współrzędną $X^n(i)$, to przypisuje jej się losową liczbę naturalną z przedziału $[1; J]$.

Istotną proponowaną w pracy innowacją dla wielokryterialnych algorytmów ewolucyjnych jest adaptacyjna zmiana prawdopodobieństwa krzyżowania oraz tempa mutacji w zależności od czasu przeszukiwań. Sheble i Britting zaobserwowali poprawienie jakości wyznaczonych rozwiązań w zagadnieniach optymalizacji jednokryterialnej, jeśli zaczęli stosować zmniejszanie prawdopodobieństwa krzyżowania i zwiększanie tempa mutacji [11].

Eksperymenty numeryczne z dobieraniem stałych wartości prawdopodobieństwa krzyżowania wykazały, że algorytm WAE+ wyznaczył najlepsze rozwiązania dla umiarkowanych wartości na poziomie 0,2. Niech T_{\max} oznacza liczbę populacji wygenerowanych podczas działania algorytmu, a t – numer populacji. Wartość prawdopodobieństwa krzyżowania powinna ulegać zmniejszaniu w miarę postępu w eksploracji przestrzeni przeszukiwań. W początkowej fazie eksploracji istotne jest znalezienie obszaru zawierającego rozwiązanie o wysokiej jakości. Dzięki intensywnie realizowanym krzyżowaniom taki obszar może zostać zidentyfikowany. Z kolei mniejsza liczba krzyżowań osobników powoduje skoncentrowanie przeszukiwań w obszarze wokół być może optimum globalnego. Wartość prawdopodobieństwa krzyżowania jest wyznaczana w następujący sposób:

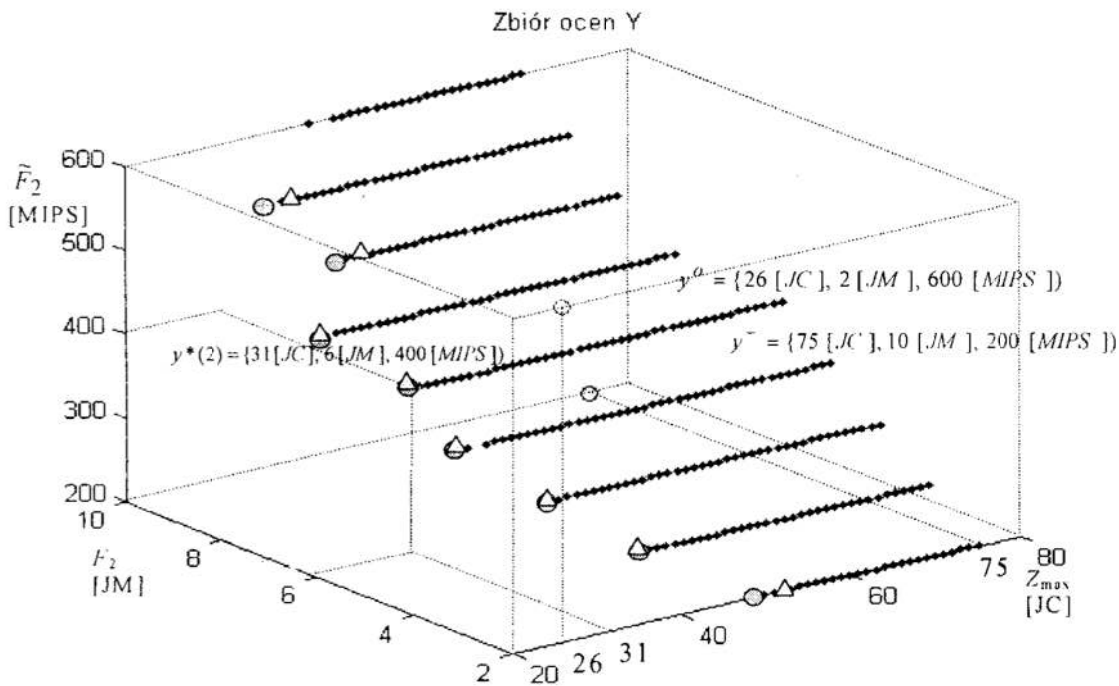
$$p_c = e^{-t/T_{\max}}. \quad (9)$$

Stałe tempo mutacji będące na poziomie $1/M$, gdzie M reprezentuje liczbę zmiennych decyzyjnych, zapewniło uzyskiwanie wyników stosunkowo wysokiej jakości dla algorytmu WAE+. Mutacja w początkowym okresie przeszukiwań nie odgrywa istotnej roli, ale kiedy niektóre osobniki populacji znajdują się w obszarze z lokalnym optimum jest *de facto* jedynym mechanizmem mogącym zapewnić wyznaczenie tego optimum. Jeżeli tempo mutacji w początkowej fazie przeszukiwań jest bliskie zeru, to w końcowym etapie przeszukiwań powinno zostać wzmocnione. Proponuje się zastosowanie następującej zależności na wartość tempa mutacji [11]:

$$p_m = e^{0,05t/T_{\max}} - 1. \quad (10)$$

6. Wyniki eksperymentów numerycznych

Algorytm WAE+ zastosowano do rozwiązania 30 instancji zagadnienia (7), w których liczba zadań była zmieniana od 2 do 60. Wyznaczone reprezentacje rozwiązań efektywnych cechowały się wysokim stopniem zbieżności do zbioru wyników optymalnych w sensie Pareto [2]. Średni błąd wyniósł 2,8 %, błąd maksymalny – 9,7 %, a w 23,5% przypadków wyznaczono cały zbiór wyników optymalnych w sensie Pareto. Na rys. 1 ukazano wyniki uzyskane za pomocą algorytmu WAE+ (trójkąty) w odniesieniu do zbioru Pareto (małe koła) w przestrzeni wielokryterialnej dla instancji o 30 zadaniach [2]. Poprzednia wersja algorytmu bez zmiany wartości prawdopodobieństwa krzyżowania oraz tempa mutacji cechowała się wyznaczaniem wyników o gorszej jakości niż Algorytm WAE+. Średni błąd wyniósł 5,7 %, co także należy uznać za dobrą ocenę algorytmu ewolucyjnego.



Rys. 1. Zbiór ocen instancji zagadnienia (7) o 30 zadaniach

Fig. 1. An evaluation set of the problem instance (7) with 30 tasks

7. Wnioski i uwagi

Wyniki uzyskane za pomocą algorytmu ewolucyjnego WAE+ dla badanych testowych problemów optymalizacji wielokryterialnej potwierdzają fakt przyśpieszenia zbieżności algorytmu w stosunku do wielokryterialnego algorytmu genetycznego. Odpowiednie kodowanie chromosomów w sposób istotny redukuje przestrzeń przeszukiwań, co wywiera istotny wpływ na jakość uzyskiwanych rozwiązań. Zmniejszanie prawdopodobieństwa krzyżowania oraz zwiększanie tempa mutacji podczas przeszukiwań poprawia jakość wyznaczonych reprezentacji niezdominowanych. Wzrost tempa mutacji jest niewielki, a jego wartość końcowa nie przekracza 0,05 dla 200 generacji. Interesującym kierunkiem dalszych badań jest weryfikacja adaptacyjnych własności algorytmów ewolucyjnych w optymalizacji wielokryterialnej.

Literatura

1. Ameljańczyk A.: *Optymalizacja wielokryterialna*. WAT, Warszawa 1986.
2. Balicki J., Kitowski Z.: *Wielokryterialny algorytm ewolucyjny do wyznaczania Pareto-optymalnych przydziałów modułów programów o sekwencyjno – równoległej strukturze interakcji*, Materiały III Krajowej Konferencji „Algorytmy Ewolucyjne i Optymalizacja Globalna”, Politechnika Warszawska, Potok Złoty, 1999, ss. 1-8.
3. Chu W. W., Lan L. M. T.: *Task allocation and precedence relations for distributed real-time systems*. IEEE Transactions on Computers, Vol. C-36, No. 6, June 1987, pp. 667-679.
4. Coulouris G., Dollimore J., Kindberg T.: *Systemy rozproszone. Podstawy i projektowanie*. WNT, Warszawa 1998.
5. Fourman M.P.: *Compaction of symbolic layout using genetic algorithms*. Proceedings of the First International Conference on Genetic Algorithms, New York 1985.
6. Goldberg D.E.: *Algorytmy genetyczne i ich zastosowania*, WNT, Warszawa 1995.
7. Michalewicz Z.: *Algorytmy genetyczne + struktury danych = programy ewolucyjne*. WNT, Warszawa 1996.
8. Osyczka A., Kuchta W., Czula R.: *Computer aided multicriterion optimization system for computationally expensive functions*. Structural Optimization, No. 8, 1994, pp. 37-41.
9. Perry P.J.: *Java – tworzenie apletów*. README, Warszawa 1996.
10. Schaffer J.D.; *Multiple objective optimization with vector evaluated genetic algorithm*, In J.J.Grefenstete (Ed.) Genetic Algorithms and Their Applications: Proc. of the First Int. Conf. on Genetic Algorithms, Hillsdale, NJ: Lawrence Erlbaum, 1985, pp.93-100.
11. Sheble G. B., Britting K.: *Refined genetic algorithm – economic dispatch example*. IEEE Transactions on Power Systems, Vol. 10, February 1995, pp. 117-124.
12. Srinivas N., Deb K.: *Multiobjective optimization using nondominated sorting in genetic algorithms*. Evolutionary Computation, Vol. 2, No. 3, 1994, pp. 221-248.
13. <http://www.specbench.org/>, Strona macierzysta The Standard Performance Evaluation Corporation, 1999.
14. <http://www.qnx.com/>, Strona macierzysta QNX Software Systems Ltd. 1999.

Streszczenie

W pracy rozważa się zagadnienie wyznaczania reprezentacji przydziałów zadań do komputerów w systemie czasu rzeczywistego. Minimalizuje się obciążenie newralgicznego komputera, koszt komputerów realizujących zadania oraz maksymalizuje się wydajność systemu. Do rozwiązania sformułowanego zagadnienia polioptymalizacji zaproponowano adaptacyjny algorytm ewolucyjny.

ADAPTIVE EVOLUTIONARY ALGORITHM FOR TASK ASSIGNMENT OPTIMIZATION IN REAL TIME SYSTEMS

Summary

In this paper, an optimization problem of the task assignment in the real time system has been considered. A workload of a bottleneck computer and an overhead cost of computers have been minimized. Moreover, a system performance has been maximized. An adaptive evolutionary algorithm for solving the established multicriteria optimization problem has been proposed.

Referat recenzowany

OPTYMALNY WYBÓR I KIERUNKI ROZWOJU ZINTEGROWANEGO SYSTEMU UZBROJENIA

1 Zintegrowany system uzbrojenia – próba definicji

Ewolucja techniki dokonuje się nie tylko poprzez doskonalenie dotychczasowych oraz powstawanie nowych narzędzi obróbki, środków transportu, mechanizmów, urządzeń pomiarowych, czujników, rejestratorów itp. Bardzo wyraźnym krokiem na drodze tej ewolucji stało się powstanie systemów technicznych. Określenie „system” stosuje się w połączeniu z nazwą handlową lub nazwą firmy. Takie nazwy spotyka się w telekomunikacji, w technice komputerowej, w technice pomiarów i rejestracji, wreszcie w technice wojskowej i robotyce. Jakie cechy tej generacji techniki stanowią o jej systemowości o tym, że można systemy techniczne traktować jako pewną całość względnie odosobnioną?

System można definiować dwojako; podając cel działania i spełniane funkcje lub też podając zbiór elementów i relacji między nimi. W pierwszej definicji obiektami systemu są procesy (wymuszenia, reakcje, decyzje i oceny). Drugi sposób polega na potraktowaniu systemu jako struktury, której obiekty są materialne.

Opis i przykłady

Czy można podać w sposób uogólniony cel i funkcje robota przemysłowego czy też systemu uzbrojenia?

Zintegrowany system uzbrojenia (ZSU) to obiekt złożony z urządzeń technicznych lub układów antropotechnicznych różniących się funkcją, działaniem, konstrukcją i technologią wytwarzania. Zintegrowany system uzbrojenia działa w swoistym środowisku. Tam poszukuje, wykrywa, rozpoznaje i obrabia (np. niszczy) właściwe sobie obiekty.

Systemy zintegrowane łączą w sobie układy (urządzenia techniczne) realizujące kolejne fazy cyklu złożonego działania w analogii do zachowań organizmów zwierzęcych.

W najprostszych przypadkach cykl ten zawiera następujące procesy:

- poszukiwanie obiektu oddziaływania (UW),
- śledzenie obiektu (US),
- sterowany ruch organu roboczego w kierunku obiektu.

Przyjęte oznaczenia:

Ω – położenie i cechy obiektu (Ω_w – po wykryciu),

Ω_p – położenie i cechy organu roboczego,

δ – uchyb śledzenia,

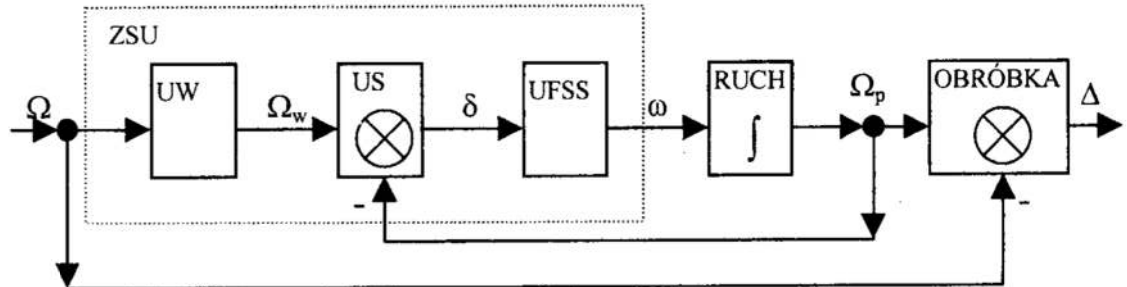
ω – sterowanie,

Δ – dokładność obróbki,

UFSS – układ formowania sygnału sterującego.

ZSU – zintegrowany system uzbrojenia.

W przypadku ZSU sterowanie jest rozbieżne, przebiega raz wewnątrz systemu ale i poza systemem. System zintegrowany może wtedy nabrać charakteru adaptacyjnego. W systemie pojawia się blok formowania sygnału adaptacji zmieniający charakterystyki filtracji i predykcji lub też współczynniki prawa sterowania albo parametry procesu obróbki.



Rys. 1. Schemat blokowy zintegrowanego systemu uzbrojenia (ZSU)

Fig. 1. Block diagram of integrated weapon's system (IWS)

Przy dominującym wpływie sterowania rozbieżnego system zintegrowany ma charakter rozgrywający. Działanie systemu jest wtedy nakierowane na minimaksowe kryterium jakości. W systemie pojawia się blok wyboru zmieniający model predykcji, prawo sterowania czy rodzaj obróbki [5]. Zintegrowane systemy uzbrojenia (ZSU) to systemy o sterowaniu rozbieżnym, adaptacyjne i rozgrywające.

2 Charakterystyka procesów składowych

Poszukiwanie i rozpoznanie

Poszukiwanie i rozpoznanie to dwa etapy procesu postrzegania. Poszukiwanie polega na możliwie najwcześniejszym stwierdzeniu obecności zbliżającego się obiektu posiadającego cechy zbliżone do pożądaných. Proces poszukiwania może posiadać charakter aktywny lub pasywny. W obu przypadkach polega on na rejestracji zaburzeń tła energetycznego. Charakter energetyczny procesu sprawia, że może być obrazowany przy pomocy pojęcia strefy wykrycia tj. takiego obszaru wokół ZSU, w którym zaburzenie spowodowane obecnością pożądanego obiektu jest możliwe do rejestracji przez elementy czujnikowe (sensory) ZSU.

Drugim etapem postrzegania jest rozpoznanie. Proces ten będzie charakteryzowany przez trzy aspekty: przestrzenny, czasowy, probabilistyczny. Proces rozpoznania ma także charakter energetyczny co wiąże się z pojęciem strefy rozpoznania. Aspekt czasowy jest związany z czasem przetwarzania informacji. Proces ten można scharakteryzować jako program podejmowania decyzji statystycznych na podstawie rejestracji i analizy anizotropowości zaburzenia tła energetycznego – stąd aspekt probabilistyczny. Układy wykrywania i rozpoznawania są często układami antropotechnicznymi i posiadają zdolności adaptacyjne, są rozgrywające lub uczące się. Obecnie szybko postępuje automatyzacja tych układów.

Śledzenie

Śledzenie jest w swej istocie procesem ciągłego pomiaru uchybu pomiędzy położeniem i cechami obiektu oddziaływania a położeniem organu roboczego i cechami pożądanymi obiektu po obróbce. Urządzenia śledzące (pomiarowe) do pomiaru ciągłego są złożonymi układami regulacji automatycznej o ograniczonych możliwościach

charakteryzujących się zwykle możliwością reakcji na uchyb w określonych granicach. Proces pomiaru ciągłego posiada także skończoną stałą czasową i czas regulacji. Może posiadać charakter losowy. Ścisłe związane z procesem śledzenia są procesy filtracji i predykcji. Są to procesy wyznaczania estymatorów parametrów losowego procesu pomiaru (śledzenia). Filtracja polega na wyznaczeniu estymatorów rozkładu wielkości mierzonej w momencie pomiaru. Predykcja jest zagadnieniem wyznaczania estymatorów rozkładu wielkości mierzonej dla przyszłej chwili w stosunku do momentu pomiaru czasu. W prostszych przypadkach występuje proces wygładzania, który da się określić jako wyznaczanie estymatorów procesu mierzonego dla interwału czasowego wcześniejszego niż chwila pomiaru. Integralną część filtru stanowi zawsze model dynamiczny obserwowanego (mierzonego) procesu. Filtry, w których istnieje możliwość zmiany parametrów tego modelu nazywamy filtrami podatnymi. Występują one w adaptacyjnych systemach sterowania.

Formowanie sygnału sterującego

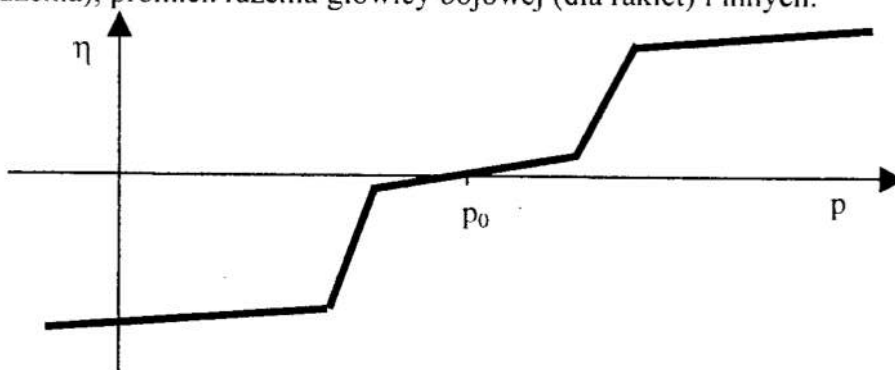
Formowanie sygnału sterującego dokonuje się poprzez techniczną realizację prawa sterowania tj. zależności przyporządkowującej sterowanie organem roboczym uchybowi pomiędzy położeniem i cechami obiektu obróbki a położeniem i stanem organu roboczego zintegrowanego systemu uzbrojenia. Prawo sterowania ustala trajektorię organu roboczego a jego wybór zależy od dynamiki układu przemieszczającego. Można je traktować jako model dynamiki organu roboczego.

3 Efektywność zintegrowanego systemu uzbrojenia

Zintegrowany system uzbrojenia (ZSU) posiada co najmniej charakter adaptacyjny a najczęściej jest to system rozgrywający, uczący się i rozwijający. Na podstawie eksperymentów symulacyjnych. systemów uzbrojenia. [1] można ocenić ich efektywność w pojedynku ogniowym z systemem antagonistycznym o analogicznym celu działania. Wskaźnik efektu pojedynku w postaci:

$$\eta = n_s - n_w, \quad (1)$$

gdzie n_s – względna liczbę sukcesów, n_w – względna liczbę porażek, jest zależny od takich charakterystyk jak: zasięg wykrycia, czas reakcji, dokładność naprowadzania (śledzenia), promień rażenia głowicy bojowej (dla rakiet) i innych.



Rys. 2. Wskaźnik efektu

Fig. 2. The effectiveness indicator

Prezentowany na rysunku 2 charakter zależności czynnikowej jest uniwersalny i charakterystyczny. Na efekt działania (pojedynku) zintegrowanego systemu uzbrojenia wpływają równocześnie i współistotnie czynniki wynikające z ograniczeń procesów:

wykrywania, śledzenia, rażenia. Cechy charakterystyczne zależności czynnikowej (rys.2) pozwalają sformułować następujące wnioski.

- 1 W przestrzeni A parametrów istnieją hiperpowierzchnie dzielące przestrzeń parametrów A na dwie półprzestrzenie: A(+) półprzestrzeń sukcesów, A(-) półprzestrzeń porażki.
- 2 Podział ten jest nieostry (rozmyty). Pomędzy wymienionymi hiperpowierzchniami istnieje obszar równoważności z systemem antagonistycznym.
- 3 Wrażliwość parametryczna efektu pojedynku ma charakter nieliniowy. Niewrażliwe parametrycznie (lub mało wrażliwe) mogą być konstrukcje nadmiernie efektywne, równoważne antagonistycznemu lub krańcowo nieefektywne.

4 Kierunki rozwoju konstrukcji zintegrowanego systemu uzbrojenia

Zintegrowany system uzbrojenia sprawdza się w pojedynku ogniowym z właściwym mu systemem antagonistycznym. Równowaga osiągnięta w takim pojedynku (w sensie statystycznym) nie jest stanem jedynym. Badany system uzbrojenia może w wyniku procesów rozwojowych (jego lub systemu antagonistycznego) dość gwałtownie przejść do dwóch innych stanów: krańcowej nieefektywności lub efektywności nadmiernej. Zdarzenie takie może być spowodowane zmianami w obrębie każdego z procesów elementarnych zachodzących w ZSU. Efekt gwałtownej zmiany sytuacji efektywnościowej badanego Zintegrowanego Systemu Uzbrojenia może zostać wywołany zmianą odległości wykrycia, dokładności rozpoznania, precyzji śledzenia, jakości prawa sterowania i wieloma innymi czynnikami charakteryzującymi poszczególne procesy składowe działania ZSU [2].

Modelowanie systemu zintegrowanego

Istnieje możliwość skutecznego modelowania wybranych procesów składających się na funkcjonowanie ZSU. Można przyjąć, że w sposób spójny można modelować procesy składowe: poszukiwania i wykrywania, śledzenia i naprowadzania, proces roboczy. W ogólności każdy z tych procesów można opisać podając (w uproszeniu):

- równanie stanu, $\frac{dx}{dt} = F(x, u, t, \alpha)$ (2)

- równanie wyjścia, $y = Y(x, u, t, \alpha)$ (3)

- ograniczenia, $R(x, t) \leq p$ (4)

gdzie: \mathbf{x} – wektor stanu, \mathbf{u} – wektor wymuszeń, α – wektor parametrów,
 \mathbf{p} – wektor ograniczeń.

Wyodrębniono w tym opisie ograniczenia. Traktując formalnie są to także parametry modelu lecz ich usytuowanie nie pozwala na efektywne stosowanie klasycznych metod wrażliwości.

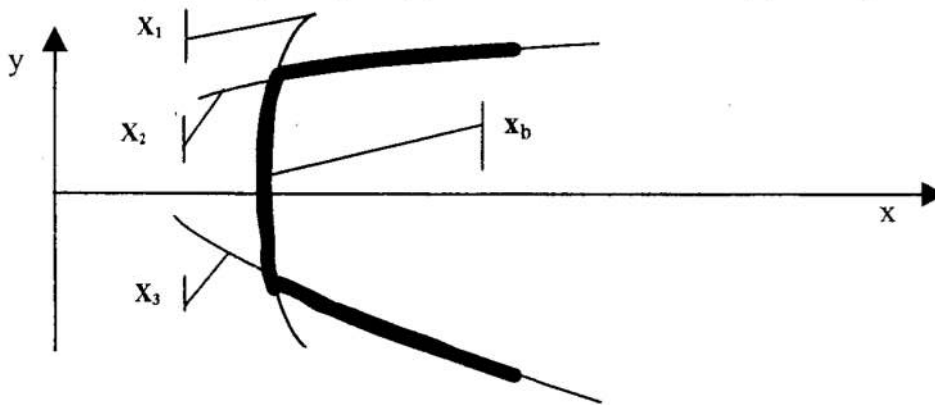
Modelując każdy z procesów składowych (poszukiwanie, śledzenie, ruch organu roboczego) można otrzymać modele analogiczne formalnie lecz całkowicie niespójne treściowo. Modele wymienionych procesów są tworzone w odmiennych przestrzeniach stanów. Przenoszenie wyników badań z jednego z nich do drugiego jest techniką skomplikowana i nieefektywną. Odmiennie są także przestrzenie parametrów i ograniczeń. Jedyną podprzestrzenią niezmienną dla wszystkich trzech wymienionych modeli jest przestrzeń wymuszeń. Dla ZSU jest to przestrzeń stanów obiektu obróbki (niszczonego celu). Przestrzeń wymuszeń jest jedyną przestrzenią umożliwiającą porównanie

parametrów ograniczających każdego z procesów składowych (poszukiwanie, śledzenie, ruch organu roboczego). Aby dokonać takiego porównania trzeba przetransponować obszar ograniczeń z przestrzeni parametrów do przestrzeni wymuszeń i to dla modelu każdego z wyróżnionych procesów składowych oddzielnie. Należy umieć rozwikłać dwie relacje modelu bazowego. Chodzi tu o relacje ograniczeń i relację więzów (składniki równań stanu). Można to sformułować następująco:

1. Określić zbiór momentów czasu $\{t_p\}$ takich, że $R(\mathbf{x}, t_p) = \mathbf{p}$.
Inaczej mówiąc określić dziedzinę modelu. (5)

2. Określić zbiór wymuszeń $\{\mathbf{u}_p\}$ takich że $\frac{d\mathbf{x}}{dt} = F(\mathbf{x}, \mathbf{u}_p, t_p)$. (6)

Tak otrzymany zbiór wymuszeń $\{\mathbf{u}_p\}$ jest obrazem ograniczeń jednego procesu składowego w przestrzeni wymuszeń. Oczywiście w realnych przypadkach podanie funkcji rozwikłującej w sposób jawny jest niemożliwe. Pozostają metody numeryczne.



Rys. 3. Przekrój strefy roboczej
Fig. 3. Section of the working space

Jeśli założyć istnienie takiej operacji lub techniki rozwikłującej równania stanu i relację ograniczeń to każdej składowej p_i wektora ograniczeń \mathbf{p} można przypisać równanie

$$F(\mathbf{u}_i(p_i)) = 0 \quad (7)$$

powierzchni ograniczającej w przestrzeni wymuszeń. Ściśle mówiąc w wybranej podprzestrzeni wymuszeń. Wybieranej pod kątem minimalnej wymiarowości wystarczającej do analizy problemu. Zwykle można się ograniczyć do przestrzeni położeń obiektu oddziaływania i w niej wyznaczać powierzchnie odpowiadające określonej realizacji pełnego wektora wymuszenia. Powierzchnie ograniczające

$$\mathbf{x}_i(p_i) = W_i[\mathbf{u}_i(p_i)] \quad (8)$$

tworzą płaty dane równaniami:

$$\mathbf{x}_b = \max\{\mathbf{x}_i(p_i)\} \text{ dla płata bliższego, } \mathbf{x}_d = \min\{\mathbf{x}_k(p_k)\} \text{ dla płata dalszego,} \quad (9)$$

gdzie kierunek osi x jest kierunkiem ruchu obiektu obróbki. Konstrukcja płata bliższego jest przedstawiona na rysunku 3.

Cel konstrukcji i jej jakość. „Metoda maksymalnego dopasowania ważonej objętości strefy roboczej”

Jakość konstrukcji traktuje się zwykle jako miarę odległości (metrykę) pomiędzy celem konstrukcji a jej realizacją. Cel konstrukcji systemu zintegrowanego można formułować

jako żądanie skutecznej (z określonym prawdopodobieństwem) obróbki przez ZSU w zamkniętym obszarze przestrzeni (w strefie roboczej) obiektów o charakterystykach przedziałami określonych. Dla zintegrowanego systemu uzbrojenia są to np.:

- współrzędne położenia obiektu $x(x,y,h)$,
- czynniki ruchu $V(V_v, V_y, V_h)$
- wymiary geometryczne,
- podatność na obróbkę i inne.

Można je przedstawić w formie dystrybuant $F_i(u_i)$.

Cel konstrukcji a właściwie jego miarę można przedstawić jako ważoną objętość strefy roboczej V_0 . (użyty symbol całkowania w sensie Lebesgue'a)

$$V_0 = \int_{\Xi} \mathbf{x}_0(\mathbf{u}) \cdot d\mathbf{F}(\mathbf{u}) \quad (10)$$

gdzie: \mathbf{x}_0 – pożądany zasięg roboczy, $\mathbf{F}(\mathbf{u})$ – wektor dystrybuant $F(u_i)$,
 Ξ – obszar zmienności wektora \mathbf{u} .

Do oceny stopnia realizacji celu konstrukcji wykorzystuje się miarę uchylenia np. w postaci:

$$J = |V_0 - V| \quad (11)$$

gdzie: J – wskaźnik jakości konstrukcji,
 V_0 – pożądana ważona objętość strefy roboczej,
 V – rzeczywista ważona objętość strefy roboczej.

Wykorzystując pojęcia płatów bliższej i dalszej granicy strefy roboczej można określić ważoną rzeczywistą objętość strefy roboczej jako:

$$V = \int_{\Xi} |\mathbf{x}_d - \mathbf{x}_n| \cdot d\mathbf{F}(\mathbf{u}) \quad (12)$$

zaś wskaźnik osiągnięcia celu konstrukcji :

$$J = \int_{\Xi} |\mathbf{x}_0 - \mathbf{x}_d + \mathbf{x}_n| \cdot d\mathbf{F}(\mathbf{u}) \text{ lub } J = \int_{\Xi} |\mathbf{x}_0 - \mathbf{x}_d + \mathbf{x}_n| \cdot f(\mathbf{u}) d\mathbf{u} \quad (13)$$

gdzie: $f(\mathbf{u})$ – gęstość rozkładu wektora \mathbf{u} .

Z analizy przedstawionego wskaźnika a zwłaszcza z faktu, że zawiera on składniki typu

$$\mathbf{x}(\mathbf{p}) = \max \left\{ \left| \mathbf{x}_i(p_i) \right| \right\} \quad (14)$$

wynika, że wielkość wskaźnika J zależy zawsze w sytuacjach granicznych od jednego parametru konstrukcyjnego determinującego wielkość strefy roboczej, podczas gdy zmiana pozostałych nie pociąga za sobą skutków w odniesieniu do wskaźnika J , lub wpływa nań słabo. Próba ekstremalizacji takiego wskaźnika da rozwiązanie bliskie celowi konstrukcji lecz bez względu na koszty ich osiągnięcia. W rzeczywistości polepszenie dowolnego parametru konstrukcyjnego pociąga za sobą koszty badawcze, projektowe czy technologiczne, które można uznać za proporcjonalne do gradientu tego parametru. Uwzględnić to można poprzez rozszerzenie postaci wskaźnika jakości konstrukcji o funkcję kary do postaci:

$$J' = \int_{\Xi} |\mathbf{x}_0 + \mathbf{x}_b - \mathbf{x}_d| f(\mathbf{u}) \cdot d\mathbf{u} + \sum_{i \in I} \lambda_i p_i \quad (15)$$

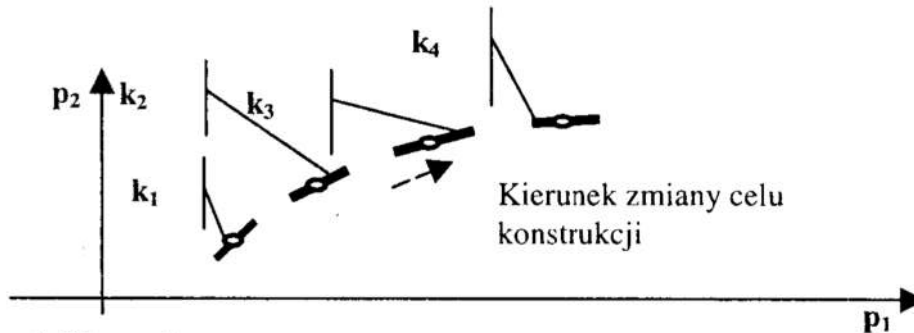
gdzie: p_i – i -ty parametr konstrukcyjny ograniczający strefę roboczą,
 λ_i – jednostkowy koszt polepszenia i -tego parametru,
 I – zbiór indeksów parametrów konstrukcyjnych ograniczających strefę roboczą.

Kierunek rozwoju konstrukcji systemu zintegrowanego

Problem wyboru konstrukcji systemu zintegrowanego można podzielić na dwa podproblemy:

1. wybór optimum lokalnego adekwatnego do celu konstrukcji,
2. wybór kierunku rozwoju konstrukcji wokół optimum lokalnego.

Ad1. **Problem pierwszy** sprowadza się do wyznaczenia wektora ograniczeń konstrukcyjnych \mathbf{p}_1 odpowiednio do celu konstrukcji danej w postaci V_{01} pożądanej objętości strefy roboczej oraz dystrybuant występowania różnych obiektów oddziaływania. Istnieją efektywne algorytmy optymalizacyjne rozwiązania tego problemu dla trójwymiarowego wektora ograniczeń. Dla większej liczby wymiarów szybko rośnie czas obliczeń i konieczna pojemność pamięci operacyjnej. [3]



Rys. 4. Kierunek rozwoju zintegrowanego systemu uzbrojenia

Fig. 4. The developmental direction's of integrated weapon's system

Ad2. **Drugi problem** sprowadza się do badania gradientu wokół każdego optimum lokalnego \mathbf{p}_1 . Jeżeli oznaczyć kierunek zmian od \mathbf{p}_1 w przestrzeni parametrów ograniczających strefę roboczą jako \mathbf{k} , to każdy wektor zmian parametrów $\Delta\mathbf{p}$ leżący na tym kierunku można określić jako:

$$\Delta\mathbf{p} = \beta \cdot \mathbf{k} \quad (16)$$

gdzie: $\Delta\mathbf{p}$ – wektor zmian parametrów ograniczających strefę roboczą,
 β – skala wektora,

Poszukiwanym kierunkiem rozwoju \mathbf{k}_r konstrukcji ZSU nazwiemy taki wybrany kierunek wokół optimum lokalnego \mathbf{p}_1 , który spełnia warunek:

$$\mathbf{k}_r : \exists_k \left[S_{kr} = \max \frac{\partial J^r(\mathbf{k}_r)}{\partial \beta} \leq S_k = \max \frac{\partial J^r(\mathbf{k})}{\partial \beta} \right] \quad (17)$$

gdzie: $\frac{\partial J^r(\mathbf{k})}{\partial \beta}$ - funkcja wrażliwości wskaźnika J w kierunku \mathbf{k} .

5 Wnioski

Zdefiniowano pojęcie zintegrowanego systemu uzbrojenia jako uogólnienie pewnej klasy systemów technicznych. Przedstawiona metoda optymalizacyjna „metoda maksymalnego dopasowania ważonej objętości strefy roboczej” wydaje się być atrakcyjna przynajmniej w odniesieniu do obiektów tej klasy.

1. Istnieje możliwość skutecznego badania ograniczeń konstrukcyjnych poprzez ich transponowanie do przestrzeni wymuszeń drogą operacji rozwikłującej równania stanu i ograniczeń.

2. Objętość tak określonego obszaru może być miarą osiągnięcia celu konstrukcji.
3. Uwzględnienie różnorodności wymuszeń prowadzi do sformułowania pojęcia ważonej objętości strefy roboczej.
4. Wszystko to pozwala określić postać wskaźnika jakości zależnego od celu konstrukcji i parametrów ograniczających.
5. Klasyczne gradientowe metody optymalizacji są skuteczne przy rozwiązywaniu takich zagadnień, metody optymalizacji neuronowej mogą poszerzyć zastosowania. [4]
6. Wskazuje się na efektywną procedurę wyznaczania globalnego kierunku rozwoju konstrukcji ZSU co pozwala precyzować pożądane właściwości systemów realizujących zmieniony cel – powiększone wymagania.

Literatura

1. Bartkiewicz S.: *Możliwości zwalczania celów niskolejących ... Cz.1 Sprawozdanie z badań teoretycznych*. Praca naukowo-badawcza. Wyższa Szkoła Oficerska Wojsk Obrony Przeciwlotniczej Koszalin 1984 (opracowanie wewnętrzne).
2. Bartkiewicz S., Ociepa Z., Madej W., Boroń S.: *Możliwości zwiększenia skuteczności artyleryjskiego zestawu przeciwlotniczego*. w: Materiały konferencyjne V Krajowej Konferencji „Automatyzacja i Eksploatacja Systemów Sterowania Tom1. Gdynia 11-13 października 1995 Akademia Marynarki Wojennej Gdynia 1995
3. Findensein W., Szymanowski J., Wierzbicki A.: *Teoria i metody optymalizacji*. PWN Warszawa 1980.
4. Osowski S.: *Sieci neuronowe w ujęciu algorytmicznym*. Wydawnictwa Naukowo-Techniczne Warszawa 1996.
5. Staniszewski R.: *Cybernetyczna teoria projektowania* Ossolineum Wrocław 1986

Streszczenie

Opracowanie zawiera sformułowanie potrzeby i propozycję postaci matematycznej wskaźnika jakości zintegrowanego systemu uzbrojenia (ZSU). Wskaźnik ten pozwala oceniać przydatność ZSU do obsługi obiektów określonej klasy i wybierać kierunek rozwoju.

THE OPTIMUM SELECTING AND THE DEVELOPMENTAL DIRECTION'S OF INTEGRATED WEAPON'S SYSTEM

Summary

The paper presents the need of indicator of effectiveness (J) of integrated weapon's system (IWS) and its mathematical form. This indicator makes it possible to choose the properties of construction IWS to serve objects of specified class.

Referat recenzowany

KOMPUTEROWE MONITOROWANIE ZNOSZENIA JAJ W HODOWLI DOŚWIADCZALNEJ

1. Wstęp

W hodowli doświadczalnej drobiu występuje czasem potrzeba dokładnej rejestracji daty i czasu zegarowego, w którym kury znoszą jaja. Dane te umożliwiają określanie indywidualnego cyklu nośności dla każdej kury, długości serii, itp. Poprzez odpowiednie selekcjonowanie i dostosowanie np. programu świetlnego, można następnie uzyskać istotny wzrost wydajności (nośności) w całej hodowli.

W referacie przedstawiono sposób rozwiązania tego problemu środkami technicznymi, na przykładach instalacji uruchomionych w trzech różnych placówkach badawczych.

Jest rzeczą oczywistą, że automatyczna rejestracja zdarzenia polegającego na zniesieniu jaja przez kurę jest niewykonalna w hodowli prowadzonej na otwartej przestrzeni lub w klatkach zbiorczych. Prezentowane tu rozwiązania nadają się więc tylko do hodowli w klatkach indywidualnych, zorganizowanych w tzw. baterie.

System rejestracji znoszonych jaj powinien umożliwić utworzenie wynikowego pliku tekstowego zawierającego podstawowe dane o każdym zdarzeniu, czyli numer klatki, datę, godzinę i minutę zniesienia jaja.

W referacie omówiono problemy związane z budową systemów automatycznej rejestracji znoszonych jaj, posługując się przykładami trzech aplikacji działających w konkretnych placówkach doświadczalnych:

- hodowla produkcyjna kur niosek z wydzieloną częścią doświadczalną obejmującą baterię 720 klatek (2 piętra × 2 rzędy × 180 klatek)
- hodowla doświadczalna przepiórek niosek, obejmująca baterię 80 klatek (5 pięter × 2 rzędy × 8 klatek)
- hodowla doświadczalna kur niosek, obejmująca 144 klatki w trzech bateriach po 48 klatek (2 piętra × 2 rzędy × 12 klatek).

Wyszczególnione powyżej 3 różne konfiguracje baterii hodowlanych mają podstawowe znaczenie z punktu widzenia projektowania sieci przewodów łączących instalację.

2. Ogólna koncepcja systemu

System będący przedmiotem referatu, niezależnie od konkretnej aplikacji, składa się z trzech podstawowych części:

- zestaw czujników reagujących na zniesione jajo
- elektroniczny układ rejestrujący sygnały z czujników, będący odmianą sterownika logicznego
- okablowanie systemowe, czyli sieć przewodów łączących czujniki ze sterownikiem.

* *Politechnika Poznańska, Katedra Automatyki, Robotyki i Informatyki ul. Piotrowo 3a, 60-695 Poznań, mb@ar-kari.put.poznan.pl*

Czujnik jest tym elementem, który decyduje o pewności działania i wiarygodności uzyskanych wyników rejestracji. Wymagana duża liczba czujników wyklucza jednak stosowanie wyrafinowanych środków technicznych. Potrzebny jest zatem tani i niezawodny czujnik stykowy, odporny na zabrudzone i agresywne środowisko, a także na działania mechaniczne (przy wymianie ptaków, czyszczeniu klatek itp.)

Okablowanie nie powinno utrudniać dostępu do klatek, będąc jednocześnie odporne na wspomniane działania mechaniczne. Schemat połączeń powinien umożliwić łatwe ułożenie przewodów z uwzględnieniem minimalizacji łącznej długości całej sieci.

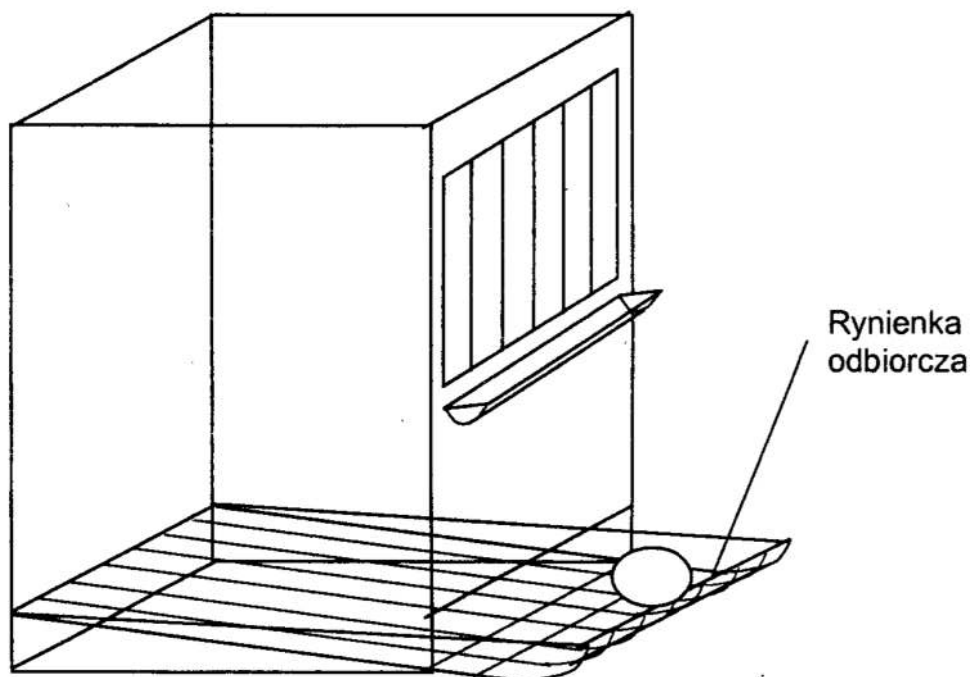
Elektroniczny układ rejestrujący jest odpowiedzialny za niezawodny i trwały zapis wszystkich zdarzeń (zniesienia jaj w danych chwilach czasowych). Będąc w istocie cyfrowym sterownikiem logicznym musi komunikować się z komputerem zewnętrznym dla umożliwienia archiwizacji i przetwarzania uzyskanych danych.

W następnym rozdziale przedstawiono bliżej problemy i ich rozwiązania, dotyczące konstrukcji poszczególnych elementów całego systemu.

3. Czujnik zniesienia jaja

Zarejestrowanie jaja w chwili zniesienia go przez kurę jest praktycznie możliwe tylko przy wykorzystaniu specyficznego kształtu klatki hodowlanej. Wykonana jest ona z drutu stalowego ocynkowanego, a jej podłoga jest pochylona w kierunku zewnętrznym tak, aby jajo stoczyło się do rynienki odbiorczej, skąd jest zabierane przez obsługę.

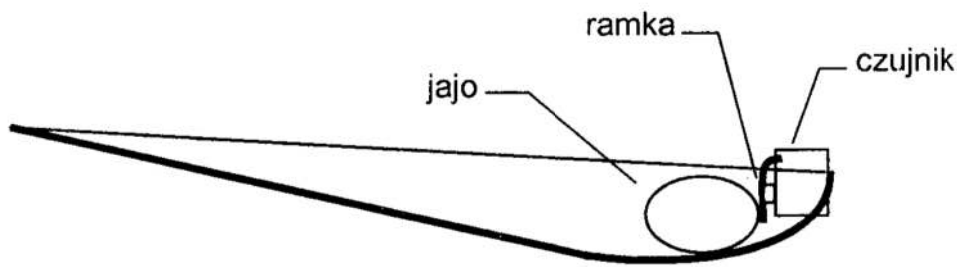
Konstrukcję klatki hodowlanej przedstawiono schematycznie na rysunku 1.



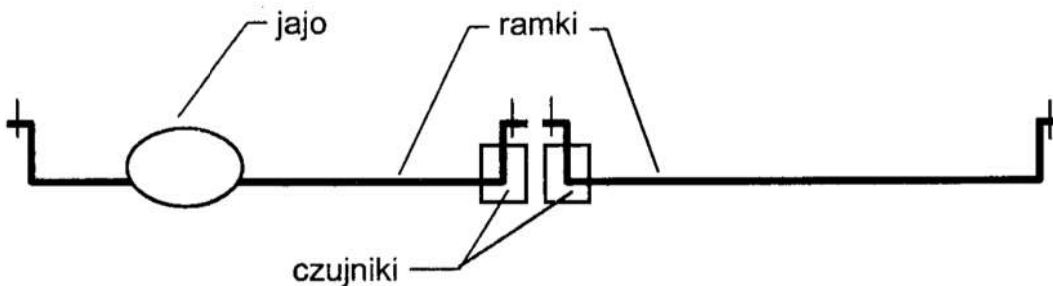
Rys. 1. Klatka hodowlana – schemat konstrukcyjny

Fig. 1. Hen cage. Construction schema

W ostatniej fazie ruchu po zniesieniu jajo uderza o pręt ograniczający rynienkę i odbija się od niego, zatrzymując się w pewnej niewielkiej odległości. Ten fakt został wykorzystany przy budowie czujnika stykowego. Jego konstrukcję oparto o typowy wyłącznik kontaktronowy, zapewniający hermetyczne oddzielenie styków od środowiska zewnętrznego. Mikrowyłączniki takie, stosowane m.in. do budowy klawiatur, wyposażono w obudowy mocujące je parami w miejscach, gdzie stykają się dwie klatki. Taki układ ułatwia późniejszy montaż przewodów, które można było prowadzić czwórkami. Mikrowyłącznik działa jednak punktowo, natomiast miejsce, gdzie może trafić jajo ma formę liniową. Każdy czujnik został więc wyposażony w ramkę obrotową przenoszącą na czujnik siłę z dowolnego punktu uderzenia jaja. Konstrukcję tę ilustrują rysunki 2 i 3.



Rys. 2. Lokalizacja czujnika
Fig. 2. Detector location



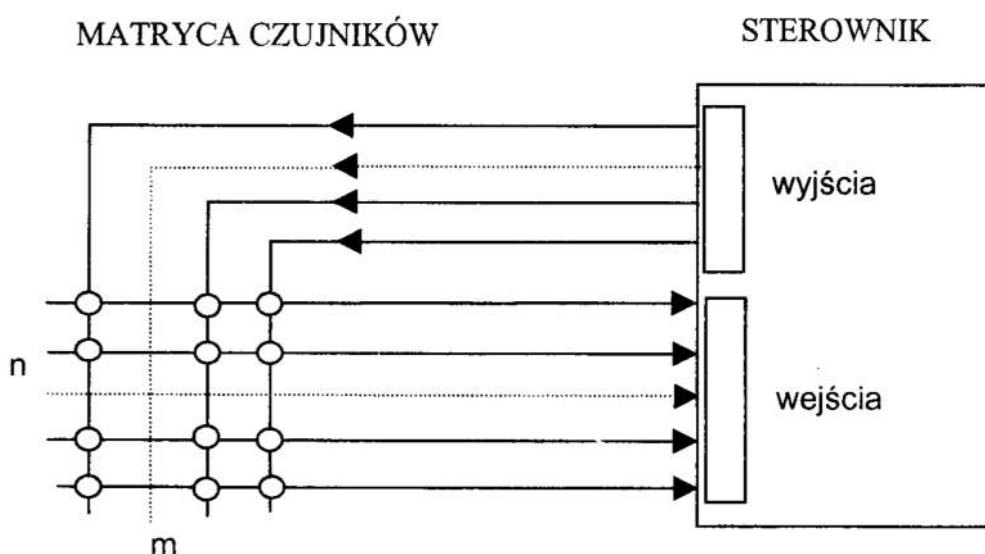
Rys. 3. Rozmieszczenie ramek i czujników w dwóch klatkach
Fig. 3. Layout of the frames and detectors in two cages

Wykorzystując typowy mikrowyłącznik kontaktronowy do budowy czujnika, należało uwzględnić właściwy dobór sprężynki powrotnej. Dla jaja kurzego oryginalna sprężynka okazała się wystarczająca, jednak dla jaja przepiórki była zdecydowanie za sztywna. Osłabienie wszystkich sprężynek dla całej baterii uzyskano poprzez usunięcie części zwojów, dobierając ich ilość doświadczalnie. Staczające się jajo powinno bowiem uruchomić styki tylko w chwili uderzenia w ramkę, natomiast po odbiciu się i ewentualnym powrocie styk musi pozostać otwarty.

4. Sterownik systemowy

Praktycznie niemożliwe jest zastosowanie sterownika o liczbie wejść odpowiadającej ilości czujników. Abstrahując nawet od rozmiarów samego sterownika, wymagałoby to prowadzenia ogromnej ilości indywidualnych przewodów tworzących wiązkę nie mieszczącą się w otoczeniu klatek. Pewnym rozwiązaniem mogłoby być tworzenie lokalnych koncentratorów, jednak wymagałoby to rozproszenia urządzeń elektronicznych w obrębie samych klatek.

Jedynym rozsądnym rozwiązaniem problemu okazało się zorganizowanie zestawu czujników w matrycę podobną do klawiatury komputerowej. Matrycę taką o liczbie wierszy n i liczbie kolumn m podłącza się do sterownika o n wejściach i m wyjściach binarnych (rys.4).



Rys.4. Matrycowy układ czujników
Fig. 4. Matrix structure of the detectors

W powyższym układzie sterownik programowo realizuje przeszukiwanie wszystkich czujników, wystawiając logiczne "0" na kolejne wyjścia i odczytując stan wszystkich wejść. Jeżeli po wystawieniu "0" na wyjściu "j" również pojawi się "0" na wejściu "i", to pobudzony czujnik związany jest właśnie z tym wyjściem i wejściem. Zakładając jednakową ilość w wejść i wyjść sterownika, można oszacować minimalną ilość przewodów p potrzebnych do podłączenia k czujników, jako $p = 2\sqrt{k}$. W praktyce po powyższym oszacowaniu dobiera się liczbę wejść n nieco różną od liczby wyjść m tak, aby spełniony był warunek $m \times n \geq k$. Liczba przewodów równa jest wówczas $p = m+n$. Z zasady działania układu matrycowego wynika, że czujniki reagują na chwilowe, krótkotrwałe zwarcie styków o czasie trwania Δt . Dla m wyjść sterownika częstotliwość przeszukiwania powinna spełniać warunek: $f \geq m/\Delta t$, czyli wystarczać dla rejestracji nawet pozornie jednocześnie znoszonych jaj. Ważne jest jednak, aby żaden czujnik nie był nigdy trwale zwarty, gdyż prowadzi to do przenoszenia się impulsów napięciowych na sąsiednie linie, co w efekcie powoduje rejestrację fałszywych zdarzeń.

Sterownik systemowy dla opisywanych aplikacji dobierany był indywidualnie. W pierwszej z nich, o największej liczbie klatek $k = 720$, praktyczne okazało się wybranie 24 wejść i 32 wyjść sterownika, ze względu na bajtową organizację portów równoległych (3 bajty na wejściu i 4 bajty na wyjściu). Liczba przewodów wyniosła zatem $p = 24 + 32 = 56$. Częstotliwość przeszukiwania $f \geq 32/0.05 = 640$ Hz. Dla uzyskania dostatecznie wysokiej częstotliwości okazało się potrzebne utworzenie tablicy przeliczeń adresów czujników na ich numery w baterii. Podobnie konieczne jest programowe wyeliminowanie efektu drgań styków.

Sterownik do tej aplikacji został zbudowany w oparciu o mikrokomputer jednoukładowy 80C51 i dodatkowe moduły we/wy równoległych 8255 [1] [2]. Zastosowano pełną izolację galwaniczną wszystkich wejść i wyjść za pomocą transoptorów ILQ 74. Sterownik wyposażony został w wyświetlacz i foliową klawiaturę alfanumeryczną.

System może pracować w dwóch trybach: 1 – “odczyt” i 2 – “zapis”.

W trybie pierwszym możliwe jest monitorowanie wszystkich klatek “on line”, z wyświetlaniem numeru klatki, w której zostało zniesione ostatnie jajko.

W trybie drugim ustawia się zegar, oraz zeruje pamięć. Sterownik posiada też wyłącznik “Blokada” stosowany np. podczas wybierania jaj.

Dane o czasie zniesienia jaj we wszystkich klatkach przechowywane są w pamięci RAM sterownika i powinny być przesyłane regularnie do komputera zewnętrznego. Komunikacja odbywa się poprzez łącze szeregowe RS232 i jest inicjowana programem RS.exe wykonywanym przez komputer. Program wywołuje się z parametrami określającymi numer portu szeregowego, napęd i ścieżkę foldera zbiorów wynikowych, oraz dwie chwile czasowe w ciągu doby, w których należy rozpocząć transmisję.

Zbiór wynikowy otrzymuje nazwę z datą i numerem transmisji. Ma format rekordu zgodnego z TP6.0, zawierającego numer klatki, minutę, godzinę, dzień, miesiąc i rok.

W pozostałych aplikacjach systemu monitorowania jaj zdecydowano się na wykorzystanie profesjonalnego, programowalnego sterownika typu MUKS (mikroprocesorowy układ kontrolno-sterujący). Wyposażony jest on w 24 wejścia i 24 wyjścia binarne z optoizolacją, wyświetlacz ciekłokrystaliczny z podświetleniem i klawiaturę 2×16 znaków. Posiada też zegar czasu rzeczywistego i podtrzymanie bateryjne pamięci RAM. Zawartość pamięci wystarcza do zarejestrowania ok. 1500 jaj.

Dla hodowli przepiórek w 80 klatkach wystarczyło użycie tylko 10 wejść i 8 wyjść binarnych, czyli 18 linii łączących. Natomiast dla hodowli kur w 3 bateriach po 48 klatek zastosowano odrębne 3 układy we/wy dla każdej baterii. Każdy układ posiada 8 wejść i 6 wyjść, czyli potrzebne są łącznie 3 przewody 14 żyłowe łączące sterownik z częścią obiektową.

Komputer zewnętrzny (nadrzędny) można podłączyć w dowolnej chwili dla ściągnięcia danych na dysk twardy poprzez łącze RS232. Z poziomu komputera nadrzędnego można również testować cały układ z wizualizacją stanu wszystkich czujników na ekranie monitora, a także wyświetlać stan specjalnych liczników programowych zliczających jaja w każdej klatce osobno w zadanym okresie czasu.

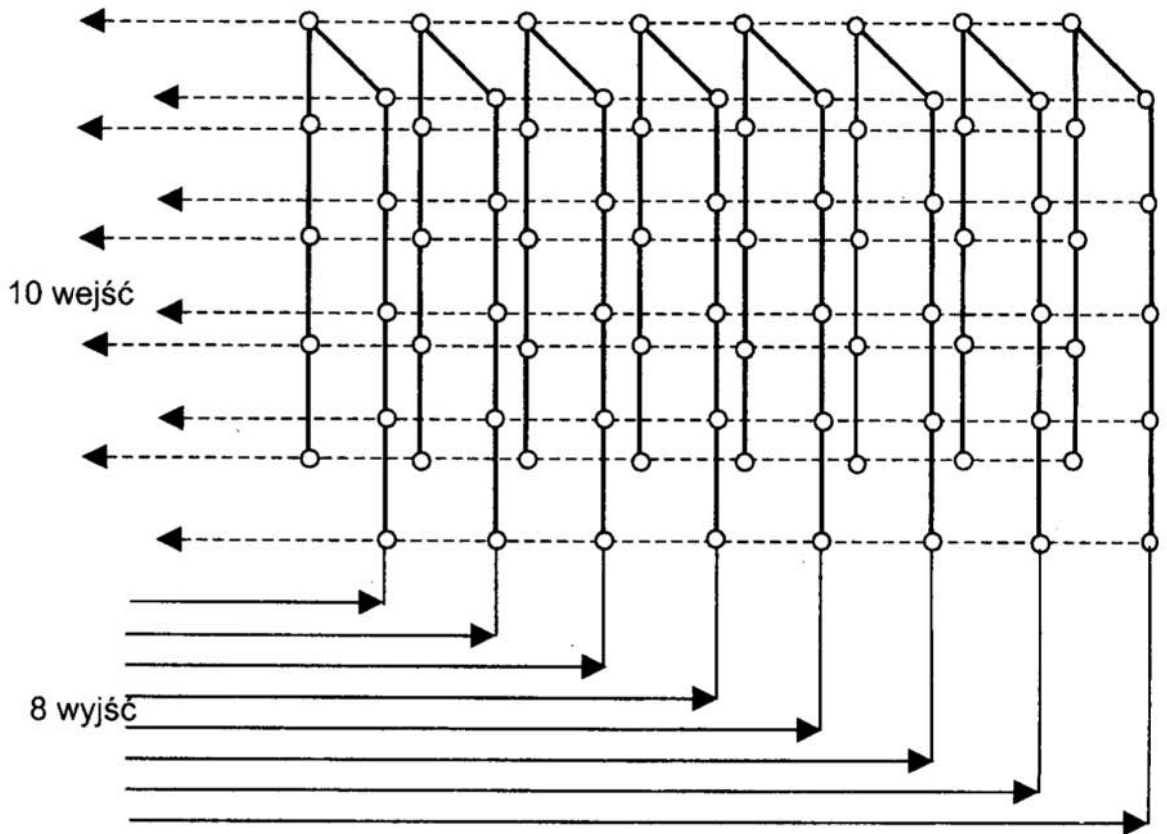
Z klawiatury własnej sterownika możliwe jest ustawianie czasu i daty, odczyt poszczególnych zdarzeń, wyświetlanie stanu wybranego licznika lub jego zerowanie, testowanie czujników, oraz blokada wszystkich wejść. Operacje te można zabezpieczyć hasłem.

Sterownik posiada też możliwość programowego nastawienia czasu "nieczułości" każdego czujnika po zarejestrowaniu jaja. Zwiększa to odporność systemu na przypadkowe uruchomienie czujnika np. przez obsługę.

5. Sieć przewodów łączących

Problem zaprojektowania odpowiedniej konfiguracji sieci przewodów łączących poszczególne czujniki między sobą, oraz formujących linię sprzęgającą całość ze sterownikiem, jest wbrew pozorom dość złożony. Wynika to ze specyficznego rozkładu przestrzennego czujników w konkretnym obiekcie (np. baterii klatek rozmieszczonych w kilku rzędach i piętrach). Rozkład ten należy przetransponować na matrycę dwuwymiarową o n wierszach i m kolumnach tak, aby zminimalizować zarówno łączną długość przewodów jak i ilość węzłów (połączeń międzyprzewodowych).

Stosunkowo łatwo dało się to rozwiązać dla małych baterii klatek z przepiórkami i kurami w hodowlach typowo doświadczalnych. Np. dla hodowli przepiórek układ przewodów wyglądał jak na rysunku 5. Tak prosty układ powstał dzięki możliwości określenia ilości wierszy odpowiadającej dokładnie ilości klatek w jednym segmencie baterii.



Rys. 5. Układ przewodów dla 80 klatek z przepiórkami

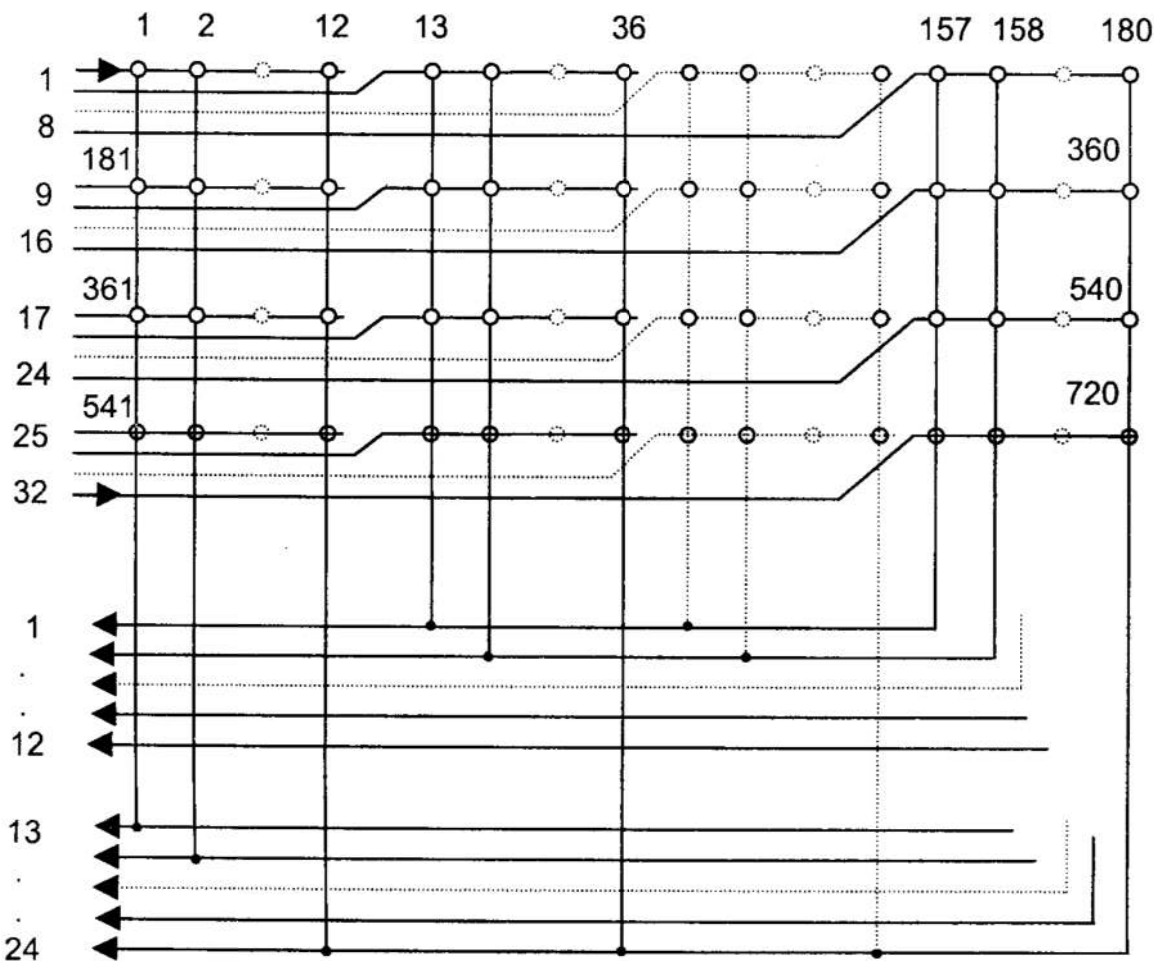
Fig. 5. Wiring layout for 80 quail cages

Zastosowanie tak przejrzystego układu było jednak niemożliwe dla hodowli produkcyjno-doświadczalnej kur w baterii liczącej 720 klatek.

Należało tutaj “zdekomponować” trójwymiarowy układ klatek $2 \times 2 \times 180$ i rozłożyć go na płaszczyźnie w matrycy 24×32 , a następnie matrycę tę z powrotem przełożyć na trójwymiarowy układ przewodów biegnących przez wszystkie klatki baterii.

Biorąc pod uwagę dużą liczbę przewodów, które trzeba prowadzić równoległe, zdecydowano się wykorzystać przewody wielożyłowe taśmowe. Jednak bezpośrednio wykorzystanie ich z częściowym rozdzieleniem poszczególnych żył okazało się możliwe znowu tylko w prostym przypadku jak na rysunku 5.

Dla baterii 720 klatek konieczne okazało się osobne prowadzenie 4 przewodów wielożyłowych wzdłuż całej baterii jako swoistej “magistrali”, do której należało podłączyć poprzeczne przewody obejmujące klatki lokalne. Rozwiązanie to przedstawiono schematycznie na rysunku 6.



Rys. 6. Układ przewodów dla 720 klatek z kurami

Fig. 6. Wiring layout for 720 hen cages

Widoczne na rysunku 6. cztery rzędy klatek po 180 sztuk odpowiadają układowi dwupiętrowemu, po 2 rzędy na każdym piętrze. Przewody pionowe przechodzą przez cztery klatki tworzące dwupiętrowy segment. Dzięki podłączeniu ich do poziomej “magistrali” uzyskano przejrzysty i oszczędny układ przestrzenny przewodów prowadzących bezpośrednio do sterownika systemowego.

6. Podsumowanie

Zagadnienia opisane w niniejszej pracy są z natury nietypowe – potrzeba rejestracji dokładnego czasu zniesienia jaja przez kurę nie występuje powszechnie. Z drugiej strony, jeśli już wystąpi, niewielu jest chętnych do rozwiązania tego problemu – z uwagi na charakter zastosowania, nie rokujący serii zamówień. Autor podjął się go niejako z przekory, zainspirowany oryginalnością zadania. Poza tym będąc sam pracownikiem naukowym w dziedzinie techniki, odczuł swoistą satysfakcję z korzyści dla nauk przyrodniczych, jakie może przynieść opisany wyżej oryginalny system komputerowy.

Literatura

1. Rydzewski A.: *Mikrokontrolery rodziny MRC-51*. WNT Warszawa 1995.
2. Misiurewicz Piotr.: *Podstawy techniki mikroprocesorowej*. WNT Warszawa 1991.

Streszczenie

W referacie przedstawiono system automatycznej rejestracji jaj w chwili ich znoszenia przez kury hodowane w klatkach. Dokładny czas zniesienia każdego jaja jest potrzebny do celów badawczych. Omówiono problemy związane z konstrukcją odpowiedniego czujnika elektromechanicznego, rejestratora cyfrowego i sieci przewodów łączących wszystkie klatki. Opisano również sposoby archiwizacji uzyskanych danych. Przedstawiono wersje projektów zrealizowanych dla różnych hodowli – od małej hodowli przepiórek w 80 klatkach do dużej hodowli kur w 720 klatkach.

EGGS MONITORING COMPUTER SYSTEM FOR A RESEARCH FARM

Summary

An automatic system of registration of eggs for a research hen farm is presented in this paper. For scientific needs, the exact laying time must be registered for each hen. The problems with construction of the appropriate electromechanical detector, digital recorder and wire net connecting all cages are discussed. The methods of archiving of the registered data are also described. Various applications of the system are presented: from a small 80-cages quail farm up to a large 720-cages hen farm.

Referat recenzowany

ZASTOSOWANIE TECHNIK PROGRAMOWANIA W LOGICE Z OGRANICZENIAMI DO ZARZĄDZANIA PRZEPŁYWEM WIELOASORTYMENTOWEJ PRODUKCJI RYTMICZNEJ

1 Wstęp

Potrzeba pełnego wykorzystania mocy produkcyjnych rodzi konieczność szybkiej wstępnej selekcji zgłaszanych zleceń. Ma to szczególne znaczenie w sytuacjach, gdy w systemie (realizującym już pewną produkcję), nie wszystkie jego dostępne moce produkcyjne są w pełni wykorzystane. Oznacza to, że dokonywany wybór winien optymalizować stopień wykorzystania zasobów systemu, gwarantując jednocześnie niezakłócony przebieg dotychczasowo realizowanej produkcji oraz terminowe zakończenie nowego zlecenia.

Rozważany w pracy problem planowania przepływu produkcji sprowadza się do wyznaczenia harmonogramu przepływu partii wyrobów w danej komórce produkcyjnej przy spełnieniu ograniczeń logistycznych wnoszonych przez strukturę systemu produkcyjnego, systemu transportu międzystanowiskowego, pojemności buforów składowania międzyoperacyjnego itp. Poszukiwane rozwiązanie winno również spełniać ograniczenia determinujące wymagania technologiczne występujące w toku realizacji danego zlecenia (następstwo operacji, terminowa realizacja zlecenia, wzajemne wykluczanie się procesów na wspólnych zasobach). Przedstawione podejście wykorzystuje mechanizmy systemu DFKI **Oz2**. System umożliwia korzystanie z technik takich jak branch and bound, itp. w połączeniu z programowalnymi strategiami przeszukiwania i dystrybucji zmiennych [4], [5].

2. Sformułowanie problemu

Model

Rozważmy system produkcyjny składający się z M maszyn, $k \in K = \{1, \dots, M\}$, w którym realizowany jest pewien asortyment wyrobów. Z każdym z W wyrobów skojarzony jest proces technologiczny P_j , $j \in J = \{1, \dots, W\}$. Składa się on z sekwencji operacji na przydzielonych maszynach:

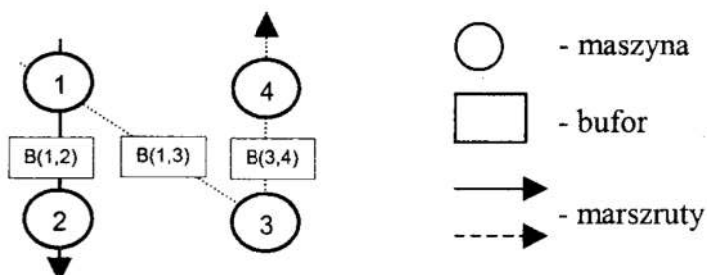
$$P_j = [m(j,1), \dots, m(j, z_j)] \quad (1)$$

gdzie: $m(j,i) \in K$ oznacza numer stanowiska wykorzystywanego w i -tej operacji procesu P_j , ($P_{j,i}$ oznacza i -tą operację procesu P_j obsługiwaną przez $m(j,i)$.) Niech $t(P_{j,1}), \dots, t(P_{j,z_j})$ oznaczają czasy obróbki części partii produkcyjnych na odpowiednich maszynach występujących wzdłuż marszruty technologicznej procesu P_j . z_j oznacza liczbę operacji różnego typu składających się na proces P_j . Przykład syste-

* Politechnika Zielonogórska, Instytut Informatyki i Zarządzania, ul. Podgórna 50, 65-246 Zielona Góra, tel. (0 68) 3282-264. e-mail: T.Borowiecki@iiz.pz.zgora.pl, S.Saniuk@iiz.pz.zgora.pl, Z.Banaszak@pz.zgora.pl

mu złożonego z czterech stanowisk $K=\{1,\dots,4\}$, trzech buforów składowania między-stanowiskowego B(1,2), B(1,3), B(3,4) przedstawia rysunek 1. B(1,2) oznacza bufor pomiędzy stanowiskami 1 i 2, $c(B(1,2))$ oznacza pojemność tego bufora.

Realizowane są tam dwa procesy zadane marszrutami, odpowiednio $1 \rightarrow 2$ oraz $1 \rightarrow 3 \rightarrow 4$. Przyjmuje się, że produkcja realizowana jest partiami, $A = \{a_1, \dots, a_w\}$, gdzie a_j jest wielkością partii produkcyjnej wyrobu skojarzonego z procesem P_j .



Rys. 1. Struktura marszrut technologicznych
Fig. 1. Production routing structure

Sformułowanie problemu

W danym systemie wytwórczym, w skład którego wchodzi obrabiarki CNC i bufory składowania międzystanowiskowego, realizowana jest jednoczesna produkcja pewnego asortymentu wyrobów. Zakładając, że moce produkcyjne systemu nie są wykorzystane w 100% (tzn. stopień wykorzystania zasobów systemu jest mniejszy od jedności) oraz dany jest zbiór zleceń produkcyjnych (scharakteryzowanych przez wielkość produkcji, marszrutę technologiczną (m.in. czasy operacji technologicznych), a także termin ukończenia zlecenia). System pracuje z ustalonym rytmem wyznaczonym przez stanowisko stanowiące wąskie gardło. Znając okres pracy systemu i termin realizacji nowowprowadzanego zlecenia zakłada się, iż istnieje dopuszczalna wielkość partii, która gwarantuje jego terminową realizację. Powstaje pytanie, czy można zrealizować nowowprowadzone zlecenie z założoną wielkością partii, biorąc pod uwagę narzucone ograniczenia wynikające ze struktury systemu i wymagań procesu technologicznego?

Przyjmuje się również ograniczenia determinujące sposób przepływu produkcji oraz ograniczenia wynikające ze struktury systemu [1]:

- wyroby są poddawane obróbce zgodnie z założoną marszrutą technologiczną,
- maszyna w danej chwili może realizować najwyżej jedno zadanie,
- maszyny przydzielane są na czas realizacji całej partii produkcyjnej wyrobu,
- operacja związana z obróbką danej partii produkcji nie może być przerwana,
- wyrób (część partii) po zakończeniu obróbki na danej maszynie trafia do buforu międzystanowiskowego, skąd podejmowany jest do dalszej obróbki pod warunkiem dostępności następczej maszyny z marszrutę technologicznej,
- bufory międzystanowiskowe mają ograniczone pojemności.

3. Zastosowane podejście

Programowanie w logice z ograniczeniami (ang. Constraint Logic Programming (CLP)) [1] w systemie **Oz2** opiera się na podstawowych mechanizmach: propagacji ograniczeń i dystrybucji zmiennych [1], [4], [5]. Propagacja ograniczeń jest efektywnym mechani-

zmem wnioskowania opartego na równoległym działaniu propagatorów wymieniających i gromadzących informacje w tzw. zbiorze wymiany ograniczeń. Dystrybucja zmiennych dzieli problem na uzupełniające się wzajemnie elementy według założonej strategii [4], [5]. Rozwiązanie problemu, jeśli takie istnieje, jest wyznaczane w kolejnych krokach propagacji i dystrybucji. Dystrybucja zmiennych prowadzi do potencjalnej eksplozji kombinatorycznej liczby rozpatrywanych podproblemów. Jej rozmiary jednak mogą być opanowane w praktyce dzięki zastosowaniu mocnych reguł propagacji ograniczeń i odpowiednio dobieranych heurystyk.

Problem szczegółowego harmonogramowania przepływu produkcji wzdłuż wewnątrzkomórkowych marszrut technologicznych należy do klasy problemów kombinatorycznych i może być modelowany przy użyciu zmiennych o skończonych domenach. *Skończona domena* jest zbiorem nieujemnych liczb całkowitych. Notacja $m \# n$ oznacza domenę będącą zbiorem liczb całkowitych $\{m, \dots, n\}$. *Ograniczenie* (w sensie CLP) ma postać logicznego predykatu. Typowymi przykładami ograniczeń występujących w klasie problemów ze zmiennymi o skończonych domenach są:

$X = 67$, $X \in 0 \# 9$, $X = Y$, $X + Y + Z < U$, $X^2 - Y^2 = Z^2$, X_1, \dots, X_9 są różnowartościowe, co najmniej (co najwyżej) n spośród X_1, \dots, X_9 , $n \in \{1, \dots, 9\}$ zmiennych przyjmuje wartość spełniającą zadane kryterium (dwa ostatnie propagatory są propagatorami symbolicznymi).

Problem ze zmiennymi o skończonych domenach składa się również ze skończonego zbioru ograniczeń, ograniczających wszystkie zmienne problemu. Jako rozwiązanie problemu poszukiwane jest takie przyporządkowanie zmiennym wartości (z ich domen), aby spełniało ono wszystkie ograniczenia problemu.

4. Planowanie przepływu produkcji

Model ograniczeń

Niech $s_{(j,i)}^p$ oznacza moment rozpoczęcia obróbki p -tej $p \in \{1, \dots, a_j\}$ części partii i -tej operacji procesu P_j prowadzonej na maszynie $m(j,i)$ (p -jest indeksem części partii używanym w odniesieniu do danego procesu P_j). P_j -ty proces produkcyjny modelowany jest sekwencją rekordów:

$$\{ s_{(j,i)}^p \# t(P_{j,i}) \# [\text{lista operacji bezpośrednio poprzedzających } P_{j,i}] \# m(j,i) \}$$

$$s_{(j,1)}^1 \# t(P_{j,1}) \# [nil] \# m(j,1)$$

...

$$s_{(j,1)}^{a_j} \# t(P_{j,1}) \# [nil] \# m(j,1)$$

$$s_{(j,2)}^1 \# t(P_{j,2}) \# [s_{(j,1)}^1] \# m(j,2)$$

...

$$s_{(j,2)}^{a_j} \# t(P_{j,2}) \# [s_{(j,1)}^{a_j}] \# m(j,2) \tag{2}$$

...

$$s_{(j,z_j)}^1 \# t(P_{j,z_j}) \# [s_{(j,z_j-1)}^1] \# m(j,z_j)$$

...

$$s_{(j,z_j)}^{a_j} \# t(P_{j,z_j}) \# [s_{(j,z_j-1)}^{a_j}] \# m(j,z_j)$$

gdzie znak # jest separatorem pól zgodnie ze składnią Oz2. W nawiasach { } ujęto komentarz objaśniający znaczenie poszczególnych pól. Znacznik *nil* oznacza, że dana lista jest pusta bo operacja jest pierwszą z marszruty.

Podobnie można zapisać wszystkie procesy występujące w systemie. Problem wzajemnego wykluczania procesów współbieżnych może być postrzegany jako problem decyzyjny, gdzie dla każdej pary procesów P_{j1} i P_{j2} współzawodniczących o dostęp do wspólnego zasobu (maszyny) należy wyznaczyć kolejność ich wykonania. Przyjmijmy, że procesy $P_{j1,i1}$ i $P_{j2,i2}$ (składające się tylko z jednej operacji, odpowiednio $i1$ i $i2$) współzawodniczą o dostęp do wspólnego zasobu $k=1$ ($m(j1,i1) = m(j2,i2) = 1$). Dla uproszczenia przyjmijmy że procesy realizowane są z wielkością partii 1. Przyjmijmy chwile, w których operacje te mogą się rozpocząć jako $s_{(j1,i1)}^1$ i $s_{(j2,i2)}^1$. Procesy zajmują maszynę odpowiednio przez $t(P_{j1,i1})$ i $t(P_{j2,i2})$ jednostki czasu. Oznacza to, że poniższy warunek musi być spełniony:

$$s_{(j1,i1)}^1 + t(P_{j1,i1}) \leq s_{(j2,i2)}^1 \quad \vee \quad s_{(j2,i2)}^1 + t(P_{j2,i2}) \leq s_{(j1,i1)}^1 \quad (3)$$

Fakt, że rozpoczęcie obróbki p -tej części partii produkcyjnej operacji $P_{j,i}$ na maszynie $m(j,i+1)$ może mieć miejsce w chwili $s_{(j,i+1)}^p$ upływającej po zakończeniu operacji rozpoczętej w chwili $s_{(j,i)}^p$ realizowanej na $m(j,i)$, oznacza, że musi zachodzić zależność:

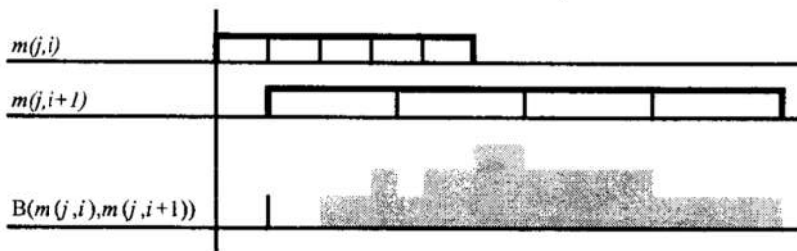
$$s_{(j,i)}^p + t(P_{j,i}) \leq s_{(j,i+1)}^p \quad (4)$$

Założenie niewywłaszczalności procesów (nie dopuszczające do przerywania wykonania partii na danej maszynie) opisuje poniższy warunek:

$$s_{(j,i)}^{p+1} = s_{(j,i)}^p + t(P_{j,i}) \quad (5)$$

Oznacza on, że momenty wprowadzania kolejnych elementów partii do maszyny, zdeterminowane są przez chwile ukończenia elementów partii bezpośrednio je poprzedzających. W celu uwzględnienia ograniczonych pojemności buforów składowania międzyoperacyjnego, wystarczy rozważyć następujące dwa przypadki:

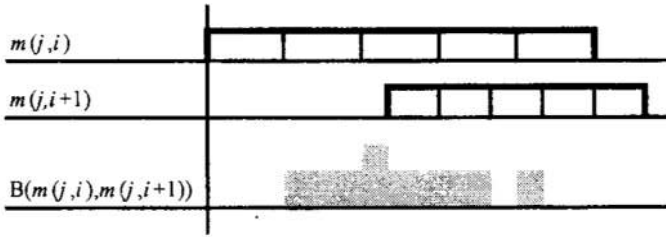
- w marszrucie technologicznej danego procesu P_j wydajność maszyny $m(j,i)$ jest wyższa od wydajności maszyny $m(j,i+1)$ (patrz rysunek2);



Rys. 2. Diagram Gantt'a przepływu produkcji pomiędzy maszynami $m(j,i)$ i $m(j,i+1)$ oraz diagram obciążenia bufora $B(m(j,i), m(j,i+1))$

Fig. 2. Gantt's chart encompassing the workflow following workstations $m(j,i)$ and $m(j,i+1)$ and capacity requirement of the buffer $B(m(j,i), m(j,i+1))$

- w marszrucie technologicznej danego procesu wydajność maszyny $m(j, i+1)$ jest wyższa od wydajności maszyny $m(j, i)$ (patrz rysunek 3).



Rys. 3. Diagram Gantt'a przepływu produkcji pomiędzy maszynami $m(j, i)$ i $m(j, i+1)$ oraz diagram obciążenia bufora $B(m(j, i), m(j, i+1))$

Fig. 3. Gantt's chart encompassing the workflow following workstations $m(j, i)$ and $m(j, i+1)$ and capacity requirement of the buffer $B(m(j, i), m(j, i+1))$

W pierwszym wypadku obserwowane jest gromadzenie się wyrobów na buforze związanym (poprzedzającym) z maszyną o niższej efektywności. Szczyt obciążenia bufora występuje w chwili zakończenia obróbki partii na maszynie o wyższej wydajności. Wymagana pojemność bufora c ze względu na proces j , może być szacowana według poniższej zależności:

$$c(B(m(j, i), m(j, i+1))) \geq \left\lfloor a_j \cdot \left(1 - \frac{t(P_{j,i})}{t(P_{j,i+1})} \right) \right\rfloor + \left\lfloor \frac{s_{(j,i+1)}^1 - s_{(j,i)}^1}{t(P_{j,i})} \right\rfloor, \quad (6)$$

gdzie $\lfloor \rfloor$ oznacza dolne zaokrąglenie wartości. W drugim przypadku, ponieważ maszyna $m(j, i+1)$ jest bardziej wydajna niż maszyna $m(j, i)$, wprowadzanie kolejnych elementów partii musi odbywać się z opóźnieniem w celu otrzymania założonego schematu przepływu. Oznacza to niebezpieczeństwo przepełnienia odpowiedniego magazynu, a w konsekwencji konieczność sprawdzania poniższego warunku:

$$\left\lfloor \frac{s_{(j,i+1)}^1 - s_{(j,i)}^1}{t(P_{j,i})} \right\rfloor \leq c(B(m(j, i), m(j, i+1))) \quad (7)$$

Z założeń modelu systemu wynika, że dane są czasy poszczególnych operacji składających się na marszrutę procesu technologicznego. Przyjmując $c(B(m(j, i), m(j, i+1)))$ za dane, zależności (6), (7) traktuje się jako ograniczenia dotyczące terminów podjęcia partii na kolejnym stanowisku, które gwarantują nieprzepełnienie bufora. Jeśli pojemności buforów są co najmniej równe wielkościom partii, to nie narzucają żadnych ograniczeń na terminy startów partii.

Przepływ produkcji gwarantujący terminową realizację zleceń produkcyjnych w stanie ustalonym posiada cykliczny charakter, zdeterminowany przez wąskie gardło systemu. Nowowprowadzony do systemu proces ma być realizowany z wyznaczoną wielkością partii wynikającą z zapotrzebowania oraz zastanej długości okresu pracy systemu. Łatwo zauważyć, że po wprowadzeniu nowego procesu maksymalna wartość długości uszeregowania dla wszystkich maszyn nie może przekroczyć wartości okresu. Dla operacji procesów współzawodniczących o dostęp do danego urządzenia, musi zatem zachodzić:

$$\max_{j \in J(k)} (s_{(j,i(j,k))}^{a_j} + t(P_{j,i(j,k)})) - \min_{j \in J(k)} (s_{(j,i(j,k))}^1) \leq T, \text{ obowiązuje dla każdego } k \quad (8)$$

gdzie: $J(k)$ - zbiór procesów o marszrutach przechodzących przez maszynę k ,
 $i(j,k)$ - numer operacji procesu j wykonywanej na maszynie k .

Tak więc, przy modelowym ujęciu zadania programowania w logice z ograniczeniami, dane wyjściowe (1), (2) oraz niżej wymienione ograniczenia stanowią określenie pewnego problemu harmonogramowania przepływu produkcji:

- pojemność każdej maszyny wynosi jeden (oznacza to, że w danej chwili tylko jedna operacja może być na niej wykonywana) (3);
- ograniczenie reguluje technologiczny porządek wykonywanych czynności (4);
- zadania produkcyjne (operacje technologiczne na partiach wyrobów nie są wyłączalne) (5);
- pojemności buforów są ograniczone (6), (7);
- okres pracy systemu nie ulega zmianie (8).

Takie ujęcie modelu zadania, gdzie zmienne decyzyjne przyjmują wartości określające chwile rozpoczęcia poszczególnych operacji, a domeny tych zmiennych wyznaczone są na podstawie oszacowań terminu realizacji poszczególnych operacji pozwala ponadto na modelowanie, często spotykanych, dodatkowych zależności wynikających np. z:

- ograniczonych pojemności magazynów międzyoperacyjnych,
- zakazu przerw pomiędzy wykonywaniem operacji dla danego wyrobu na kolejnych maszynach,
- dopuszczalności przerw o ograniczonej długości pomiędzy operacjami na kolejnych maszynach dla danego wyrobu,
- kryteriów kosztowych wynikających z przyjętego sposobu obciążania maszyn.

Model tak skonstruowany nie wymaga stosowania dodatkowych zmiennych zerojedynkowych (jak to ma miejsce przy podejściu za pomocą np. programowania całkowitoliczbowego). Informacje o tym, która operacja może zostać przydzielona do jakiej maszyny w danej chwili niesie ze sobą domena zmiennej. Domena zmiennej poddawana jest propagacji ograniczeń, które wyrażają sposób zachowania systemu. Dzięki temu że ograniczenia zapisywane są za pomocą propagatorów języka wysokiego poziomu (duża elastyczność i intuicyjność zapisu), możliwe jest swobodne tworzenie modeli problemów uwzględniających różne (dotąd często odrębnie rozpatrywane) zagadnienia, wynikające z interakcji podsystemów transportu, magazynowania międzystanowiskowego, wymagań technologicznych procesu produkcyjnego, wskaźników kosztu, itp.

Dzięki takiemu ujęciu modelu możliwe było wykorzystanie standardowych procedur harmonogramowania zaimplementowanych w języku **Oz2**, korzystających z efektywnych mechanizmów propagacji ograniczeń oraz pewnych heurystyk wywodzących się z badań operacyjnych. W języku **Oz2** istnieje możliwość swobodnego programowania strategii przeszukiwania i dystrybucji zmiennych, co można dodatkowo wykorzystać w celu zwiększenia efektywności poszukiwania rozwiązań (wykorzystując pewne specyficzne właściwości danego problemu).

Model zagadnienia harmonogramowania niepodzielnych zadań, o ustalonym czasie trwania na przydzielonych zasobach o jednostkowej pojemności został rozszerzony na przypadek przepływu partii wyrobów o ustalonym schemacie z uwzględnieniem ograni-

czonych pojemności buforów przystanowiskowych. Zagadnienie to było rozpatrywane w przykładowym systemie, przy założeniu cyklicznego (rytmicznego) charakteru realizowanego w nim przepływu produkcji.

Oprogramowanie CLP umożliwia łatwe programowanie takich problemów na drodze zastąpienia poszczególnych zmiennych danymi rzeczywistymi. W szczególności zaś, umożliwiając rozwiązywanie problemów planowania wykonawczego, stanowić może cenne uzupełnienie oprogramowania klasy MRP II [2].

Przykład

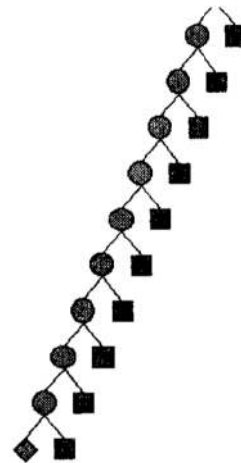
Rozważmy system złożony z czterech maszyn $K = \{1,2,3,4\}$ oraz buforów składowania międzyoperacyjnego $B(1,2), B(1,3), B(3,4)$ o pojemnościach odpowiednio 2, 4, 3. Niech proces P_1 realizowany będzie partiami zawierającymi 4 elementy. Czy możliwe jest uruchomienie dodatkowego procesu P_2 , realizowanego partiami zawierającymi 4 elementy? Procesy P_1 i P_2 są znane:

$$P_1 = [1, 2], \quad P_2 = [1, 3, 4]$$

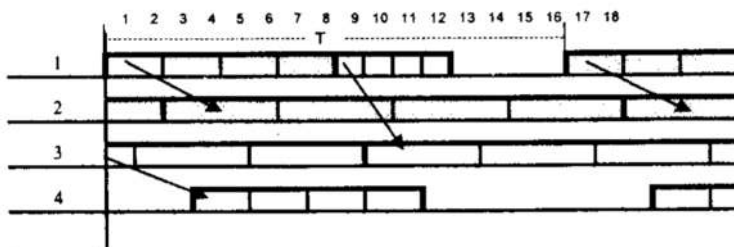
$$t(P_{1,1}) = 2, t(P_{1,2}) = 4 \quad t(P_{2,1}) = 1, t(P_{2,2}) = 4, t(P_{2,3}) = 2$$

Program zawiera tekstualne specyfikacje ograniczeń (3) - (8). Proces przeszukiwania przestrzeni rozwiązań ilustruje fragment pełnego drzewa przeszukiwań, przedstawiony na rysunku 4, gdzie wierzchołki typu kółka oznaczają wybór, wierzchołki typu romb oznaczają rozwiązanie dopuszczalne, wierzchołki typu kwadrat oznaczają brak rozwiązania.

Wierzchołki oznaczające brak rozwiązania wiążą się z sytuacjami, w których któreś z narzuconych ograniczeń nie może być spełnione, np. założona pojemność bufora jest zbyt mała. Przykład rozwiązania dopuszczalnego, w którym procesy P_1 i P_2 spełniają zadane ograniczenia przedstawia rysunek 5. W przedstawionym rozwiązaniu okres przebiegu ustalonego wynosi $T = 16$.



Rys. 4. Fragment drzewa poszukiwań
Fig. 4. A part of the search tree



Legend:

■ -workflow P_1 , □ -workflow P_2

Rys. 5. Diagram Gantt'a przepływu produkcji podczas realizacji procesu P_1 i P_2
Fig. 5. Gantt's chart of workflows P_1 and P_2 following of the all constraints

Literatura

1. Borowiecki T., Banaszak Z.: *A constraint logic programming approach for AGV flow control. Proc. of the Workshop on Constraint Programming for Decision and Control.* Ed. J. Figwer, June 28, Gliwice, Poland 1999, pp.41-48.
2. Landvater D., Gray C.: *MRP II Standard System. A Handbook for Manufacturing Software Survival.* Oliver Wight Limited Publications Inc., Essex, 1995.
3. Schollaert F.: *Money Resource Planning – MRP-III.* Library Albert I, Brussels, 1994.
4. Schulte Ch., Smolka G., Wurtz J.: *Finite Domain Constraint Programmig in Oz. DFKI OZ documentation series.* German Research Center for Artificial Intelligence, Stuhlsaltzenhausweg 3, D-66123 Saarbrucken, Germany, 1998.
5. Henz M., Muller M.: *Tips on Oz. DFKI OZ documentation series.* German Research Center for Artificial Intelligence, Stuhlsaltzenhausweg 3, D-66123 Saarbrucken, Germany, 1997.

Streszczenie

Planowanie przepływu produkcji wiąże się z koniecznością podejmowania szeregu decyzji, m.in. w zakresie porcjowania, marszrutowania i harmonogramowania. W stosowanych podejściach, do rozwiązywania poszczególnych problemów cząstkowych wykorzystywane są odpowiednie techniki programowania matematycznego. Sekwencja uzyskanych w ten sposób rozwiązań optymalnych zadań cząstkowych nie gwarantuje optymalności poszukiwanego rozwiązania globalnego. W pracy przedstawiono metodykę sprawdzania, czy do harmonogramu ustalonej produkcji rytmicznej o danym okresie powtarzalności można dodatkowo włączyć nowy proces o danej marszrucie, czasach operacji i wielkości partii, przy czym wykorzystuje się język Oz2. Programowanie zadań w tym języku sprowadza się do specyfikacji ograniczeń. Poszukiwanie rozwiązania oparte jest na metodzie przeszukiwania z propagacją ograniczeń. Efektywność proponowanego podejścia zilustrowana została na przykładzie hipotetycznego systemu jednoczesnej produkcji rytmicznej.

APPLICATION OF CONSTRAINT LOGIC PROGRAMMING FOR REPETITIVE PRODUCTION FLOW MANAGEMENT

Summary

Production planning consists of many decision-making activities including such issues as batch sizing, production routing and jobs scheduling. Methods of mathematical programming are used in order to resolve the above-mentioned problems separately. However, separately local problems resolution does not guarantee a global solution obtained is an optimal one. In this paper an approach to scheduling, based on Oz2 language is proposed. A task-programming resume in constraints specification using a kind of textual specification. A method of CLP is employed as to refine a solution. Efficiency of the approach proposed is illustrated on an example of a repetitive manufacturing system.

Referat recenzowany

KSZTAŁTOWANIE KONSTRUKCJI METODĄ OPTYMALNEGO ROZKŁADU RODZAJU MATERIAŁU

1 Wprowadzenie

Optymalne kształtowanie konstrukcji ma coraz szersze zastosowanie w projektowaniu wspomaganym komputerowo wykorzystywanym w produkcji. Problem optymalizacji kształtu jest problemem nie tylko opisu granicy konstrukcji ale również problemem wymiarowania i rozkładu rodzaju materiału przy zastosowaniu materiałów złożonych.

W niektórych warunkach pracy i wynikających z nich obciążeniach konstrukcji, jako kryterium wytrzymałościowe stosuje się wielkość nagromadzonej potencjalnej energii deformacji. Potencjalna energia deformacji charakteryzuje nam również uogólnioną sztywność konstrukcji.

Dla konstrukcji z materiałów izotropowych liniowo sprężystych całkowita jednostkowa potencjalna energia deformacji jest określona wyrażeniem [2]:

$$u = \frac{1}{2E} \left[\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x) \right] + \frac{1+\mu}{E} (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2), \quad (1)$$

natomiast jednostkową energię deformacji objętościowej oraz odkształcenia postaciowego przedstawiają wzory:

$$u_v = \frac{1-2\mu}{6E} (\sigma_x + \sigma_y + \sigma_z)^2, \quad (2)$$

$$u_p = \frac{1+\mu}{6E} \left[(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2 \right] + \frac{1+\mu}{E} (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2), \quad (3)$$

gdzie: $\sigma_x, \sigma_y, \sigma_z$ i $\tau_{xy}, \tau_{yz}, \tau_{zx}$ – odpowiednie naprężenia w konstrukcji, E, μ – stałe Younga i Poissona.

Całkowita energia konstrukcji będzie równa całce:

$$U_c = \int_V u dV = \int_V (u_v + u_p) dV. \quad (4)$$

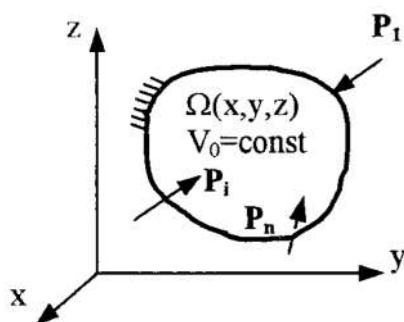
W wyrażeniach (1) – (3) występują stałe E i μ . Rzeczywiste materiały konstrukcyjne charakteryzują się określonymi stałymi modułami sprężystości Younga E i współczynnikami Poissona μ , przy innych stałych warunkach. Przy optymalizacji kształtu możemy posłużyć się stałymi E i μ , znajdując ich optymalne rozłożenie w przestrzeni konstrukcyjnej. Następnie możemy przejść do problemu poszukiwania kształtu w zadanej przestrzeni konstrukcyjnej transformując teoretycznie optymalny rozkład rodzaju materiału w objętości na kształt konstrukcji.

*Politechnika Koszalińska, ul. Raclawicka 15-17, 75-620 Koszalin, E-Mail: fligiel@lew.tu.koszalin.pl

Takie podejście do poszukiwania konstrukcji umożliwia poszukiwanie kształtu konstrukcji bez zmiany wymiarowania i umożliwia wybór (zmianę) materiału konstrukcyjnego w elementach objętości konstrukcji nie odpowiadających zadanym warunkom wytrzymałościowym. Dla rzeczywistych obiektów konstrukcyjnych ze względów funkcjonalnych nie zawsze optymalizacja kształtu poprzez wymiarowanie jest dostatecznie prosta do realizacji, gdyż istnieją stałe określone wymiary wynikające z własności funkcjonalnych i użytkowych konstrukcji. Dla swobodnych powierzchni konstrukcyjnych optymalizacja poprzez wymiarowanie jest stosunkowo łatwa do przeprowadzenia.

2 Synteza konstrukcji w zadanej przestrzeni konstrukcyjnej

Zadanie syntezy konstrukcji w zadanej przestrzeni konstrukcyjnej podzielimy na dwa etapy. W pierwszym etapie znajdziemy optymalny rozkład własności materiałowych w zadanej objętości. W drugim etapie natomiast przeprowadzimy transformację własności materiałowych na rzeczywiste własności użytych materiałów konstrukcyjnych i wymiarowanie. W wyniku tak przeprowadzanej optymalizacji otrzymamy zwymiarowaną konstrukcję o wymaganych (zadanych) własnościach materiałowych spełniającą przyjęte kryteria optymalizacji.



Rys. 1. Przestrzeń konstruowania

Fig. 1. Design space

Pierwszy etap zadania syntezy sformułujemy w następujący sposób (rys. 1): w zadanej przestrzeni konstrukcyjnej $\Omega(x,y,z)$ należącej do R^3 , znaleźć najbardziej sztywną konstrukcję przy stałej objętości $V_0 = \text{const}$ i umownej sprężystości określającej rozkład materiałów złożonych w konstrukcji $E_h = \int_{V_0} E dV_0 = \text{const}$, przy niezmiennych

warunkach obciążenia i podparcia konstrukcji. Jako kryterium uogólnionej sztywności przyjmujemy wielkość potencjalnej energii deformacji U_c nagromadzonej w objętości V_0 .

Całkowita potencjalna energia deformacji będzie określona:

$$U_c = \int_{V_0} \left(\frac{1}{2E} [\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x)] + \frac{1+\mu}{E} (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2) \right) dV_0. \quad (5)$$

W taki sam sposób otrzymamy energię odkształcenia objętościowego i postaciowego z zależności (2) i (3).

Ograniczenia zapiszemy w postaci:

$$Q_E = \int_{V_0} E dV_0 - E_V = 0, \quad E_V = \sum_{j=1}^k \int_{V_j} E_{0j} dV_{0j} = \text{const}, \quad V_0 = \sum_{j=1}^k V_{0j} = \text{const}, \quad (6)$$

gdzie: k – liczba różnorodnych rzeczywistych materiałów mających zastosowanie w konstrukcji.

Funkcjonał Lagrange'a uwzględniający (5) i (6) będzie następujący [1]:

$$L_E = U_c + \lambda Q_E, \quad (7)$$

gdzie: λ – współczynnik Lagrange'a.

Z warunku ekstremum (7):

$$\frac{\partial L_E}{\partial E} = \frac{\partial U_c}{\partial E} + \lambda \frac{Q_E}{\partial E} = 0, \quad (8)$$

z uwzględnieniem ograniczeń (6), po odpowiednich przekształceniach otrzymamy zależność na określenie ekstremalnych wartości modułów Younga, dla każdej r -tej iteracji:

$$E_r = \frac{\sum_{j=1}^k \int_{V_j} E_{0j} dV_{0j} \cdot \left(\sqrt{[\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu_{0j}(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x)] + 2(1+\mu_{0j})(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2)} \right)_{r-1}}{\int_{V_0} \left(\sqrt{[\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu_{0j}(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x)] + 2(1+\mu_{0j})(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2)} \right)_{r-1} dV_0}, \quad (9)$$

gdzie: E_{0j} , μ_{0j} – początkowe wartości.

Obliczenia zatrzymujemy, gdy różnica $E_r - E_{r-1} < \varepsilon$ (ε – mała zadana wartość), w każdym punkcie przestrzeni $\Omega(x,y,z)$ jest spełniona.

W drugim etapie należy dokonać transformacji optymalnego rozłożenia rodzaju materiału w przestrzeni konstrukcyjnej $\Omega(x,y,z)$ na wymiary konstrukcji i rzeczywisty rodzaj materiału konstrukcyjnego przyjętego ze względów funkcjonalnych i użytkowych.

Potencjalna energia deformacji dla r -tej iteracji wynosi:

$$U_r = \int_{V_0} \left(\frac{1}{2E_r} [\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu_{0j}(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x)]_r + \frac{1+\mu_{0j}}{E_r} (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2)_r \right) dV_0. \quad (10)$$

Ponieważ, konstrukcja z poszukiwanymi nowymi wymiarami ma mieć taką samą sztywność jak z optymalnie rozłożonym rodzajem materiału, to energie deformacji konstrukcji powinny być sobie równe.

Potencjalna energia deformacji dla konstrukcji z poszukiwanymi optymalnymi wymiarami powinna być równa:

$$U_w = \sum_{j=1}^k \int_{V_{wj}} \left(\frac{1}{2E_{0j}} [\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - 2\mu_{0j}(\sigma_x\sigma_y + \sigma_y\sigma_z + \sigma_z\sigma_x)]_r + \frac{1+\mu_{0j}}{E_{0j}} (\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2)_r \right) dV_{wj}. \quad (11)$$

W zależności (11) podcałkowe naprężenia w każdym punkcie przestrzeni konstrukcyjnej nie odpowiadają rzeczywistym przy zmienionych wymiarach i/lub zastosowanemu rzeczywistemu materiałowi konstrukcyjnemu, dlatego do dalszych obliczeń przyjmujemy funkcję różnicy potencjalnych energii deformacji:

$$f(V_j, E_j) = U_r - U_w \quad (12)$$

Ograniczenia (6) przekształca się do postaci:

$$Q_w = \int_{V_0} E_i dV_r - E_V = 0 \quad (13)$$

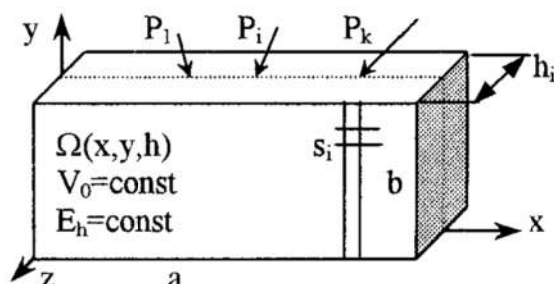
Aby wyznaczyć funkcję rozłożenia wymiarów w przestrzeni konstrukcyjnej $\Omega(x,y,z)$ posłużymy się funkcjonałem Lagrange'a:

$$L = f(V_j, E_j) + \lambda Q_w, \quad (14)$$

z którego wyznaczymy zależność na obliczenie poszukiwanych transformowanych wymiarów konstrukcji. Zależność na obliczenie poszukiwanych wymiarów będzie funkcją rodzaju rozpatrywanego zadania wynikającego z własności funkcjonalnych i użytkowych konstrukcji.

3 Przykład syntezy konstrukcji

Jako przykład rozpatrzmy dwuwymiarową przestrzeń konstrukcyjną $\Omega(x,y,h)$, gdzie h będzie grubością konstrukcji.



Rys. 2. Dyskretna przestrzeń konstruowania

Fig. 2. Discrete design space

Dokonując dyskretyzacji przestrzeni $\Omega(x,y,h)$ na n elementów (rys. 2), wyrażenia (1), (2) i (3) przyjmą postać:

$$U = \sum_{i=1}^n \left[\frac{1}{2E_i} (\sigma_{xi}^2 - 2\mu_i \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2) + \frac{1+\mu_i}{E_i} \tau_{xyi}^2 \right] V_{0i}, \quad (15)$$

$$U_v = \sum_{i=1}^n \frac{1-2\mu_i}{6E_i} (\sigma_{xi} + \sigma_{yi})^2 V_{0i}, \quad (16)$$

$$U_p = \sum_{i=1}^n \left[\frac{1+\mu_i}{3E_i} (\sigma_{xi}^2 - \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2) + \frac{1+\mu_i}{E_i} \tau_{xyi}^2 \right] V_{0i}, \quad (17)$$

gdzie: $V_{0i} = s_i h_{0i}$ – objętość wydzielonej i -tej przestrzeni o powierzchni s_i i grubości początkowej h_{0i} z zadanej przestrzeni $\Omega(a,b,h)$.

Uwzględniając w (7) i (8) wyrażenie na energię (15) oraz ograniczenia:

$$Q_E = \sum_{i=1}^n E_{0ji} V_{0i} - E_V = 0, \quad E_V = \sum_{i=1}^n E_i V_{0i} = \text{const}, \quad V_0 = \sum_{i=1}^n V_{0i} = \text{const}, \quad (18)$$

gdzie: E_{0ji} , μ_{0ji} – stałe zadanego rodzaju materiału konstrukcyjnego, dostaniemy zależność na określenie stałych Younga dla pierwszego etapu optymalizacji:

$$E_{ir} = \frac{\sum_{i=1}^n E_{0ji} s_i h_{0i} \cdot \sqrt{(\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2 + 2(1 + \mu_{0ji}) \tau_{xyi}^2)_{r-1}}}{\sum_{i=1}^n s_i h_{0i} \sqrt{(\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2 + 2(1 + \mu_{0ji}) \tau_{xyi}^2)_{r-1}}}. \quad (19)$$

W drugim etapie optymalizacji zależność (12) będzie określona w następujący sposób:

$$f(V_j, E_j) = \sum_{i=1}^n \left[\frac{1}{2E_{ir}} (\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2)_r + \frac{1 + \mu_{0ji}}{E_{ir}} (\tau_{xyi}^2)_r \right] V_{0i} - \\ - \sum_{i=1}^n \left[\frac{1}{2E_{0ji}} (\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2)_r + \frac{1 + \mu_{0ji}}{E_{0ji}} \tau_{(xyi)}^2 \right] V_i, \quad (20)$$

gdzie: $V_i = s_i h_i$ – objętość wydzielonej i -tej transformowanej przestrzeni o powierzchni s_i i poszukiwanej grubości h_i .

Ograniczenie (13) zapiszemy:

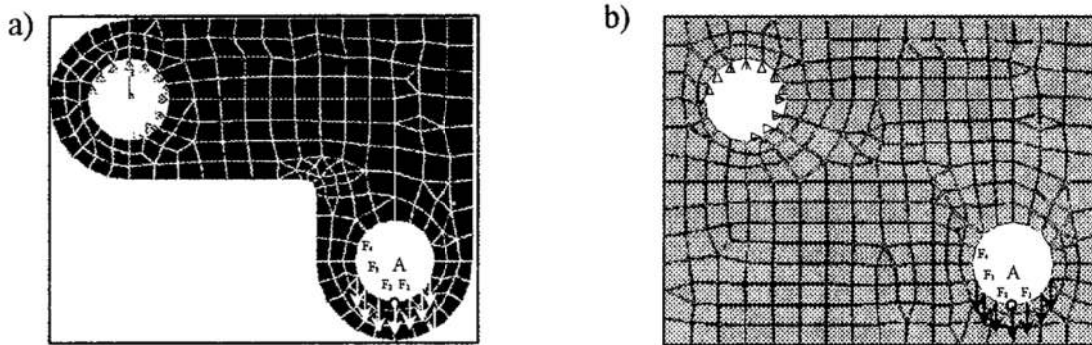
$$Q_w = \sum_{i=1}^n E_{ir} V_{0i} - E_V = 0, \quad E_V = \sum_{i=1}^n E_{0ji} V_i = \text{const}, \quad V_0 = \sum_{i=1}^n V_i = \text{const}. \quad (21)$$

Przyjmując, $R_i = \sqrt{[\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2 + 2(1 + \mu_{0ji}) \tau_{xyi}^2]_r} h_i$, oraz uwzględniając (20) i (21) dostaniemy zależność na określenie konwertowanych grubości z stałych Younga E_{ir} wyznaczonych w pierwszym etapie optymalizacji:

$$h_i = \frac{\sum_{i=1}^n E_{ir} s_i h_{0i} \cdot \sqrt{[\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2 + 2(1 + \mu_{0ji}) \tau_{xyi}^2]_r}}{\sum_{i=1}^n E_{0ji} s_i \sqrt{[\sigma_{xi}^2 - 2\mu_{0ji} \sigma_{xi} \sigma_{yi} + \sigma_{yi}^2 + 2(1 + \mu_{0ji}) \tau_{xyi}^2]_r}}. \quad (22)$$

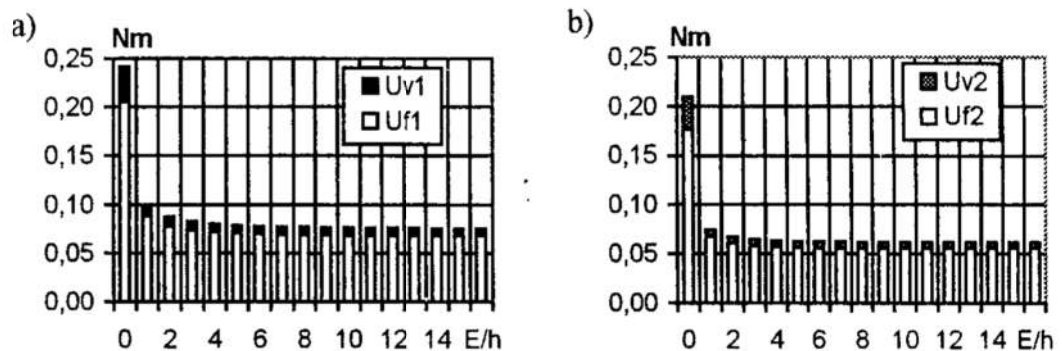
Zgodnie z powyższą metodyką przeprowadzono optymalizację dwóch przestrzeni konstrukcyjnych pokazanych na rysunku 3a i b. Do obliczeń wykorzystano metodę elementów skończonych. Konstrukcję podzielono na elementy skończone i obciążono siłami $F_1=800$ N, $F_2=760$ N, $F_3=550$ N, $F_4=200$ N. Początkowa grubość (startowa) elementów była podebrana tak, aby objętość $V_0 = 1,3682 \cdot 10^{-4} \text{ m}^3$ była taka sama dla każdej konstrukcji. Siatka węzłów i elementów oraz obciążenia i punkty

zamocowania pokazane są na rysunku 3a i 3b. Program obliczeniowy napisano w języku Turbo-Pascal.



Rys. 3. Modele obliczeniowe
Fig. 3. Mathematical models

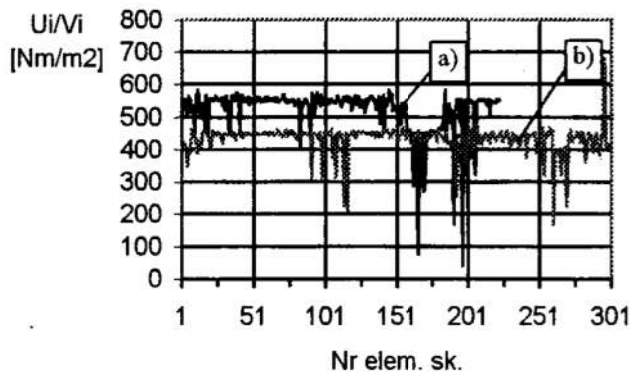
Na rysunku 4a i 4b przedstawiono zmiany sumy energii deformacji postaciowej i objętościowej wraz z numerem iteracji. Ostatnia kolumna E/h przedstawia energie konstrukcji z transformowanymi grubościami. Jak można zauważyć, praktycznie transformacja modułów Younga E na grubości h nie wywołuje zmian całkowitej energii jak i jej składowych, chociaż występuje brak koincydencji pomiędzy grubościami i naprężeniami we wzorze (22). Rozkład zredukowanych naprężeń w 15 iteracji dla wariantu a i b jest mocno zróżnicowany. Naprężenia zredukowane dla wariantu a zmieniają się od 0,035 Pa do 259,698 MPa, a dla wariantu b od 41 Pa do 307,718 MPa. Natomiast naprężenia zredukowane dla transformowanych grubości przyjmują wartości, odpowiednio, od 4,466 Pa do 16,611 MPa i od 8,942 Pa do 18,156 MPa. Z powyższego wynika, że część materiału jest niewykorzystana pod względem wytrzymałościowym i możliwe jest zastosowanie materiału o małej wytrzymałości. W konstrukcjach zoptymalizowanych wartości naprężeń są wyrównane.



Rys. 4. Potencjalna energia deformacji $U_c = U_v + U_p$, odkształcenia objętościowego U_v i postaciowego U_p

Fig. 4. Energy of deformation $U_c = U_v + U_p$, strain energy of volume change U_v , strain energy of distortion U_p

Wartość jednostkowej potencjalnej energii deformacji w elementach skończonych również jest wyrównana względem konstrukcji wyjściowej, co można zauważyć na rysunku 5a i b.

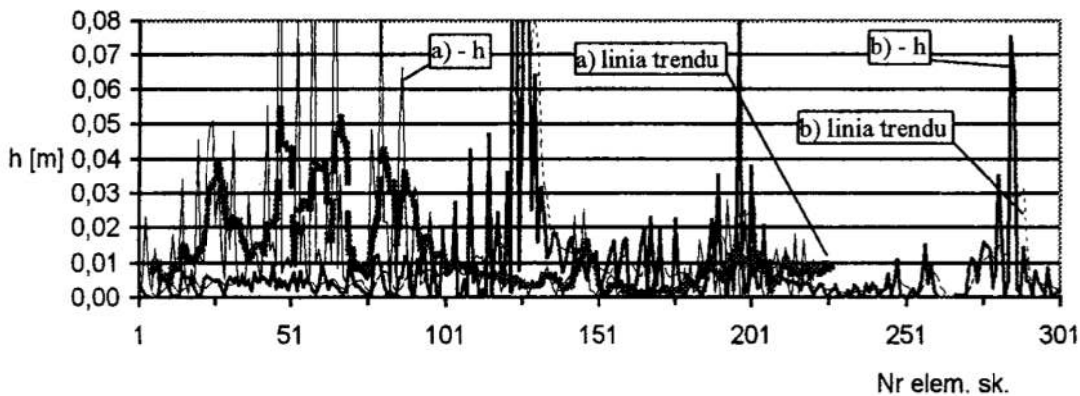


Rys. 5. Jednostkowa potencjalna energia

Fig. 5. Unitary energy of deformation

Jest to zgodne z definicją najbardziej sztywnej konstrukcji, w której jednostkowa potencjalna energia deformacji (potencjał) jest taki sam w każdym punkcie przestrzeni konstrukcyjnej [3].

Dla punktu A na rysunku 3a i 3b przemieszczenia zmniejszyły się, odpowiednio, dla osi x - 292% i 334%, a dla osi y - 312% i 334%, w odniesieniu do wielkości wyjściowych, tj., konstrukcja jest około trzykrotnie sztywniejsza niż wyjściowa.



Rys. 6. Grubości elementów

Fig. 6. Thicknesses of elements

Porównując optymalne kształty wariantu a i b można zauważyć, że mniejsza względna jednostkowa potencjalna energia deformacji występuje w obu wariantach w tych samych obszarach. W wariantcie b większy jest obszar małych wymiarów grubości w prawym górnym obszarze niż w wariantcie a. Jest to wynik przejścia obciążenia siłami wewnętrznymi przestrzeni materiałowej, która nie występuje w wariantcie a. Generalnie grubości elementów w wariantcie a są większe niż w b i ich średnia zmiany jest większa, co wynika z szerokości linii trendu (rys. 6).

Literatura

1. Fligiel M.: *Modelowanie konstrukcji z uwzględnieniem dymorfizmu potencjalnej energii deformacji*. Materiały XVII Ogólnopolskiej Konferencji „Poliptymalizacja i Komputerowe Wspomaganie Projektowania” Koszalin – Mielno, 1999, s. 64-71.
2. Timoshenko S.P., Goodier J.N.: *Theory of elasticity*. 3rd ed., New York, McGraw-Hill Book Co., Inc., 1970.
3. Wasiutyński Z.: *O kształtowaniu wytrzymałościowym*. Akademia Nauk Technicznych, Warszawa, 1939.

Streszczenie

W pracy rozpatruje się optymalizację kształtu konstrukcji w zadanej przestrzeni konstrukcyjnej metodą optymalnego rozkładu materiału z transformacją na wielkości wymiarowe. Optymalizacja odbywa się w dwóch etapach. W pierwszym etapie znajduje się optymalne własności materiałowe konstrukcji. W drugim etapie dokonuje się transformacji własności materiałowych na wymiary z uwzględnieniem rzeczywistych własności materiałowych konstrukcji. Proces poszukiwania optymalnego kształtu odbywa się w sposób iteracyjny z podziałem przestrzeni konstrukcyjnej na elementy skończone. Na dwóch przykładach o takich samych warunkach zamocowania i obciążenia pokazano poszukiwanie kształtu konstrukcji.

FORMING CONSTRUCTION BY METHOD OF OPTIMAL DISTRIBUTION OF SORT OF MATERIAL

Summary

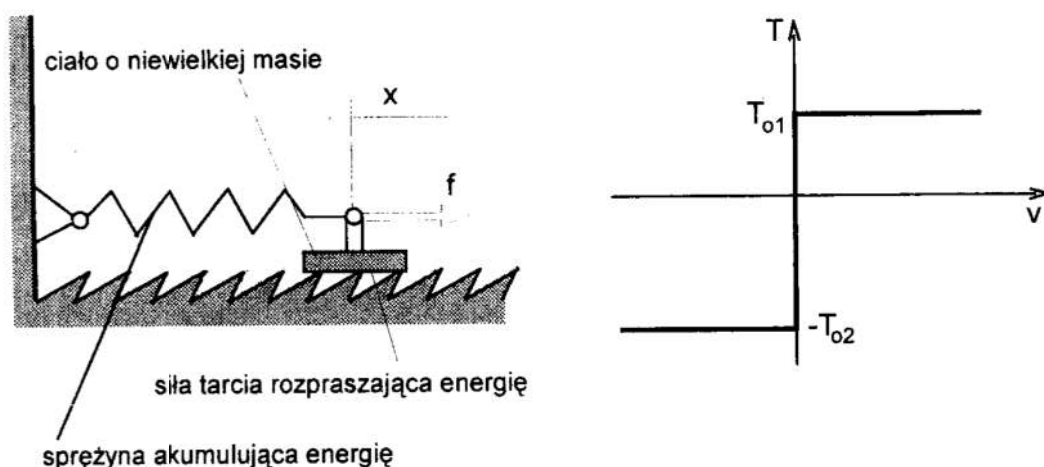
Subject matter of the paper is optimisation of design shape in a given structural space with a method of optimal distribution of material and a transformation into dimensional quantities. Optimisation advances in two stages. In the first stage optimal material properties of the design are to be determined. In the second stage takes place the transformation of material properties into dimensions taking the real material properties of the design into consideration. The process of searching the optimal shape is iterative with division of the structural space into finite elements. The searching design shape is exemplified by two designs in the same conditions of fixing and loading

Referat recenzowany

DYNAMIKA UKŁADU SPRĘŻYSTO-CIERNEGO

1. Wprowadzenie

Układ sprężysto-cierny jest układem mechanicznym, w którym energia jest akumulowana w potencjalnej postaci, a jej rozpraszanie wywołują siły tarcia suchego. Zasadniczym zadaniem dynamicznej analizy takiego układu jest wyznaczenie uogólnionych prędkości układu. Pokażemy, że przedstawione zadanie można sformułować w postaci zagadnienia optymalizacyjnego. Rozważana tu problematyka stanowi przykład wykorzystania prezentowanego w pracy [3] ogólnego sformułowania wariacyjnego zadania mechaniki.



Rys. 1. Schemat elementarnego układu sprężysto-ciernego

Fig. 1. Schematic diagram of an elementary elastic system with friction.

Na rysunku 1 pokazano elementarny przykład ilustrujący rozpatrywaną tematykę. Układ ten składa się ze sprężyny akumulującej energię oraz ciała o niewielkiej masie ($m \approx 0$) ślizgającego się po podłożu. Zadanie sprowadza się do wyznaczenia prędkości ciała, na które działa znana zmieniająca się w czasie siła f . Ruch ciała określa równanie równowagi sił

$$kx + T = f ,$$

oraz cierna charakterystyka siły tarcia pokazana na rysunku 1.

W ogólnym przypadku matematyczny opis ciernej charakterystyki układu ma postać dwóch następujących relacji, które były prezentowane w pracy [3]

$$T \in \Theta, v \in -\mathcal{K}(T, \Theta), \quad (1)$$

gdzie

$T \in \mathbb{R}^N$ – wektor uogólnionej siły tarcia;

* Instytut Pojazdów Politechniki Warszawskiej, ul. Narbutta 84, 02-524 Warszawa

** Instytut Matematyczny PAN, ul. Śniadeckich 8, 00-950 Warszawa

$v \in \mathbb{R}^N$ – wektor uogólnionej prędkości;
 $\Theta \subset \mathbb{R}^N$ – wypukły zbiór dopuszczalnych wartości siły tarcia;
 $\mathcal{K}(T, \Theta) \subset \mathbb{R}^N$ – zbiór zdefiniowany następująco

$$\mathcal{K}(T, \Theta) := \left\{ v \in \mathbb{R}^N : v^T(\eta - T) \geq 0, \forall \eta \in \Theta \right\}. \quad (2)$$

Zbiór $-\mathcal{K}(T, \Theta)$ określa prędkości, które mogą wystąpić w układzie wtedy, gdy działa siła tarcia określona wektorem T .

W rozpatrywanym przykładzie opis zbioru Θ ma bardzo prostą postać

$$\Theta := \left\{ T \in \mathbb{R}^1 : -T_{02} \leq T \leq T_{01} \right\}, \quad (3)$$

natomiast wykres zamieszczony na rysunku 1 ilustruje postać związków (1).

W dalszych rozważaniach pokażemy jak z warunku równowagi sił oraz opisu cierniej charakterystyki układu (1) wynika równanie określające uogólnione prędkości układu.

2. Opis układu sprężysto–ciernego

Rozważamy układ mechaniczny, którego konfigurację opisuje wektor współrzędnych uogólnionych $X \in \mathbb{R}^N$, a energię zaakumulowaną w układzie określa wyrażenie

$$E_p(X) := \frac{1}{2} X^T K X, \quad (4)$$

gdzie

$K \in \mathbb{R}^{N \times N}$ – macierz sztywności układu; zakładamy, że jest ona symetryczna i ściśle dodatnio określona.

Uogólnioną siłę tarcia $T \in \mathbb{R}^N$ określamy zbiorem dopuszczalnych wartości $\Theta \subset \mathbb{R}^N$. Przyjmujemy, że zbiór Θ jest wypukły oraz zawiera zero, tzn. $0 \in \Theta$.

Jak już wspomniano cierną charakterystykę układu opisujemy za pomocą relacji podanych we wzorze (1).

3. Zadanie dynamiki

Pełny opis zadania dynamiki rozpatrywanego układu zawiera równanie równowagi sił oraz relacje określające cierną charakterystykę

$$K \cdot X(t) + T(t) = F(t), \quad (5a)$$

$$T(t) \in \Theta, \quad \dot{X}(t) \in -\mathcal{K}(T(t), \Theta), \quad (5b)$$

gdzie oprócz wcześniej przedstawionych symboli występuje

$F(t) \in \mathbb{R}^N$ – funkcja opisująca znany przebieg siły działającej na układ.

W pracy [3] pokazano, że zadanie (5) można sformułować w równoważnej wariacyjnej postaci. W tym celu definiujemy funkcjonal dysypacji

$$D(v, t) := v^T \cdot [KX(t) - F(t)] + \sigma_{\Theta}(v), \quad (6a)$$

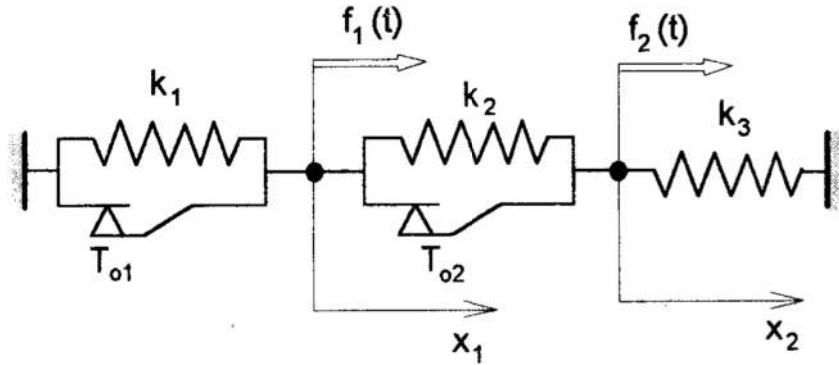
jeśli

$$\sigma_{\Theta}(v) := \sup_{\xi \in \Theta} \xi^T v. \quad (6b)$$

Równoważne sformułowanie zadania (5) w wariacyjnej postaci jest następujące

$$\dot{X}(t) = \arg \min_{v \in \mathbb{R}^N} D(v, t). \quad (7)$$

Podane rozważania zilustrujemy opisem przykładowego układu sprężysto-ciernego.

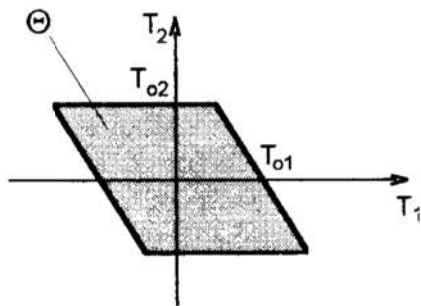


Rys. 2. Schemat układu rozważanego w przykładzie 1

Fig. 2. Schematic diagram of the system considered in Example 1

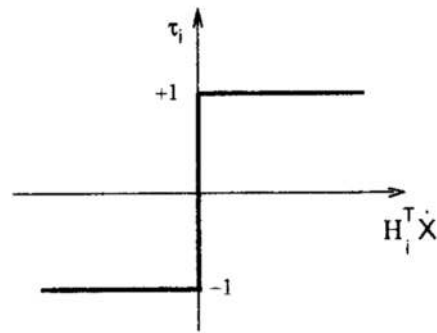
PRZYKŁAD 1. Sformułować zadanie dynamiki układu, którego schemat pokazano na rysunku 2. Na początku opiszemy wektory współrzędnych i sił zewnętrznych oraz wyznaczmy macierz sztywności.

$$(P1) \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad F = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad K := \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_1 + k_3 \end{bmatrix}.$$



Rys. 3. Wykres zbioru Θ opisanego wzorze (P2)

Fig. 3. Plot of the set Θ described by formula (P2)



Rys. 4. Wykres relacji (P5)

Fig. 4. Plot of the relation (P5)

Zbiór dopuszczalnych sił tarcia jest określony następująco

$$(P2) \quad \Theta := \left\{ T \in \mathbb{R}^2 : T = H_1 T_{o1} \tau_1 + H_2 T_{o2} \tau_2 ; \tau_1, \tau_2 \in [-1, +1] \right\}, \quad H_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

gdzie za pomocą τ_1, τ_2 oznaczono mnożniki sił tarcia, natomiast dodatnie liczby T_{01}, T_{02} określają graniczne wartości tych sił w suwakach układu z rysunku 2.

Na rysunku 3 pokazano wykres zbioru Θ wyznaczony według wzoru (P2).

Dla rozważanego układu sformułowanie zadania (5) przybiera postać

$$(P3) \quad KX + T = F,$$

$$(P4) \quad T = H_1 T_{01} \tau_1 + H_2 T_{02} \tau_2, \quad \tau_1, \tau_2 \in [-1, +1],$$

$$(P5) \quad \tau_i \in \partial \left| H_i^T \dot{X} \right|, \quad i = 1, 2,$$

gdzie ∂ oznacza subgraniczkę.

Wykres relacji (P5) zamieszczono na rysunku 4. Na podstawie powyższych warunków należy ustalić wartości czterech niewiadomych: mnożników $\tau_1, \tau_2 \in [-1, +1]$ oraz prędkości $\dot{x}_1, \dot{x}_2 \in \mathbb{R}^1$. Przytoczone wyżej warunki umożliwiają jednoznaczne wyznaczenie niewiadomych tylko wtedy, gdy rozwiązanie ma postać:

$$(P6) \quad \tau_1, \tau_2 \in (-1, +1) \text{ oraz } \dot{x}_1 = \dot{x}_2 = 0;$$

w pozostałych przypadkach z podanego opisu zadania nie można wyznaczyć jednoznacznego rozwiązania.

Opisana w przykładzie 1 niejednoznaczność rozwiązania zadania dynamiki wynika stąd, iż funkcjonal dyssypacji dla układu sprężysto-cierne nie jest ściśle wypukły, to znaczy nie spełnia warunku (patrz [1])

$$D(\lambda v^{(1)} + (1-\lambda)v^{(2)}) < \lambda \cdot D(v^{(1)}) + (1-\lambda) \cdot D(v^{(2)}), \quad \forall v^{(1)}, v^{(2)} \in \mathbb{R}^N, \quad \forall \lambda \in [0, 1],$$

jeśli $D: \mathbb{R}^N \rightarrow \mathbb{R}^1$ funkcja dyssypacji według wzoru (6a).

Zatem w ogólnym przypadku na podstawie opisu zadania (5) nie można uzyskać jednoznacznego rozwiązania względem $\dot{X}(t) \in \mathbb{R}^N$. Jednoznaczne rozwiązanie uzyskuje się tylko wtedy, gdy jego postać jest następująca [2]

$$T(t) \in \text{Int } \Theta, \quad \dot{X}(t) = 0, \quad (8)$$

gdzie $\text{Int } \Theta$ oznacza wnętrze zbioru Θ .

Zbiór niejednoznacznych rozwiązań zadania (5) opisują następujące relacje [2]

$$T(t) \in \text{Fr } \Theta, \quad \dot{X}(t) \in -\mathcal{K}(T, \Theta) \quad (9)$$

jeśli $\text{Fr } \Theta$ oznacza brzeg zbioru Θ .

Aby wyznaczyć jednoznaczne rozwiązanie sformułujemy dodatkowe warunki umożliwiające wybór rozwiązania ze zbioru opisanego we wzorze (9).

4. Uzupełniony opis zadania dynamiki

Podstawę do sformułowania uzupełniających warunków stanowi następstwo różniczkowe równania równowagi sił (5a) oraz następstwo różniczkowe relacji opisanej wzorem (5b); następstwa te analizowano w pracy [2]. Stąd otrzymujemy dodatkowe warunki pozwalające jednoznacznie wyznaczyć prędkości układu

$$K\dot{X} + \dot{T} = \dot{F}, \quad (10a)$$

$$\dot{T} \in \mathcal{D}\Theta(T), \quad (10b)$$

$$\dot{X}(t) \in -\mathcal{K}(\dot{T}, \mathcal{D}\Theta(T)), \quad (10c)$$

jeśli opis różniczkowego następstwa zbioru $\Theta \subset \mathbb{R}^N$ ma postać

$$\mathcal{D}\Theta(T) := \left\{ \dot{T} \in \mathbb{R}^N : \liminf_{h \rightarrow 0^+} \text{dist} \left(\dot{T}, \frac{\Theta - T}{h} \right) = 0 \right\}, \quad (11)$$

natomiast dist jest funkcją odległości punktu od zbioru

$$\text{dist}(T, \Theta) := \inf_{\xi \in \Theta} \|\xi - T\|_2. \quad (12)$$

Zbiór $\mathcal{D}\Theta(T)$ wyznacza dopuszczalne prędkości zmian siły tarcia $\dot{T} \in \mathbb{R}^N$

Problem opisany we wzorach (10) jest wygodniej przedstawić w zmienionej postaci

$$\begin{cases} K^{-1}\dot{T} = K^{-1}\dot{F} + \rho, & (\rho := -\dot{X}), \\ \dot{T} \in \mathcal{D}\Theta(T), \quad \rho \in \mathcal{K}(\dot{T}, \mathcal{D}\Theta(T)). \end{cases} \quad (13)$$

W pracy [3] pokazano, że powyższe warunki są równoważne następującemu wariacyjnemu sformułowaniu tego zadania

$$\dot{T} = \arg \min_{\xi \in \mathcal{D}\Theta(T)} \left\{ \frac{1}{2} \zeta^T K^{-1} \zeta - \zeta^T K^{-1} \dot{F} \right\}. \quad (14)$$

W przytoczonym wyżej wariacyjnym sformułowaniu minimalizowany funkcjonal jest ściśle wypukły, gdyż macierz K jest z założenia dodatnio określona, a ponadto zbiór $\mathcal{D}\Theta$ jest wypukły. Wymienione cechy gwarantują jednoznaczne rozwiązanie zadania (13) względem \dot{T} oraz $\rho = -\dot{X}$.

5. Pełny opis zadania dynamiki układu sprężysto-cierne

Na pełny opis zadania dynamiki składają się:

– warunki równowagi i opis zbioru sił tarcia

$$KX + T = F, \quad T \in \Theta, \quad (15a)$$

przy czym powyższe relacje powinny być spełnione w każdej chwili;

– zestaw relacji określających prędkość zmiany siły \dot{T} oraz współrzędnych \dot{X}

$$\begin{cases} K^{-1}\dot{T} = K^{-1}\dot{F} + \rho, \\ \dot{T} \in \mathcal{D}\Theta(T), \rho \in \mathcal{K}(\dot{T}, \mathcal{D}\Theta(T)), \end{cases} \quad (15b)$$

$$\dot{X} = -\rho. \quad (15c)$$

Jeżeli skorzystamy z wariacyjnego opisu zadania (14), to zestaw warunków (15) można przedstawić również w wariacyjnej postaci

$$\begin{cases} KX + T = F, \quad T \in \Theta, \end{cases} \quad (16a)$$

$$\begin{cases} \dot{T} = \arg \min_{\xi \in \mathcal{D}\Theta(T)} \left\{ \frac{1}{2} \zeta^T K^{-1} \zeta - \zeta^T K^{-1} \dot{F} \right\}, \end{cases} \quad (16b)$$

$$\begin{cases} \dot{X} = K^{-1}(\dot{F} - \dot{T}). \end{cases} \quad (16c)$$

Podane wyżej sformułowania zadania dynamiki układu sprężysto-ciernego umożliwiają wyznaczenie funkcji X , określającej zmiany współrzędnych wywołane działaniem siły F .

Metodę wyznaczania rozwiązania zadania (15) określa następujący algorytm. Na początku sprawdzamy czy wektory $X^0 := X(0)$ $T^0 := T(0)$ opisujące początkowy stan układu spełniają warunki przytoczone we wzorze (15a). Następnie na podstawie relacji ze wzoru (15b) lub (16b) ustalamy wartość prędkości zmiany siły tarcia \dot{T} , a prędkość uogólnioną \dot{X} wyznaczamy ze wzoru (15c). Poszukiwany przebieg funkcji X oraz T uzyskujemy stosując numeryczną metodę wyznaczania rozwiązania rozważanego wyżej problemu różniczkowego.

Podstawowym zadaniem w opisanym wyżej algorytmie jest wyznaczenie wartości prędkości zmiany siły tarcia \dot{T} . W tym celu należy ustalić postać zbioru $\mathcal{D}\Theta(T)$, oraz rozwiązać problem (15b) lub (16b). Proponowany sposób rozwiązania tego zadania przedstawiono na przykładzie układu, którego schemat zamieszczono na rysunku 2. Dla uproszczenia opisu układu przyjmiemy, że $k_1 = k_2 = k_3 = k$. Rozważymy układ w dwóch sytuacjach, w których stan układu jest następujący $X_1 = X_2 = 0$, $T_2 = 0$, $F_2 = 0$, $\dot{F}_1 > 0$, $\dot{F}_2 = 0$; a ponadto zakładamy, że w pierwszej sytuacji powstałej w chwili t_1 działają siły $F_1 = T_1 = \frac{1}{2} T_{01}$, natomiast w drugim przypadku w chwili t_2 siły te wynoszą $F_1 = T_1 = T_{01}$. W pierwszej opisanej chwili wartość siły $T(t_1) \in R^2$ jest taka, że wektor $T(t_1)$ należy do wnętrza zbioru Θ pokazanego na rysunku 3; wtedy zgodnie z warunkiem (8) mamy $\dot{X}(t_1) = 0$, a na podstawie równania (10a) ustalamy $\dot{T}(t_1) = \dot{F}(t_1)$. Uzyskany rezultat można otrzymać również ze wzorów (15b), gdyż w opisanej sytuacji mamy $\mathcal{D}\Theta(T(t_1)) = R^2$ oraz $\rho = 0$.

Rozpatrzmy teraz drugą sytuację, w której siła $T(t_2) \in \mathbb{R}^2$ należy do brzegu zbioru Θ ; siłę tarcia w rozważanej chwili opisuje wektor $T = \begin{bmatrix} T_{01} \\ 0 \end{bmatrix}$. Dla tej siły tarcia opis zbioru $\mathcal{D}\Theta(T)$ ma postać

$$\mathcal{D}\Theta(T) := \{ \dot{T} \in \mathbb{R}^2 : \dot{T}_1 + \dot{T}_2 \leq 0 \} = \{ \dot{T} \in \mathbb{R}^2 : G^T \dot{T} \geq 0 \}, \quad G := \begin{bmatrix} -1 \\ -1 \end{bmatrix}. \quad (17)$$

Rozwiązując zadanie (15b) lub zadanie optymalizacyjne (16b) wyznaczamy $\dot{T} \in \mathbb{R}^2$. W pracy [3] pokazano, że dla uzyskanego opisu zbioru $\mathcal{D}\Theta$ zadanie (15b) można przedstawić następująco: wyznaczyć $\lambda \in \mathbb{R}^1$ i $\dot{T} \in \mathbb{R}^2$ spełniające warunki

$$\begin{cases} K^{-1} \dot{T} = K^{-1} \dot{F} + G \lambda, & (18a) \\ \lambda = [\lambda - \alpha G^T \dot{T}]^+, \quad [\xi]^+ := \begin{cases} \xi, & \text{gdy } \xi > 0 \\ 0, & \text{gdy } \xi < 0 \end{cases} & (18b) \end{cases}$$

jeśli α jest dowolną liczbą dodatnią.

Z zestawu równań (18) po wyeliminowaniu zmiennej \dot{T} otrzymujemy równanie względem mnożnika $\lambda \in \mathbb{R}^1$

$$\lambda = [\lambda - \alpha (G^T \dot{F} + G^T K G \lambda)]^+. \quad (19)$$

Stąd po uwzględnieniu opisu macierzy $K \in \mathbb{R}^{2 \times 2}$ oraz wektora $G \in \mathbb{R}^2$ i przyjęciu

$$\alpha := \frac{1}{G^T K G} = \frac{1}{2k} > 0$$

otrzymujemy wzór na mnożnik λ

$$\lambda = \left[\frac{\dot{F}_1 + \dot{F}_2}{2k} \right]^+.$$

Po ustaleniu wartości mnożnika λ bez trudu obliczamy wektor \dot{T} oraz prędkości $\dot{X} := -G \lambda$. W rozpatrywanej chwili t_2 dla $\dot{F}_1 > 0$, $\dot{F}_2 = 0$ otrzymujemy

$$\dot{T} = \frac{1}{2} \dot{F}_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \dot{X} = \frac{\dot{F}_1}{2k} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Przytoczone wyżej wyniki można bez trudu zinterpretować na rysunku układu (rys.2).

6. Zakończenie

Rozważany układ sprężysto-cierny służy do analizy ruchu układów mechanicznych, których inercyjne cechy mogą być pominięte. Układy takie zwane reologicznymi są wykorzystywane do modelowania struktur sprężysto-dyssypacyjnych stosowanych w różnego rodzaju urządzeniach amortyzujących. Za jego pomocą mogą też być

modelowane właściwości sprężysto-dyssypacyjne wielu rzeczywistych materiałów. Przedstawiony tu matematyczny opis dynamiki posłużył do opracowania komputerowej metody wyznaczania odkształceń układu sprężysto-ciernego, powodowanych działaniem zmiennego obciążenia.

Literatura

1. Duvaut G., Lions J.L.: *Les inéquations en Mécanique et Physique*. Dunod, Paris 1972.
2. Grzesikiewicz W.: *Dynamika układów mechanicznych z więzami*. WPW, Warszawa 1990.
3. Grzesikiewicz W., Wakulicz A.: *Nieładkie zadanie w mechanice*. Zeszyty Naukowe Wydziału Mechanicznego Nr 26, Politechnika Koszalińska 1999.
4. Panagiotopoulos P.D.: *Inequality Problems in Mechanics and Applications*. Birkhäuser, Stuttgart 1985.

Streszczenie

Rozważamy układ mechaniczny, w którym jest akumulowana energia w postaci potencjalnej, a jej rozpraszanie powodują siły tarcia suchego. Zasadniczym zadaniem rozpatrywanego zagadnienia jest wyznaczenie uogólnionych prędkości układu. W tym celu są formułowane warunki równowagi układu i konstytutywny opis tarcia suchego oraz wynikające z tych opisów następstwa różniczkowe. Pokazujemy, że sformułowanie rozpatrywanego zadania można przedstawić w postaci wariacyjnej. Rozważania są ilustrowane przykładem.

DYNAMICS OF ELASTIC SYSTEM WITH DRY FRICTION

Summary

We consider a mechanical system in which energy is accumulated in the potential form and forces of dry friction cause dissipation. The main problem for such a system is to determine its generalised velocities. To this end the conditions of equilibrium and a constitutive description of dry friction are formulated. Then, the differential successions resulting from that description are derived. We show that the formulation of the problem in question can be presented in a variational form. The mathematical description of the system is illustrated with an example.

Referat recenzowany

DOPASOWANIE PRĘDKOŚCI OBROTOWEJ WRZECIONA W CELU PRZECIWDZIAŁANIA DRGANIOM NARZĘDZIE-PRZEDMIOT

1 Wstęp

Poszukiwanie sposobów kontrolowanego przebiegu zjawisk dynamicznych obiektów mechanicznych stanowiło przedmiot wielu badań w odniesieniu do obrabiarek skrawających do metali oraz realizowanych na nich procesach obróbkowych [2], [6], [7]. Jako zjawisko dynamiczne o istotnym znaczeniu uznano drgania względne narzędzie-przedmiot. Przy określonych warunkach ich rozwój może doprowadzić do utraty stabilności oraz do powstania drgań samowzbudnych typu *chatter* [11], których zasadniczymi przyczynami są: sprzężenia wewnętrzne w masowo-sprężysto-dyssypacyjnym (MSD) układzie obrabiarka-uchwyt-przedmiot obrabiany-narzędzie skrawające (OUPN) oraz efekt regeneracji śladu [9], [13].

Problem zapobiegania powstawaniu drgań *chatter* próbowano rozwiązać za pomocą: regulacji prędkości obrotowej wrzeciona z dopasowaniem do częstotliwości drgań własnych układu [12], sterowania pulsacją prędkości obrotowej [1], [3] regulacji prędkości obrotowej wrzeciona *on-line* z dopasowaniem do optymalnej wartości kąta przesunięcia fazowego między dwoma kolejnymi przejściami narzędzia [8] oraz ustawicznego zwiększania prędkości skrawania [10].

2 Nadzorowanie drgań narzędzie-przedmiot w warunkach stacjonarnych

Jeżeli w procesie skrawania:

- bierze udział stała liczba ostrzy (tzw. skrawanie pełne);
- przyjmujemy założenie o równomiernej podziałce ostrzy narzędzia;
- prędkość obrotowa wrzeciona $n=n_0=\text{const}$;
- zmienne w czasie położenie ostrzy zastępujemy ich położeniem uśrednionym;
- pominiemy wpływ ruchu posuwowego przedmiotu na dynamikę układu,

to otrzymamy wówczas model stacjonarny procesu frezowania przy wymuszeniu okresowym siłą skrawania. Stanowi on wygodne uproszczenie modelu niestacjonarnego, gdyż umożliwia poszukiwanie optymalnej prędkości obrotowej wrzeciona o stałej wartości, przeciwdziałającej drganiom typu *chatter*.

Rozważmy równanie dynamiki układu stacjonarnego, które przyjmuje postać [4]:

$$\left\{ \mathbf{M}s^2 + \mathbf{L}s + \mathbf{K} - \left[\mathbf{D}_p(s) + s\mathbf{D}_v(s) + \mathbf{D}_o(s)e^{-sT} \right] \right\} \mathbf{q}(s) = \mathbf{f}(s), \quad (1)$$

gdzie:

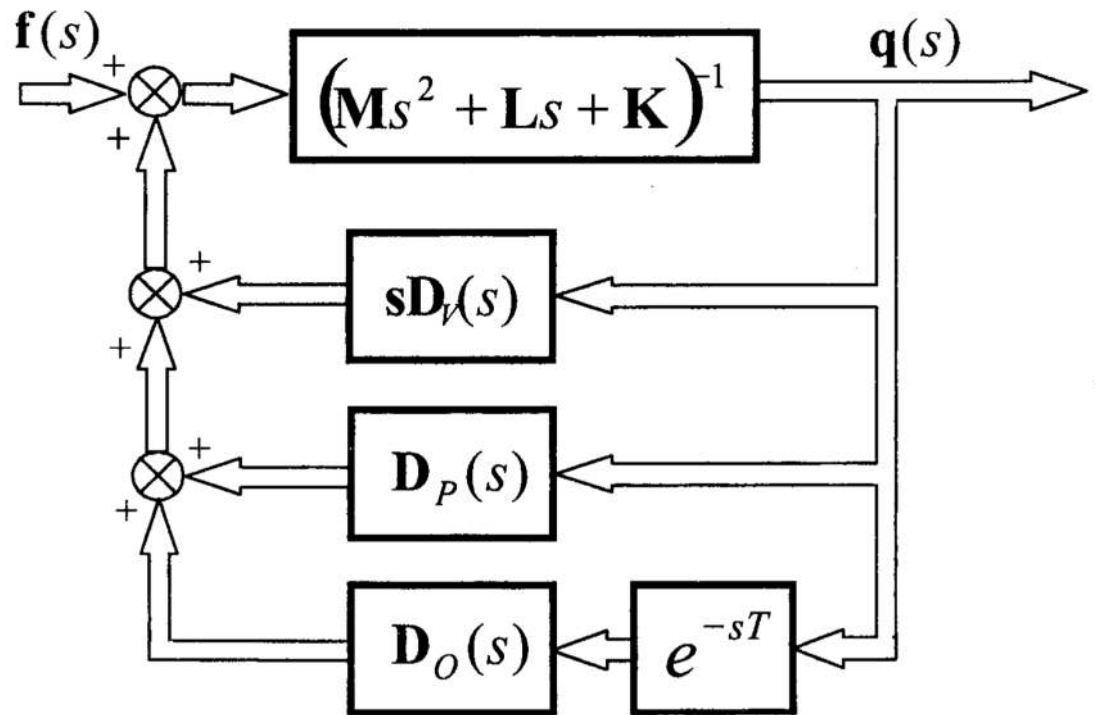
- $\mathbf{D}_p(s)$ – macierz sprzężeń zwrotnych przez przemieszczenie,
- $\mathbf{D}_v(s)$ – macierz sprzężeń zwrotnych przez prędkość,
- $\mathbf{D}_o(s)$ – macierz opóźniających sprzężeń zwrotnych,
- s – operator zespolony.

* Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk, e-mail: kkalinski@sunrise.pg.gda.pl

** Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk, e-mail: tkuchars@sunrise.pg.gda.pl

*** Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk, e-mail: ssawiak@sunrise.pg.gda.pl

Zależność (1) zilustrowano schematem blokowym (rys. 1).



Rys. 1. Schemat blokowy modelu stacjonarnego procesu frezowania czółowego
Fig. 1. Block-diagram of stationary model of the face milling process

Występujące we wzorze (1) macierze przyjmują postać [4]:

$$\mathbf{D}_p(s) \equiv \mathbf{D}_p = - \sum_{l=1}^{i_j} \mathbf{T}_l^T \mathbf{D}_{pl} \mathbf{T}_l, \quad (2)$$

$$\mathbf{D}_v(s) \equiv \mathbf{0}, \quad (3)$$

$$\mathbf{D}_o(s) \equiv \mathbf{D}_o = \sum_{l=1}^{i_j} \mathbf{T}_l^T \mathbf{D}_{ol} \mathbf{T}_l, \quad (4)$$

zaś wektor sił uogólnionych:

$$\mathbf{f}(s) = \sum_{l=1}^{i_j} \mathbf{T}_l^T \mathbf{F}_l^0(s). \quad (5)$$

Transformacja widmowa równania (1) daje zależność:

$$\sum_{\alpha=0}^{M-1} \left(-\mathbf{M}\omega_{\alpha}^2 + \mathbf{L}j\omega_{\alpha} + \mathbf{K} - \mathbf{D}_p - \mathbf{D}_o e^{-j\omega_{\alpha} \frac{60}{zn}} \right) \mathbf{q}(j\omega_{\alpha}) = \sum_{\alpha=0}^{M-1} \mathbf{f}(j\omega_{\alpha}). \quad (6)$$

Wektor $\mathbf{q}(j\omega_{\alpha})$ jest zespolonym wektorem przemieszczeń uogólnionych odpowiadających częstości ω_{α} , moduł jego składowej $q_i(j\omega_{\alpha})$ – amplitudą drgań w kierunku współrzędnej uogólnionej i , zaś argument – kątem przesunięcia fazowego.

Rozważmy drgania okresowe z udziałem składowych harmonicznnych o częstościach kołowych $\omega_0, \omega_1, \omega_2, \dots, \omega_{M-1}$. Poszukujemy takiej wartości prędkości obrotowej n , która zminimalizuje amplitudy drgań odpowiadające poszczególnym częstościom. Wprowadzamy funkcję celu:

$$F(n) = \sum_{\alpha=0}^{M-1} (\mathbf{q}^{0\alpha})^T \boldsymbol{\rho}_{\alpha} \mathbf{q}^{0\alpha} \quad (7)$$

przy ograniczeniu:

$$n_{min} \leq n \leq n_{max}, \quad (8)$$

gdzie:

$\boldsymbol{\rho}_{\alpha}$ – macierz współczynników wagowych,

$$\mathbf{q}^{0\alpha} = \text{col}(|q_i^{0\alpha}(j\omega_{\alpha})|),$$

co stanowi sformułowanie problemu nadzorowania drgań w dziedzinie częstotliwości. Równania (7) i (8) definiują ponadto zagadnienie optymalizacji statycznej z ograniczeniami nierównościami, które rozwiązuje się stosując metodę bezgradientową *Hooke'a-Jeevesa* w połączeniu z funkcją kary *Heaviside'a*.

Dopasowanie prędkości obrotowej wrzeciona do optymalnego kąta przesunięcia fazowego – podejście uogólnione

Jeżeli podczas procesu skrawania dominują drgania *chatter* o częstości $\omega_{\alpha} = 2\pi f_{\alpha}$, to ciąg optymalnych prędkości obrotowych wyznacza się z zależności [4]:

$$\frac{zn_{\alpha}}{60} = \frac{f_{\alpha}}{0.25 + k}, \quad (9)$$

gdzie:

- f_{α} – zaobserwowana częstotliwość drgań *chatter*,
- n_{α} – optymalna prędkość obrotowa, odpowiadająca drganiom o częstotliwości f_{α} ,
- z – liczba ostrzy,
- k – całkowita liczba dodatnia ($k=0, 1, \dots$).

Rezultat powyższy jest uogólnieniem warunku Liao-Young'a [8] w odniesieniu do układów o skończonej liczbie stopni swobody i jest zgodny z wynikami badań charakterystyk częstotliwościowych *Nyquista* w celu analizy wibrostaibilności [13]. Jednakże, umożliwia on dobór optymalnej prędkości obrotowej z uwzględnieniem wpływu wyłączenie jednego rezonansu *chatter*. Zastosowanie opisanego sposobu wymaga zatem

wyodrębnienia rezonansu *chatter* o częstotliwości f_α , którego amplituda dominuje w widmie.

Dopasowanie prędkości obrotowej wrzeciona – podejście zmodyfikowane

W trakcie stacjonarnego procesu skrawania obserwuje się ustalone drgania okresowe, których składowe harmoniczne mają częstotliwości ω_α , $\alpha=1, \dots, M-1$. Wyznaczono pracę sił skrawania wszystkich „aktywnych” ostrzy narzędzia, wykonaną w przedziale czasu odpowiadającym pełnej liczbie okresów drgań o częstotliwościach ω_α , $\alpha=1, \dots, M-1$. Czas ten wyznaczamy z zależności:

$$t_{max} = LCM\left(\frac{2\pi}{\omega_\alpha}\right), \quad \alpha = 1, \dots, M-1, \quad (10)$$

gdzie:

$LCM(.)$ – najmniejsza wspólna wielokrotność wyrażenia $(.)$.

Praca wykonana w czasie t_{max} przyjmie zatem postać [4]:

$$L = \sum_{\alpha=0}^{M-1} (\mathbf{q}_{k-T_i}^{0\alpha})^T \left[\underbrace{\sum_{l=1}^{i_l} \mathbf{T}_l^T \mathbf{D}_{Ol}^T \mathbf{T}_l}_{\mathbf{D}_O^T} \pi \sin\left(2\pi f_\alpha \frac{60}{zn}\right) \cdot t_{max} \right] (\mathbf{q}_k^{0\alpha}). \quad (11)$$

Problem dopasowania optymalnej prędkości obrotowej wrzeciona polega na poszukiwaniu minimum wyrażenia (11) przy ograniczeniu (8). Nie można w tym przypadku sformułować tak prostego warunku, jakim jest wzór (9). Jednakże, przyjmując oznaczenia:

$$\mathbf{e}_\alpha \equiv - \sum_{l=1}^{i_l} \underbrace{\mathbf{T}_l^T \mathbf{D}_{Ol}^T \mathbf{T}_l}_{\mathbf{D}_O^T} \pi \sin\left(2\pi f_\alpha \frac{60}{zn}\right) \cdot t_{max},$$

$$\mathbf{q}^{0\alpha} \equiv \mathbf{q}_k^{0\alpha} \equiv \mathbf{q}_{k-T_i}^{0\alpha},$$

zagadnienie sprowadzamy do omówionego problemu poszukiwania optymalnej prędkości obrotowej wrzeciona w warunkach stacjonarnych (wzory (7) i (8)).

Opisane podejście zmodyfikowane, w odróżnieniu od podejścia uogólnionego, uwzględnia jednocześnie większą liczbę rezonansów drgań *chatter* w widmie oraz drgania obserwowane w różnych kierunkach.

3 Analiza możliwości wdrożenia strategii na rzeczywistym obiekcie drgającym

Opisane strategie dopasowania optymalnej prędkości obrotowej wrzeciona obrabiarki można zrealizować:

- w sposób manualny za pomocą standardowego układu sterowania obrabiarki,
 - jako sterowanie w układzie zamkniętym (tryb *on-line*).
- W obu przypadkach proponuje się następującą procedurę nadzorowania drgań:
- obserwacja drgań za pomocą pomiaru drgań poprzecznych wirującego narzędzia we wzajemnie prostopadłych kierunkach;
 - analiza widmowa oraz identyfikacja rezonansów o częstotliwościach f_α ;

- dobór prędkości obrotowej n_α zgodnie z warunkiem (9) w przypadku podejścia uogólnionego, albo
- optymalizacja statyczna (dobór prędkości obrotowej n) wyrażenia (11) w przypadku podejścia zmodyfikowanego;
- symulacja komputerowa modelu obliczeniowego, albo realizacja procesu skrawania na rzeczywistej obrabiarce.

Strategia wykazuje zalety w porównaniu z dotychczasowym sformułowaniem *Liao-Younga* [8], gdyż:

- warunek *Liao-Younga*, wyprowadzony dla układu o jednym stopniu swobody, uwzględnia jedynie dynamiczną zmianę grubości warstwy skrawanej. Jego uogólnienie umożliwiło rozszerzenie zakresu rozważań w odniesieniu do układów o skończonej liczbie stopni swobody oraz analizę bardziej skomplikowanych przypadków dynamiki skrawania;
- podstawą wdrożenia opisanych strategii jest precyzyjny pomiar przemieszczeń narzędzia w dwóch wzajemnie prostopadłych kierunkach, za pomocą czujników zbliżeniowych. *Liao* i *Young* proponowali pomiar sił skrawania z wykorzystaniem kosztownych siłomierzy, zazwyczaj instalowanych na stole. Jest to sposób niepraktyczny w odniesieniu do nowoczesnych obrabiarek (np. VMC FADAL), w przypadku których drgania stołu są prawie niezauważalne. Ponadto, małe przekroje warstwy skrawanej (rzędu $0.03 - 0.05 \text{ mm}^2$) generują siły skrawania o niskim poziomie, co wymaga stosowania czulszej oraz bardziej precyzyjnej aparatury pomiarowej;
- warunek *Liao-Younga* dotyczył układu dynamicznego obrabiarki o stałej w czasie konfiguracji.

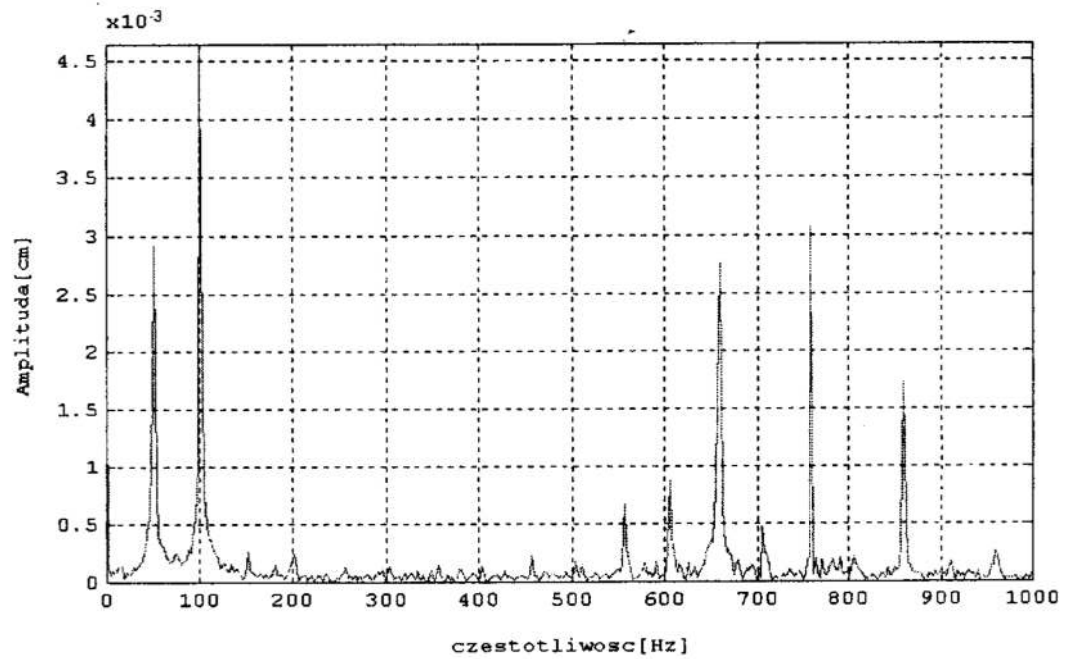
4 Badania doświadczalne

Opracowano i wdrożono metodykę pomiaru [5], konkurencyjną w odniesieniu do stosowanych dotychczas technik pomiarowych z wykorzystaniem siłomierzy instalowanych na stole frezerskim czy na imaku nożowym, akcelerometrów oraz czujników optycznych. Jej istota polega na pomiarze bezstykowym czujnikami zbliżeniowymi drgań poprzecznych (przemieszczeń) wirującego narzędzia (freza trzpieniowego) w dwóch wzajemnie prostopadłych kierunkach: prostopadłym do posuwu oraz równoległym do posuwu. Wykonano stanowisko badawcze oraz uruchomiono w języku MATLAB pakiet programów niezbędnych do obsługi pomiarów oraz analizy wyników. Dzięki temu uzyskano możliwość:

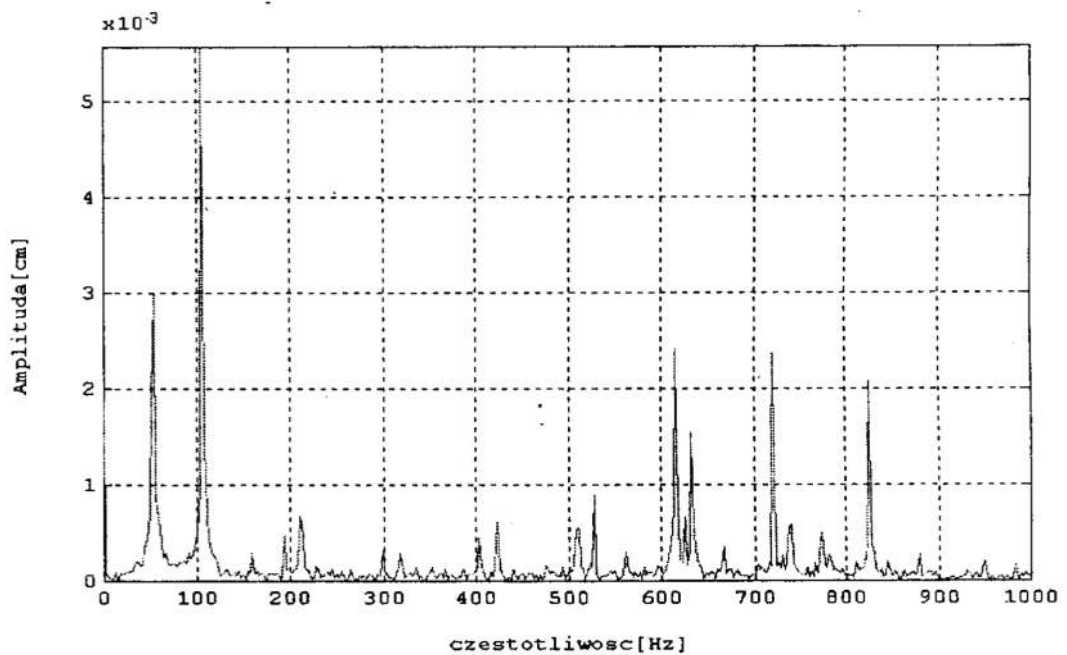
- badania własności dynamicznych napędu głównego obrabiarki przy skokowej zmianie zadanych prędkości obrotowych;
- pomiaru prędkości obrotowej wrzeciona, spełniającej wymagania eksperymentu, z dokładnością znacznie przewyższającą możliwości oferowanych tachometrów;
- analizy widmowej przebiegów czasowych drgań poprzecznych narzędzia w procesie frezowania czołowego w celu określenia parametrów drgań *chatter*;
- doboru optymalnych prędkości obrotowych, przeciwdziałających zaobserwowanym drganiom *chatter*, na podstawie warunków wynikających z zastosowanej strategii nadzorowania drgań.

W celu przeprowadzenia badań wykorzystano następującą aparaturę i oprogramowanie:

- 2 systemy pomiarowe typu 3300 firmy Bently Nevada. W skład każdego z nich wchodzi: sonda wiroprowadowa (zbliżeniowa) – czułość: $7.87 \text{ mV}/\mu\text{m}$ oraz proksymitor;
- specjalny uchwyt mocujący sondy, instalowany na tulei wrzecionowej;
- system kondycjonowania sygnałów (wzmacniacz 4-kanalowy z filtrami);
- komputerowy system pomiaru drgań z kartą pomiarową 500 kHz;



Rys. 2. Widmo amplitudowe drgań narzędzia w kierunku posuwu podczas obróbki brązu BA1032 188 HB. Nominalna prędkość obrotowa $n_0=3000$ obr/min
 Fig. 2. Amplitude spectrum of tool vibration along feed direction during machining of bronze BA1032 188 HB. Nominal spindle speed $n_0=3000$ rev/min



Rys. 3. Widmo amplitudowe drgań narzędzia w kierunku posuwu podczas obróbki brązu BA1032 188 HB. Dopasowana prędkość obrotowa $n_0=3145$ obr/min
 Fig. 3. Amplitude spectrum of tool vibration along feed direction during machining of bronze BA1032 188 HB. Matched spindle speed $n_0=3145$ rev/min

- komputerowy system analizy danych wraz z pakietem autorskich programów obliczeniowych, przygotowanych w języku MATLAB.

W przypadku stosowania niewielkich wzmocnień (znikomy wpływ zakłóceń pomiarowych) błąd pomiaru przemieszczeń wyniósł 0.5 – 1.0 % [5].

Należy zauważyć, że wspomniane wcześniej próby pomiarowe sił skrawania nie znalazły zastosowania w opisywanych w pracy eksperymentach, gdyż:

- z uwagi na małe w badanych przypadkach przekroje warstwy skrawanej, poziom sił skrawania jest niewielki, co wymagałoby stosowania siłomierzy o dużej czułości;
- z uwagi na przewidywane częstotliwości drgań *chatter* siłomierze powinny mieć pasmo przenoszenia do ok. kilkudziesięciu kHz;
- badania siłomierzowe są bardzo kosztowne. Przykładowo: cena średniej klasy siłomierza trójskładowego firmy Kistler wynosi ok. 100.000 PLN. Koszt aparatury pomiarowej (bez komputerów i oprogramowania) wykorzystanej w pracy jest znacznie niższy;
- konsekwencją obecności siłomierza w układzie obrabiarki jest dodatkowe podatne ogniwo, które zniekształca własności dynamiczne badanego obiektu.

Próba pomiarowa: Frezowanie rowka w próbce z brązu BA1032 188 HB o wymiarach 200×60×60 frezem trzpieniowym typu NOMA 206.016W–W w warunkach eksploatacyjnych na nowoczesnym centrum frezerskim VMC FADAL. Nominalna prędkość obrotowa wrzeciona: $n_0=3000$ obr/min, głębokość frezowania $a_p=0.3$ mm, posuw na ostrze $f_z=0.15$ mm. W trakcie badań wyznaczano przebiegi czasowe drgań z częstotliwością próbkowania 10 kHz, a następnie – widma amplitudowe w paśmie od 0 do 1000 Hz.

Na podstawie obserwacji amplitudowego widma drgań narzędzia w kierunku posuwu (rys. 2) stwierdzono pojawienie się drgań *chatter*, czemu towarzyszyły rezonanse o częstotliwościach $f_{c1}=659.18$ Hz, $f_{c2}=759.28$ Hz i $f_{c3}=859.38$ Hz, oraz amplitudach odpowiednio: $q_1^0=0.028$ mm, $q_2^0=0.031$ mm i $q_3^0=0.017$ mm. Z uwagi na istnienie kilku rezonansów o zbliżonych amplitudach, prędkość optymalną wyznaczono stosując jej dopasowanie według podejścia zmodyfikowanego. W rezultacie, skrawanie z wyznaczoną prędkością $n=3145$ obr/min doprowadziło do zmian wartości amplitud rezonansów w widmie (rys. 3) do wartości $q_1^0=0.024$ mm, $q_2^0=0.024$ mm i $q_3^0=0.021$ mm przy nowych wartościach częstotliwości: $f_{c1}=615.23$ Hz, $f_{c2}=720.21$ Hz i $f_{c3}=825.20$ Hz.

Eksperyment przeprowadzony na maszynie w trakcie procesu obróbkowego potwierdził rezultaty wcześniejszych symulacji komputerowych [4], [5] w zakresie możliwości proponowanych w pracy strategii dopasowania prędkości obrotowej wrzeciona w celu przeciwdziałania drganiom narzędzie–przedmiot.

Literatura

1. Altintas Y., Chan P.K.: *In-Process Detection and Suppression of Chatter in Milling*. Int. J. Mach. Tools Manufact. 32, 3, 329–347, 1992.
2. Choudhury S.K., Goudimenko N.N., Kudinov V.A.: *On-line control of machine tool vibration in turning*. Int. J. Mach. Tools Manufact. 37, 6, 801–811, 1997.
3. Jemielniak K., Widota A.: *Suppression of self-excited vibration by the spindle speed variation method*. Int. J. Mach. Tool Des. Res. 24, 3, 207–214, 1984.
4. Kaliński K.: *On one method of the tool-workpiece vibration control during cutting process*. Postępy Technologii Maszyn i Urządzeń, 23, 3, 17–42, 1999.

5. Kaliński K., Kucharski T., Sawiak S.: *Wdrażanie strategii nadzorowania drgań narzędzie–przedmiot z dopasowaniem prędkości obrotowej wrzeciona obrabiarki do optymalnego kąta przesunięcia fazowego*. Mat: Konferencja Naukowo–Techniczna MECHANIKA '99 „Nauka i Praktyka” Gdańsk 25–26 listopada 1999, Tom 2, 126–131, Gdańsk 1999.
6. Kawaji S., Tanaka Y., Matsunaga N., Sasaoka T.: *Suppression of drilling chatter by thrust force control*. Proc.: The Fourth International Conference on Motion and Vibration Control, MOVIC'98, August 25–28 1998, ETH Zurich, Switzerland, Volume 1, 115–120.
7. Kopač J., Pogačnik M.: *Theory and practice of achieving quality surface in turn milling*. Int. J. Mach. Tools Manuf. 37, 5, 709–715, 1997.
8. Liao Y.S., Young Y.C.: *A new on–line spindle speed regulation strategy for chatter control*. Int. J. Mach. Tools Manuf. 36, 5, 651–660, 1996.
9. Marchelek K.: *Dynamika obrabiarek*. Wyd. 2. WNT, Warszawa 1991.
10. Schulz H.: *State of the art of high–speed machining*. Proc.: 1st French and German Conference on HIGH SPEED MACHINING, Metz–France 17–18 June 1997, 1–7.
11. Shi H.M., Tobias S.A.: *Theory of finite amplitude machine tool instability*. Int. J. Mach. Tool Des. Res. 24, 1, 45–69, 1984.
12. Smith S., Tlustý J.: *Update on High–Speed Milling Dynamics*. Trans. of the ASME Journal Engineering for Industry 112, May 1990, 142–149.
13. Tomków J.: *Wibrostanilność obrabiarek*. WNT, Warszawa 1997.

Streszczenie

Praca poświęcona jest nadzorowaniu drgań narzędzie – przedmiot podczas procesów obróbkowych w kolejnych przejściach ostrzy narzędzia. Opracowano zmodyfikowaną strategię dopasowania prędkości obrotowej wrzeciona z wykorzystaniem metod optymalizacji statycznej w warunkach stacjonarnych. Przeprowadzono badania doświadczalne proponowanej strategii nadzorowania na nowoczesnym centrum frezerskim VMC FADAL. Ich rezultaty potwierdziły przewidywania wcześniejszych symulacji komputerowych.

MATCHING THE SPINDLE SPEED IN ORDER TO COUNTERACT THE TOOL–WORKPIECE VIBRATION

Summary

The paper is devoted to the tool–workpiece vibration surveillance during machining in subsequent passes of tool edges. A modified strategy of matching spindle speed using static optimisation methods in steady conditions is developed. An experimental investigation of the surveillance strategy is performed on contemporary milling centre VMC FADAL. The obtained results confirmed prediction of former computer simulations.

Referat recenzowany

CHARAKTERYSTYKI ZASTĘPCZE WAŁECZKÓW PRZY OBLICZANIU NOŚNOŚCI ŁOŻYSK WIĘNCOWYCH METODĄ ELEMENTÓW SKOŃCZONYCH

1 Wstęp

Do obliczania nośności statycznej łożysk tocznych wieńcowych coraz częściej stosuje się metodę elementów skończonych [1], [2], [3]. Pozwala ona na zbudowanie modeli łożysk, dzięki którym możliwe jest uwzględnienie w obliczeniach deformacji pierścieni łożyskowych oraz całej konstrukcji maszyny roboczej, czego nie zapewniają analityczne metody obliczania nośności łożysk oparte na założeniu nieodkształcalności pierścieni łożyskowych. Związane jest z tym zagadnienie modelowania części tocznych. Z uwagi na duży stopień komplikacji modelu i uniknięcie zbędnego modelowania wielu węzłów stykowych, dobre wyniki daje zastąpienie części tocznych elementami prętowymi. Pociąga to za sobą konieczność wyznaczenia odpowiednich charakterystyk tych elementów, tak by zapewniały one właściwy stan deformacji łożyska przy obciążeniu go zadaniem obciążeniem. Przeprowadzane były już próby opracowania modeli części tocznych zarówno dla kulek [4] jak i wałeczków [5]. Jednak doświadczenia autora nabyte w czasie modelowania łożysk wieńcowych wskazują na potrzebę rewizji uprzednio zaproponowanego modelu i sposobu symulacji części tocznych przy obliczaniu nośności łożysk wieńcowych za pomocą metody elementów skończonych.

W poniższym opracowaniu przedstawiono analizę strefy styku wałeczka z bieżnią łożyska wieńcowego oraz nowy sposób określania charakterystyki elementu modelującego wałeczek.

2 Analityczne obliczenia wałeczków łożyskowych

Modelowanie rozkładu obciążenia wewnętrznego w łożyskach tocznych wieńcowych przy analitycznym obliczaniu nośności łożysk zakłada ciągły rozkład obciążenia wzdłuż obwodu łożyska [6]. W uproszczonym opisie zagadnienia kontaktowego przyjmuje się związek pomiędzy obciążeniem części tocznej a wywołaną przezeń deformacją (zbliżeniem pierścieni łożyska) w postaci:

$$\eta = cP^u \quad (1)$$

gdzie:

u – wykładnik potęgowy, c – stała zależna od wymiarów wałeczka i rodzaju materiałów stykających się elementów.

Według [7] odkształcenie wałeczka, a tym samym wzajemne zbliżenie bieżni łożyska określa zależność:

$$\eta = \frac{4(1-\nu^2)P}{\pi E} \frac{1}{l} \left[\frac{1}{3} + \ln \frac{d}{\sqrt{\frac{4(1-\nu^2)P}{E} \frac{1}{l}}} \right] \quad (2)$$

* Politechnika Częstochowska, Instytut Mechaniki i Podstaw Konstrukcji Maszyn,
ul. Dąbrowskiego 73, 42-200 Częstochowa, E-mail: ludwik@matinf.pcz.czyst.pl

Jeżeli aproksymować ją wzorem (1) to, dla wałeczka o średnicy $d = 20$ mm i długości czynnej $l = 18$ mm otrzymuje się $c = 7,73968 \times 10^{-6}$, $u = 0,885378$. W dotychczasowych opracowaniach dotyczących analitycznego obliczania nośności łożysk wieńcowych [6] przyjmuje się $u = 0,9$, a np. w pracy [8] $u = 0,95$. Wartość wykładnika u określa stopień nieliniowości charakterystyki, wartość stałej c nie ma w metodzie analitycznej tak dużego znaczenia, gdyż istotna jest tu wzajemna proporcja pomiędzy deformacjami w poszczególnych rzędach łożyska. Wynika stąd, że w zakresie obliczeń analitycznych występują w różnych źródłach pewne rozbieżności co do opisu charakterystyki wałeczka łożyskowego.

Graniczne obciążenie wałeczka P_{\max} oblicza się w oparciu o przyjęte kryterium wytrzymałościowe. W budowie obliczeniowych modeli analitycznych łożyska przyjmuje się, za Palmgrenem [9], kryterium względnych odkształceń plastycznych wałeczka:

$$\frac{\delta_{pl dop}}{d} = 0,0002 \quad (3)$$

co daje maksymalną siłę, jaką można obciążyć wałeczek [10]:

$$P_{dop} = 123,62 f_H^2 dl \quad (4)$$

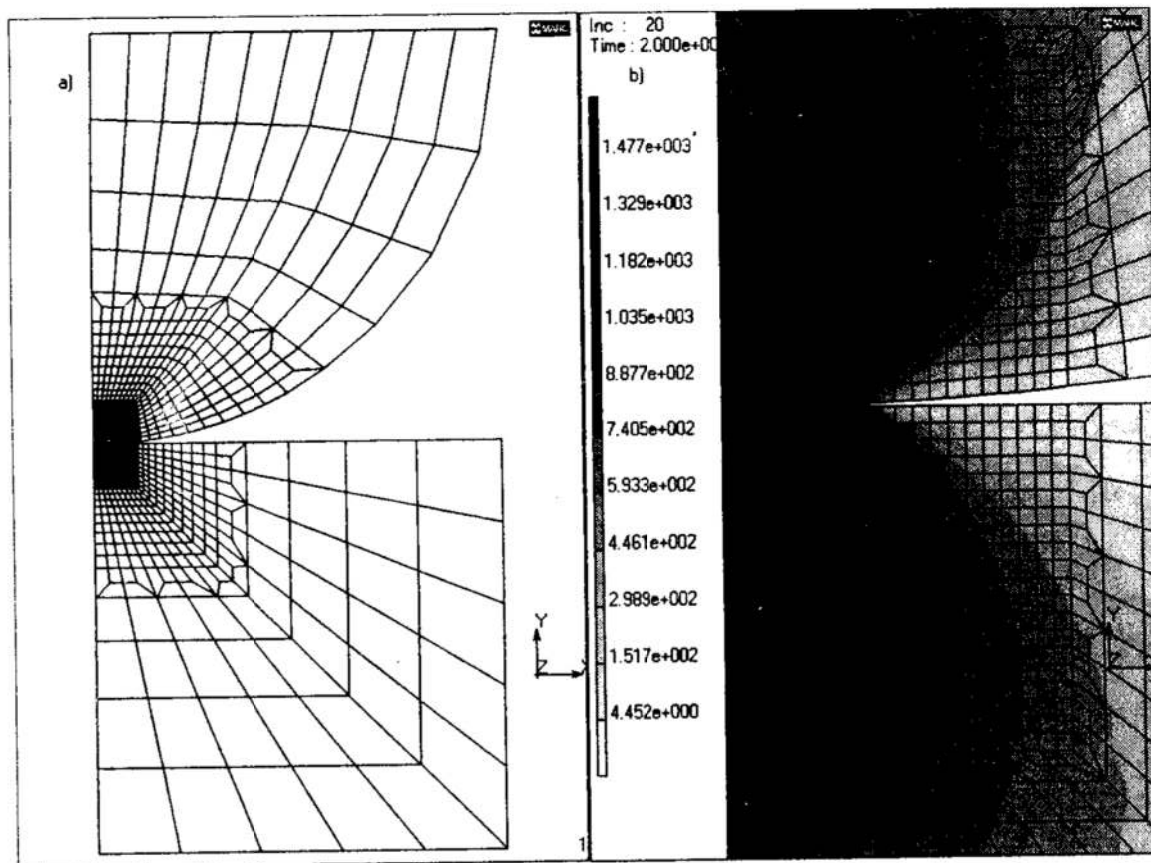
gdzie:

d, l – średnica i długość wałeczka w mm,

f_H – współczynnik twardości określony wzorem: $f_H = (HV/750)^2$

3 Analiza styku wałeczka z bieżnią z wykorzystaniem metody elementów skończonych

Przedstawione wyżej zależności były sformułowane przy założeniu kontaktu wałeczka z płaszczyzną sprężystą bez określania rozmiarów bieżni łożyska. Jest to podstawowa wada modeli analitycznych. Modelowanie strefy styku za pomocą elementów skończonych daje konstruktorowi znacznie większe możliwości niż metody analityczne [11]. Poniżej przedstawiono założenia upraszczające, które uwzględniono przy budowie modelu dyskretnego strefy styku wałeczka z bieżniami łożyska wieńcowego. Przyjęto, że bieżnie łożyska są płaskie (to założenie jest spełnione w łożysku trzyczęściowym, a w łożyskach krzyżowych wpływ krzywizny stożkowych bieżni łożyska na deformacje w strefie styku można pominąć, z uwagi na duże rozmiary średnicy tocznej łożyska w stosunku do średnicy wałeczków [12]), pominięto wpływ podziałki obwodowej wałeczków, z uwagi na specyfikę dyskretyzacji pierścieni łożyska oraz zaniedbano spiętrzenia naprężeń na krańcach wałeczków co pozwoliło przyjąć płaski model strefy styku. To ostatnie założenie wprowadza pewien błąd do charakterystyk strefy styku, jednak przyjęto je z uwagi na lokalny charakter spiętrzenia naprężeń oraz coraz częstsze stosowanie w łożyskach wieńcowych wałeczków korygowanych zmniejszających krawędziowe koncentracje naprężeń. Analiza modeli strefy styku z uwzględnieniem koncentracji naprężeń dla różnych postaci korekcji wałeczków będzie przedstawiona w kolejnej pracy autora. Przyjęto natomiast, że w modelowaniu strefy styku będzie uwzględniana grubość pierścieni łożyskowych. Wynika to z istoty zaproponowanego w niniejszej pracy sposobu definiowania charakterystyk wałeczków, dla którego istotne jest nie tyle odkształcenie samego wałeczka, lecz deformacja całego węzła łożyskowego, rozumiana jako zbliżenie tych powierzchni pierścieni łożyska, które stykają się z powierzchniami osadczymi konstrukcji wsporczej. Podstawową wersję modelu dyskretnego strefy styku wałeczka z bieżnią łożyska wieńcowego pokazano na rysunku 1. Model dotyczy wałeczka o średnicy $d = 20$ mm i długości czynnej $l = 18$ mm. Przy obciążeniu go siłą $P_{\max} = 35,204$ kN zbliżenie pierścieni obliczone z wykorzystaniem wzoru (4) wyniesie $\eta = 0,0823$ mm.



Rys. 1. Model strefy styku wałeczka z bieżnią: a) siatka modelu, b) stan naprężeń zastępczych w strefie styku

Fig. 1. Contact zone model of roller and roller bearing race: a) mesh of a model; b) equivalent stress distribution in the contact zone

Przyjęto, że obydwie bieżnie łożyska są jednakowe, co pozwoliło zastosować symetrię względem średnicy wałeczka równoległej do osi x i w połączeniu z warunkiem symetrii zadania względem osi y poddać dyskretyzacji $1/4$ przekroju poprzecznego wałeczka.

Model zbudowano przy użyciu programu MARC [13], do budowy siatki modelu użyto 4-węzłowych elementów o następujących funkcjach kształtu [14]:

$$u_x = a_0 + a_1x + a_2y + a_3xy \quad u_y = b_0 + b_1x + b_2y + b_3xy \quad (5)$$

Z uwagi na lokalny charakter bezpośredniej strefy styku zastosowano nierównomierną podziałkę siatki modelu. W bezpośredniej strefie styku zarówno w wałeczku, jak i bieżni wyodrębniono fragmenty modelu o rozmiarach 1×1 mm i podzielono je na 400 elementów każdy, następnie stopniowo zwiększano rozmiary elementów do uzyskania siatki jak na rysunku 1. Na węzły siatki nałożono następujące warunki brzegowe: węzły w osi symetrii y modelu mogą tylko przemieszczać się wzdłuż osi y , wynika to z warunku symetrii, podobnym ograniczeniom podlegają węzły na krawędzi bieżni równoległej do osi symetrii y . Węzły dolnej krawędzi bieżni mają odebrane wszystkie stopnie swobody, dodatkowo zastosowano związanie wszystkich węzłów na krawędzi średnicowej wałeczka równoległej do osi x , tak by ich przemieszczenia w kierunku y były jednakowe, w celu zapewnienia warunku symetrii wynikającego z zastosowania jednakowych bieżni łożyska.

Zagadnienie kontaktowe potraktowano jako styk dwóch ciał odkształcalnych. W wałeczku i bieżni wyodrębniono pewną liczbę elementów stanowiących strefę styku. Dodatkowo w centralnym punkcie styku wprowadzono pomocnicze ciało nieodkształcalne w postaci odcinka prostej równoległej do osi y , z uwagi na wymagania programu MARC (zastępuje ono odebranie stopnia swobody x węzłowi w centralnym punkcie styku). Ogólna liczba elementów wynosi 1420. Do krawędzi średnicowej wałeczka przyłożono obciążenie ciągłe o wartości wynikającej z maksymalnej siły obciążającej wałeczek i jego rozmiarów. Zadanie zdefiniowano jako przypadek płaskiego stanu odkształcenia i rozwiązano dla 20 kroków obciążenia.

Model z rysunku 1 nazwano A1, na jego wzór zbudowano drugi model A2 o dwukrotnie większej grubości bieżni. Dla tych modeli obliczono charakterystyki deformacji pierścieni i wałeczka. W tabeli 1 podano wyniki obliczeń w postaci wartości wykładników potęgowych u oraz stałych c uzyskanych przy aproksymacji uzyskanych charakterystyk funkcją (1) a także maksymalne zbliżenie pierścieni dla siły $P = 36$ kN. Wartości wykładnika u świadczą o stopniu nieliniowości modelu.

Tabela 1. Wyniki obliczeń modeli strefy styku

Table 1. Calculation results of a contact zone models

Model	η_{\max} [mm]	węzeł łożyskowy		wałeczek		bieżnia	
		$c \times 10^6$	u	$c \times 10^6$	u	$c \times 10^6$	u
A1	0.07795	7.0234	0.8884	3.5087	0.8921	3.5189	0.8846
A2	0.08500	7.0306	0.8965	3.5085	0.8921	3.5275	0.9006

Otrzymane wyniki wskazują, że dla analizowanych modeli charakterystyka wałeczka praktycznie nie ulega zmianie, różnice w wartościach zbliżenia pierścieni wynikają wyłącznie z odmiennego stanu deformacji bieżni, co uzasadnia przyjęte założenie o uwzględnianiu grubości bieżni.

4 Określanie charakterystyki elementu prętowego

Z uwagi na to, że pierścienie łożyskowe są zbudowane z bryłowych elementów 3D, w węzłach łączących siatkę pierścieni z elementami symulującymi części toczne dochodzi do lokalnych zakłóceń stanu deformacji pierścieni. Dlatego bardzo ważne jest aby element prętowy zastępujący wałeczek lub kulkę pozwalał na możliwie wierne odwzorowanie stanu przemieszczeń zachodzących w miejscach styku części tocznych z bieżniami łożyska, powinien on zapewniać:

- uzyskanie odpowiedniej siły odpowiadającej sile docisku bieżni do wałeczka,
- zapewnienie nieliniowego charakteru deformacji w strefie styku,
- uzyskanie zbliżenia pierścieni o tej samej wartości co dla rzeczywistego styku wałeczka z bieżniami łożyska.

Powyższe warunki uzyskuje się poprzez odpowiednie określenie charakterystyki wałeczka. Na charakterystykę tę składają się przekrój obliczeniowy i związek pomiędzy siłą w pręcie, a jego deformacją.

Przekrój obliczeniowy oblicza się z zależności, którą zaproponowano w [4]:

$$A_t = \frac{P_{\max} d}{\eta_{\max} E} \quad (6)$$

Całkowite odkształcenie η_c w strefie styku jest sumą odkształcenia wałeczka η_w i bieżni łożyska η_b :

$$\eta_c = \eta_w + \eta_b \quad (7)$$

W pracy [5] przyjęto, że charakterystykę wałeczka określa się na podstawie jego deformacji η_w w modelu testowym. Można zatem określić typową charakterystykę dla wałeczka o określonej średnicy d . W pracy tej przedstawiono również analizę dyskretyzacji siatek pierścieni łożyskowych i wskazano na najlepszy sposób budowy siatki z uwagi na zachowanie wymagań wymienionych na wstępie rozdziału. Doświadczenia uzyskane w czasie modelowania łożysk wieńcowych wskazują jednak, że nie zawsze udaje się zbudować siatkę spełniającą powyższe zalecenia, a co za tym idzie obliczenia mogą być obciążone błędami. W celu uniknięcia tej niedogodności proponuje się inny sposób określania charakterystyki elementu prętowego modelującego wałeczek, a mianowicie:

1. Należy zbudować dyskretny model obliczeniowy strefy styku wałeczka z bieżniami łożyska, np. za pomocą zaproponowanego powyżej sposobu modelowania (model typu A). Model ten powinien w miarę możliwości jak najlepiej odwzorowywać rzeczywisty kształt pierścieni łożyska
2. Wynikiem obliczeń takiego modelu powinna być charakterystyka odkształceń całkowitych w strefie styku.
3. Następnie należy zbudować model strefy styku z wykorzystaniem rzeczywistej siatki, za pomocą której jest zbudowany model pierścieni łożyska i obciążyć element prętowy siłą docisku wałeczków do bieżni (element prętowy może mieć charakterystykę liniową). Model ten nazwano T.
4. Wynikiem obliczeń tego modelu powinna być charakterystyka odkształceń bieżni łożyska bez części odkształceń wynikającej z deformacji pręta.
5. W ostatnim etapie obliczeń należy sporządzić charakterystykę elementu prętowego jako różnicę deformacji całkowitych η_c z modelu strefy styku i deformacji bieżni η_{bT} z modelu pierścieni łożyskowych. Tę charakterystykę nazwano charakterystyką zastępczą, gdyż nie jest ona wynikiem deformacji żadnego z elementów strefy styku.

W typowym przypadku charakterystykę zastępczą elementu prętowego można więc sporządzić obliczając wartości fikcyjnych deformacji η_t elementu prętowego ze wzoru:

$$\eta_t = \eta_c - \eta_{bT} \quad (8)$$

Powyższa zależność jest słuszna, gdy pracę wałeczka w łożysku można traktować jako symetryczną, to znaczy, że obydwie bieżnie łożyska mają w przybliżeniu takie same rozmiary przekroju poprzecznego (przede wszystkim chodzi o grubość). Jeżeli przekroje poprzeczne poszczególnych bieżni są odmiennie podzielone na elementy, lub bieżnie różnią się grubością wówczas punkt 3 i 4 powyżej sformułowanego sposobu postępowania przy określaniu charakterystyki elementu prętowego należy powtórzyć dla każdej z bieżni oddzielnie. Otrzyma się wtedy dwie wartości deformacji bieżni η'_{bT} i η''_{bT} , a deformacje elementu prętowego obliczać się będzie ze wzoru:

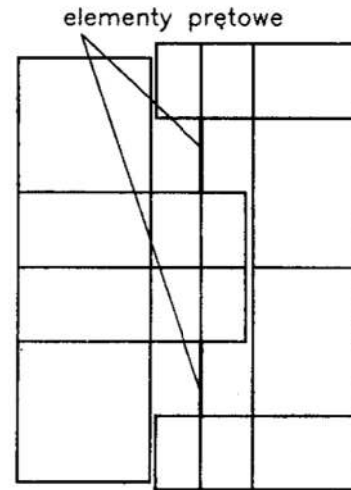
$$\eta_t = \eta_c - \frac{\eta'_{bT} + \eta''_{bT}}{2} \quad (9)$$

Jeżeli ze wzoru (8) lub (9) otrzyma się ujemną wartość deformacji elementu prętowego oznacza to, że siatka pierścieni łożyskowych jest nieprawidłowa (zbyt drobna) i należy zastosować inny sposób dyskretyzacji pierścieni. Obliczenia wynikowe dają zastępczą nieliniową charakterystykę elementu zastępującego wałeczek.

5 Przykład obliczeniowy

Zaproponowaną metodę zastosowano do modelu pierścieni łożyskowych łożyska trzyrzędowego, którego siatkę w przekroju poprzecznym pokazano na rysunku 2. Model strefy styku (T1) z elementem prętowym pokazano na rysunku 3a, jest on fragmentem

siatki pierścienia łożyska z rysunku 2. W rozpatrywanym łożysku trzyczęściowym podziałka między wałeczkami wynosi ok. $1,2 d$, szerokość bieżni ok. $1,1 d$, a grubość bieżni pierścienia zewnętrznego jest równa d . Przekrój obliczeniowy elementu prętowego obliczony ze wzoru (6) wynosi $85,93 \text{ mm}^2$. Model T1 jest wycinkiem stanowiącym $\frac{1}{4}$ podziałki bieżni łożyskowej, zatem do obliczeń przyjęto odpowiednio $\frac{1}{4}$ wymaganego przekroju i $\frac{1}{4}$ obciążenia wałeczka, czyli 9 kN . Do poszczególnych płaszczyzn modelu przyłożono odpowiednie więzy wynikające bądź z symetrii układu, bądź z podparcia bieżni analogicznie jak dla modelu dyskretnego strefy styku omówionego powyżej. Wyniki przeprowadzonych obliczeń pozwoliły na określenie charakterystyk elementów prętowych modelujących wałeczki.



Rys. 2. Siatka przekroju poprzecznego łożyska trzyczęściowego

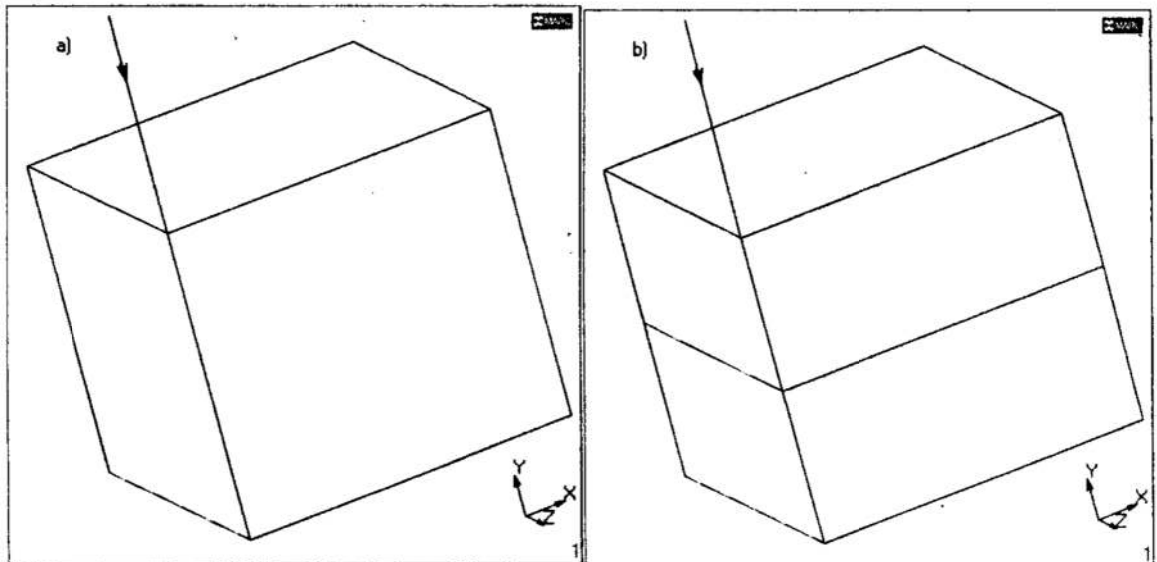
Fig. 2. A cross section mesh of three-row bearing

Dla porównania, jak wpływa na charakterystykę zastępczą wałeczka sposób dyskretyzacji pierścienia łożyskowego przeprowadzono obliczenia dla modelu T2 (rys. 3b).

Na rysunku 4 pokazano charakterystyki całkowitego zbliżenia pierścieni modelu testowego strefy styku (modelu A2), deformacji bieżni modeli zastępczego i charakterystyki zastępcze elementów prętowych dla modelu T1 (rys. 4a) i T2 (rys. 4b). Otrzymane charakterystyki można wyrazić zależnościami:

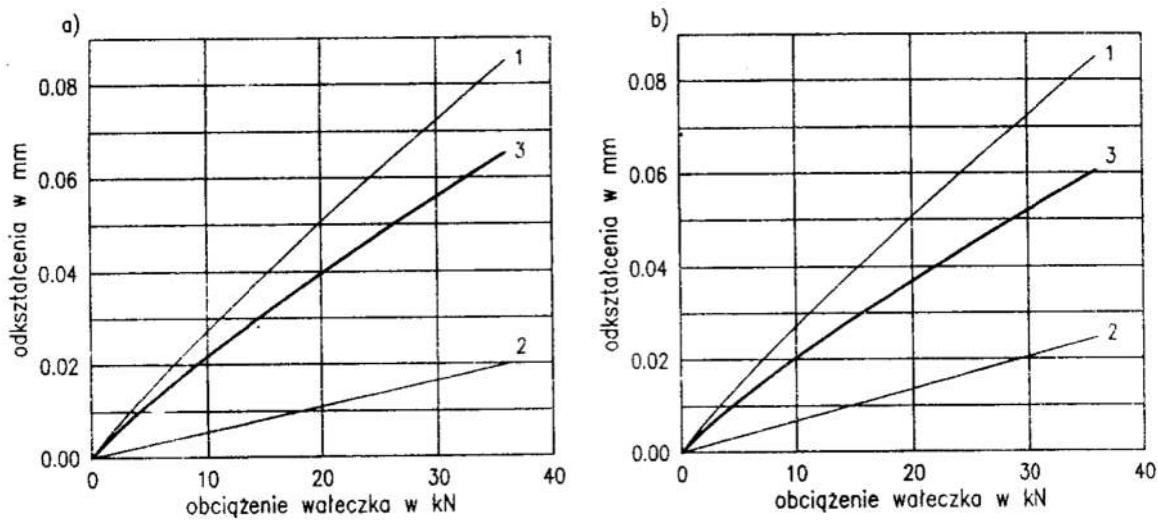
$$\eta = 7.1422 \cdot 10^{-6} P^{0.87006} \quad (\text{dla modelu T1}) \quad (10)$$

$$\eta = 7.2503 \cdot 10^{-6} P^{0.86149} \quad (\text{dla modelu T2}) \quad (11)$$



Rys. 3 Siatka modelu T1 (a) i modelu T2 (b)

Fig. 3. A mesh of model T1 (a) and model T2 (b)



Rys. 4. Charakterystyki modeli T1(a) i T2(b): 1- zbliżenie pierścieni dla modelu A2, 2- deformacja bieżni, 3- charakterystyka zastępcza elementu prętowego
 Fig. 4. Characteristics of model T1(a) and T2(b): 1- mutual displacement of bearing rings, 2- race deformation, 3- equivalent characteristic of truss element

5 Uwagi końcowe

Przeprowadzona analiza deformacji w strefie styku wałeczka z bieżniami łożyska wieńcowego oraz wyniki przykładowych obliczeń pozwalają stwierdzić, że:

1. Zbudowany model dyskretny strefy styku, mimo pewnych uproszczeń, pozwolił na analizę nieliniowości charakterystyki węzła łożyskowego bieżnia – wałeczek – bieżnia. Otrzymane wyniki mogą być również wykorzystane w obliczeniach analitycznych, gdyż źródła literaturowe różnią się w ocenie stopnia nieliniowości w strefie styku wałeczka z bieżniami łożyska.
2. Zaproponowany sposób określania charakterystyki elementu prętowego symulującego oddziaływanie wałeczka pozwala na dostosowanie go do założonej siatki pierścieni łożyskowych, inaczej niż w dotychczasowej metodzie, która określała jedynie ogólne warunki, jakie musi spełnić siatka pierścieni łożyskowych. Oczywiście warunki te nadal są aktualne.
3. Konieczna jest rozszerzona analiza strefy styku umożliwiająca badanie rozkładu deformacji wzdłuż tworzącej wałeczka uwzględniająca koncentrację naprężeń na jego końcach. Zagadnienie to będzie przedmiotem kolejnego opracowania autora.

Literatura

1. Zienkiewicz O.C.: *Metoda elementów skończonych*. Arkady, Warszawa 1972.
2. Bathe K.J.: *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, Inc. Englewood Cliffs 1982.
3. Rusiński E.: *Metoda elementów skończonych. System COSMOS/M*. Wydawnictwa Komunikacji i Łączności, Warszawa 1994.
4. Dziurski A., Kania L., Mazanek E.: *Modelowanie części tocznych w łożyskach wieńcowych za pomocą elementów prętowych*. XI Konferencja „Metody i środki projektowania wspomaganego komputerowo”. Warszawa 1997.
5. Dziurski A., Kania L., Mazanek E.: *Modelowanie wałeczków łożyskowych w łożyskach tocznych wieńcowych za pomocą elementów skończonych*. Zeszyty Naukowe

Wydziału Mechanicznego Politechniki Koszalińskiej nr 23. Wydawnictwo Uczelniane Politechniki Koszalińskiej 1998.

6. A. Dziurski, L. Kania, E. Mazanek: *Charakterystyki łożysk wieńcowych*. Przegląd Mechaniczny, z.20 Vol. 49 (1990).
 7. Orłóś Z.: *Wytrzymałość materiałów*. Poradnik inżyniera mechanika t. I. WNT, Warszawa 1968.
 8. Eschmann P., Hasbargen L., Weigand K.: *Die Wälzlagerpraxis*. Oldenburg Verlag, München 1978.
 9. Palmgren A.: *Łożyska toczne*. PWT, Warszawa 1951.
 10. A. Dziurski, L. Kania, E. Mazanek: *Łożyska wieńcowe krzyżowo-wałeczkowe*. Przegląd Mechaniczny nr 11 Vol. 42 (1983).
 11. Smolnicki T.: *Nieliniowe modele układu bieżnia – kula – bieżnia do wyznaczania rozkładu nacisków w wielkowymiarowym łożysku tocznym*. Przegląd Mechaniczny, z.5-6 Vol.58 (1999).
 12. Dziurski A., Kania L., Mazanek E.: *Modelowanie strefy styku kulki z bieżniami łożysk tocznych wieńcowych*. XVII Sympozjon Podstaw Konstrukcji Maszyn. Lublin-Nałęczów 1995.
 13. *Marc Volume A. Theory and User Information*. Marc ARC, Palo Alto 1997.
 14. *Marc Volume B. Element Library*. Marc ARC, Palo Alto 1997.
- Pracę wykonano w ramach projektu badawczego Nr 7 T07C 001 12 finansowanego przez Komitet Badań Naukowych w latach 1997-1999.

Streszczenie

W łożyskach tocznych wieńcowych, przy obliczaniu ich nośności statycznej za pomocą metody elementów skończonych, części toczne zastępuje się elementami prętowymi. Elementy te powinny mieć odpowiednią charakterystykę, tak by zapewnić niezmienną deformację pierścieni łożyska pod wpływem przyłożonego obciążenia w stosunku do rzeczywistych odkształceń łożyska. W referacie przedstawiono wyniki przeprowadzonych obliczeń modelu strefy styku i na ich podstawie zaproponowano nowy tryb określania zastępczych charakterystyk elementów prętowych uwzględniający sposób dyskretyzacji pierścieni łożyskowych oraz zilustrowano go przykładem obliczeniowym.

EQUIVALENT CHARACTERISTICS OF RIMMED BEARING ROLLERS FOR LARGE DIAMETER BEARINGS CAPACITY CALCULATIONS WITH THE USE OF FEM

Summary

When the carrying capacity of rimmed ring bearings is calculated with the use of FEM rolling elements are replaced by truss elements. The truss elements should have proper characteristic to ensure relation between deformation and loading the same as for the real roller in the bearing. In the paper results of calculation for a contact zone model are shown on the base of them a new mode of designation of equivalent characteristics for truss elements, which takes into account bearing ring discretization. The work is illustrated with calculational example.

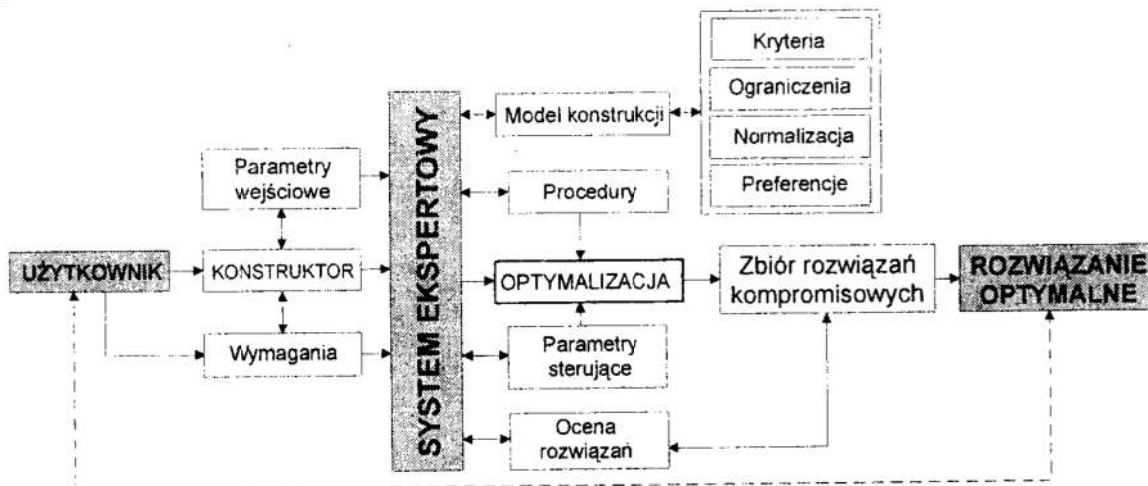
Referat recenzowany

OPTYMALIZACJA WIELOKRYTERIALNA TRÓJWARSTWOWEJ PANELI WALCOWEJ PRZY OBCIĄŻENIACH ZŁOŻONYCH

1 Wstęp

Trójwarstwowe cienkościenne konstrukcje powłokowe charakteryzują się korzystnym stosunkiem masy do przenoszonego obciążenia. Ich wadą jest podatność na utratę stateczności oraz trudności w produkcji i montażu. Walory użytkowe konstrukcji powłokowych można w znaczący sposób zwiększyć poprzez odpowiednie ukształtowanie warstwy wypełniającej. Konstrukcje takie stanowią interesujące zagadnienie dla optymalizacji wielokryterialnej. Ze względu na złożoność modelu teoretycznego opisującego zachowanie się konstrukcji wielowarstwowych oraz złożoność procesu optymalizacji wektorowej celowe wydaje się włączenie w proces projektowania konstrukcji systemu ekspertowego.

W pracy przedstawiono wyniki obliczeń numerycznych dla dwukryterialnej optymalizacji trójwarstwowej paneli walcowej, otrzymanych za pomocą programu komputerowego wykorzystującego bazę wiedzy i mechanizmy wnioskowania systemu ekspertowego. Ogólny schemat programu przedstawiono na rysunku 1. Program znajduje się na etapie szczegółowego opracowywania [3]. Otrzymane wyniki potwierdzają prawidłowość przyjętych założeń. Procedury optymalizacyjne programu pracują z ciągłymi i dyskretnymi zbiorami zmiennych decyzyjnych (metoda Hooke'a-Jeevesa z funkcją kary, przegląd zupełny, symulowane wyzarzanie). Każda z procedur posiada rozbudowany moduł sterowania parametrami. Wybór rozwiązania optymalnego oparty jest o zbiór funkcji preferencji.



Rys. 1. Schemat programu komputerowego wspomaganego systemem ekspertowym
Fig. 1. Diagram of computer program aided by expert system

* Instytut Techniki, Wyższa Szkoła Pedagogiczna w Zielonej Górze, ul. Wojska Polskiego 69, 65-762 Zielona Góra, E-mail: renik@asia.aw.wsp.zgora.pl

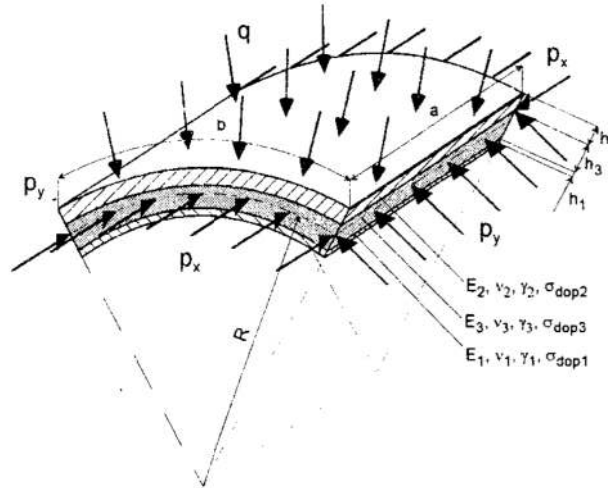
** Instytut Mechaniki Stosowanej, Politechnika Poznańska, ul. Piotrowo 3, 60-965 Poznań, E-mail: Marian.Ostwald@put.poznan.pl

Praca jest kontynuacją badań związanych z optymalnym projektowaniem konstrukcji trójwarstwowych. Optymalizację skalarną trójwarstwowej paneli walcowej (z masą jako kryterium) przedstawiono w pracy [4]. W pracach [6] oraz [8] przedstawiono wyniki optymalizacji wielokryterialnej płyt trójwarstwowych.

2 Model dwukryterialnej optymalizacji paneli walcowej

Model otwartej trójwarstwowej paneli walcowej przedstawiono na rysunku 2. Panela składa się z dwóch cienkich warstw nośnych (okładzin) o grubościach h_1 i h_2 , wykonanych z materiału o wysokich własnościach wytrzymałościowych. Pomiędzy okładzinami znajduje się grubsza warstwa wypełniająca (rdzeń) o grubości h_3 , wykonana z materiału o znacznie niższych własnościach wytrzymałościowych.

W modelu paneli założono, że materiał okładzin ma własności sprężyste izotropowe. Rdzeń wykonany jest z materiału o module sprężystości wielokrotnie niższym niż okładziny. Rdzeń jest nieściśliwy w kierunku normalnym do powierzchni środkowej powłoki. Dla konstrukcji trójwarstwowej przyjęto przemieszczeniowy model zgodny z tzw. hipotezą linii łamanej, zakładającą liniowy charakter rozkładu naprężeń ścinających rdzeń wzdłuż jego grubości. Przyjęto również, że odkształcenia paneli są liniowo sprężyste. Brzegi powłoki są swobodnie podparte i mają przepony łączące ze sobą poszczególne warstwy.



Rys. 2. Model trójwarstwowej paneli walcowej
Fig.2. Model of sandwich cylindrical panel

Zadanie optymalizacji dwukryterialnej z wykorzystaniem koncepcji optimum Pareto sformułowano następująco:

$$\tilde{Q}(\bar{x}) = [Q_1(\bar{x}), Q_2(\bar{x})] = (1 - \text{WAGA}) \cdot Q_1(\bar{x}) + \text{WAGA} \cdot Q_2(\bar{x}) \rightarrow \text{minimum}$$

gdzie: $\bar{x} = [h_1, h_2, h_3]$ - zbiór zmiennych decyzyjnych,

$Q_1(\bar{x}), Q_2(\bar{x})$ - kryteria optymalizacyjne,

WAGA - współczynnik wagi kryteriów, $0 \leq \text{WAGA} \leq 1,0$.

Podstawowym warunkiem w optymalnym projektowaniu konstrukcji jest sformułowanie kryteriów optymalizacyjnych, zmiennych decyzyjnych oraz zbioru warunków ograniczających. Dla konstrukcji powłokowych najczęściej stosowanym kryterium jest kryterium minimalnej masy konstrukcji, posiadające pewien sens ekonomiczny. Zgodnie z przyjętymi oznaczeniami (rys. 2), pierwsze kryterium ma postać:

$$Q_1(\bar{x}) = a \cdot b (h_1 \gamma_1 + h_2 \gamma_2 + h_3 \gamma_3) \text{ [kg]} \rightarrow \text{minimum,}$$

gdzie γ_i - gęstość materiału dla i-tej warstwy.

Drugim kryterium jest żądanie zapewnienia konstrukcji maksymalnej sztywności, wyrażającej się minimalnymi ugięciami powłoki. Ze względu na brak wiarygodnych zależności opisujących ugięcia, w pracy przyjęto uproszczony model postępowania, wyrażający się w sformułowaniu warunku minimum podatności konstrukcji na odkształcenia [5]. Dla otwartej paneli walcowej warunek ten zdefiniowany jest w następujący sposób:

$$Q_2(\bar{x}) = \frac{1}{D(\bar{x})} [1/\text{MNm}] \rightarrow \text{minimum},$$

gdzie: $D(\bar{x}) = \frac{E_n h_1 h_2 (h_1 + 2h_3 + h_2)^2}{4(h_1 + h_2)(1 - \nu_n^2)} \left[\frac{1}{\text{MNm}} \right]$ - sztywność powłoki na zginanie [5],

$E_n = E_1 = E_2$, moduły Younga dla okładzin, $\nu_{1,2}$ - liczby Poissona dla okładzin. W podobny sposób przyjęto kryterium optymalizacyjne dla zamkniętych powłok walcowych [5], [7].

Zmiennymi decyzyjnymi są grubości warstw powłoki h_i ($i = 1, 2, 3$). Zbiór warunków ograniczających sformułowano następująco:

1) Dopuszczalne obciążenie krytyczne jest większe od osiowej siły P_x wywołującej ściskanie:

$$P_{\text{kr dop}} = \frac{P_{\text{kryt}}^{\text{lin}}}{\alpha \cdot n} \geq P_x,$$

gdzie górne obciążenie krytyczne $P_{\text{kryt}}^{\text{lin}}$ [MN/m] wyznaczane jest jako najmniejszy dodatni pierwiastek równania algebraicznego [9]:

$$A_1 \cdot \tilde{P}_x^4 + A_2 \cdot \tilde{P}_x^3 + A_3 \cdot \tilde{P}_x^2 + A_4 \cdot \tilde{P}_x + A_5 = 0,$$

$$[\tilde{P}_x] = 1, \quad P_{\text{kryt}}^{\text{lin}} = \frac{B \cdot (h_1 + h_2 + h_3)^2}{a \cdot b} \tilde{P}_x, \quad B = \sum_{i=1}^{i=3} B_i = \sum_{i=1}^{i=3} \frac{E_i h_i}{1 - \nu_i^2},$$

A_1, A_2, A_3, A_4 i A_5 - współczynniki zależne od fizycznych i geometrycznych parametrów paneli [9]. Współczynnik $\alpha = 1,4$ [2] uwzględnia wpływ ugięć początkowych (imperfekcji) na wartość siły krytycznej, $n = 1,25$ jest współczynnikiem bezpieczeństwa.

2) Naprężenia normalne od siły P_x wywołującej ściskanie oraz od ciśnienia q w warstwach nośnych nie przekraczają wartości dopuszczalnych:

$$\sigma_{\text{redi}} = \sqrt{(\sigma_{x_i}(P_x, q))^2 + (\sigma_{y_i}(P_x, q))^2 - (\sigma_{x_i}(P_x, q))(\sigma_{y_i}(P_x, q))} \leq \sigma_{\text{dop}}.$$

3) Ugięcie środka paneli [1] nie może przekraczać wartości określonych przez konstruktora:

$$w = \frac{5}{384} \frac{q \cdot a^4}{D(\bar{x})} \cdot C \leq w_{\text{dop}},$$

gdzie:

$$C = \left(1 + \frac{48 \cdot k}{5\pi^2} \right) - \frac{1536}{5\pi^5 \cdot \text{ch}(\pi/2\lambda)} \left(1 + k + \frac{\pi \cdot \text{th}(\pi/2\lambda)}{4\lambda} \right), \quad k = \frac{E_{1,2}}{(1 - \nu_{1,2}^2)} \frac{\pi^2 \cdot h_1 h_2 h_3}{G \cdot a^2 (h_1 + h_2)}.$$

- 4) Ze względu na ważność wzorów na obliczanie obciążeń krytycznych spełniony musi być warunek [9]:

$$\frac{R}{h_1 + h_2 + h_3} \geq 30.$$

Warunek ten wynika z założeń poczynionych przy rozwiązywaniu zagadnienia stateczności paneli.

- 5) Grubości warstw nośnych spełniają warunki zgodności z normami. Grubość wypełniacza zmienia się z dokładnością do 1 mm. Ze względów technologicznych i konstrukcyjnych przyjęto więc następujące ograniczenia:

$$(a) \quad \begin{array}{l} 0,1 \text{ mm} \leq h_1, h_2 \leq 3,0 \text{ mm}, \\ 1 \text{ mm} \leq h_3 \leq 50 \text{ mm}. \end{array} \quad \text{oraz} \quad (b) \quad \begin{array}{l} 0,1 \text{ mm} \leq h_1, h_2 \leq 4,0 \text{ mm}, \\ 1 \text{ mm} \leq h_3 \leq 60 \text{ mm}. \end{array}$$

W początkowej fazie obliczeń uwzględniono dodatkowo szereg ograniczeń związanych z naprężeniami stycznymi w wypełniaczu oraz z ważnością wzorów na naprężenia i ugięcia (w sumie brano pod uwagę 20 warunków ograniczających). Po analizie wyników, biorąc pod uwagę przyjęty w pracy model paneli z tzw. wypełniaczem lekkim, w pracy uwzględniono 11 warunków ograniczających.

3 Obliczenia numeryczne

Do obliczeń numerycznych przyjęto następujące dane liczbowe:

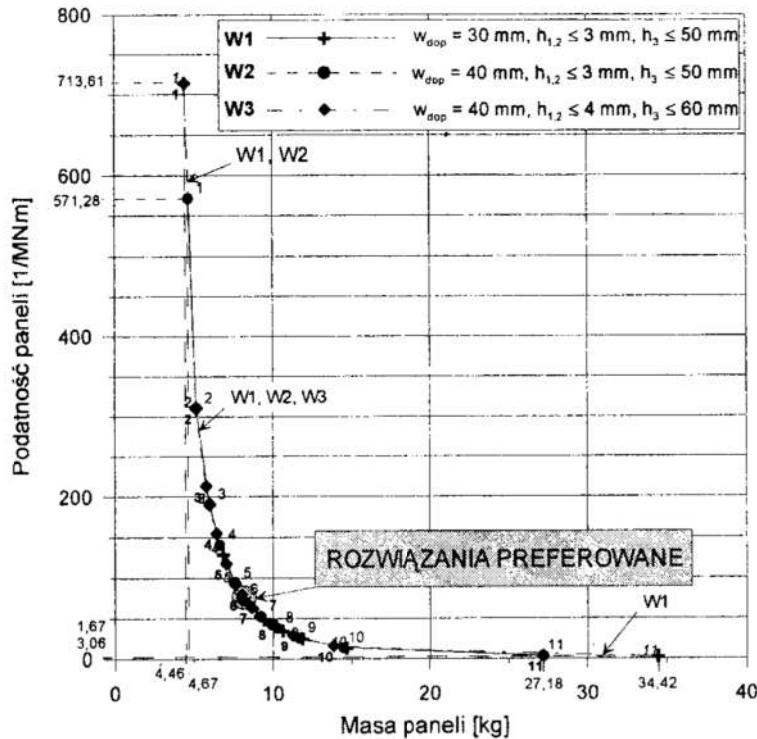
- warstwy nośne wykonane są z duralu PA6, dla którego $E_{1,2}=7,06 \cdot 10^4$ MPa, $\nu_{1,2}=0,3$, $\gamma_{1,2}=2780$ kg/m³, $\sigma_{dop}=0,75 R_e=195$ MPa,
- warstwa wypełniająca wykonana jest z piankowego tworzywa sztucznego, dla którego $E_3=53$ MPa, $\nu_3=0$, $\gamma_3=210$ kg/m³,
- promień krzywizny powierzchni środkowej paneli $R=2000$ mm,
- długość i szerokość paneli $a=b=1,0$ m,
- osiowa siła ściskająca $P_x=0,1$ MN/m,
- ciśnienie zewnętrzne $q=0,01, 0,02, 0,03, 0,04, 0,05$ oraz $0,10$ MPa,
- dopuszczalne ugięcie środka płyty $w_{dop}=40$ mm oraz $w_{dop}=30$ mm.

Do obliczeń przyjęto trzy modele paneli:

- W1 – warunki ograniczające 1 – 5a, $w_{dop} = 40$ mm, (model podstawowy),
- W2 – warunki ograniczające 1 – 5a, $w_{dop} = 30$ mm,
- W3 – warunki ograniczające 1 – 5b, $w_{dop} = 40$ mm.

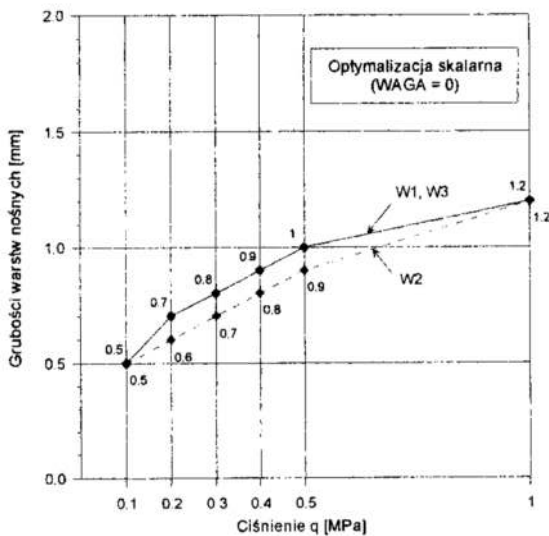
W stosunku do modelu podstawowego W1, model W2 zakłada ostrzejsze warunki sztywnościowe, natomiast model W3 poszerza warunki geometryczne.

Wyniki obliczeń numerycznych w postaci zbiorów rozwiązań Pareto optymalnych dla wszystkich wariantów pokazano na rysunku 3. Punkty nr 1 (WAGA=0) prezentują wyniki optymalizacji skalarnej z masą jako kryterium, punkty nr 11 (WAGA=1) optymalizacji skalarnej z podatnością jako kryterium. Wszystkie punkty oznaczone numerami od 2 do 10 przedstawiają rozwiązania optymalizacji dwukryterialnej (WAGA=0,1÷0,9). Najsilniejsze ograniczenie brzegu Pareto wystąpiło dla modelu W2. Rozwiązania preferowane wyznaczone były za pomocą funkcji dystansowych (z $p = 1, 2$) oraz metody min-max. W większości przypadków rozwiązania preferowane wyznaczone były przez wszystkie trzy funkcje preferencji dla parametru WAGA = 0,5 (punkt nr 6 na rys. 3).



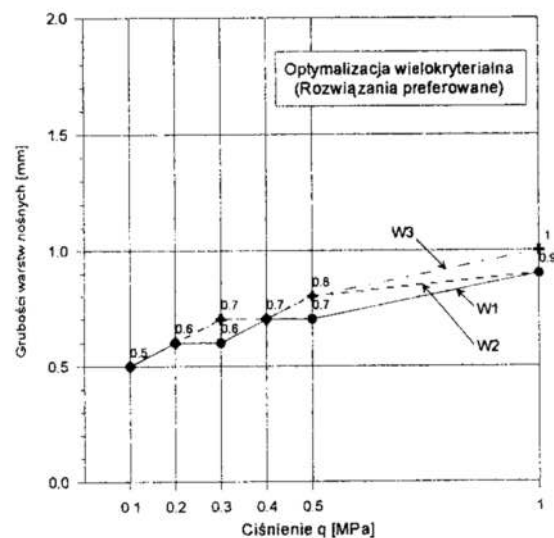
Rys. 3. Zbiory rozwiązań Pareto- optymalnych dla różnych modeli, $P_x=0,1 \text{ MN/m}$, $q=0,01 \text{ MPa}$

Fig. 3. Pareto optimal solutions sets for different models, $P_x=0,1 \text{ MN/m}$, $q=0,01 \text{ MPa}$



Rys. 4. Grubości warstw nośnych, optymalizacja skalarna

Fig. 4. Thicknesses of the layers, scalar optimization



Rys. 5. Grubości warstw nośnych, optymalizacja wielokryterialna

Fig. 5. Thicknesses of the layers, multicriteria optimization

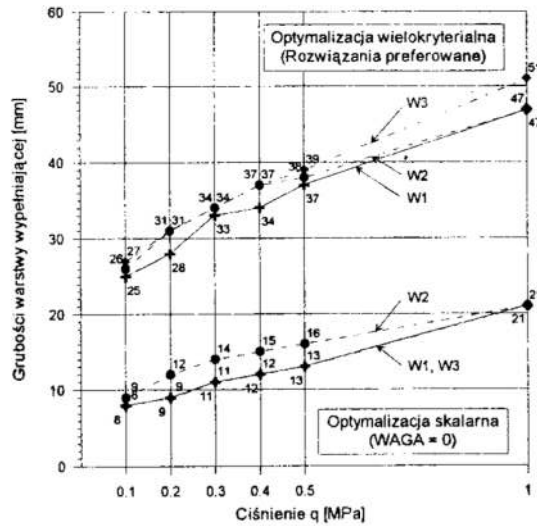
Optymalne grubości warstw nośnych dla różnych modeli paneli poddanych obciążeniom złożonym przedstawiono w tabeli 1. Są to rozwiązania otrzymane z dyskretnych zbiorów zmiennych decyzyjnych za pomocą procedur MESP (przeгляд zupełny) i SW (symulowane wyzarcane) – rozwiązania mają więc znaczenie praktyczne. Dla rozważanego modelu czas obliczeń w procedurze MESP był znacznie krótszy niż dla procedury SW.

W ostatniej kolumnie porównano masy poszczególnych rozwiązań w odniesieniu do modelu W1. Zmiany grubości warstw paneli dla poszczególnych modeli w funkcji ciśnienia q przedstawiono na rysunkach 4, 5 oraz 6.

Aktywnymi warunkami ograniczającymi dla $WAGA = 0, 0,1$ oraz $0,2$ były warunki wytrzymałościowe. W modelu W2 dla $q = 0,01 \div 0,03$ MPa uaktywniło się także ograniczenie na w_{dop} . Ze względu na stosunkowo niską wartość siły ściskającej P_x warunek stateczności w żadnym przypadku nie był aktywny. Dla współczynnika $WAGA = 1$ aktywnymi warunkami były ograniczenia w zakresie parametrów geometrycznych. W modelu W3 aktywny był także warunek ograniczający wartość R/h_i .

Tabela 1. Rozwiązania preferowane dla $P_x = 0,1$ MN/m dla różnych modeli
Table 1. Preferred solutions for $P_x = 0,1$ MN/m for different models

q MPa	Model	Grubości optymalne [mm]			Q ₁	Q ₂	P _{kryt} ^{dop}	σ _{1,2}	$\frac{Q_1^{(W1)}}{Q_1^{(W2)}}$ $\frac{Q_1^{(W1)}}{Q_1^{(W3)}}$
		h ₁	h ₂	h ₃	kg	1/MNm	MN/m	MPa	
0,01	W1	0,5	0,5	25,0	8,0300	79,2896	0,3921	129,12	1,0
	W2	0,5	0,5	26,0	8,2400	73,4184	0,4043	127,88	1,026
	W3	0,5	0,5	27,0	8,4500	68,1760	0,4168	126,73	1,052
0,02	W1	0,6	0,6	28,0	9,2160	52,5271	0,4304	128,20	1,0
	W2	0,6	0,6	31,0	9,8460	43,0270	0,4681	123,63	0,819
	W3	0,6	0,6	31,0	9,8460	43,0270	0,4681	123,63	0,819
0,03	W1	0,6	0,6	33,0	10,2660	38,0572	0,4325	141,08	1,0
	W2	0,7	0,7	34,0	11,0320	30,5851	0,5020	119,39	0,8037
	W3	0,7	0,7	34,0	11,0320	30,5851	0,5020	119,39	0,8037
0,04	W1	0,7	0,7	34,0	11,0320	30,5851	0,4462	136,12	1,0
	W2	0,7	0,7	37,0	11,6620	25,9111	0,4806	130,72	0,8472
	W3	0,7	0,7	37,0	11,6620	25,9111	0,4806	130,72	0,8472
0,05	W1	0,7	0,7	37,0	11,6620	25,9111	0,4326	146,17	1,0
	W2	0,8	0,8	38,0	12,4280	21,4049	0,4896	126,04	0,8261
	W3	0,8	0,8	39,0	12,6380	20,3428	0,5006	124,38	0,7851
0,10	W1	0,9	0,9	47,0	14,8740	12,4840	0,3356	148,40	1,0
	W2	0,9	0,9	47,0	14,8740	12,4840	0,3356	148,40	1,0000
	W3	1,0	1,0	51,0	16,2700	9,5337	0,3738	126,80	0,7639



Rys. 6. Grubość warstwy wypełniającej, optymalizacja skalarna i wielokryterialna
 Fig. 6. Thicknesses of the core, scalar and multicriteria optimization

4 Wnioski

Zastosowanie optymalizacji wielokryterialnej pozwoliło na uzyskanie rozwiązań uwzględniających aspekt ekonomiczny konstrukcji (minimum masy) z aspektem użytkowym (maksymalna odporność na odkształcenia). Rozwiązania są kompromisem pomiędzy tymi sprzecznymi celami. W rozważanym modelu paneli wykorzystano program umożliwiający otrzymanie wyników o praktycznym znaczeniu (zgodnych na przykład z obowiązującymi normami) oraz rozpatrywanie problemu poprzez wyznaczanie aktywnych warunków ograniczających. Analiza wyników pozwala więc na ocenę wpływu poszczególnych ograniczeń na rozwiązania optymalne i stanowić może wytyczne dla konstruktorów projektujących cienkościennie konstrukcje powłokowe.

Analiza otrzymanych rozwiązań pokazuje, że zmiana warunków ograniczających może przyczynić się do poprawienia optymalnych parametrów konstrukcji. Problem zmiany warunków ograniczających jest związany z preferencjami użytkownika oraz rozwiązaniami możliwymi do zrealizowania przez konstruktora i wymaga dalszych badań w tym zakresie.

Praca umożliwiła zweryfikowanie pewnych założeń teoretycznych pozwalających na włączenie do procesu optymalnego projektowania konstrukcji systemu ekspertowego, za pomocą którego zarówno użytkownik jak i konstruktor będzie miał możliwość podejmowania odpowiednio umotywowanych decyzji.

Praca realizowana w ramach grantu Politechniki Poznańskiej, Instytut Mechaniki Stosowanej, nr PB 21-005/2000 BW.

Literatura

1. Aleksandrov A. J., Briukker M. P., Kurszin A. M., Prusakov A. P.: *Pasczet triehslojnych panelej*. Oborongiz, Moskwa 1960.
2. Bushnell D.: *PANDA-2 - program for minimum weight design of stiffened, composite, locally buckled panels*. Computers & Structures, vol. 25, No. 4, pp. 469-605, 1987.
3. Kasperska R., Ostwald M.: *System wspomagający optymalizację wielokryterialną*. WSP Zielona Góra, Studia i Materiały XLVIII, zeszyt 1, s. 19-28, 1999.

4. Ostwald M., Sekulski Z.: *Dobór optymalnych grubości warstw trójwarstwowej paneli walcowej poddanej osiowemu ściskaniu*. Zeszyty Naukowe Politechniki Śląskiej, seria Mechanika, z. 91, 1989, str. 223-228.
5. Ostwald M.: *Optymalne projektowanie trójwarstwowych konstrukcji powłokowych*. Politechnika Poznańska, Rozprawy nr 290, Poznań 1993.
6. Ostwald M.: *Optymalizacja wektorowa płyt trójwarstwowych*. Zeszyty Naukowe Wydziału Mechanicznego, Nr 19, WSI Koszalin, 1995, str. 147-154.
7. Ostwald M.: *Multicriteria optimization of cylindrical sandwich shells under combined loads*. Structural Optimization, vol. 12, no. 2/3, October 1996, pp. 159-166.
8. Ostwald M.: *Multicriterion optimization of sandwich plates*. Proceedings of the Second World Congress of Structural and Multidisciplinary Optimization, May 26-30 1997, Zakopane, Poland, pp. 699-704.
9. Sekulski Z.: *Nieliniowe zagadnienie stateczności trójwarstwowej paneli walcowej przy obciążeniu złożonym*. Archiwum Budowy Maszyn, tom XXXI, z. 1-2, 1984.

Streszczenie

W pracy sformułowano model dwukryterialnej optymalizacji trójwarstwowej cienkościennej paneli walcowej poddanej obciążeniu złożonemu. W modelu przyjęto dwa kryteria optymalizacyjne. Pierwszym kryterium jest masa paneli, drugim jej podatność zdefiniowana jako odwrotność sztywności na zginanie. Podatność jest jakościową miarą odkształcalności paneli. Zmiennymi decyzyjnymi są grubości warstw. Warunkami ograniczającym są: stateczność konstrukcji, warunki wytrzymałościowe, warunki określające ważność stosowanych modeli teoretycznych oraz warunki technologiczno-konstrukcyjne. Zadanie rozwiązano w oparciu o koncepcję optimum Pareto z wykorzystaniem programu komputerowego zawierającego system ekspertowy.

MULTICRITERIA OPTIMIZATION OF SANDWICH CYLINDRICAL PANELS UNDER COMBINED LOADS

Summary

In the work, the bicriterial optimization model of the sandwich, thin-walled cylindrical panel under combined loads is presented. The two objective functions are taken into account. The weight is the first objective, and the flexibility is the second one. The flexibility is defined as the converse of the panel bending rigidity and it represents some qualitative measure of the panel deformability. The thicknesses of the layers are taken into account as the design variables. The constraints include a stability condition, the stress conditions, the importance of the theoretical models and constraints connected with technological and constructional requirements. The optimization problem is solved with the help of the Pareto concept of optimality and by the means of the computer program, which includes elements of the expert system.

Referat recenzowany

ROZWIĄZANIE PROBLEMU TYPU *Picus Viridis* W ŚRODOWISKU MATLAB® SIMULINK®

1 Wprowadzenie

W budowie maszyn szeroko stosowane są siłowniki pneumatyczne (Rys. 1). Zapewniają one liniowy ruch elementu roboczego z prędkościami rzędu 0.5–2, m/s. W przemyśle samochodowym, meblarskim współpracuje ze sobą niekiedy kilkadziesiąt siłowników. Ponieważ napęd pneumatyczny jest napędem energochłonnym istotnym staje się poprawny dobór elementów układu [3], [4].

Racjonalne zastosowanie napędu wymaga oceny jego współpracy z napędzanym mechanizmem. Może to być realizowane na drodze badań eksperymentalnych na obiekcie rzeczywistym (prototypie) lub na drodze badań symulacyjnych (z wykorzystaniem modeli komputerowych). W dobie rozwoju technik komputerowych oczywistym staje się postulat komputerowej optymalizacji konstrukcji [3], [4]. Prowadzenie komputerowych badań, szczególnie badań optymalizacyjnych wymaga szybkich modeli obiektów.



Rys.1 Siłownik pneumatyczny firmy Camozzi

Fig.1 Pneumatic cylinder of Camozzi

W Katedrze Systemów Sterowania Politechniki Koszalińskiej prowadzone są prace nad komputerowym wspomaganie projektowania układów pneumatycznych [3], [4]. Jednym z kierunków tych prac jest poszukiwanie szybkich modeli komputerowych. Problemem, który należy rozwiązać w przypadku budowy modelu dynamiki siłownika pneumatycznego (Rys. 1) jest model dojścia tłoka siłownika do ścianki (pokrywy) i jego postoju na ściance pod działaniem zmiennego obciążenia. Należy zaznaczyć, że celem tego modelowania nie jest analiza dojścia do ścianki i ewentualnego uderzenia w nią, a tylko zapewnienie poprawnego zachowania się modelu komputerowego w procesie symulacji.

Inspiracją niniejszej pracy był referat profesora W. Grzesikiewicza [1] zaprezentowany na XV Konferencji „Poliptymalizacja i CAD” w 1997 roku w Mielnie. Omówiono w nim układ mechaniczny w postaci dziecięcia (*Picus Viridis*), którego ruch był ograniczony (występowały więzy jednostronne) i uwzględniane było tarcie suche. Wskazano na możliwości przyspieszenia obliczeń komputerowych.

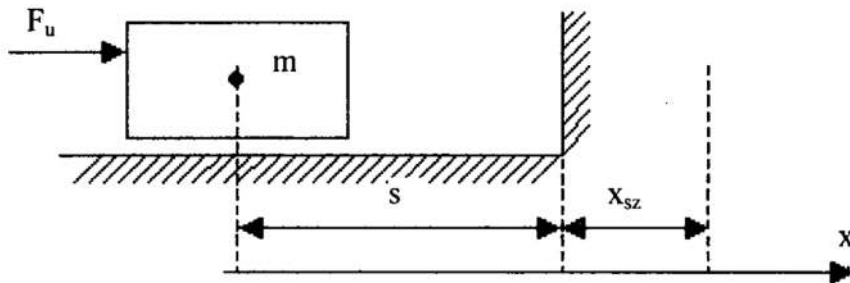
W przypadku siłownika pneumatycznego mamy również do czynienia z ograniczeniem (w postaci ścianki) ruchu. W Katedrze Systemów Sterowania Politechniki Koszalińskiej do symulacji wykorzystuje się środowisko MatLab® Simulink®. Dotychczas stosowany model komputerowy siłownika pneumatycznego wymagał dużych nakładów obliczeniowych dla okresu postoju tłoka na ściance oraz doboru parametrów ścianki dla konkretnego siłownika.

Zdecydowano się zaimplementować przedstawione w [1], [2] rozwiązanie problemu i porównać szybkość działania nowej procedury z uzyskiwanymi dotychczas wynikami.

* Katedra Systemów Sterowania, Politechnika Koszalińska

2 Model matematyczny

Analizie poddano (Rys. 2) element o masie m znajdujący się pod działaniem siły użytecznej F_u zmiennej w czasie. Pominięto wszelkie opory ruchu (tarcie, opory powietrza itp.). Po pokonaniu drogi s (skoku siłownika) element ma się zatrzymać. Dopuszczalne jest zatrzymanie tłoka w zakresie x_{sz} co w przypadku siłownika oznacza tak zwaną objętością szkodliwą. Przekroczenie x_{sz} spowoduje błędy w modelu siłownika.



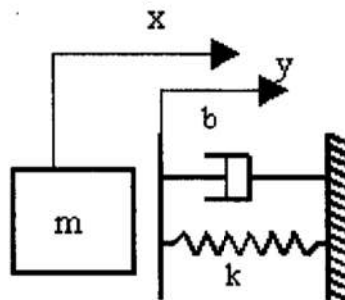
Rys. 2 Schemat analizowanego układu
Fig. 2 Scheme of analysed system

Analizie poddano dwa podejścia:

- Ścianka ma właściwości sprężysto–lepkie (Rys. 3);
- Wyznaczana jest reakcja [1], [2] zatrzymująca poruszający się element wewnątrz ścianki (Rys.4).

Model sprężysto–lepkiej ścianki

Do zatrzymania układu wykorzystano model podatnej ścianki, składającej się ze sprężyny o sztywności k oraz tłumika wiskotycznego o tłumieniu b (Rys.3). Układ taki jest prosty do implementacji komputerowej, wymaga jednak odpowiedniego dobrania wartości parametrów sprężystości k i lepkości b ścianki praktycznie dla każdego siłownika. Wadą takiego podejścia są duże nakłady obliczeniowe szczególnie, w przypadku źle dobranych wartości parametrów ścianki.



Rys. 3 Schemat sprężysto–lepkiej ścianki
Fig. 3 Scheme of viscoelastic wall

Ruch masy m z uwzględnieniem sprężysto–lepkiej ścianki (Rys. 3) opisuje równanie:

$$m \cdot \ddot{x} = F_u - R \quad (1)$$

gdzie:

F_u – siła wymuszająca,

R – reakcja ścianki,

$$R \in \begin{cases} \{0\}, & \text{gdy } -x + s > 0, \text{ nie ma styku} \\ R^+, & \text{gdy } -x + s \leq 0, \text{ jest styk} \end{cases} \quad (2)$$

$$R^+ \in \begin{cases} \{0\}, & \text{gdy } -\dot{x} > 0, \text{ masa } m \text{ odchodzi od styku} \\ b \cdot \dot{y} + k \cdot y, & \text{gdy } -\dot{x} \leq 0, \text{ masa } m \text{ trwa w styku} \end{cases} \quad (3)$$

x – droga,

$$y = (x - s), \dot{y} = \dot{x}$$

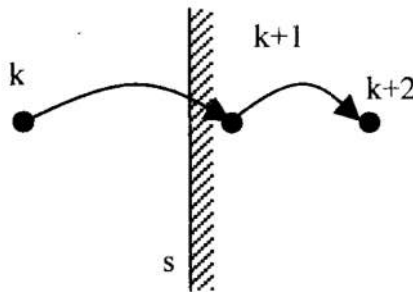
s – odległość początkowa,

k – stała sprężyny,

b – stała tłumika.

Wyznaczanie reakcji ścianki

W rozwiązaniu wzorowanym na [1] przyjmuje się, iż w chwili wejścia (punkt $k+1$ na rysunku 4) elementu w ściankę ($x \geq s$) wyznaczona zostaje reakcja R o wartości powodującej uzyskanie w następnym kroku całkowania (punkt $k+2$) zerowej wartości prędkości.



Rys. 4 Kolejne kroki całkowania

Fig. 4 Stoppage in wall

Opis matematyczny ruchu masy m (Rys. 2) przedstawia równanie (1). Reakcja w analizowanym przypadku (Rys. 4) opisana jest zależnością (2) a jej wartość R^+ wyznaczana jest z równań:

$$R^+ \in \begin{cases} \frac{m \cdot \dot{x}}{\Delta t} + F_u, & \text{gdy } \dot{x} > 0, \text{ uderzenie w ściankę} \\ F_u, & \text{gdy } \dot{x} = 0 \wedge F_u > 0, \text{ masa } m \text{ trwa w styku} \\ \{0\}, & \text{gdy } \dot{x} = 0 \wedge F_u \leq 0, \text{ masa } m \text{ odchodzi od styku} \end{cases} \quad (5)$$

gdzie:

Δt – przyrost czasu (krok całkowania).

Wyhamowanie poruszającej się masy powinno się odbyć w przyjętej przestrzeni szkodliwej (x_{sz} na rysunku 2), wynikającej z parametrów konstrukcyjnych siłownika. Warunek ten, w przypadku symulacji siłownika pneumatycznego ogranicza maksymalną wartość kroku całkowania.

3. Wyniki testów

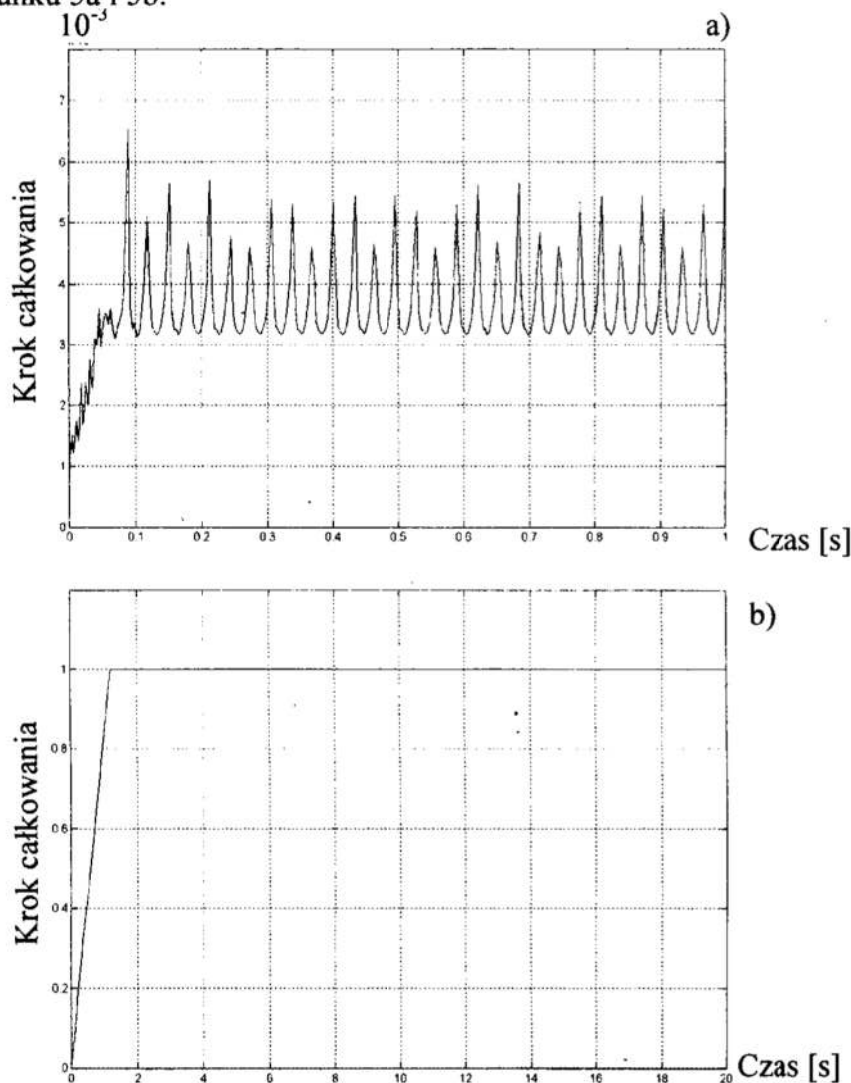
Modele komputerowe obu omówionych wyżej wariantów zbudowano w środowisku MatLab Simulink i przeprowadzono testy. Testowanie wykonano dla całkowania metodą Runge–Kutty rzędu 4 ze stałym i ze zmiennym krokiem całkowania. Wykorzystywane były trzy rodzaje siły wymuszającej F_u :

- siła o charakterze stałym;
- siła o charakterze zmiennym (sinusoidalnie o różnej częstotliwości, losowo);
- siła o charakterze skoku jednostkowego.

Analizowano wpływ parametrów całkowania oraz parametrów układu na przebieg procesu zatrzymania i postój masy na ścianie. Poniżej przedstawiono wybrane wyniki.

Postój na ścianie przy zmiennym obciążeniu

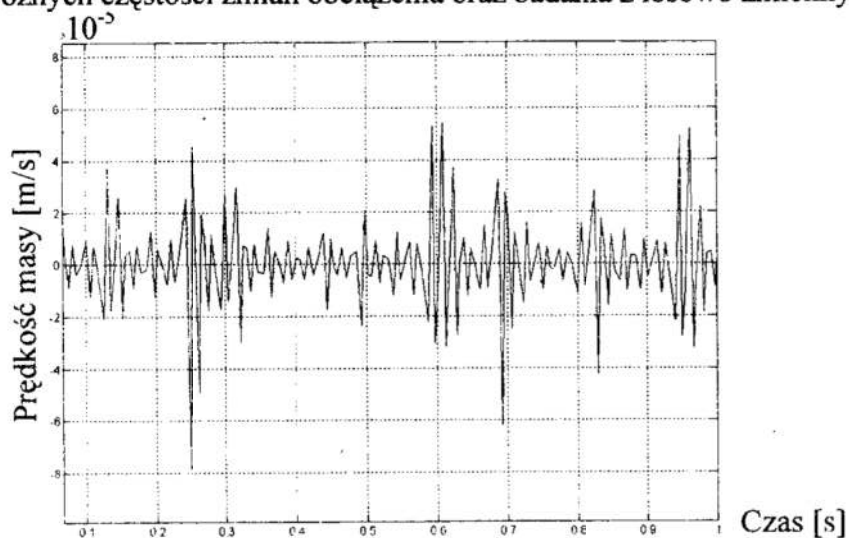
W ramach testów przeanalizowano postój masy ($m = 4$ kg) na ścianie przy działaniu sinusoidalnie zmiennego obciążenia ($F_u = 90\text{--}110$ N). Parametry całkowania przyjęto odpowiednio: krok minimalny 10^{-9} s, krok maksymalny 1 s. Wartość kroku pokazano na rysunku 5a i 5b.



Rys. 5 Zmiana kroku całkowania: a) ścianka lepko–sprężysta, b) liczona reakcja
Fig. 5 Value of integration step as a function of time: a) viscoelastic wall b) calculated reactive force

Na rysunku 6 pokazano zmiany prędkości dla ścianki lepko-sprężystej i stałego wymuszenia ($F_u = 100$ N).

Z zaprezentowanych wyników (Rys. 5 i Rys. 6) widać, iż w przypadku postoju masy na ścianie wyliczanie reakcji pozwala na całkowanie modelu z maksymalnym krokiem. Na przebieg procesu obliczeń nie ma wpływu zmienność obciążenia. Potwierdziły to badania dla różnych częstości zmian obciążenia oraz badania z losowo zmiennym obciążeniem.



Rys. 6 Zmiany prędkości dla ścianki sprężysto-lepkiej przy stałym wymuszeniu
Fig. 6 Velocity variation for viscoelastic wall and constant input function

Zderzenia ze ścianką

W ramach testów przeanalizowano również ruch masy ($m = 4$ kg) pod działaniem zmiennego obciążenia ($F_u = -50$ – 250 N). Masa poruszała się również poza ścianką (Rys. 7c). Parametry całkowania przyjęto odpowiednio: krok minimalny 10^{-9} s, krok maksymalny 1 s. Wartość kroku pokazano na rysunku 7a i 7b.

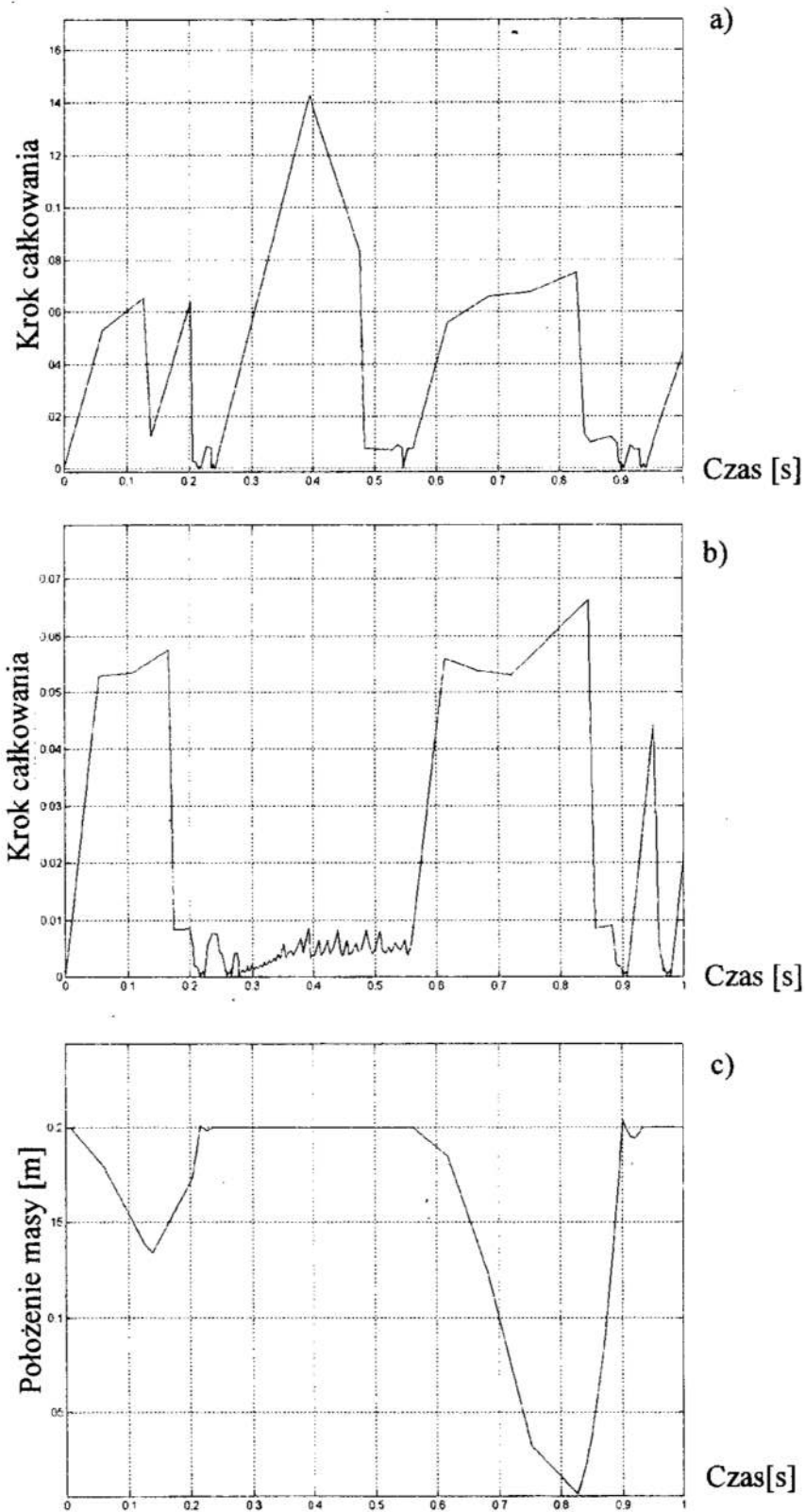
Widać (Rys 7), że dla okresu postoju tłoka na ścianie ($t = 0.2$ – 0.6 rysunek 7c) krok całkowania w przypadku wyznaczania reakcji jest znacznie większy (Rys. 7a) niż dla modelu sprężysto-lepkiej ścianki (Rys. 7b).

4. Uwagi i wnioski

Analizując otrzymane wyniki stwierdzić można, że zastosowanie nowej procedury:

1. zmniejszyło nakłady obliczeniowe w sytuacji postoju masy m (tłoka siłownika) na ścianie,
2. utrzymało dokładność obliczeń dla modelu zasadniczego,
3. uniezależniło parametry modelu ścianki od parametrów obiektu (siłownika).

Należy również zaznaczyć, że zaproponowane rozwiązanie wynika ze środowiska, w którym implementowano model siłownika.



Rys. 7 Układu pod zmiennym obciążeniem: a) krok całkowania dla obliczanej reakcji b) krok całkowania dla ścianki lepko-sprężystej c) położenie masy
 Fig. 7 System under changeable load: a) value of integration step calculated reactive force, b) value of integration step for viscoelastic wall c) position variation

Literatura

1. Grzesikiewicz W.: *Dynamika układów typu Picus Viridis*. Materiały XV Ogólnopolskiej Konferencji „Poliptymalizacja i Komputerowe Wspomaganie Projektowania” – Streszczenia. Koszalin–Mielno 1997 s.17
2. Grzesikiewicz W.: *Dynamika układów mechanicznych z więzami*. PWN, Warszawa 1979
3. Grymek Sz., Kiczowski T.: *Komputerowy moduł wspomagania obliczeń i doboru elementów pneumatycznych z katalogu*. Materiały XI Konferencji „PNEUMA'98”. Wydawnictwo Uczelniane Politechniki Koszalińskiej. Koszalin 1998r. s. 15–23.
4. Grymek Sz., Kiczowski T.: *Obliczenia i dobór elementów pneumatycznych wspomaganie komputerowo*. XII Konferencja nt „Metody i Środki Projektowania Wspomagane Komputerowo” Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 1999r. s.101–106.

Streszczenie

Zastosowanie CAD w projektowaniu napędów pneumatycznych wymaga stworzenia szybkich modeli komputerowych elementów pneumatycznych. Jednym z problemów, które należy rozwiązać jest opis zatrzymania i postoju tłoka siłownika pneumatycznego na ścianie.

W pracy przedstawiono dwa sposoby opisu tego problemu. Przedstawiono również wyniki symulacji w środowisku MatLab/Simulink.

SOLUTION OF *Picus Viridis* TYPE PROBLEM IN MATLab® SIMULINK® ENVIRONMENT

Summary

Creation of fast computer models of pneumatic elements is needed for a use of CAD technique in a designing of pneumatic driving systems. Description of a stoppage (or a stay) of a piston on the pneumatic cylinder wall is one of the task to be solved.

Two means of description of the problem as well as some simulation results in the MatLab/Simulink environment are presented in this paper.

Referat recenzowany

WIRTUALNA ORGANIZACJA PRAC PROJEKTOWYCH PRZY WYKORZYSTANIU DYNAMICZNYCH DOKUMENTÓW

1. Ewolucja systemów CAD

Projektowanie jest jednym z najbardziej złożonych procesów, których wykonywanie może być wspomagane przez wyspecjalizowane systemy informacyjne – systemy tworzące zintegrowane środowisko komputerowego wspomaganie projektowania (CAD). Dostępne dziś systemy CAD mają na celu wspomaganie prac projektowo-konstrukcyjnych usprawniając przede wszystkim zadania obliczeniowe, kreślarskie, symulacyjne itp. Globalizacja światowych rynków zbytu i idące za nią coraz głębsze współdziałanie różnych przedsiębiorstw rozproszonych po całym świecie, powoduje, że konieczna jest współpraca wielu projektantów, konstruktorów z różnych miejsc świata nad jednym wspólnym projektem. Nowoczesne środowisko informacyjne CAD ma więc do zrealizowania nowe zadania – koordynację procesu projektowo-konstrukcyjnego oraz integrację cząstkowych modułów projektu w globalnym, rozproszonym i zdecentralizowanym środowisku informacyjnym.

Mimo, iż rozwiązania dotyczące integracji działań w projektowaniu nie są tak rozwijane, jak ma to np. miejsce w zarządzaniu czy produkcji [3], możliwe staje się stworzenie nowego, otwartego modelu środowiska wspomaganie prac projektowych i konstruktorskich. Wydaje się, że stajemy u progu kolejnej, szczególnie istotnej fazy rozwoju systemów CAD [5].

Do chwili obecnej pakiety CAD były budowane na podstawie dość „sztywnej” wizji procesu projektowo-konstrukcyjnego i to dla ściśle określonej klasy obiektów. Z tego względu system CAD (wspomagający konstruktora czy projektanta we wszystkich fazach projektu) był mało elastyczny – nie był nastawiony na rozwiązywanie różnych klas projektów/konstrukcji. Wraz z pojawieniem się nowych realiów światowego rynku, pojawia się zapotrzebowanie na systemy, w których procesy projektowo-konstrukcyjne nie są na stałe „zaszyte” w systemie, przeciwnie – powinna istnieć możliwość ich rekonfigurowania w celu dopasowania do konkretnego zadania. W tak rozumianym systemie projektant sam tworzy sobie konfigurację systemu CAD, korzystając z gotowych, dostępnych modułów systemu (jak ma to miejsce np. w systemie CACSD dla automatyki, czy SPICE dla elektroniki) i pewnych przesłanek, jak taki proces powinien wyglądać. Wszystko to musi się odbywać w rozproszonym i zdecentralizowanym globalnym środowisku informacyjnym. Aby zrealizować taką koncepcję systemu, konieczne są badania teoretyczne w dziedzinie projektowania systemów oraz tworzenie i rozwijanie odpowiednich narzędzi [1].

2. Proces projektowo–konstrukcyjny w wirtualnym biurze projektowym

Proces projektowo-konstrukcyjny – bez względu na dziedzinę, której dotyczy (mechanika, architektura, elektronika, informatyka, itd.) – składa się z sekwencyjnego (w dzisiejszych czasach zazwyczaj iteracyjnego i inkrementacyjnego) wykonywania pewnych czynności (rys. 1), do których zaliczyć można: określenie wymagań, analiza problemu, projektowanie, tworzenie modelu lub konstrukcja prototypu, symulacja lub testowanie, dystrybucja (w etapie końcowym). Nowością – zaczerpniętą z projektowania np. systemów informatycznych, lecz możliwą do zastosowania w innych dziedzinach – jest iteracyjność polegająca na wielokrotnym wykonywaniu tych samych czynności (cykli)

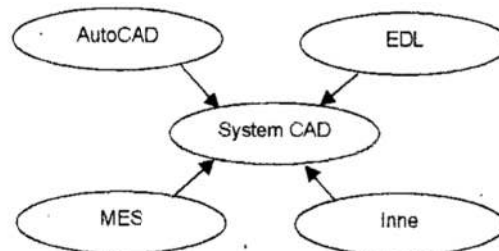
* Politechnika Zielonogórska, Wydział Elektryczny, Instytut Informatyki i Elektroniki, ul. Podgórna 50, 65-001 Zielona Góra, tel. (068) 3282421, fax (068) 3244733, e-mail: A.Kolopienczyk@pz.zgora.pl

oraz inkrementacja, polegająca na ulepszaniu lub-rozbudowywaniu projektu/konstrukcji o nowe elementy w każdym cyklu.



Rys. 1. Iteracyjne wykonywanie kolejnych etapów procesu projektowo-konstrukcyjnego
Fig. 1. Iterative performing of steps of the engineering design process

Każdy z wymienionych wyżej etapów w procesie projektowo-konstrukcyjnym stanowi osobny proces, na zakończenie którego składać się może wykonanie wielu różnych czynności, wspomaganych przez różne narzędzia. Oprócz kompletnych systemów komputerowych CAD/CAM, dostępne są inne programy do wspomaganie wybranych etapów projektów inżynierskich [4]: CAE (Computer Aided Engineering) - wspomaganie prac inżynierskich w procesie projektowo-konstrukcyjnym, CAP (Computer Aided Planing) - planowanie czynności w procesie wytwórczym, CAT (Computer Aided Testing) - badania doświadczalne, symulacyjne i prototypowe, CAQ (Computer Aided Quality Control) - kontrola i zapewnienie jakości w cyklu powstawania wyrobu, CAS (Computer Aided Styling) - kształtowanie ostatecznej formy wyrobu. Każda z takich aplikacji posiada swoje własne moduły, odpowiedzialne za wykonywanie obliczeń, prezentację wyników, dostęp do baz danych, wspomaganie decyzji, itd. System CAD powinien integrować wszystkie niezależne do tej pory aplikacje wspomagające prace inżynierskie tak, aby wspomagać projektanta/konstruktora we wszystkich fazach złożonego procesu projektowo-konstrukcyjnego (rys. 2).

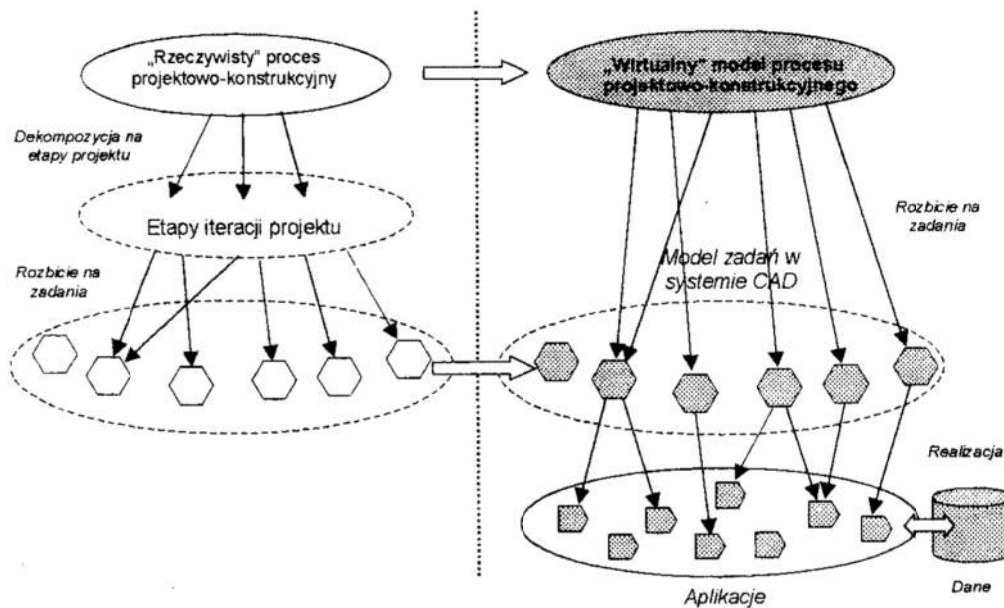


Rys. 2. System CAD integrujący wyspecjalizowane aplikacje wspomagające różne prace inżynierskie
Fig. 2 CAD system integrates specialised applications of engineering activities aid

Idąc dalej, wykonywanie tak rozumianego procesu projektowo-konstrukcyjnego, powinno być wspomagane przez system CAD na etapie koordynacji rozproszonych prac, wymiany informacji pomiędzy projektantami. System taki gromadzić powinien ponadto w jednym i spójnym repozytorium wszystkie dane związane z projektem – zarówno same dane projektowe, jak i dodatkowe uwagi, komentarze dotyczące wszystkich etapów procesu (specyfikacja wymagań i założeń, uwagi analityczne i projektowe, komentarze dotyczące prototypów, modeli i ich testowania itp.).

Tak rozumiany system informacyjny staje się elementem scalającym wszystkie czynności projektantów/konstruktorów, tworząc pewnego rodzaju „wirtualne” biuro projektowe. Komputerowe środowisko informacyjne staje się podstawowym, jeśli nie jedynym, medium, dzięki któremu kontaktują się osoby biorące udział w projekcie.

W opisywanym środowisku wspomagającym prace projektowo-konstrukcyjne możemy zauważyć, że zintegrowany system CAD stanowi pewien model przenoszący proces ze świata „rzeczywistego”, operującego na konkretnych bytach i czynnościach, w świat wirtualny, reprezentowany przez pewien model rzeczywistego procesu, w którym zadania są opisane w sposób komputerowy, rozbite na pewne aplikacje wspomagające konkretne czynności. Mapowanie z tak rozumianego świata rzeczywistego do wirtualnego zachodzi na dwóch etapach – na etapie modelowania poszczególnych zadań oraz na etapie projektowania procesu składającego się z tych zadań (rys. 3).



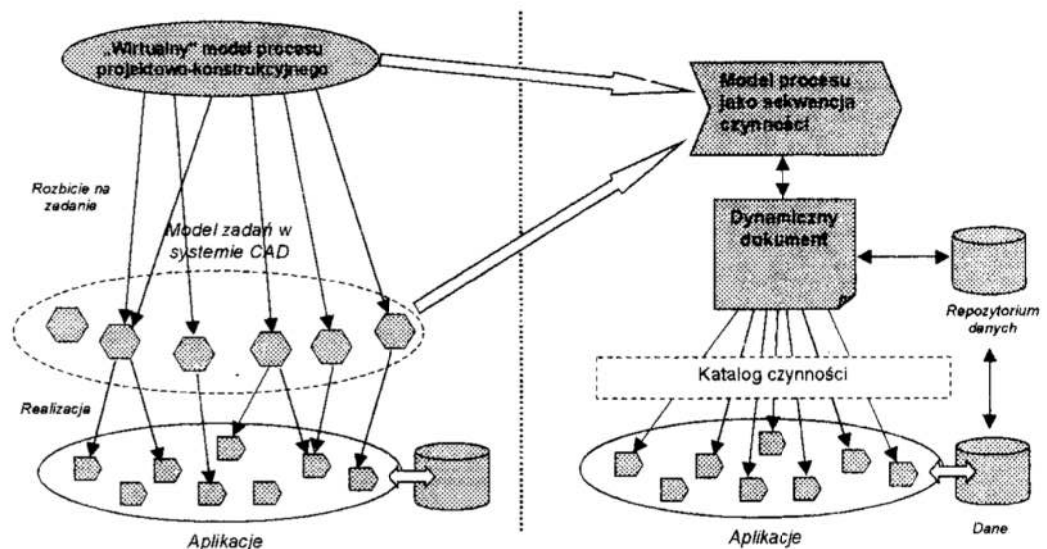
Rys. 3. Transformacja rzeczywistego procesu w wirtualny model
Fig. 3. Migration from real process into virtual process model

3. Dynamiczne dokumenty jako element integrujący prace projektowo-konstrukcyjne

Modelowanie procesu projektowo-konstrukcyjnego niesie za sobą konieczność stworzenia mechanizmów umożliwiających integrację aplikacji oraz elastyczne tworzenie i nadzór nad wykonywaniem procesów w rozproszonym i zdecentralizowanym środowisku informacyjnym wirtualnego biura projektowego. Szukamy więc narzędzi umożliwiających wspomaganie wykonywania rozproszonych procesów projektowo-konstrukcyjnych.

Integracja aplikacji wspomagających prace inżynierskie jest konieczna do stworzenia systemu CAD – dzięki niej możemy tworzyć spójny i zautomatyzowany proces projektowo-konstrukcyjny. Integracja taka powinna przy tym spełniać warunki: (1) powinna zostać zachowana integralność procesu projektowo-konstrukcyjnego, (2) możliwe powinno być nadzorowanie przebiegiem procesu; (3) sam proces powinien być wysoko konfigurowalny, dostosowany do aktualnych wymagań projektanta/konstruktor; (4) możliwa powinna być praca w sieciowym, rozproszonym i zdecentralizowanym i środowisku informacyjnym rozległych sieci komputerowych.

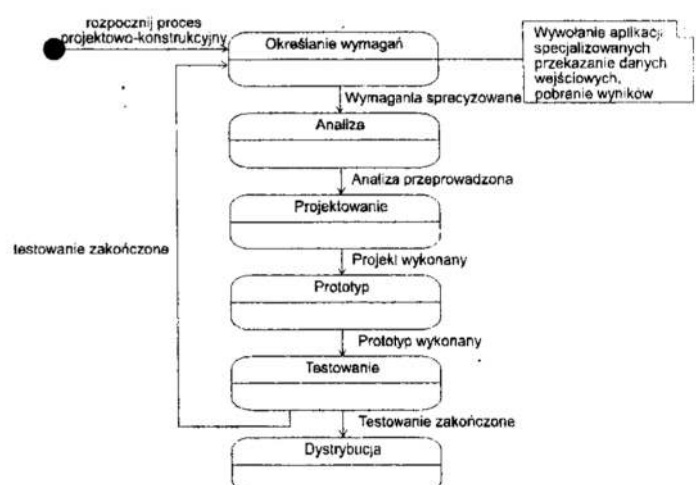
Proponuje się wykorzystanie wirtualnych, dynamicznych dokumentów [2] do integracji systemu CAD (Rys. 4). Dynamiczny dokument jest implementacją modelu procesu. Dla prac projektowo-konstrukcyjnych model samego procesu zdekomponowany na czynności stanowi źródło do powstania opisu procesu dla wirtualnego dokumentu (por. [6]). Wirtualny dokument pozwala następnie na przemapowanie czynności na wykonanie konkretnych zadań przy wykorzystaniu dostępnych narzędzi wspomagających pracę inżynierskie. W końcowym etapie dane przetwarzane przez aplikacje gromadzone są w jednym wspólnym repozytorium, stanowiącym część wirtualnego dokumentu. Dokument wirtualny zawiera: 1) dane projektu – zhierarchizowane drzewo kolejnych etapów projektu, rozbitych na elementy rozwijane w różnych modułach systemu CAD, zintegrowanych w jedną całość w dokumencie, 2) informacje o sposobie integracji modułów oraz zależnościach i wpływie modyfikacji jednego modułu na inne moduły projektu, 3) informacje o przebiegu całego procesu projektowo-konstrukcyjnego, rozbitego na poszczególne etapy.



Rys. 4. Realizacja procesu przy wykorzystaniu wirtualnych dokumentów
Fig. 4. Process realisation using virtual documents

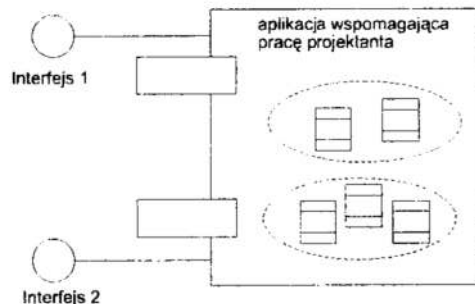
Wirtualny dokument przejmuje ponadto kontrolę nad całym procesem (rys. 5), umożliwi pracę w środowisku rozproszonym, każda wykorzystywana aplikacja może się znajdować na innym komputerze, przyłączonym do wspólnej sieci komputerowej.

Istotnym zagadnieniem jest sposób integracji aplikacji wspomagających pracę inżynierskie. W obiektowym podejściu, każda aplikacja wspomagająca pracę projektanta może być traktowana jako komponent, składający się



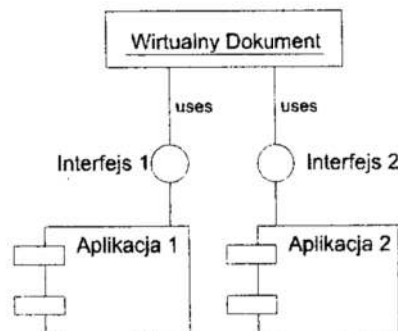
Rys. 5. Model procesu
Rys. 5. Process model

z kolaboracji pewnych obiektów zawierających metody (zadania obliczeniowe, optymalizacyjne, edytorskie, itd.), oraz pewne pola (przetwarzane dane, stałe informacje „zaszyte” wewnątrz aplikacji) (Rys. 6, notacja UML). Każda z aplikacji wyposażona jest także w pewne interfejsy (w zależności o rodzaju aplikacji implementacja interfejsu może być zrealizowana poprzez mechanizmy COM lub OLE, operacje na plikach, dostęp do współdzielonych baz danych, eksportowanie i importowanie, wraz z ewentualną konwersją, danych specyficznych dla poszczególnych aplikacji).



Rys. 6. Ogólny obiektowy model aplikacji
Fig. 6. General object model of an application

Wirtualne dokumenty stanowią łącze pomiędzy aplikacjami (rys. 7), które – przy wykorzystaniu interfejsów – przesyłają dane pomiędzy aplikacją a repozytorium danych dokumentu, zarządzając przepływem procesu.



Rys. 7. Wirtualny dokument jako łącznik pomiędzy aplikacjami
Fig. 7. Virtual document as a link between applications

Integralność danych oraz integracja aplikacji nadzorowana jest przez mechanizmy dokumentów wirtualnych obsługujących proces. To właśnie poprzez dokumenty przekazywane są dane do aplikacji, a następnie pobierane są wyniki, które zostają umieszczone we współdzielonych, rozproszonych bazach danych. Możliwe jest więc wykorzystanie różnych, niezależnych do tej pory aplikacji w jednym procesie.

Rozwiązanie pozwala na tworzenie dynamicznych, stale zmieniających się procesów w rozproszonym, heterogenicznym środowisku informacyjnym opartym o technologie komputerowe. System wykorzystując te same podstawowe elementy składowe jest zdolny do swej ciągłej reorganizacji.

4. Podsumowanie

Zintegrowane środowisko CAD może się stać narzędziem nie tylko wspomagającym prace bezpośrednio związane z projektem, lecz także nadzorować cały przebieg procesu

projektowo-konstrukcyjnego. Mechanizmy wirtualnych, dynamicznych dokumentów pozwalają na praktyczną realizację takiego nowego podejścia.

Dynamiczne dokumenty mogą stanowić spoiwo pomiędzy wyspecjalizowanymi aplikacjami wspomagającymi prace inżynierskie, umożliwiając tym samym tworzenie systemów CAD na bazie już istniejących aplikacji. Dynamiczne dokumenty nie są same w sobie ośrodkiem integracji systemu, pozwalają natomiast prowadzić nadzór nad całym procesem projektowo-konstrukcyjnym, automatycznie kierując prace do kolejnych etapów.

W artykule przedstawiono wstępne rozważania możliwości zastosowania wirtualnych, dynamicznych dokumentów do tworzenia nowoczesnych, otwartych systemów CAD. Dalsze prace mają na celu głębszą analizę problemu oraz próbne implementacje zgodnie z tworzonym modelem środowiska CAD.

Literatura

1. Kierzkowski Z., Kołopieńczyk A., Włodarczyk Z., *Możliwości tworzenia środowiska informacyjnego wspomagania prac zespołowych w projektowaniu*, XVI Ogólnopolska konferencja Polioptymalizacja i Komputerowe Wspomaganie Projektowania, materiały konferencyjne, Politechnika Koszalińska, Mielno 1998.
2. Kierzkowski Z., Kołopieńczyk A., *Virtual Documents Modelling in the Process Oriented Decentralized Management Systems*, Międzynarodowa Konferencja CIM'99, WNT, Zakopane 1999.
3. Kierzkowski Z., Kołopieńczyk A., *Tworzenie środowiska obiegu informacji w postaci elektronicznej w systemach informatycznych zarządzania*, IV Konferencja Komputerowe Systemy Wielodostępne KSW'98, Bydgoszcz 1998.
4. Sileikis W., Grzechca W., *Systemy PDM*, XVI Ogólnopolska konferencja Polioptymalizacja i Komputerowe Wspomaganie Projektowania, Politechnika Koszalińska, Mielno 1998.
5. Tarnowski W., *Polioptymalizacja jako środek integracji w nowoczesnych systemach CAD*, XVI Ogólnopolska konferencja Polioptymalizacja i Komputerowe Wspomaganie Projektowania, Politechnika Koszalińska, Mielno 1998.
6. Kołopieńczyk A.: *UML Specification Of Virtual Documents For Software Development*. ISICIS XIV, proceedings of Fourteenth the International Symposium on Computer and Information Sciences, Kusadasi 1999.

Streszczenie

W artykule przedstawiono sposób tworzenia nowoczesnych systemów CAD w oparciu o wirtualne, dynamiczne dokumenty. Takie systemy pozwolą na tworzenie „wirtualnych pracowni projektanckich”, w których proces projektowo-konstrukcyjny jest nadzorowany przez systemy informacyjne, umożliwiając integrację aplikacji i danych w globalnym, rozproszonym i zdecentralizowanym komputerowym środowisku informacyjnym.

VIRTUAL ORGANISATION OF DESIGN ACTIVITIES USING DYNAMIC DOCUMENTS

Summary

In this paper a new role of integrated CAD systems is introduced. A method of modelling and building of such systems is shown. The presented method is based on virtual documents, as a medium for the integration of design activities in distributed, decentralised information systems of “virtual design office”.

Referat recenzowany

METODA OSZACOWANIA RYZYKA W SYSTEMACH TECHNICZNYCH Z WYKORZYSTANIEM ANALIZY PRZEDZIAŁOWEJ

1. Wprowadzenie

Staranne modelowanie probabilistyczne i systematyczna analiza ryzyka mają coraz większe znaczenie w zarządzaniu bezpieczeństwem eksploatacji obiektów technicznych. Proces przeprowadzania analizy ryzyka z ogólną propozycją dotyczącą metod analizy jakościowej i metod szacowania wskaźników ryzyka został zarysowany w ustanowionej niedawno normie PN-IEC 300-3-9 [8]. Norma ta zawiera również uwagi dotyczące zakresu i stopnia szczegółowości (dogłębności) analizy ryzyka w kolejnych fazach życia obiektu technicznego. Ma to związek z celem analizy ryzyka i jakością jej przeprowadzania. Wymaga się w tej normie, aby oszacowania wskaźników ryzyka zawierały ocenę niepewności.

Według normy PN-IEC 300-3-9 *analiza ryzyka w systemach technicznych jest systematycznym stosowaniem dostępnych informacji do zidentyfikowania zagrożenia i do oszacowania ryzyka dla osób, populacji, mienia lub środowiska*. W normie tej przytoczono również definicje takich pojęć jak: system, ocena ryzyka, oszacowanie ryzyka, zarządzanie ryzykiem. Ryzyko rozumie się dalej jako możliwość wystąpienia niepożądanych zdarzeń (sekwencji awaryjnych) powodujących określone konsekwencje. Do wyznaczania poziomu ryzyka można stosować różne miary. Możliwość wyraża się zwykle poprzez prawdopodobieństwo wystąpienia zdarzenia lub jego częstość, to jest prawdopodobieństwo na jednostkę czasu. Ze względu na olbrzymią liczbę potencjalnych zdarzeń awaryjnych w złożonym obiekcie przemysłowym i możliwych poziomów ich konsekwencji zasadne jest grupowanie w analizie ryzyka zdarzeń i konsekwencji w kategorie.

Niniejsza praca dotyczy problematyki reprezentacji i oceny niepewności w modelowaniu i szacowaniu ryzyka związanego z eksploatacją systemu technicznego, stwarzającego potencjalnie zagrożenie dla ludzi, mienia i środowiska. Wyniki oszacowania miary ryzyka z oceną niepewności mają kluczowe znaczenie w zarządzaniu bezpieczeństwem i sterowaniu ryzykiem tych systemów [6].

2. Źródła danych niezawodnościowych i reprezentacja niepewności

Jednym z najważniejszych i najtrudniejszych zadań w modelowaniu probabilistycznym i analizie ryzyka jest zgromadzenie danych niezawodnościowych. Dane te są parametrami w cząstkowych modelach probabilistycznych i modelu ryzyka systemu technicznego. Od rzetelności ich zgromadzenia zależy wiarygodność uzyskanych wyników. Ma to podstawowe znaczenie dla podejmowania decyzji bazujących na tych

* Politechnika Gdańska, Wydział Elektrotechniki i Automatyki, G. Narutowicza 11/12, 80-952 Gdańsk
Email: kazkos@elv.pg.gda.pl, mkwies@elv.pg.gda.pl

modelach. Problem polega na tym, że z powodu niedostatków organizacyjnych, a także obiektywnych trudności w gromadzeniu danych z eksploatacji rzeczywistych obiektów (mała ich liczba lub system dopiero projektowany) oraz znanych ograniczeń w korzystaniu z istniejących baz danych niezawodnościowych o charakterze ogólnym, analityk zmuszony jest do korzystania z różnych źródeł danych o uszkodzeniach i zdarzeniach awaryjnych takich jak:

1. Źródłowe zbiory danych o charakterze ogólnym o uszkodzeniach obiektów technicznych w zestawieniach branżowych lub o charakterze ogólnym;
2. Dane historyczne dotyczące awarii w instalacjach przemysłowych danego typu;
3. Przyspieszone badania niezawodności elementów i modułów elektronicznych;
4. Oszacowania miar niezawodności obiektów danej kategorii na podstawie danych statystycznych uszkodzeń obiektów podobnych;
5. Zestawienia uaktualnionych miar probabilistycznych wyznaczonych na podstawie danych o uszkodzeniach zaistniałych w konkretnym obiekcie złożonym;
6. Dane historyczne dotyczące błędów popełnionych przez człowieka działającego w instalacjach przemysłowych;
7. Dane eksperymentalne dotyczące błędów popełnionych przez operatora podczas treningu i testów na symulatorze obiektu technicznego;
8. Metody i poradniki wspomagające predykcyjne oszacowanie niezawodności dla różnych warunków środowiskowych.

Tak więc analityk staje przez zadaniem opracowania bazy danych dla projektu analizy ryzyka danego systemu technicznego korzystając z różnych źródeł informacji. W wymienionych powyżej źródłach informacji niepewność co do parametrów modeli probabilistycznych (intensywność uszkodzeń, częstość zdarzenia inicjującego stan nienormalny, prawdopodobieństwo niezadziałania na żądanie, prawdopodobieństwo błędu człowieka itd.) reprezentuje się na różne sposoby stosując: przedziały liczbowe, określenia lingwistyczne, zdefiniowane rozkłady probabilistyczne (najczęściej: równomierny, trójkątny, logarytmo-normalny), kwantyle rozkładów probabilistycznych (niekiedy bez podania typu rozkładu).

Należy podkreślić, że modele probabilistyczne i model ryzyka danego systemu mają zwykle charakter predykcyjny. Ważne jest dlatego, aby uwzględnić ich specyfikę funkcjonalną instalacji w tym systemie, uwarunkowania jakościowe w procesie jego projektowania i eksploatacji oraz warunki środowiskowe. W tej sytuacji zdecydowano, aby niepewność związaną z modelami probabilistycznymi i oszacowaniami ryzyka przedstawiać za pomocą rozkładu możliwości, który łączy w pewnym sensie opisane powyżej podejścia reprezentowania niepewności. Jak wiadomo stosowanie rozkładów możliwości jest szczególnie uzasadnione w przypadku modeli predykcyjnych [4]. Dla innych sposobów reprezentacji niepewności konieczne jest przeprowadzenie odpowiedniej transformacji na rozkład możliwości. W następnym rozdziale opisano dwie metody transformacji rozkładu prawdopodobieństwa na rozkład możliwości.

3. Transformacja z rozkładu prawdopodobieństwa na rozkład możliwości

Istnieje wiele metod transformacji z rozkładu prawdopodobieństwa na rozkład możliwości i odwrotnie. Metody te zależą od typu rozwiązywanego problemu, związanego z daną transformacją.

W niniejszej pracy omawia się dwa rodzaje transformacji, a mianowicie klasyczną już transformację wprowadzoną przez Dubois i Prade'a [1] opartą o koncepcję niedokładnego prawdopodobieństwa oraz transformację Klira [4] zachowującą niepewność danych podczas jej dokonywania.

Podejście Dubois i Prade'a

Podejście do teorii możliwości zaproponowane przez Dubois i Prade'a [1] opiera się na koncepcji nierównych prawdopodobieństw dla zdarzeń elementarnych oraz niedokładnego prawdopodobieństwa [9], [2]. Autorzy przyjmują założenie, że eksperyment statystyczny może być zniekształcony i w takiej sytuacji prawdopodobieństwa przyporządkowane zdarzeniom elementarnym nie koniecznie muszą być równe. W konsekwencji wprowadza się dwie miary niepewności w miejsce miary prawdopodobieństwa, a mianowicie potrzeby oraz możliwości. Ta ostatnia jest zgodna z miarą możliwości wprowadzoną przez Zadeha [11]. Obydwie miary mogą być obliczone z rozkładu możliwości [1].

W oparciu o przedstawioną ideę można zdefiniować transformację prawdopodobieństwo – możliwość i odwrotnie. Załóżmy zbiór zdarzeń elementarnych $X = \{x_i; i = 1, 2, \dots, n\}$ oraz uporządkowany rozkład prawdopodobieństwa $p = (p_1, p_2, \dots, p_n)$, $(p_1 \geq p_2 \geq \dots \geq p_n)$, gdzie $p_i = \text{Pro}(\{x_i\})$, $\sum_{i=1}^n p_i = 1$ oraz Pro jest miarą prawdopodobieństwa. Załóżmy ponadto zbiór zagnieżdżonych zdarzeń $A_i = \{x_1, x_2, \dots, x_i\}$, gdzie $A_0 = \emptyset$. Wówczas rozkład możliwości $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ można obliczyć zgodnie z [1]:

$$\forall i \quad \pi_i = \sum_{j=1}^n \min(p_i, p_j), \quad (1)$$

podczas gdy transformacja odwrotna wyrazi się zależnością:

$$\forall i = 1, \dots, n \quad p_i = \sum_{j=i}^n \frac{1}{j} (\pi_i - \pi_{j+1}); \quad \pi_{n+1} = 0. \quad (2)$$

Podejście Klira

Załóżmy, że mamy uporządkowane rozkłady prawdopodobieństwa i możliwości odpowiednio $p = (p_1, p_2, \dots, p_n)$ oraz $\pi = (\pi_1, \pi_2, \dots, \pi_n)$. Klir [4] proponuje następującą transformację z rozkładu prawdopodobieństwa na rozkład możliwości:

$$\pi_i = \left(\frac{p_i}{p_1} \right)^\beta, \quad i = 1, 2, \dots, n, \quad (3)$$

gdzie współczynnik β powinien być obliczony zgodnie z równaniem bilansu niepewności [4]:

$$H(p) = N(\pi) + S(\pi), \quad 0 < \beta < 1, \quad (4)$$

gdzie $H(p)$ jest entropią Shanona [10], $N(\pi)$ U-niepewnością [3] oraz $S(\pi)$ funkcją dysonansu [4]. Suma $N(\pi) + S(\pi)$ wyraża całkowitą niepewność związaną z rozkładem możliwości π i może być traktowana jako uogólnienie entropii Shanona dla przypadku posybilistycznego. Transformacja odwrotna może być obliczona zgodnie z zależnością:

$$p_i = \frac{\pi_i^{1/\beta}}{\sum_{k=1}^n \pi_k^{1/\beta}}, \quad i = 1, 2, \dots, n. \quad (5)$$

Zakładając, że dany jest rozkład prawdopodobieństwa p , obliczenie β sprowadza się do rozwiązania następującego równania:

$$f(\beta) = \sum_{i=1}^n p_i \log_2 p_i + \sum_{i=2}^n \left(\left(\frac{p_i}{p_1} \right)^\beta - \left(\frac{p_{i+1}}{p_1} \right)^\beta \right) \log_2 \frac{i^2}{\sum_{j=1}^i \left(\frac{p_j}{p_1} \right)^\beta} = 0, \quad (6)$$

które może być rozwiązane za pomocą dowolnej procedury numerycznej do rozwiązywania równań nieliniowych, ponieważ funkcja $f(\beta)$ jest monotoniczna. Więcej szczegółów dotyczących transformacji można znaleźć w pracy [4]. Opisane metody transformacji zostały zrealizowane w module programu komputerowego do wspomaganie analizy ryzyka złożonych systemów przemysłowych.

4. Obliczenia przedziałowe

Zakłada się następującą postać rozmytego przedziału prawdopodobieństwa:

$$\tilde{a} = (l, \underline{a}, \bar{a}, u), \quad (7)$$

gdzie l oraz u są odpowiednio wartościami dolnymi i górnymi, natomiast \underline{a} i \bar{a} są odpowiednio dolnymi i górnymi wartościami modalnymi przedziału rozmytego \tilde{a} . W przypadkach szczególnych otrzymujemy - (i) rozmyte prawdopodobieństwo dla $\underline{a} = \bar{a}$, (ii) prawdopodobieństwo przedziałowe, gdy $l = \underline{a}$ i $\bar{a} = u$ oraz (iii) prawdopodobieństwo punktowe, gdy $l = \underline{a} = \bar{a} = u$.

W celu wykonania operacji algebraicznych na przedziałach rozmytych można wykonać obliczenia na wybranym skończonym zbiorze ich α -poziomów (przedziałów liczbowych), a następnie wykorzystać analizę przedziałową [7] oraz w końcu złożyć rozmyty przedział wynikowy. Załóżmy, że dwie liczby rozmyte \tilde{a} i \tilde{b} mają odpowiednio α -obciążenia $a^\alpha = [a_l^\alpha, a_r^\alpha]$ i $b^\alpha = [b_l^\alpha, b_r^\alpha]$. W celu wykonywania operacji na

α -poziomach dla $\alpha \in [0,1]$ wykorzystuje się podstawowe operacje algebraiczne na dodatnich przedziałach liczbowych, zdefiniowane następująco [7]:

$$\tilde{c} = \tilde{a} + \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha + b_l^\alpha, a_r^\alpha + b_r^\alpha), \quad (8)$$

$$\tilde{c} = \tilde{a} - \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha - b_r^\alpha, a_r^\alpha - b_l^\alpha), \quad (9)$$

$$\tilde{c} = \tilde{a} \cdot \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha \cdot b_l^\alpha, a_r^\alpha \cdot b_r^\alpha), \quad (10)$$

$$\tilde{c} = \tilde{a} / \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha / b_r^\alpha, a_r^\alpha / b_l^\alpha), \quad (11)$$

Łatwo zauważyć, że operacje na przedziałach liczbowych sprowadzają się do operacji na ich dolnych i górnych kresach (Rys. 1) i mogą być łatwo rozszerzone do operacji na wektorach i macierzach z rozmytymi elementami.

W trakcie wykonywania bardziej złożonych operacji na przedziałach rozmytych często dana zmienna występuje częściej niż raz w danym wyrażeniu. Powoduje to nieuzasadniony wzrost rozmycia wyniku. W takich sytuacjach celowe wydaje się stosować arytmetykę rozmytą z ograniczeniami [5] która zapobiega wzrostowi niepewności. Wówczas zakładając, że rozważane wyrażenie jest monotoniczne ze względu na każdą ze zmiennych, operacje odejmowania i dzielenia (9) i (11) należy zastąpić następująco zdefiniowanymi operacjami:

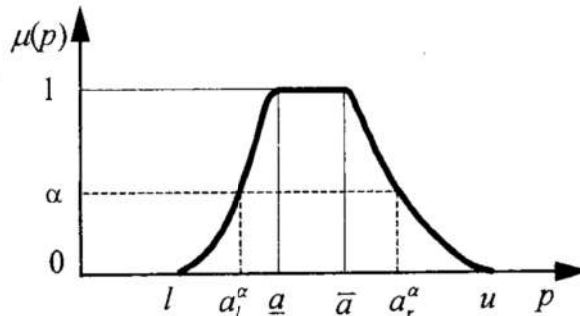
$$\tilde{c} = \tilde{a} - \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha - b_l^\alpha, a_r^\alpha - b_r^\alpha), \quad (12)$$

$$\tilde{c} = \tilde{a} / \tilde{b} : \forall \alpha (c_l^\alpha, c_r^\alpha) = (a_l^\alpha / b_l^\alpha, a_r^\alpha / b_r^\alpha), \quad (13)$$

Z praktycznego punktu widzenia celowe wydaje się wykonywanie dodatkowo operacji na wartościach średnich modalnych:

$$a = \frac{a + \bar{a}}{2}. \quad (14)$$

Należy jednak podkreślić, że dla operacji mnożenia i dzielenia otrzymane wyniki dla wartości modalnych będą się różniły od wartości modalnych wyników obliczeń. Różnica zależy od długości nośników i przedziałów modalnych operandów [6]. W przypadkach praktycznych spotykanych w analizach ryzyka różnica ta mieści się zwykle w zakresie kilkunastu procent.



Rys. 1. Rozmyty przedział prawdopodobieństwa

Fig. 1. Fuzzy interval probability

We wspomnianym programie komputerowym końcowe indeksy ryzyka są obliczane i agregowane z wykorzystaniem podejścia opartego na α -poziomach z wykorzystaniem zależności (8)-(11) oraz możliwością wykorzystania zależności (12) i (13).

5. Oszacowanie ryzyka z wykorzystaniem analizy przedziałowej

Ryzyko rozumie się często jako kombinację niepewności i straty/ szkody [6]. W kontekście modelowania probabilistycznego i oceny systemów technicznych ryzyko definiuje się za pomocą zbioru trójek, z których każda zawiera: zdarzenie, prawdopodobieństwo (lub częstość) oraz konsekwencje (niekorzystny skutek):

$$\mathfrak{R}^p = \{ \langle s_i, p_i, n_i \rangle \} \quad (15)$$

lub:

$$\mathfrak{R}^w = \{ \langle s_i, f_i, n_i \rangle \}, \quad (16)$$

gdzie: s_i oznacza zdarzenie awaryjne zaistniałe według i -tego scenariusza, czyli co może się stać (i -ta sekwencja zdarzeń), f_i jest częstością i -tego scenariusza, a n_i oznacza jedną lub więcej miar skutków powstałych w wyniku przebiegu awarii według danego scenariusza, w tym strat ekonomicznych. Na podstawie tego zbioru trójek można wyznaczyć różne wskaźniki ryzyka lub wykreślić odpowiednie zależności funkcyjne ułatwiające ocenę i porównywanie ryzyka dla różnych systemów technicznych.

Jak wspomniano oszacowanie ryzyka powinno być powiązane z oceną niepewności. Proponuje się dlatego zastosować oszacowania przedziałowe tych miar. Miarę ryzyka wyznacza się na podstawie zdefiniowanych wektorów częstości (dla kategorii zdarzeń) i macierzy zawierających prawdopodobieństwa warunkowe zdarzeń [6]:

$$\tilde{f}_{SA} = \tilde{f}_{ZI} \tilde{p}(SA|ZI), \quad (17)$$

$$\tilde{f}'_{KS} = \tilde{f}_{SA} \tilde{p}(KS'|SA), \quad (18)$$

gdzie:

- \tilde{f}_{ZI} – wierszowy wektor częstości wyróżnionych kategorii inicjatorów zdarzeń nienormalnych/ awaryjnych [a^{-1}],
- $\tilde{p}(SA|ZI)$ – macierz ($I \times J$) prawdopodobieństw zdarzeń warunkowych, że i -ty inicjator przyczyni się do wystąpienia j -tej kategorii stanów awaryjnych obiektu (przypadków awarii),
- \tilde{f}_{SA} – wierszowy wektor częstości wystąpienia wyróżnionych kategorii stanów awaryjnych obiektu (przypadków awarii),
- $\tilde{p}(KS'|SA)$ – macierz ($J \times K$) prawdopodobieństw zdarzeń warunkowych, że j -ty stan awaryjny obiektu spowoduje k -tą kategorię skutków (strat) l -tego rodzaju,
- \tilde{f}'_{KS} – wierszowy wektor częstości wystąpienia wyróżnionych kategorii skutków (strat) l -tego rodzaju,

Niech wielkość skutków (strat) l -tego rodzaju określa wektor kolumnowy \tilde{n}'_{KS} , przy czym skutek ma wymiar [jedn. skutku l -tego rodzaju]. Skutkami mogą być obrażenia,

wypadki, straty materialne itp. wyrażone w odpowiednich jednostkach. Przedziałowa miara ryzyka (rozkład możliwości miary ryzyka dla strat l -tego rodzaju) będzie więc:

$$\tilde{R}^l = \tilde{f}_{KS}^l \tilde{n}_{KS}^l = \sum_k \tilde{f}_{KS_k}^l \tilde{n}_{KS_k}^l \quad (19)$$

Gdyby policzyć wskaźnik ryzyka na podstawie wartości modalnych składowych wektorów \tilde{f}_{KS}^l i \tilde{n}_{KS}^l przez zastosowanie wzoru (19) uzyskałoby się modalną (punktową) wartość ryzyka

$$R^l = \sum_k f_{KS_k}^l n_{KS_k}^l \quad (20)$$

która może się nieznacznie różnić od wartości modalnej ryzyka obliczonej na podstawie przedziału rozmytego \tilde{R}^l , wyznaczonego na podstawie wzoru (19), co zasygnalizowano w rozdziale 3 niniejszego referatu. Uzyskane wartości przedziałowe ryzyka oraz wartość modalna używa się do oceny i porównywania ryzyka z wartościami lub przedziałami kryterialnymi.

Przedstawiona metoda jest w trakcie implementacji w systemie komputerowym wspomagania analizy ryzyka SKWAR [6].

6. Uwagi i wnioski

Modelowanie probabilistyczne i oszacowanie ryzyka eksploatacji złożonego obiektu przemysłowego przeprowadza się w warunkach niepewności. W związku ze szczupłością dostępnych danych niezawodnościowych analityk zmuszony jest korzystać z różnych źródeł informacji, w tym oszacowań dokonywanych przez ekspertów, w których stosuje się różne sposoby reprezentacji niepewności. W referacie proponuje się podejście integrujące różne metody reprezentacji niepewności na rozkład możliwości w postaci liczb lub przedziałów rozmytych. Zarysowano dwie metody transformacji rozkładu prawdopodobieństwa na rozkład możliwości: Dubois i Prade'a [1], [2] oraz Klira [5]. Do badania propagacji niepewności w modelu ryzyka zaproponowano metodę wykorzystującą oszacowania przedziałowe i arytmetykę rozmytą. Aby zapobiegać nadmiernej propagacji niepewności zaproponowano stosowanie arytmetyki rozmytej z ograniczeniami. Zaleca się pewną ostrożność w wyznaczaniu i interpretowaniu wartości modalnych miar ryzyka, gdyż uzyskane wartości mogą różnić się zależnie od przyjętej metody obliczeniowej. W przypadkach praktycznych spotykanych w analizach ryzyka różnica ta mieści się zwykle w zakresie kilkunastu procent

Literatura

1. Dubois, D. & Prade, H.: *Unfair coins and necessity measures: towards a possibilistic interpretation of histograms*. Fuzzy Sets and Systems, Vol. 10, 1983:15-20.
2. Dubois, D. & Prade, H.: *Fuzzy sets and statistical data*. European J. Operational Research, Vol. 25, 1986: 345-356.
3. Higashi, M. & Klir, G.J.: *On the notion of distance representing information closeness: Possibility and probability distributions*. Intern. J. General Systems, Vol. 9 No. 2, 1983: 103-115.
4. Klir, G.J.: *Developments in Uncertainty-Based Information*. In: *Advances in Computers* (M.C. Yovits Ed.). New York: Academic Press 1993: 255-322.

5. Klir, G.: *Fuzzy arithmetic with requisite constraints*. Fuzzy Sets and Systems Vol. 91, 1997: 165-175.
6. Kosmowski K.T., Downarowicz O., Duzinkiewicz K, Kwiesielewicz M: *Metody i narzędzia analizy ryzyka eksploatacji złożonych obiektów przemysłowych*. Opracowanie wykonane w ramach Strategicznego Programu Rządowego SPR-1, zadanie badawcze nr 02.6.11; Politechnika Gdańska, listopad 1999.
7. Moore, R.: *Interval Analysis*. Englewood Cliffs: Prentice Hall, 1996.
8. PrPN-IEC 300-3-9: *Analiza ryzyka w systemach technicznych* (IEC 300-3-9 z roku 1995). Norma PN-IEC 300-3-9 ustanowiona w Polsce w czerwcu 1999.
9. Shafer, G.: *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.
10. Shannon C.E.: *The mathematical theory of Communication*. Bell System Technical Journal, Vol. 27, 1948: 379-423, 623-656.
11. Zadeh, L.A.: *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets and Systems Vol. 1, No. 1, 1978: 3-28.

Streszczenie

Praca dotyczy problematyki reprezentacji i oceny niepewności w modelowaniu probabilistycznym i szacowaniu ryzyka związanego z eksploatacją systemu technicznego. Przedstawia się źródła danych niezawodnościowych oraz charakteryzuje się podstawowe reprezentacje związane z ich niepewnością. Ze względu na różny jej charakter wprowadza się transformacje z postaci probabilistycznej na postać posybilistyczną, proponując następnie wykonywanie dalszych obliczeń na przedziałach rozmytych, co sprowadza się do analizy przedziałowej. Przedstawiono metodę oszacowania ryzyka z wykorzystaniem zaproponowanej metodyki.

RISK ASSESSMENT METHOD IN TECHNICAL SYSTEMS USING INTERVAL ANALYSIS

Summary

The paper concerns uncertainty representation and evaluation in probabilistic modeling and risk estimation in exploitation of a technical system. Sources of reliability data are presented and their basic uncertainty representations are pointed-out. Considering a different character of the data uncertainty probability-possibility transformations are used in order to get an uniform set of data. Next fuzzy interval arithmetic led to interval calculations is used. Finally a method of risk estimation is presented basing on the approach proposed.

Referat recenzowany

EFEKTYWNE METODY WYZNACZENIA BRZEGU PARETO I PUNKTU SALUKWADZE

1 Wstęp

Szereg zadań optymalizacyjnych spotykanych w praktyce charakteryzuje się koniecznością uwzględnienia wielu na ogół sprzecznych kryteriów [3]. Prowadzi to do tzw. zadań polioptymalizacyjnych, przy rozwiązywaniu których korzysta się z koncepcji Pareto, Salukwadze [6], skalaryzacji, a także z metody NBI (Normal Boundary Intersection) – nowego podejścia zaproponowanego przez I.Dasa i J. Dennisa [1, 2].

Celem niniejszej pracy jest przedstawienie tych koncepcji na prostych przykładach numerycznych. W pracy w zadaniu optymalizacji dwukryterialnej wyznaczono numerycznie brzeg Pareto i rozwiązanie optymalne w sensie Salukwadze metodą skalaryzacyjną i metodą NBI. Okazuje się, że najbardziej popularna metoda skalaryzacyjna ma poważną wadę, która utrudnia często dokładne wyznaczenie brzegu Pareto i leżącego na nim punktu Salukwadze. Lepszą od niej jest metoda NBI, która poprzez prawie równomierne rozłożenie punktów na brzegu Pareto dokładniej go przybliża, a także daje lepsze możliwości oszacowania położenia punktu Salukwadze.

W pracy zasygnalizowano również ważny problem równoważności zadań: optymalizacji w sensie Pareto, skalaryzacyjnego oraz zadania NBI. W ogólnej sytuacji wspomniane zadania nie są równoważne.

2 Optimum w sensie Pareto

Niech Q będzie niepustym zbiorem rozwiązań dopuszczalnych w R^n , $f_i : R^n \rightarrow R$, $i = 1, \dots, s$ są danymi funkcjami, a $F = [f_1, \dots, f_s]^T$ jest wektorowym wskaźnikiem jakości.

Rozważmy następujące zadanie optymalizacji wektorowej.

Zadanie (P)

Znaleźć x^0 takie, że

$$\text{Pareto } \min_{x \in Q \cup U(x^0)} F(x) = F(x^0)$$

gdzie $U(x^0)$ jest pewnym otoczeniem punktu x^0 .

* Instytut Informatyki, Uniwersytet Śląski, ul. Będzińska 60, 41-200 Sosnowiec, kotarski@gate.math.us.edu.pl

** Instytut Informatyki, Uniwersytet Śląski, ul. Będzińska 60, 41-200 Sosnowiec, pyszny@us.edu.pl

Definicja 1 Punkt $x^0 \in R^n$ nazywa się globalnym (lokalnym) Pareto optimum dla Zadania (P) wtedy i tylko wtedy, gdy $x^0 \in Q$ i nie istnieje $x \neq x^0$ takie, że $x \in Q$ ($Q \cap U(x^0)$) i spełnione są $f_i(x) \leq f_i(x^0)$ z co najmniej jedną nierównością ostrą dla pewnego $i, 1 \leq i \leq s$.

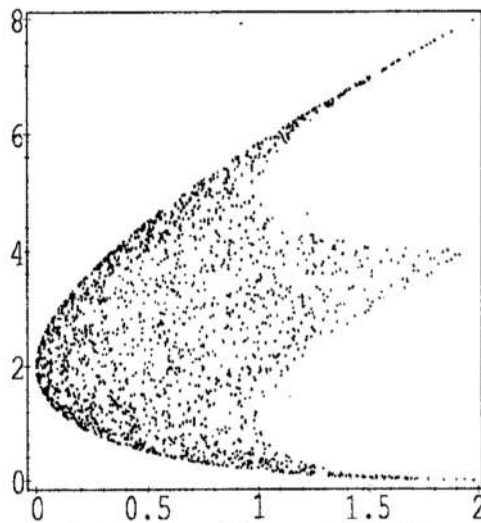
Definicja 2 Zbiór $Y := \{y = F(x); x \in Q\} \subset R^s$ nazywa się przestrzenią kryterialną.

Y jest zatem obrazem zbioru rozwiązań dopuszczalnych poprzez funkcje wektorową F.

Definicja 3 Iloczyn kartezjański $Y_E := Y_1 \times Y_2 \times \dots \times Y_s \subset R^s$, gdzie $Y_i = f_i(Q), i=1, \dots, s$ nazywa się rozszerzoną przestrzenią kryterialną.

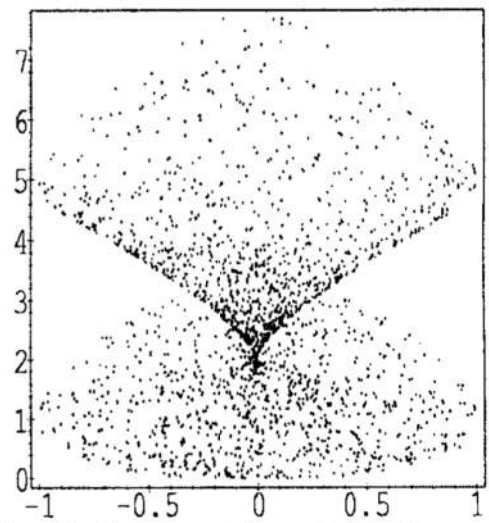
W prostokątnym układzie współrzędnych z osiami f_1, \dots, f_s obrazy rozwiązań optymalnych w sensie Pareto leżą na tzw. brzegu Pareto, który jest częścią brzegu przestrzeni kryterialnej Y.

W szczególnych przypadkach analitycznie można wyznaczyć przestrzeń kryterialną Y i brzeg Pareto. Na ogół jest to jednak niemożliwe. W przykładach podanych poniżej numerycznie wyznaczono przestrzeń kryterialną Y i brzeg Pareto.



Rys. 1. Przestrzeń kryterialna i brzeg Pareto dla przykładu 1

Fig. 1. Criterial space and Pareto boundary for example 1



Rys. 2. Przestrzeń kryterialna i brzeg Pareto dla przykładu 2

Fig. 2. Criterial space and Pareto boundary for example 2

Przykład 1 Niech

$$f_1(x_1, x_2) = x_1^2 + x_2^2$$

$$f_2(x_1, x_2) = (1-x_1)^2 + (1-x_2)^2$$

$$i Q = \{(x_1, x_2) \in R^2 : -1 \leq x_1, x_2 \leq 1\}$$

Przykład 2 Niech

$$f_1(x_1, x_2) = x_1^2 - x_2^2$$

$$f_2(x_1, x_2) = (1 - x_1)^2 + (1 - x_2)^2$$

$$i Q = \{(x_1, x_2) \in R^2 : -1 \leq x_1, x_2 \leq 1\}$$

W przykładach tych punkty z kwadratu $-1 \leq x_1, x_2 \leq 1$ były generowane losowo. Następnie przez $F = [f_1, f_2]^T$ otrzymano ich obrazy, które są elementami przestrzeni kryterialnej Y . Biorąc dostatecznie dużą liczbę punktów (w przykładach 2000) można otrzymać dostatecznie dobrą aproksymację przestrzeni kryterialnej Y . Znając kształt tejże przestrzeni dysponujemy wstępną informacją o brzegu Pareto.

Kształt przestrzeni kryterialnej Y w ogólnym przypadku jest trudny do przewidzenia. Wiadomo tylko, iż zbiór wypukły Q poprzez odwzorowanie liniowe przechodzi w zbiór wypukły, dając wypukłą przestrzeń kryterialną.

3 Optimum w sensie Salukwadze

Rozwiązań optymalnych w sensie Pareto dla zadania (P) na ogół jest nieskończenie wiele. Powstaje dlatego dylemat, które z rozwiązań Pareto optymalnych winno być wybrane jako najlepsze rozwiązanie. Jedną z metod rozwiązania tego problemu zaproponował Salukwadze [6]. Według jego propozycji należy najpierw niezależnie każdą funkcję f_i minimalizować na zbiorze Q . Punkt $F^* = [f_1^*, \dots, f_s^*]^T \in Y_E$ nazywa się punktem utopijnym lub idealnym. Punkt leżący na brzegu Pareto, który jest najbliższy (w sensie wybranej metryki np. euklidesowej) punktowi utopijnemu nazywa się punktem Salukwadze. Punkt Salukwadze jest obrazem rozwiązania optymalnego w sensie Salukwadze poprzez funkcję wektorową F . Rozwiązanie optymalne w sensie Salukwadze realizuje zatem pewien kompromis – spośród punktów optymalnych w sensie Pareto wybiera się takie rozwiązanie, które leży najbliżej punktu idealnego.

4 Metoda skalaryzacyjna

Z zadaniem (P) związane jest następujące zadanie minimalizacji skalarnej:

Zadanie (S)

Znaleźć x^0 takie, że

$$\sum_{i=1}^s \lambda_i f_i(x^0) = \min_{x \in Q \cap U(x^0)} \sum_{i=1}^s \lambda_i f_i(x)$$

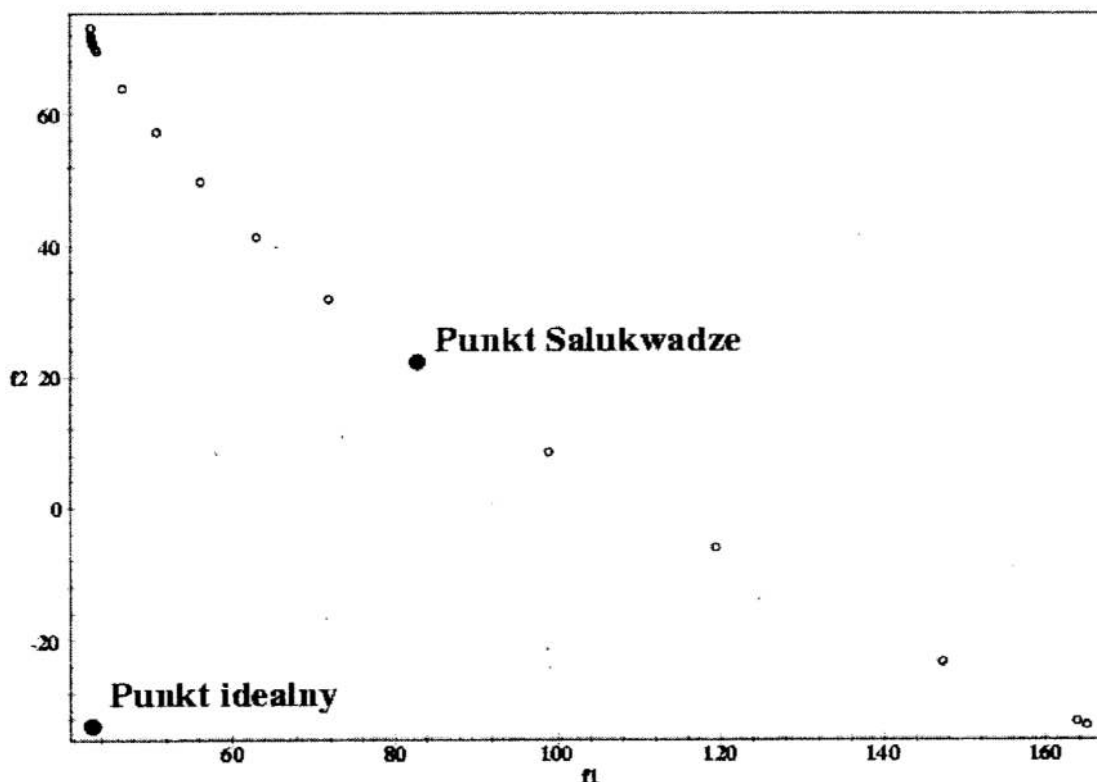
$$\text{gdzie } \lambda_i > 0, i = 1, \dots, s, \sum_{i=1}^s \lambda_i = 1.$$

Zadanie (P) przy pewnych dodatkowych założeniach jest równoważne z zadaniem (S) w następującym sensie: jeśli x^0 jest rozwiązaniem Pareto optymalnym zadania (P), to

x^0 jest również rozwiązaniem optymalnym zadania (S) przy pewnych ustalonych $\lambda_i > 0, i = 1, \dots, s, \sum_{i=1}^s \lambda_i = 1$, i odwrotnie.

Warunki, przy których zadania (P) i (S) są równoważne podano np. w [4, 5]. Rozwiązując zadanie (S) przy różnych λ_i można otrzymać punkty leżące na brzegu Pareto. Niestety na ogół równomierny rozkład parametrów λ_i nie prowadzi do równomiernego rozkładu punktów na brzegu Pareto. Wynika to z niejednostajnej parametryzacji brzegu Pareto przy pomocy λ_i . Aby zatem uzyskać dobre przybliżenie brzegu Pareto należy znaleźć wstępnie taki nierównomierny rozkład λ_i , który zapewni równomierny rozkład punktów na brzegu Pareto. Znalazienie takiego rozkładu λ_i w ogólnym przypadku jest niemożliwe.

Stosując metodę skalaryzacyjną z równomiernym rozkładem λ otrzymano brzeg Pareto dla przykładu 3, w którym wyraźnie widoczny jest nierównomierny rozkład punktów na brzegu Pareto (rys. 3).



Rys. 3. Brzeg Pareto otrzymany metodą skalaryzacyjną

Fig. 3. Pareto boundary obtained by scalarization method

Przykład 3 Rozwiązać zadanie optymalizacji dwukryterialnej:

$$\text{Pareto min} \begin{bmatrix} x_1^2 + 5x_2^2 + x_3^2 + x_4 + x_5 + x_6 \\ x_1 - 2x_2 - 3x_3 + 4x_4 + 5x_5 + x_7 \end{bmatrix}$$

przy ograniczeniach:

$$\left\{ \begin{array}{rcl} -2x_1 + 3x_2 + x_3 & & -9 = 0 \\ x_1 + x_2 & + x_4 & -8 = 0 \\ 3x_1 - 2x_2 & & + x_5 - 9 = 0 \\ & & -x_6 - 5 \leq 0 \\ & & -x_7 - 2 \leq 0 \\ & & x_6 + x_7 + 4 \leq 0 \\ & & x_1, \dots, x_5 \geq 0 \end{array} \right.$$

Wyniki obliczeń przedstawiono na rysunku 3.

5 Metoda NBI [1, 2]

Niech $f_i(x_i^*) = \min_{x \in Q} f_i(x_i)$, $i = 1, \dots, s$, oraz oznaczmy przez $F_i^* = F(x_i^*) =$

$$[f_1(x_i^*), \dots, f_s(x_i^*)]^T.$$

Następnie oznaczmy przez Φ macierz $s \times s$, w której i -ta kolumna ma postać $F_i^* - F^*$ i utwórzmy następujący zbiór.

$$PW := \left\{ \Phi\beta : \beta \in R^s, \sum_{i=1}^s \beta_i, \beta_i > 0, i = 1, \dots, s \right\}$$

$\Phi\beta$ reprezentuje punkt należący do PW . Niech $\hat{\eta}$ oznacza jednostkowy wektor normalny do zbioru PW wskazujący początek układu współrzędnych. Wtedy $\Phi\beta + t\hat{\eta}$ dla $t \in R$ reprezentuje punkty leżące na tej normalnej.

Punkt przecięcia normalnej z brzegiem Y dla ustalonego β najbliższy początkowi układu współrzędnych jest globalnym rozwiązaniem następującego zadania:

Zadanie (NBI)

Znaleźć (x^0, t^0) takie, że

$$\max_{(x, t)}$$

przy ograniczeniach:

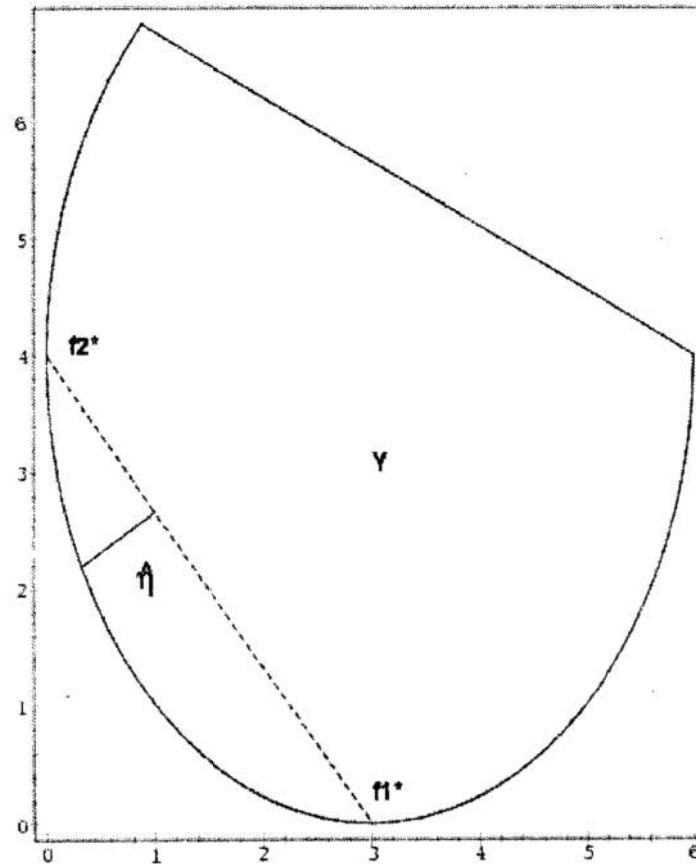
$$\Phi\beta + t \cdot \hat{\eta} = F(x) \text{ oraz } x \in Q \cap U(x^0)$$

$$\text{gdzie } \beta \in R^s, \sum_{i=1}^s \beta_i = 1, \beta_i > 0, i = 1, \dots, s$$

Rozważania powyższe w R^2 ilustruje rysunek 4. Zaznaczono na nim normalną do odcinka $\langle f_1^*, f_2^* \rangle$ odpowiadającą pewnemu $\beta \in R^2$ o dodatnich składowych, która przecinając się z brzegiem przestrzeni kryterialnej Y wyznacza na nim punkt brzegu Pareto. Punktowi temu, który jest obrazem x^0 poprzez odwzorowanie $[f_1, f_2]^T$ odpowiada wartość parametru t^0 . Zmieniając równomiernie β otrzymuje się rodzinę równoległych normalnych do odcinka $\langle f_1^*, f_2^* \rangle$, które przecinając się z brzegiem Y wyznaczają zbiór punktów prawie równomiernie rozłożonych na brzegu Pareto.

Dla zadań wielokryterialnych z dwoma wskaźnikami jakości ($s=2$), przy pewnych dodatkowych założeniach zadanie (S) i (NBI) są równoważne w następującym sensie:

jeśli przy pewnych λ_i , x^0 jest optymalnym rozwiązaniem zadania (S), to x^0 jest również rozwiązaniem optymalnym zadania (NBI) przy pewnych β_i i odwrotnie.



Rys. 4. Idea metody NBI

Fig. 4. The idea of NBI method

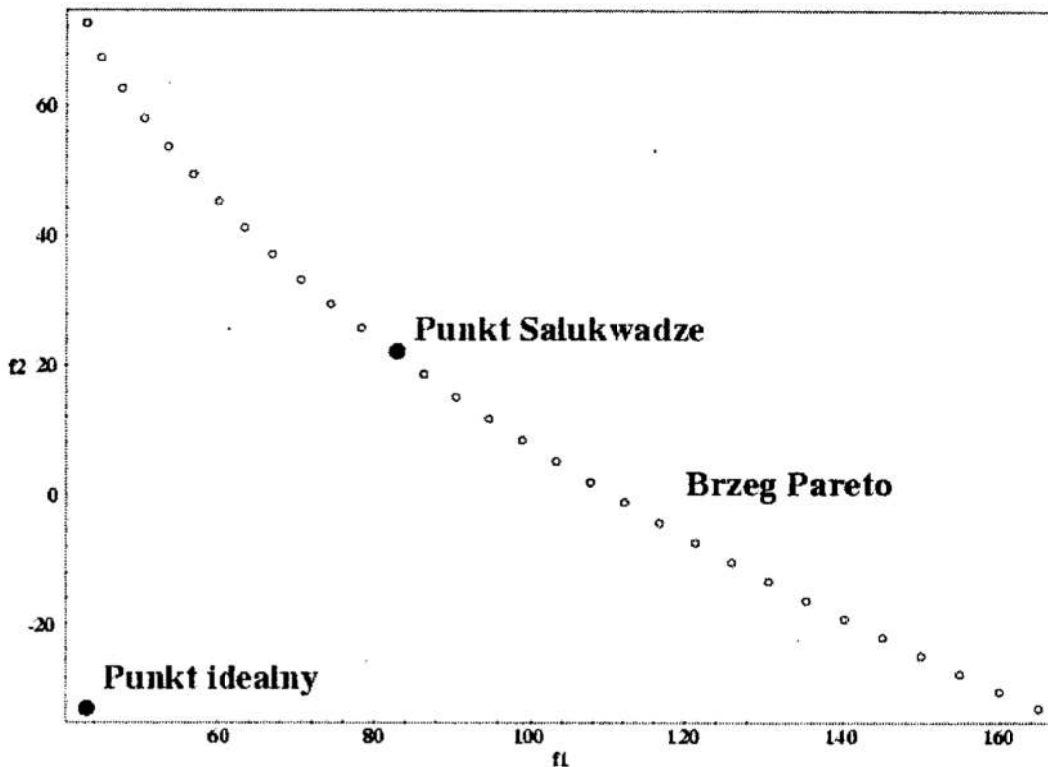
Zależności między zadaniem (S) i (NBI) podano w pracy [2]. Uwzględniając wyniki [2] i [5] można dla $s=2$ wskazać klasę zadań wielokryterialnych, dla których zadania (S), (P) i (NBI) są równoważne.

Okazuje się, że rozwiązanie zadania NBI prowadzi do prawie równomiernego rozkładu punktów na brzegu Pareto. Dzięki tej własności metoda NBI jest lepsza niż metoda skalaryzacyjna, gdyż daje dokładniejsze przybliżenie brzegu Pareto.

Zadanie z przykładu 3 zostało rozwiązane również przy pomocy metody NBI. Wyniki przedstawiono na rysunku 5.

Wyraźnie widoczny jest równomierny rozkład punktów na brzegu Pareto. Dlatego łatwo jest wyznaczyć w tym przypadku położenie punktu Salukwadze. W zadaniu tym punkt idealny ma współrzędne: $[42.8393, -33.0000]^T$, przy czym pierwszej składowej odpowiada punkt $[0.0000, 1.8929, 3.3214, 6.1071, 12.7857, -5.0000, -1.8709]^T$ a dla drugiej odpowiednio $[5.0000, 3.0000, 10.0000, 0.0000, 0.0000, -5.0000, -2.0000]^T$. Dla normy euklidesowej wyznaczono również punkt Salukwadze $[82,2115; 22,2259]^T$ oraz rozwiązanie optymalne w sensie Salukwadze

$$x_s = [2,3647; 2,3830; 6,5804; 3,2523; 6,6719; -5,0000; -2,0000]^T.$$



Rys. 5. Brzeg Pareto otrzymany metodą NBI

Fig. 5. Pareto boundary obtained by NBI method

6 Uwagi końcowe

Rozważanie zaprezentowane w pracy mogą być uogólnione na problemy optymalizacji wektorowej w przestrzeniach nieskończenie wymiarowych. W szczególności w [5] były rozpatrywane zadania sterowania optymalnego dla systemów o parametrach rozłożonych z wektorowym wskaźnikiem jakości. Obliczenia numeryczne dla przykładu 3 były wykonane w oparciu o implementację metody NBI w Matlabie 5.0, w której dodano opcję wyznaczenia punktu Salukwadze i rozwiązania optymalnego w sensie Salukwadze. Implementacja ta jest dostępna na stronie internetowej: <http://www.owl.net.rice.edu/~indra/NBI/homepage.html>. Kształt przestrzeni kryterialnych w przykładach 1, 2 uzyskano przy pomocy programu Maple.

Literatura

- [1] Dennis J., Das I.: *A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriterial optimization problems*. Structural Optimization, 14, No. 1, (1997), 63-69.
- [2] Dennis J., Das I.: *Normal-Boundary Intersection – a new method for generating the Pareto surface in nonlinear – multicriterial optimization problems*. SIAM J. Optimiz, 8, No. 3, (1989), 631-657.
- [3] Górecki H.: *Optymalizacja systemów dynamicznych*. PWN, Warszawa, 1993.
- [4] Kotarski W.: *Characterization of Pareto optimal points in problems with multi-equality constraints*. Optimization, 20, (1989), 93-106.

- [5] Kotarski W.: *Some problems of optimal and Pareto optimal control for distributed parameter systems*. Zeszyty naukowe Uniwersytetu Śląskiego, No. 1667, Katowice 1997, 1-93.
- [6] Salukwadze M. E.: *Zadaci wiektorowej optymalizacji w teorii optymalnego uprząwienia*. Miecniereba, Tbilisi, 1975.

Streszczenie

W pracy omówiono dwie znane metody wyznaczania brzegu Pareto i punktu Salukwadze dla zadań optymalizacji wielokryterialnej, a mianowicie metodę skalaryzacyjną i nową metodę NBI (Normal Boundary Intersection). Druga z tych metod daje równomierny rozkład punktów na brzegu Pareto, w przeciwieństwie do metody skalaryzacyjnej. Metodę NBI rozszerzono o opcję wyznaczenia punktu Salukwadze i rozwiązania optymalnego w sensie Salukwadze.

ON EFFECTIVE METHODS OF DETERMINING OF PARETO BOUNDARY AND SALUKWADZE POINT

Summary

Two known methods for determining Pareto boundary and Salukwadze point for multicritical optimization problems, namely scalarization method and a new NBI method (Normal Boundary Intersection) are discussed. The second method ensures an even spread of points on Pareto boundary contrary to scalarization method. NBI method has been equipped with the possibility of determining of Salukwadze point and optimal solution in the Salukwadze sense.

Referat recenzowany

METODA OKREŚLANIA PODOBIENSTWA STATKÓW W PROCESIE DOBORU UKŁADÓW AUTOMATYKI SYSTEMÓW ENERGETYCZNYCH

1 Wprowadzenie

Jedną z ogólnych zasad rozwiązywania problemów polega na wykorzystaniu doświadczenia zdobytego przy realizacji w przeszłości podobnych problemów. Odnosi się to także do zadań projektowych.

W procesie projektowania statków, szczególnie w jego fazach początkowych, wykorzystanie informacji dotyczących projektów już wykonanych jest szeroko stosowane. Jest to realizowane zarówno poprzez nagromadzone doświadczenie projektanta, jak również na drodze bezpośredniego wykorzystania dokumentacji projektowej.

W dziedzinie sztucznej inteligencji zasygnalizowane podejście rozważane jest w sposób sformalizowany w postaci metodologii zwanej wnioskowaniem na podstawie przypadków (ang. Case-Based Reasoning – CBR) [1–4].

W referacie przedstawione zostało zastosowanie tej metodologii do rozwiązania jednego z zadań wykonywanych w ramach prac nad systemem z bazą wiedzy dla wspomaganie projektowania układów automatyki okrętowej [5].

Wnioskowanie na podstawie przypadków sprowadza się do 4-etapowego cyklu następujących czynności:

- wyszukanie w bazie danych przypadków rozwiązań zrealizowanych podobnych do przypadku rozwiązywanego,
- wykorzystanie znalezionej odpowiedzi wraz z ewentualną jej adaptacją,
- weryfikacja opracowanego rozwiązania,
- zapamiętanie uzyskanego rozwiązania w bazie danych (przypadków).

Podstawowym problemem w zastosowaniach wnioskowania na podstawie przypadków jest zdefiniowanie metody określania podobieństwa przypadku rozwiązywanego i poszczególnych przypadków zrealizowanych. Ogólnie metoda taka bazuje na wartościach pewnych cech przypadków. Cechy te mogą mieć różnorodny charakter: liczbowy, binarny, tekstowy, obiektowy i inny. Metoda określania podobieństwa winna uwzględniać jak najwięcej cech oraz ich różnorodność i w wyniku dawać jednoznaczną miarę podobieństwa.

W zadaniu wspomaganie projektowania automatyki siłowni statku podobieństwo odnosi się do cech charakteryzujących siłownie statków zbudowanych. Zakłada się bowiem, że rozwiązania dotyczące automatyki są uwarunkowane pewnymi, wytypowanymi cechami siłowni statku. Ze względu na dużą liczbę uwzględnianych cech, podobieństwo statków określane jest w rozbiciu na pewne grupy cech.

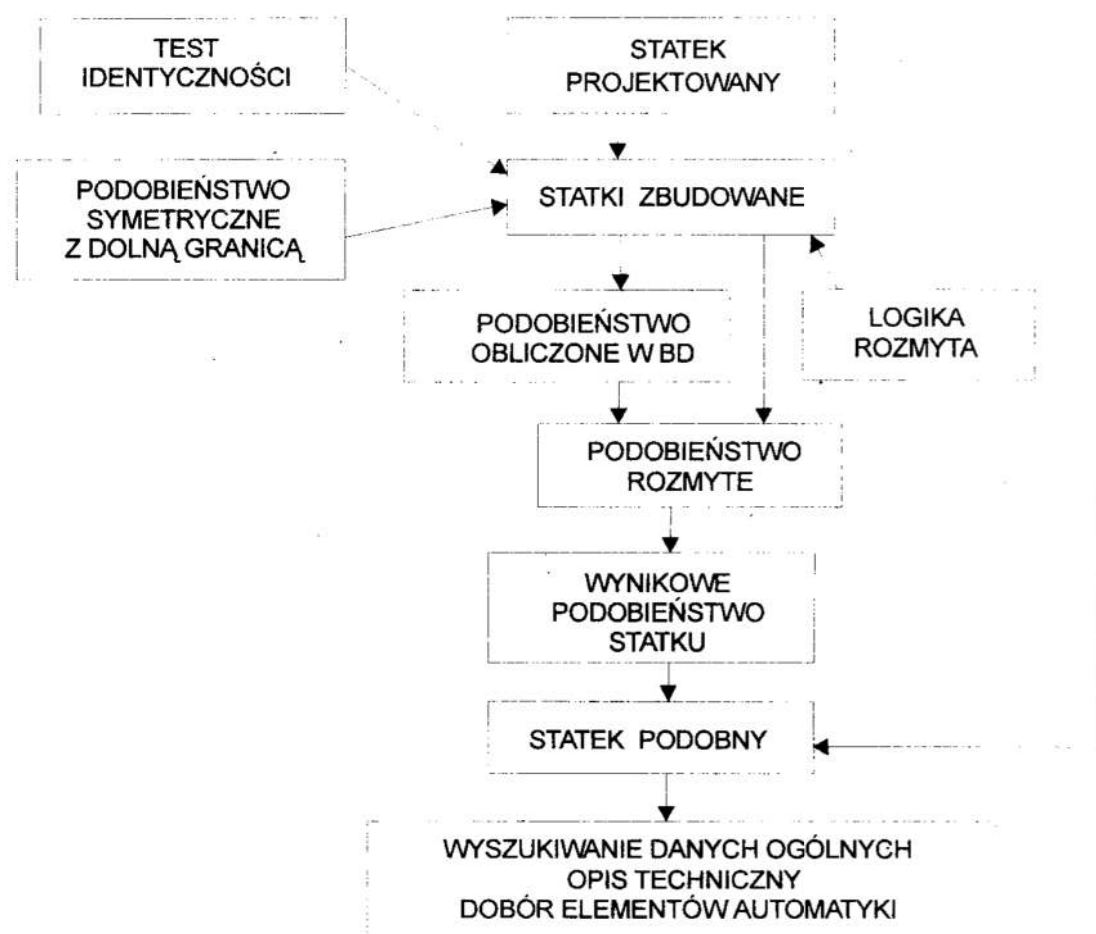
* *Politechnika Gdańska, 80-952 Gdańsk, ul. Narutowicza 11/12,
E-mail: zkowal@sparc10.ely.pg.gda.pl*

Ze względu na ograniczone możliwości stosowania logiki rozmytej w systemie Exsys – tylko do zmiennych o wartościach rzeczywistych, w opracowanym algorytmie obliczania miary podobieństwa zastosowano trzy metody: sprawdzanie identyczności parametrów, wykorzystanie funkcji symetrycznego podobieństwa z dolną granicą oraz wnioskowanie w oparciu o logikę rozmytą.

Ostatnia z tych metod wykonywana jest w systemie ekspertowym Exsys, a pozostałe dwie w bazie danych Access. Zarówno w obliczeniach podobieństw częściowych, jak również w obliczeniach podobieństwa sumarycznego wprowadzono współczynniki wagowe, umożliwiające różnicowanie ważności poszczególnych parametrów oraz wyników podobieństw częściowych.

2 Ogólna charakterystyka algorytmu określania podobieństwa statków

Opracowany algorytm dla systemu wspomagania projektowania układów automatyki statku, służy do porównywania i wyszukiwania z bazy danych statków podobnych do statku projektowanego z ustaleniem stopnia ich podobieństwa. Znajduje on zastosowanie zarówno do wyszukiwania danych ogólnych o statku, jak również do tworzenia opisu technicznego oraz doboru elementów automatyki do punktów kontrolno-pomiarowych (rys. 1).



Rys. 1. Uproszczony schemat algorytmu obliczania podobieństwa statków
Fig. 1. Simplified diagram of algorithm for ship's similarity determination

Opracowany algorytm oparty został na podejściu systemowym, co oznacza obliczanie podobieństwa poszczególnych systemów statku projektowanego i zbudowanych tj. podobieństwa ogólnego statku, napędu głównego, elektrowni oraz wybranych instalacji np. zęzowej lub paliwa. Badanie podobieństwa obejmuje wybrane parametry pochodzące z tabeli STATKI ZBUDOWANE takie jak:

dane ogólne: typ statku, wyporność, liczba kontenerów chłodzonych, liczba samochodów, towarzystwo klasyfikacyjne, klasa automatyki;

napęd główny (NG): liczba silników głównych (SG), typ SG, moc SG, prędkość obrotowa SG, liczba śrub napędowych, rodzaj śrub napędowych, liczba przekładni;

elektrownia: liczba zespołów prądotwórczych (ZP1), typ ZP1, moc ZP1, prędkość obrotowa ZP1, liczba ZP2, typ ZP2, moc ZP2, prędkość obrotowa ZP2, typ prądnicy wałowej;

instalacje:

paliwa: liczba zaworów paliwa, liczba pomp paliwa, liczba wirówek, liczba filtrów;

zęzowa: liczba zaworów zęzowych, liczba pomp zęzowych.

Podobieństwa te obliczane są przy zastosowaniu następujących metod:

Test identyczności stosowany jest w odniesieniu do parametrów tekstowych, takich jak: typ statku, towarzystwo klasyfikacyjne, klasa automatyki, typ SG, rodzaj śrub, typ ZP1 i typ ZP2, typ prądnicy wałowej oraz niektórych parametrów liczbowych, jak na przykład: liczba śrub, liczba przekładni oraz liczba ZP2. Jeśli odpowiednie wartości tabeli STATEK PROJEKTOWANY i STATKI ZBUDOWANE są identyczne, w tabeli roboczej wpisywane jest 1, jeśli różne – 0.

Symetryczne podobieństwo z dolną granicą stosowane jest w odniesieniu do parametrów liczbowych o nieco większych zakresach wartości takich jak: liczba SG, liczba ZP1 oraz parametrów wybranych instalacji tj: liczby pomp paliwa, liczby wirówek i filtrów paliwa – dla instalacji paliwa, a liczby pomp zęzowych – dla instalacji zęzowej. W metodzie tej podobieństwo dla wybranego parametru określa się według zależności:

$$S(p) = 1 - \frac{|p_1 - p_2|}{\max(p_1, p_2) - pd} \quad (1)$$

gdzie:

p_1 – wartość parametru statku projektowanego,

p_2 – wartość parametru statku zbudowanego,

pd – dolna granica wartości parametru.

Logika rozmyta stosowana jest do wybranych parametrów liczbowych o wartościach rzeczywistych i nieporównywalnych skalach takich jak: wyporność statku, liczba kontenerów chłodzonych, moc i prędkość obrotowa silnika głównego, zespołów prądotwórczych oraz liczba zaworów wybranych instalacji (paliwa i zęzowej). Metoda ta pozwala określić miarę podobieństwa jednocześnie dla dowolnej liczby parametrów.

Realizacja dwu pierwszych metod w bazie danych polega na obliczeniu podobieństw pojedynczych parametrów, a następnie sum ważonych podobieństw cząstkowych dla: danych ogólnych, napędu głównego, elektrowni, wybranych instalacji i podobieństwa całego statku.

Fragment przykładowych obliczeń podobieństwa cząstkowego dla napędu głównego przedstawiono w tabeli 1. Statek projektowany oznaczono identyfikatorem XXX. W obliczeniach tych przyjęto dla poszczególnych kolumn następujące wagi: 0,2, 0,3, 0,1, 0,2 i 0,2.

Tabela 1. Dane NG statków zbudowanych wraz z ich podobieństwem
Table 1. The date of ships builded main propulsion with its similarities

Statek	Liczba SG	Typ SG	Liczba śrub	Rodzaj śrub	Liczba przekł.	Podob. NG
B191	1	6L70 MC	1	stała	1	1
B222	2	6L46	1	nast.	1	0,4
B369	1	6RTA 62-R1	1	stała	2	0,6
B501	4	8ZAL 40 S	2	nast.	2	0,05
B683	1	5RTA 62 U	1	stała	0	0,6
B684	1	5S 60 MC	1	stała	2	0,6
...						
XXX	1	6L70 MC	1	stała	1	

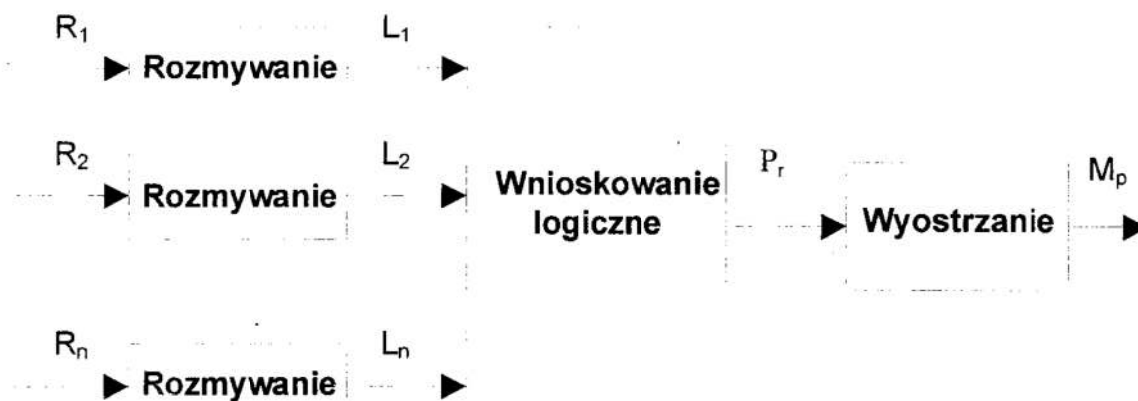
Obliczone podobieństwa w systemie bazy danych przekazywane są do systemu eksperckiego Exsys, gdzie poddawane są rozmyciu wraz z parametrami, których podobieństwo obliczane jest bezpośrednio przy zastosowaniu logiki rozmytej.

Po dokonaniu obliczeń podobieństwa w systemie Exsys do bazy danych przekazywane są, za pośrednictwem pliku tekstowego, wynikowe maksymalne podobieństwa cząstkowe: danych ogólnych, napędu głównego, elektrowni, wybranych instalacji wraz z identyfikatorami odpowiadających im statków oraz maksymalne podobieństwo sumaryczne statku. Na ich podstawie system bazy danych wyszukuje dane tego statku jako statku podobnego. Przy projektowaniu automatyki w obrębie pojedynczych systemów (instalacji) mogą być wykorzystywane statki określone przez maksymalne podobieństwa cząstkowe tych systemów.

3 Obliczenia podobieństwa z wykorzystaniem logiki rozmytej

W systemie ekspertowym Exsys Developer realizowane są obliczenia dotyczące tylko parametrów o wartościach typu liczbowego. Do parametrów tych zaliczono także podobieństwa cząstkowe w poszczególnych grupach podziału, określane w bazie danych. Do określania podobieństwa w poszczególnych grupach parametrów zastosowano metodę opartą na logice rozmytej (zbiorach rozmytych). Metodę tą zastosowano do różnic pomiędzy wartościami atrybutów statku budowanego i takich samych atrybutów poszczególnych statków zbudowanych, a także do podobieństw cząstkowych.

Metoda zbiorów rozmytych pozwala przejść od wartości liczbowych do wartości lingwistycznych (wyrażeń języka naturalnego) [6], za pomocą których w łatwy sposób można realizować wnioskowanie logiczne. Przejście takie uniezależnia od znaczenia i skali wartości liczbowych rozważanych atrybutów. Ogólny schemat realizacji omawianej metody określania podobieństwa przedstawiono na rysunku 2.



Rys. 2. Schematyczna prezentacja metody zbiorów rozmytych

Fig. 2. Diagram of fuzzy logic method

Na rysunku tym wprowadzono następujące oznaczenia:

R_1, R_2, \dots, R_n – różnice wartości liczbowych rozważanych atrybutów,

L_1, L_2, \dots, L_n – zbiory określeń lingwistycznych związane z przyjętymi zbiorami rozmytymi określonymi odpowiednio na wartościach R_1, R_2, \dots, R_n ,

P_r – podobieństwo rozmyte,

M_p – wynikowa liczbową miarą podobieństwa.

Wartość wynikowa liczbowej miary podobieństwa M_p zależy od (rys. 2):

- przyjętych funkcji przynależności, za pomocą których realizowane jest rozmywanie wartości wejściowych,
- definicji reguł, według których odbywa się wnioskowanie,
- sposobu realizacji wyostrzania.

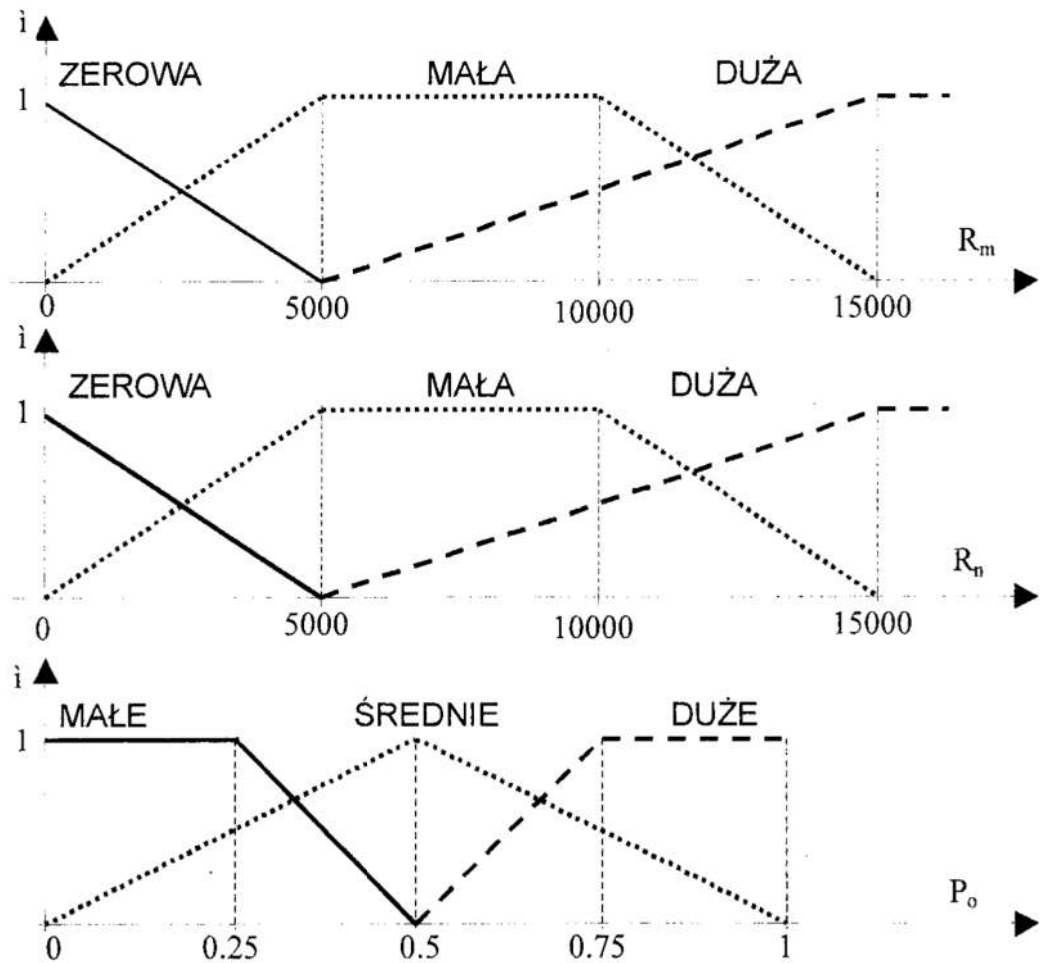
Przykładowo, w grupie napędu głównego określanie podobieństwa statków w systemie eksperckim Exsys bazuje na trzech następujących parametrach liczbowych:

- wartości bezwzględnej różnicy mocy silnika głównego statku projektowanego i statku zbudowanego (oznaczonej przez R_m),
- wartości bezwzględnej różnicy prędkości obrotowej SG statku projektowanego i statku zbudowanego (R_n),
- podobieństwie cząstkowym napędu głównego (określanym w bazie danych) i przypisanym do statku zbudowanego (P_o).

Do realizacji procesu rozmywania wartości tych parametrów zastosowano funkcje przynależności przedstawione na rysunku 3.

Przyjęte nazwy zbiorów rozmytych w dalszych rozważaniach traktowane są jako wartości lingwistyczne przypisane pytaniom w systemie Exsys. Pytania te służą do realizacji procesu wnioskowania.

W celu zmniejszenia liczby reguł wnioskowanie realizowane jest w dwu etapach. W pierwszym etapie kojarzone są wartości pytań określonych na różnicach R_m i R_n (oznaczone przez QR_m i QR_n). W drugim etapie uwzględnia się wyniki pierwszego etapu (związane z pytaniem Q_l) i wartości pytania określonego na zmiennej P_o (QP_o). Przykładowe zależności przyjęte w regułach systemu eksperckiego w obydwu etapach przedstawiono w tabelach 2 i 3.



Rys. 3. Funkcje przynależności dla parametrów napędu głównego
 Fig. 3. Membership functions for parameters of main propulsion

Tabela 2. Zależności logiczne przyjęte w pierwszym etapie wnioskowania

Table 2. Logical dependences in the first step of inference

Wartości pytań		
QR _m	QR _n	Q ₁
ZEROWA	ZEROWA	DUŻE
ZEROWA	MAŁA	DUŻE
MAŁA	ZEROWA	DUŻE
ZEROWA	DUŻA	ŚREDNIE
MAŁA	MAŁA	ŚREDNIE
DUŻA	ZEROWA	ŚREDNIE
MAŁA	DUŻA	MAŁE
MAŁA	MAŁA	MAŁE
DUŻA	DUŻA	MAŁE

Tabela 3. Zależności logiczne przyjęte w drugim etapie wnioskowania

Table 3. Logical dependences in the second step of inference

Wartości pytań		
Q ₁	QP _o	Q ₁₁
DUŻE	DUŻE	DUŻE
DUŻE	ŚREDNIE	DUŻE
ŚREDNIE	DUŻE	DUŻE
DUŻE	MAŁE	ŚREDNIE
ŚREDNIE	ŚREDNIE	ŚREDNIE
MAŁE	DUŻE	ŚREDNIE
ŚREDNIE	MAŁE	MAŁE
MAŁE	ŚREDNIE	MAŁE
MAŁE	MAŁE	MAŁE

System Exsys traktuje stopnie przynależności jako tzw. współczynniki pewności (ang. confidence values). Wnioskowanie, poprzez odpalanie odpowiednich reguł, powoduje obliczanie wartości współczynników pewności konkluzji reguł na podstawie wartości współczynników pewności przesłanek. Wartości wynikowych pytań uzyskują więc w procesie wnioskowania określone wartości współczynników pewności. W oparciu o te wartości można realizować wyostrzanie, w wyniku którego otrzymuje się jedną określoną wartość zmiennej liczbowej. Przykładowo, dla omawianej grupy napędu głównego, wyostrzanie realizowane jest według następującej zależności:

$$z = w_1c_1 + w_2c_2 + w_3c_3, \quad (2)$$

gdzie:

z – wartość liczbową podobieństwa NG (zawarta w przedziale $[0,1]$),

c_1, c_2, c_3 – wartości współczynników pewności odpowiednio dla wartości: DUŻE, ŚREDNIE, MAŁE, uzyskane w procesie wnioskowania,

w_1, w_2, w_3 – wagi o wartościach nieujemnych, $w_1 + w_2 + w_3 = 1$.

Identycznie jak dla napędu głównego określane są miary podobieństwa dla pozostałych grup – ogólnej, elektrowni i instalacji siłownianych. W oparciu o te miary cząstkowe wyliczane jest podobieństwo sumaryczne (jako suma ważona).

Cały proces obliczeniowy powtarzany jest dla każdego statku zbudowanego znajdującego się w odpowiedniej tablicy bazy danych. Głównym wynikiem obliczeń jest identyfikator statku o największej wartości podobieństwa sumarycznego – uznany za najbardziej podobny do statku projektowanego.

4 Podsumowanie

Wnioskowanie na podstawie przypadków stanowi stosunkowo nowe podejście do analizy oraz projektowania baz danych i baz wiedzy. Polega na generowaniu rozwiązań nowych problemów przez zaadaptowanie rozwiązań, które były zastosowane w przeszłości w podobnych sytuacjach. Obszar zastosowań tej metody jest szeroki i może obejmować np. diagnozowanie, klasyfikację, wspomaganie podejmowania decyzji, planowanie, projektowanie itp. Do tej pory powstało wiele komercyjnych systemów CBR, również dotyczących problematyki morskiej np. system dla projektowania koncepcyjnego statku – BASCON IV [3].

Istotnym elementem związanym z formalną reprezentacją przypadków jest zdefiniowanie sposobu ich porównania i wyszukiwania. Współczesne systemy baz danych stosują algorytmy wyszukiwania oparte na spełnieniu określonych w pytaniu warunków logicznych w logice dwuwartościowej, co może okazać się nie wystarczające dla systemów opartych na analizie przypadków.

Opracowany algorytm oparty został na metodzie wnioskowania na podstawie przypadków z zastosowaniem logiki rozmytej do porównywania i wyszukiwania statków podobnych, co umożliwi ustalenie właściwej miary podobieństwa statku projektowanego do statków zbudowanych. Algorytm ten zastosowano do wybranych parametrów charakteryzujących system energetyczny statku.

Literatura

1. Watson I: *Case-Based Reasoning is a methodology not a technology*. Knowledge-Based Systems, 12 (1999).
2. Zbroja S.: *Case-Based Reasoning w bazach danych i bazach wiedzy - wybrane aspekty formalnej reprezentacji przypadków*. II Krajowa Konferencja Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim. Kraków, 25-27 października 1999.
3. Lee D., Lee K. H.: *An approach to case-based system for conceptual ship design assistant*. Expert Systems with Applications, 16 (1999).
4. Suh M. S. i inni: *A case-based expert system approach for quality design*. Expert Systems with Applications, 15 (1998).
5. Kowalski Z., Zieliński S., Piotrowski J.: *Zastosowanie oprogramowania Resolver w systemie eksperckim wspomagania projektowania układów automatyki okrętowej*. Materiały XVI Ogólnopolskiej Konferencji Polioptymalizacja i Komputerowe Wspomaganie Projektowania. Politechnika Koszalińska, Koszalin 1998.
6. Yager R. R., Filev D. P.: *Podstawy modelowania i sterowania rozmytego*. WNT, Warszawa 1999.

Streszczenie

Przedstawiono metodę określania podobieństwa statku projektowanego do statków już zbudowanych. Metoda wykorzystuje wnioskowanie na podstawie przypadków. W zależności od cech parametrów charakteryzujących statki w metodzie wykorzystuje się: sprawdzenie identyczności, funkcję symetrycznego podobieństwa i zbiory rozmyte. W realizacji metody wykorzystuje się oprogramowanie bazy danych MS Access oraz system ekspercki Exsys Developer.

METHOD OF THE SHIPS SIMILARITY DETERMINATION DURING SHIP AUTOMATION DESIGN PROCESS

Summary

The paper presents a case-based reasoning method used to the similarity determination between a new ship and the ships from database. Subject to the types of the characteristic parameters the method use different techniques - identity identification, symmetrical similarity function and fuzzy logic. MS Access database software and Exsys Developer expert system were used in calculation according the method.

Referat recenzowany

METODA OCENY RÓWNOWAŻNOŚCI MODELI OPROGRAMOWANIA OBIEKTOWEGO

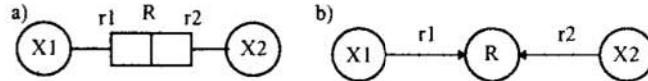
1 Rola oceny równoważności modeli oprogramowania obiektowego

W technikach obiektowych etap *analizy* jest związany z tworzeniem modeli programów obiektowych opartych na metodach wywodzących się z metody ER (*Entity Relationship*) [1], [2], [4]. W pracy do budowy modelu obiektowego wybrano metodę ORM (*Object Role Modelling*), która wywodzi się z metody ER i pozwala na budowę szczegółowego modelu klas i obiektów, syntezę wszystkich zależności funkcyjnych i niefunkcyjnych występujących w problemie [1], [3].

Równoważność schematów konceptualnych, związana jest z ich normalizacją, czyli usuwaniem redundancji i anomalii, występujących np. podczas modyfikacji danych [2], [3]. Normalizacja schematu odbywa się za pomocą jego transformacji, zachowującej równoważność schematu wejściowego i wyjściowego. W prowadzonych pracach celem badania równoważności modeli jest ocena wpływu różnych elementów zastosowanych do budowy równoważnych modeli np. polimorfizmu i specjalizacji na jakość oprogramowania. W niniejszej pracy równoważność schematów bada się za pomocą grafu funkcjonalnego diagramu ORM, opisanego w rozdziale 2. W kolejnych rozdziałach podano: zasady identyfikacji strukturalnej (3), zasady równoważności schematów i podstawy algorytmu do jej badania (4), przykład zastosowania opracowanej metody (5).

2 Wprowadzenie do metody modelowania ORM

Podstawowym elementem modelu ORM jest typ relacji oraz typ obiektowy. Relacja jest nie tylko podzbiorem produktu kartezjańskiego, lecz odwzorowaniem, określającym własności połączeń między elementami relacji. Typ relacji R , zwany typem faktu, stanowi połączenie między rolami $r1$ i $r2$, jakie obiekty $X1$ i $X2$ odgrywają w relacji R .



Rys. 1. a) Relacja R ; b) graf funkcjonalny diagramu z 1a

Fig. 1. a) Relationship R ; b) the functional graph of a model of the figure 1a

Definicja modelu ORM jest następująca [1], [3]:

Model ORM = (SI, Populacja, Ograniczenia), gdzie struktura informacyjna SI jest opisana jako zbiór elementarnych typów faktu nad zbiorem typów obiektowych:

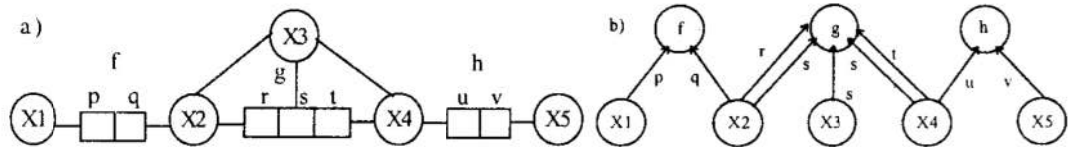
$$SI = (TOb, TWart, TFakt, TRol, Fakt, Baza, Spec, SpecRole, WRole, Pol, PolRole, Korzen).$$

Ze względu na ramy artykułu podano krótkie wyjaśnienia dotyczące elementów modeli ORM: $TWart$ - typ atrybutowy reprezentowany przez konkretne wartości (np. liczby); abstrakcyjne typy obiektowe TA ; skończony zbiór ról $TRol$, zbiór $TFakt$ wynikający z podziału zbioru $TRol$; elementy zbioru $TFakt$ określane jako typy faktu, spełniające zależność $TWart \cap TFakt \cap TA = \emptyset$; funkcja $Fakt: TRol \rightarrow TFakt$ określająca typ faktu w zależności od roli, która należy do niego, zdefiniowana jako $Fakt(p) = f \Leftrightarrow p \in f$; typy obiektowe (zbiór niepusty) $TOb = TA \cup TWart \cup TFakt$, funkcja $Baza: TRol \rightarrow TOb$ określająca typ obiektowy pełniący daną rolę; funkcja wyznaczająca pozostałe role należące do tego samego typu faktu $Inne_Role: \varphi(TRol) \rightarrow \varphi(TRol)$, relacje częściowego porządku: dziedziczenia typu specjalizacja („w dół”) - $Spec$ oraz dziedziczenia uogólniającego („do góry”) - Pol ; funkcja wyznaczająca praprzodka drzewa dziedziczenia $Korzen: TOb \rightarrow TOb$.

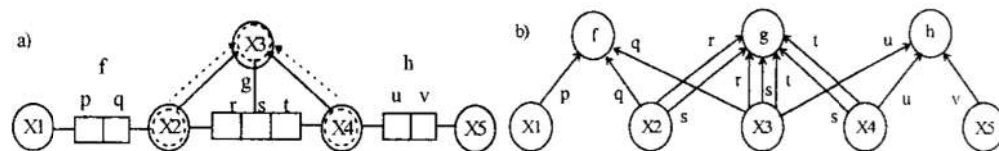
* Instytut Cybernetyki Technicznej Politechniki Wrocławskiej, ul. Janiszewskiego 11/17

E-MAIL: ZOSIA@ICT.PWR.WROC.PL

Graf funkcjonalny jest odwzorowaniem izomorficznym struktury informacyjnej i jest dwudzielnym grafem skierowanym, w którym zbiory wierzchołków: $V1$ tworzą typy faktów, $V2$ tworzą typy obiektowe, natomiast zbiór krawędzi tworzą role łączące wierzchołki z $V1$ z wierzchołkami z $V2$. Zbiór ról zawiera role wynikające z relacji specjalizacji i polimorfizmu. Na rysunku 1a przedstawiono diagram ORM binarnego typu faktu R i graf funkcjonalny tego diagramu na rysunku 1b [1], [3]. Na rysunkach 2b i 3b pokazano, przyjętą w tej pracy, zmodyfikowaną postać grafów funkcjonalnych przykładowych diagramów ORM z rysunków 2a i 3a, uwidaczniających role typów obiektowych, jakie wynikają z reguł specjalizacji ($WRole$) i polimorfizmu ($PolRole$) [3].



Rys. 2. a) Relacja Spec; b) grafy funkcjonalne diagramu z rysunku 2a
Fig. 2. a) Relationship Spec; b) the functional graphs of model of the figure 2a



Rys. 3 a) Relacja Pol; b) graf funkcjonalny diagramu z rysunku 3a
Fig. 3 a) Relationship Pol; b) the functional graph of a diagram of the figure 3a

Zbiór *Populacja* stanowi zbiór populacji *Pop* typów obiektowych tworzących strukturę informacyjną *SI*, określonych za pomocą ograniczeń ze zbioru *Ograniczenia* [3].

3 Identyfikacja strukturalna

Łączenie populacji ze strukturą informacyjną przebiega w procesie identyfikacji strukturalnej, gdzie kluczową rolę odgrywają ograniczenia *unikatowe* oraz *totalne*. Ograniczenie *totalne* oznacza, że populacja typów obiektowych związanych ze zbiorem τ danych ról ($total(\tau)$) jest równa całej populacji związanej funkcją *Baza* z każdą z tych ról. Ograniczenie *unikatowe* oznacza, że populacja typów obiektowych związanych ze zbiorem τ danych ról ($unikat(\tau)$) tylko raz wystąpi jako *Baza* z każdą z tych ról i wyznacza zależności funkcjonalne dla populacji modelu.

Niech r jest wyrażeniem relacyjnym i niech zbiory ról: τ, σ należą do wyrażenia r , czyli $\sigma, \tau \subseteq Schemat(r)$, gdzie r jest wyrażeniem relacyjnym, a $Schemat(r)$ jest wyrowadzany z typów faktów należących do wyrażenia. Wówczas, jeżeli populacja ról ze zbioru σ jest funkcjonalnie zależna od populacji ról ze zbioru τ , to mamy $Pop/\tau \rightarrow \sigma$. Zbiór ról τ , który należy do wyrażenia r ($\tau \subseteq Schemat(r)$), wtedy nazywamy kluczami kandydującymi, jeżeli $\tau \rightarrow Schemat(r)$ i τ jest minimalnym zbiorem, wg [1]:

$$\forall_{\sigma} [\tau \rightarrow Schemat(r)] \Rightarrow \sigma \subseteq \tau.$$

Rozróżnia się dwa przypadki wyznaczania ścieżek łączących role należące do zbioru ról klucza kandydującego (zbioru ról identyfikujących τ):

- jeżeli zbiór τ spełnia $\tau \subseteq f$, oznacza to, że rola łącząca też należy do typu faktu f ,
- jeżeli zbiór ról τ należy do więcej niż jednego typu faktu, należy wyznaczyć ścieżkę przebiegającą przez wszystkie typy obiektowe połączone przez typy faktów, zawierających role unikatowe ze zbioru τ . Definicja tego połączenia ma postać:

$$Połączenie(\tau) \equiv |Fakty(\tau)| > 1 \Rightarrow \exists_{p \in \tau} [\exists_{q \in \tau \setminus Fakt(p)} [(Fakt(p) \setminus \tau) = (Fakt(q) \setminus \tau)] \wedge Połączenie(\tau \setminus Fakt(p))], \text{ gdzie } Fakty(\tau) = \{Fakt(p) / p \in \tau\}.$$

W identyfikacji strukturalnej zakłada się, że każdy typ obiektowy jest dostępny w sposób jednoznaczny dopiero na podstawie właściwości struktury informacyjnej. Aby każdy typ obiektowy był rozpoznawalny, każdy typ atrybutowy musi wystąpić

w ograniczeniach *total*, czyli $\forall_{x \in TWart} \exists_{p \in TRol} \exists_{total(\tau) \in Ograniczenia} [Baza(p) = x \wedge p \in \tau]$, oraz wszystkie typy obiektowe są identyfikowane, jeżeli $\forall_{x \in TA} [JestIdF(x)]$.

Predykat *JestIdF* jest zdefiniowany [1], [3], jeżeli *x*:

1. jest typem atrybutowym, zawsze jest prawdziwa zależność *JestIdF(x)*
2. jest typem zbiorowym lub sekwencyjnym, to mamy *JestIdF(Elt(x))*
3. jest typem abstrakcyjnym, wtedy zachodzą następujące przypadki:
 - 3.1. jest podtypem i spełnia relację *x Spec y* lub *x Pol y* i *y Spec x* - wtedy *x* spełnia *JestIdF(x)*, gdy spełnione są zależności: *Korzen(x) ≠ x* oraz *JestIdF(Korzen(x))*, oraz istnieje unikatowa reguła specjalizacji lub polimorfizmu. Identyfikacja typów *Korzen(x)* przebiega według punktu 3.2
 - 3.2. jest typem elementarnym $x \in TE$, wtedy *x* jest identyfikowalny, jeżeli istnieje zbiór ról τ , który spełnia *JestIdF(\tau)*, czyli możliwa jest złożona identyfikacja

$\forall_{p \in \tau} [JestIdF(Baza(p))]; unikat(\tau); \forall_{f \in Fakty(\tau)} \exists_{p \in KoRole(x, \tau) \cap f} [unikat(p) \wedge total(p)]$,
gdzie $KoRole(x, \tau) = \{p \in \cup (Fakty(\tau) \setminus \tau) / Baza(p) = x\}$, czyli wszystkie role łączące identyfikowany typ obiektowy z rolami identyfikującymi ze zbioru τ są objęte ograniczeniami typu *total* i *unikat*.

Na podstawie zasad 1-3 zdefiniowano ścieżkę obejmującą role należące do zbioru identyfikującego τ oraz role łączące [3] i obejmującą tylko typy atrybutowe:

$$RoleIdF(x, \tau) = \{p / p \in Fakty(\tau) \wedge Połączenie(\tau) \wedge \forall_{p \in \tau} [p \in KoRole(x, \tau) \wedge unikat(p) \wedge total(p)]\}$$

Funkcja *AtrybutyIdF*: $\varphi(TOb, TRole) \rightarrow \varphi(TWart)$ wyznacza zbiór typów atrybutowych dla danego zbioru ról identyfikujących τ dla ścieżki *RoleIdF*:

$$AtrybutyIdF(x, \tau) = \{y / y \in TWart \wedge \exists_{p \in RoleIdF(x, \tau)} [Baza(p) = y]\}.$$

W przypadku identyfikacji złożonej danego typu abstrakcyjnego *x* można wyznaczyć wypadkowy zbiór identyfikujący, który jest sumą zbiorów identyfikujących τ_i wszystkich typów obiektowych abstrakcyjnych x_i biorących udział w identyfikacji złożonej. Zbiory identyfikujące τ_i spełniają następującą zależność: istnieje przynajmniej jedna rola *p* w τ_i oraz rola łącząca *q* w zbiorze *RoleIdF(x_{i+1}, \tau_{i+1})* taka, że spełniona jest między nimi zależność *Baza(q) = Baza(p)*. Suma zbiorów identyfikujących również spełnia warunki wyznaczone przez zależność *Połączenie(\tau)*. Oznacza to, że [3]

$$SumaIdF(x, (\tau_i)_{i \in n}) = \{p \in \cup_{i=1..n} \tau_i / \exists_{p \in RoleIdF(x_i, \tau_i)} \exists_{q \in KoRole(x_{i+1}, \tau_{i+1})} [Baza(p) = Baza(q)]\}$$

Dla $i=1$ zbiór identyfikujący jest określony jako *SumaIdF(x, \tau) = \tau*. Stąd ścieżka

$$RoleIdF'(x, (\tau_i)_{i \in n}) = \{p / p \in Fakty(SumaIdF(x, (\tau_i)_{i \in n})) \wedge Połączenie(SumaIdF(x, (\tau_i)_{i \in n})) \wedge \forall_{p \in SumaIdF(x, (\tau_i)_{i \in n})} [p \in KoRole(x, SumaIdF(x, (\tau_i)_{i \in n}))]\}.$$

Zbiór atrybutów identyfikujących jest teraz zdefiniowany jako

$$AtrybutyIdF''(x, (\tau_i)_{i \in n}) = \{y / y \in TWart \wedge \exists_{p \in RoleIdF'(x, (\tau_i)_{i \in n})} [Baza(p) = y]\}$$

Proces identyfikacji strukturalnej polega więc na wyznaczeniu za pomocą ścieżki identyfikującej *RoleIdF* lub *RoleIdF''* zbiorów jedno- lub wieloelementowego *AtrybutyIdF* lub *AtrybutyIdF''* typów atrybutowych.

4 Metoda oceny równoważności modeli ORM

Schematy konceptualne [3]

$SK_1 = (SI_1, Ograniczenia_1, POP_1)$, $SK_2 = (SI_2, Ograniczenia_2, POP_2)$
jeżeli mają identyczne zbiory typów atrybutowych

$TWart_1 \equiv TWart_2$, gdzie $TWart_1 \subset SI_1$ i $TWart_2 \subset SI_2$,

są równoważne:

- E1. w sensie zachowania zależności funkcyjnych, gdy można zbudować wyrażenia relacyjne zawierające zbiory ról wyznaczające wszystkie zależności funkcyjne między typami atrybutowymi

$$\exists_{i=1..n, \tau_i \in Ograniczenia_1, \sigma_i \in TRol_1} \exists_{j=1..m, \tau_j \in Ograniczenia_2, \sigma_j \in TRol_2} \\ \{ \{ BazyAtr(\tau_i) \rightarrow BazyPdAtr(\sigma_j) \} \equiv \{ BazyAtr(\tau_j) \rightarrow BazyPdAtr(\sigma_i) \} \}.$$

Zbiory ról σ_i, τ_i oraz σ_j, τ_j są powiązane funkcyjnie nad strukturą informacyjną SI_1 oraz SI_2 czyli $\tau_i \rightarrow^{SI_1} \sigma_i, \tau_j \rightarrow^{SI_2} \sigma_j$, a zbiory typów atrybutowych jako dziedziny zależności funkcyjnej $BazyAtr$ oraz przeciwdziedziny $BazyPdAtr$ są zdefiniowane

$$BazyAtr(r) = \{ y / y \in TWart \wedge \exists_{q \in r} [Baza(q) = y] \vee \exists_{x \in TA, RoleIdF(x, r)} [y \in AtrybutyIdF(x, r)] \\ \vee \exists_{x \in TA, r = SumaF(x, (\tau_i)_{i \in n})} [y \in AtrybutyIdF''(x, (\tau_i)_{i \in n})] \} \text{ oraz} \\ BazyPdAtr(r) = \{ y / y \in TWart \wedge \exists_{q \in r, unikat(q) \wedge total(q)} [y \in Bazy(Inne_Role(q))] \\ \vee \exists_{q \in r, unikat(q) \wedge total(q), x \in Bazy(Inne_Role(q)), x IdF \tau} [y \in AtrybutyIdF(x, \tau)] \\ \vee \exists_{q \in r, unikat(q) \wedge total(q), x \in Bazy(Inne_Role(q)), SumaIdF(x, (\tau_i)_{i \in n})} [y \in AtrybutyIdF''(x, (\tau_i)_{i \in n})] \}$$

Przy definicji dziedziny i przeciwdziedziny zależności funkcyjnych typy abstrakcyjne są zastępowane typami atrybutowymi należącymi do zbiorów identyfikujących $AtrybutyIdF$ i $AtrybutyIdF''$ (odpowiednio przy identyfikacji prostej i złożonej).

E2. w sensie zachowania danych, gdy mają identyczne populacje określone zależnościami funkcyjnymi Pop_W_1 i Pop_W_2 , czyli $Pop_W_1 = Pop_W_2$, gdzie $Pop_W_1 \subseteq POP_1$ i $Pop_W_2 \subseteq POP_2$.

E3. w sensie zachowania zależności niefunkcyjnych, gdy można zbudować wyrażenia relacyjne wynikające z ograniczeń wyznaczających identyczne zależności niefunkcyjne między typami atrybutowymi

$$\exists_{i=1..n, \tau_i, \sigma_i \in Ograniczenia_1} \exists_{j=1..m, \tau_j, \sigma_j \in Ograniczenia_2}$$

$$\{ \{ BazyAtr(\sigma_j), BazyAtr(\tau_i) \} \equiv \{ BazyAtr(\sigma_i), BazyAtr(\tau_j) \} \}$$

oraz populacja tych zbiorów typów atrybutowych jest identyczna $Pop_O_1 = Pop_O_2$, gdzie $Pop_O_1 \subseteq POP_1$ i $Pop_O_2 \subseteq POP_2$.

E4. Muszą być spełnione warunki 1, 2, 3, aby schematy konceptualne były równoważne.

Opracowana metoda umożliwiła sprawdzenie równoważności modeli M_1 i M_2 w sensie E4. Badanie zależności funkcyjnych opiera się na twierdzeniu o przechodności zależności funkcyjnych zdefiniowanych dla grafów funkcjonalnych [3].

Podstawy formalne algorytmu sprawdzania równoważności modeli [3]

1. Określenie identyczności zbiorów typów atrybutowych polega na sprawdzeniu równości zbiorów typów atrybutowych w zbiorze WO grafu funkcjonalnego
2. Sprawdzenie równoważności w sensie E1 jest oparte na wyszukiwaniu podgrafów częściowych zawierających ścieżkę łączącą typy atrybutowe stanowiące dziedzinę i przeciwdziedzinę zależności funkcyjnych. Każdy typ abstrakcyjny należący do ścieżki jest zastępowany odpowiednio przez zbiory $AtrybutyIdF$ lub $AtrybutyIdF''$, które są dołączane do dziedziny $BazyAtr$ lub przeciwdziedziny $BazyPdAtr$ odpowiadających im zależności funkcyjnych. Ścieżka zawiera zbiór ról łączących, w tym role wychodzące z kolejnych, połączonych wierzchołków grafu funkcjonalnego (typy obiektowe) należących do zbiorów $unikat$ i $total$ (tw. o przechodności zależności funkcyjnych). Związki funkcjonalne wyznaczają połączenie typu ścieżka_f łącząca atrybuty powiązane funkcyjnie nad strukturą informacyjną, równe $\text{ścieżka}_f(\tau) = \delta_{S(\tau)}(\langle \rangle / \langle f \in F_{akty}(\tau), f \rangle)$. Formuła selekcji $S(\tau)$ równa $S(\tau) = \bigvee_{p \in \tau, q \in Inne_Role(p)} \exists_{r \in \tau, pl} [Baza(r) = Baza(q)]$ określa zbiór ról łączących zbiory ról $RoleIdF$ lub $RoleIdF''$ lub unikatowych. Dla jednego typu faktu $\text{ścieżka}_f(\tau) = f$. Jeżeli role należące do τ wchodzące są objęte ograniczeniami $total$ i $unikat$, są dołączane do dziedziny kolejnej zależności funkcyjnej, a ich przeciwdziedziną są typy atrybutowe należące do dziedziny poprzedniej zależności.
3. Sprawdzenie równoważności w sensie E2 polega na wyznaczeniu zbioru typów bazowych ról należących do badanego ograniczenia i następnie zbioru typów atrybutowych identyfikujących typy abstrakcyjne należące do tego zbioru:

$$\text{Zbiór_}A(\sigma_i, \tau_i) = \{ \text{BazyAtr}(\sigma_i) \cup \text{BazyAtr}(\tau_i) \text{ dla } (\sigma_i, \tau_i \in \text{RolePodzbiór}) \vee \\ (\sigma_i, \tau_i \in \text{RoleRozłączność}) \vee (\sigma_i, \tau_i \in \text{RoleRówność}) \vee (\sigma_i, \tau_i \in \text{RoleCzęstotliwość}) \}$$

Sprawdzenie ograniczeń zbiorowych polega na wyznaczeniu rodziny indeksowanej zbiorów $\text{Zbiór_}A(\sigma_i, \tau_i)$ typów atrybutowych należących do każdej z dwóch sekwencji τ_i, σ_i ról danego ograniczenia, stąd mamy $\text{Zbiór_}B = (\text{Zbiór_}A(\sigma_i, \tau_i))_{i \in n}$.

Zawartość rodziny zbiorów $\text{Zbiór_}B$ w porównywanych schematach konceptualnych jest identyczna, jeżeli oba schematy są równoważne.

4. Identyczności populacji w obu schematach bada się przez porównanie populacji typów atrybutowych powiązanych funkcyjnie i niefunkcyjnie. Zakłada się, że populacja danego typu abstrakcyjnego x może być zastąpiona populacją typów atrybutowych identyfikujących AtrybutyIdF lub $\text{AtrybutyIdF}'$:

$$\text{Pop_}a(x) = \{y \mid \exists_{p \in \text{AtrybutyIdF}(x, \tau)} [\text{Pop}(p) = y] \vee \forall_{q \in \text{AtrybutyIdF}'(x, \tau_i)_{i \in n}} [\text{Pop}(q) = y]\}.$$

Populacja każdego typu faktu f łączącego typy abstrakcyjne jest równa

$$\text{Pop_}f(f) = \{x \mid \forall_{p \in \text{abstr}(f)} [x[p] \in \text{Pop_}a(\text{Baza}(p))] \wedge \\ \forall_{q \in \text{konkr}(f)} [x[q] \in \text{Pop}(\text{Baza}(q))]\}.$$

Populacja wyrażenia relacyjnego, którym jest ścieżka f jest więc równa

$$\text{Pop}(\text{ścieżka_}f(\sigma)) = \bigvee_{f \in \text{Fakty}(\sigma)} \text{Pop_}f(f).$$

Populacja wypadkowa jest złączeniem tych populacji

$$\text{Pop_}W((\sigma_i)_{i \in n}, \bigvee_{i=1..n} \text{ścieżka_}f(\sigma_i)) = \{t \mid \forall_{\sigma_i \subseteq \text{RoleUnikat} \cdot \text{RoleTotal}} [t[\sigma_i] \in \text{Pop}(\text{ścieżka_}f(\sigma_i))]\}.$$

Populacja każdego zbioru typów obiektowych objętych zależnościami niefunkcyjnymi (ilościowymi i zbiorowymi) jest wyznaczona jako

$$\text{Pop}(\text{Zbiór_}B) = (\text{Pop}(\text{Zbiór_}A(\sigma_i, \tau_i))_{i \in n}), \text{ gdzie}$$

$$\text{Pop}(\text{Zbiór_}A(\sigma_i, \tau_i)) = \{y \mid \forall_{f \in \sigma_i, p \in \tau_i} [y \in \text{Pop_}f(f) \vee y \in \text{Pop}_f(p)]\}$$

5 Przykład zastosowania metody badania równoważności

Metodę zastosowano do oceny dwóch rozwiązań M_1 i M_2 problemu sporządzania *Rachunku*, przedstawionych w postaci diagramów ORM na rysunkach 4-7 [3].

5.1. Zależności funkcyjne modelu M_1 (a) i modelu M_2 (b) są identyczne (tabela 1):

a) $\text{AtrybutyIdF}(\text{Produkt_}1, \tau_1 = \text{Produkt_}2, \tau_1 = \text{Produkt_}3, \tau_1) = \{\text{Nazwa_}P, \text{Cena}\},$

$$\tau_1 = \{\pi_{af1}, \pi_{af2}\},$$

$$\text{AtrybutyIdF}(\text{Zakup}, \tau_2 = \{\text{Numer zakupu}\}, \tau_2 = \{\pi_{bf11}\},$$

$$\text{AtrybutyIdF}(\text{Rachunek_}1, \tau_3 = \text{Rachunek_}2, \tau_3) = \{\text{Nazwa_}R\}, \tau_3 = \{\pi_{bf13}\},$$

$$\text{ścieżka_}f(\pi_{bf3}) = f3 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Gatunek},$$

$$\text{ścieżka_}f(\pi_{bf4}) = f4 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Podatek},$$

$$\text{ścieżka_}f(\pi_{af8}) = f8 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Numer zakupu},$$

$$\text{ścieżka_}f(\pi_{bf8}) = f8 \Rightarrow \text{Numer zakupu} \rightarrow \text{Nazwa_}P, \text{Cena},$$

$$\text{ścieżka_}f(\pi_{bf12}) = f12 \Rightarrow \text{Numer zakupu} \rightarrow \text{Ilość},$$

$$\text{ścieżka_}f(\pi_{bf14}, \in^s \text{Zakupy}) = \delta_{\pi_{af14} = \in^s \text{Zakupy}} f14 \bigvee_{\in \text{Zakupy}} \Rightarrow \text{Nazwa_}R \rightarrow \text{Numer zakupu},$$

$$\text{ścieżka_}f(\pi_{bf18}) = f18 \Rightarrow \text{Nazwa_}R \rightarrow \text{Płatność}.$$

b) $\text{AtrybutyIdF}(\text{Produkt_}1, \tau_1 = \text{Produkt_}2, \tau_1 = \text{Produkt_}3, \tau_1) = \{\text{Nazwa_}P, \text{Cena}\},$

$$\tau_1 = \{\pi_{af1}, \pi_{af2}\},$$

$$\text{AtrybutyIdF}(\text{Zakup}, \tau_2 = \text{Zakup_}P, \tau_2 = \text{Zakup_}N, \tau_2) = \{\text{Numer zakupu}\}, \tau_2 = \{\pi_{bf14}\}$$

$$\text{AtrybutyIdF}(\text{Rachunek_}1, \tau_3 = \text{Rachunek_}2, \tau_3) = \{\text{Nazwa_}R\}, \tau_3 = \{\pi_{bf16}\},$$

$$\text{ścieżka_}f(\pi_{bf3}) = f3 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Gatunek},$$

$$\text{ścieżka_}f(\pi_{bf4}) = f4 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Podatek},$$

$$\text{ścieżka_}f(\pi_{af8}) = f8, \text{ścieżka_}f(\pi_{bf12}) = f12 \Rightarrow \text{Nazwa_}P, \text{Cena} \rightarrow \text{Numer zakupu},$$

$$\text{ścieżka_}f(\pi_{bf8}, \pi_{af8}) = f8, \text{ścieżka_}f(\pi_{bf12}) = f12 \Rightarrow \text{Numer zakupu} \rightarrow \text{Nazwa_}P, \text{Cena}$$

$$\text{ścieżka_}f(\pi_{bf15}) = f15 \Rightarrow \text{Numer zakupu} \rightarrow \text{Ilość},$$

$$\text{ścieżka_}f(\pi_{bf17}, \in^s \text{Zakupy}) = \delta_{\pi_{af17} = \in^s \text{Zakupy}} f17 \bigvee_{\in \text{Zakupy}} \Rightarrow \text{Nazwa_}R \rightarrow \text{Numer zakupu},$$

$$\text{ścieżka_}f(\pi_{bf21}) = f21 \Rightarrow \text{Nazwa_}R \rightarrow \text{Płatność}.$$

5.2. Populacja modelu $M_1(a)$ i modelu $M_2(b)$ jest identyczna:

Nazwa $P=\{A,B\}$, Cena $\bar{=}\{10,20\}$, Gatunek $\bar{=}\{1,2\}$, Podatek $\bar{=}\{7\%, 22\%\}$,

Numer zakupu $=\{1, 2\}$, Nazwa $R = \{101, 102\}$, Płatność $= \{Przekaz, Gotówka\}$,

- a) $Pop(\text{ścieżka } f(\pi_b f3)) = \{\{B, 10, 2\}\}$; $Pop(\text{ścieżka } f(\pi_b f4)) = \{\{A, 20, 7\%\}$,
 $Pop(\text{ścieżka } f(\pi_a f8)) = \{\{A, 20, 1\}, \{B, 10, 2\}\}$; $Pop(\text{ścieżka } f(\pi_a f8)) = \{\{1, A, 20\}, \{2, B, 10\}\}$,
 $Pop(\text{ścieżka } f(\pi_a f12)) = \{\{1, 15\}, \{2, 5\}\}$; $Pop(\text{ścieżka } f(\pi_a f14, \epsilon^S_{Zakupy})) = \{\{101, 1\}, \{101, 2\}\}$
 $Pop(\text{ścieżka } f(\pi_b f18)) = \{\{101, Przekaz\}\}$,
 $Pop_W = \{\{101, Przekaz, 1, 15, A, 20, 7\%\}, \{101, Przekaz, 2, 5, B, 10, 2\}\}$,
- b) $Pop(\text{ścieżka } f(\pi_b f3)) = \{\{B, 10, 2\}\}$; $Pop(\text{ścieżka } f(\pi_b f4)) = \{\{A, 20, 7\%\}$,
 $Pop(\text{ścieżka } f(\pi_a f8)) = \{\{A, 20, 1\}\}$; $Pop(\text{ścieżka } f(\pi_a f12)) = \{\{B, 10, 2\}\}$,
 $Pop(\text{ścieżka } f(\pi_a f8)) = \{\{1, A, 20\}\}$; $Pop(\text{ścieżka } f(\pi_a f12)) = \{\{2, B, 10\}\}$,
 $Pop(\text{ścieżka } f(\pi_a f15)) = \{\{1, 15\}, \{2, 5\}\}$; $Pop(\text{ścieżka } f(\pi_a f17, \epsilon^S_{Zakupy})) = \{\{101, 1\}, \{101, 2\}\}$
 $Pop(\text{ścieżka } f(\pi_b f21)) = \{\{101, Przekaz\}\}$,
 $Pop_W = \{\{101, Przekaz, 1, 15, A, 20, 7\%\}, \{101, Przekaz, 2, 5, B, 10, 2\}\}$,

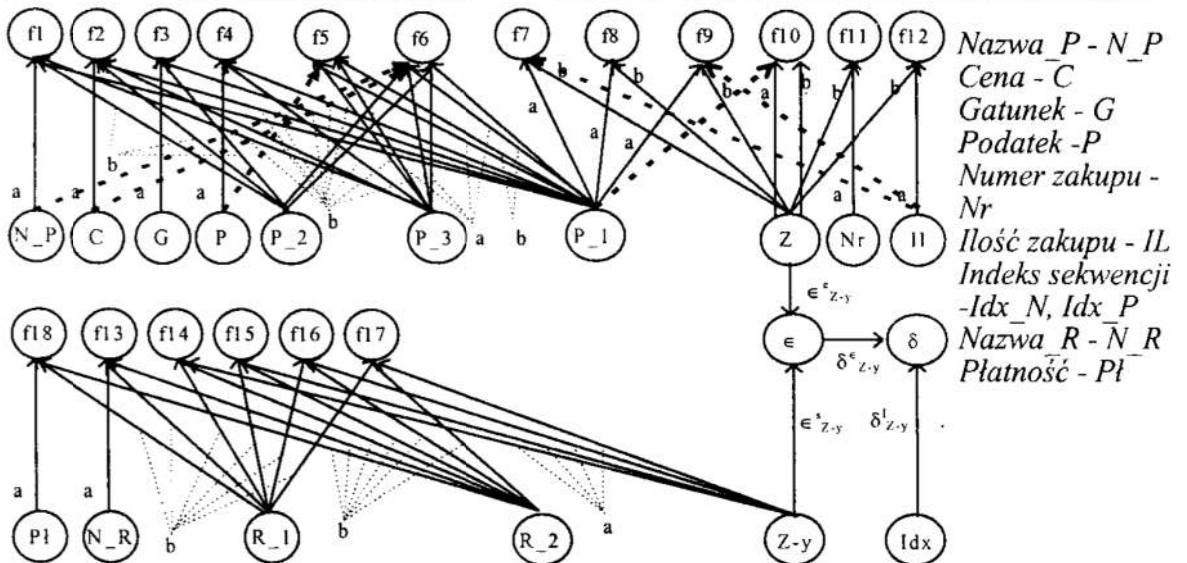
5.3 Ograniczenia zbiorowe modelu $M_1(a)$ i modelu $M_2(b)$ są identyczne:

- a) $\sigma_1 = \{\pi_b f9, \pi_a f9\}$, $\tau_1 = \{\pi_b f8, \pi_a f8\}$, $\delta_1 = \{\pi_b f7, \pi_a f7\}$,
 $\sigma_2 = \{\pi_b f4, \pi_a f4\}$, $\tau_2 = \{\pi_b f1, \pi_a f1\}$, $\delta_2 = \{\pi_b f2, \pi_a f2\}$,
 $\sigma_3 = \{\pi_b f3, \pi_a f3\}$, $\tau_3 = \{\pi_b f1, \pi_a f1\}$, $\delta_3 = \{\pi_b f2, \pi_a f2\}$,
 $Zbiór_A(\sigma_1, \tau_1) = Zbiór_A(\sigma_1, \delta_1) = \{\text{Numer zakupu, Nazwa } P, \text{Cena}\}_1$,
 $Zbiór_A(\sigma_2, \tau_2) = Zbiór_A(\sigma_2, \delta_2) = Zbiór_A(\sigma_3, \tau_3) = Zbiór_A(\sigma_3, \delta_3) = \{\text{Nazwa } P, \text{Cena}\}_2$,
 $Zbiór_B = \{\{\text{Numer zakupu, Nazwa } P, \text{Cena}\}_1, \{\text{Nazwa } P, \text{Cena}\}_2, \{\text{Nazwa } P, \text{Cena}\}_3\}$,
 $Pop(Zbiór_B) = \{\{1, A, 20\}, \{2, B, 10\}\}$.
- b) $\sigma_1 = \{\pi_a f7, \pi_b f7\}$, $\tau_1 = \{\pi_a f8, \pi_b f8\}$, $\delta_1 = \{\pi_a f9, \pi_b f9\}$, $\sigma_2 = \{\pi_b f12, \pi_a f12\}$, $\tau_2 = \{\pi_b f13, \pi_a f13\}$,
 $Zbiór_A(\sigma_1, \tau_1) = Zbiór_A(\tau_1, \delta_1) = Zbiór_A(\sigma_2, \tau_2) = \{\text{Numer zakupu, Nazwa } P, \text{Cena}\}_1$,
 $Zbiór_B = \{\{\text{Numer zakupu, Nazwa } P, \text{Cena}\}_1, \{\text{Numer zakupu, Nazwa } P, \text{Cena}\}_2\}$,
 $Pop(Zbiór_B) = \{\{1, A, 20\}, \{2, B, 10\}\}$.

Tab. 1. Zależności funkcjonalne w modelu M_1 i M_2

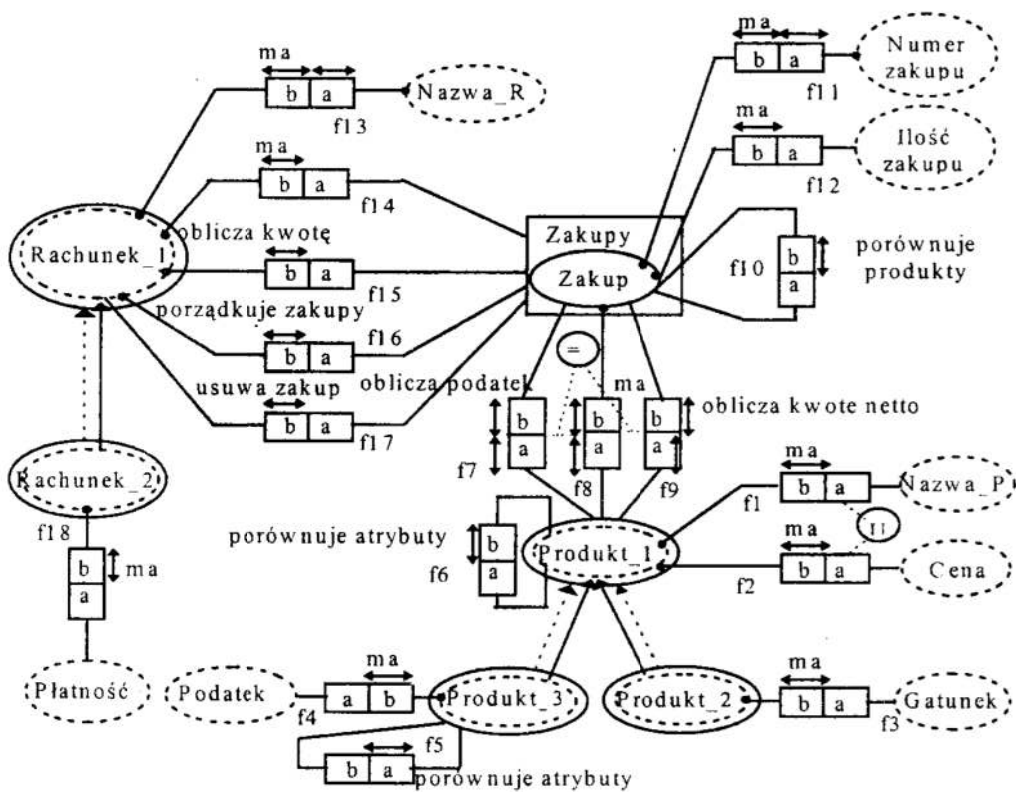
Tab.1. The functional dependences of the models M_1 i M_2

Lp.	dziedzina(M_1, M_2)	przeciwdziedzina(M_1)	przeciwdziedzina (M_2)
1	N, P, C	G, P, Nr	G, P, Nr
2	Nr	N, P, C, Nr, IL, Idx	$N, P, C, IL, Idx, N=Idx, P=Idx$
3	N, R	$Nr, Pł$	$Nr, Pł$

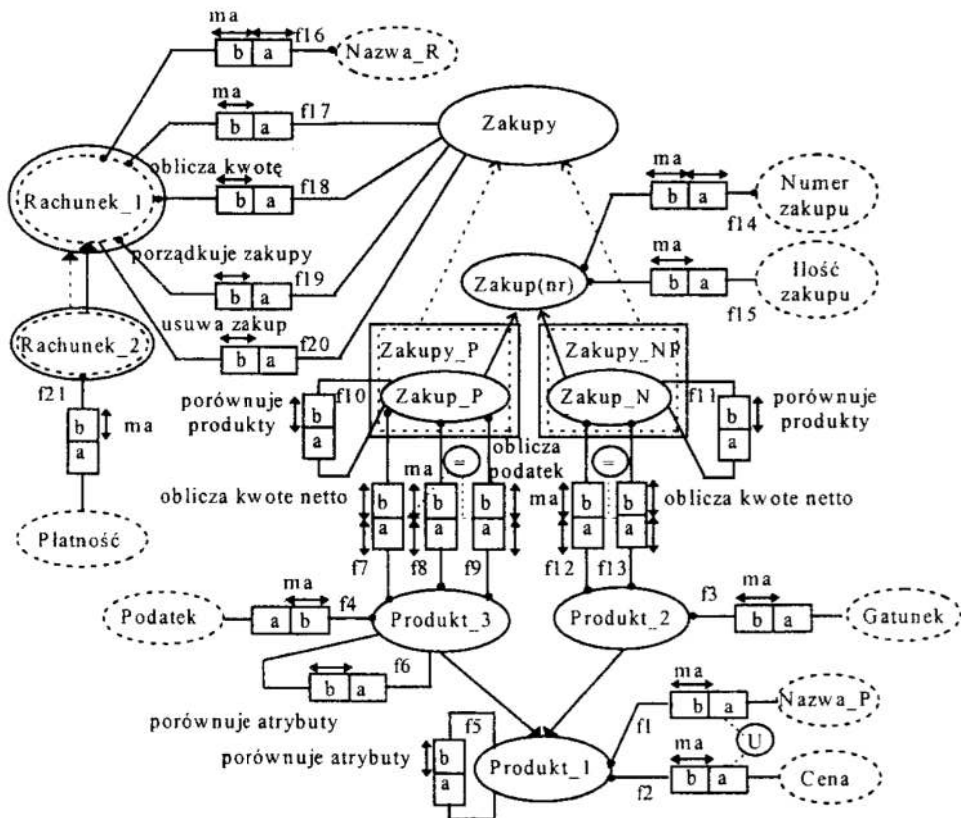


Rys. 4. Graf funkcjonalny M_1 (rysunek 5)

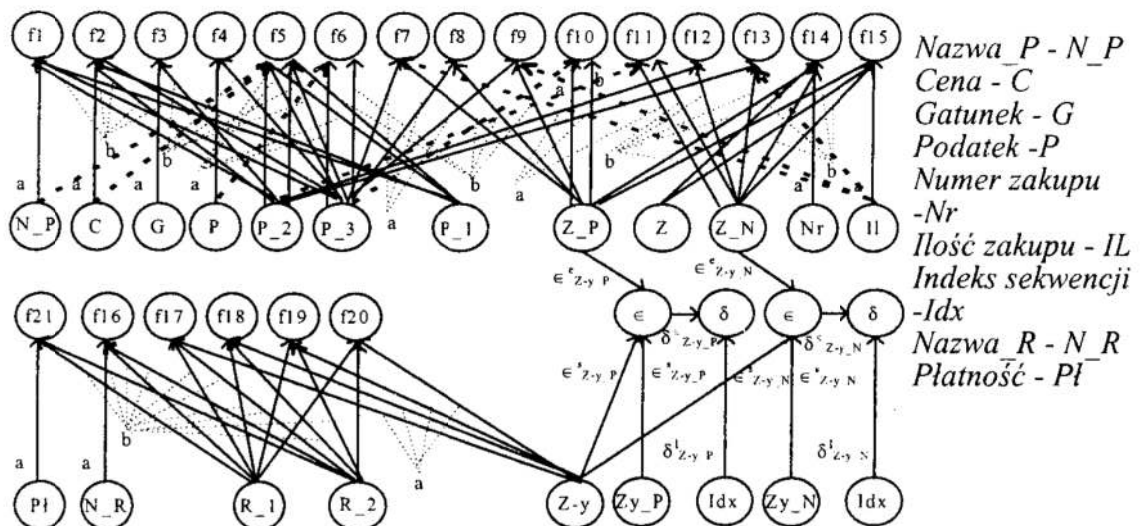
Fig. 4. functional graph M_1 (figure 5)



Rys. 5. Model ORM - M₁
 Fig. 5. Model ORM - M₁



Rys. 6. Model ORM - M₂
 Fig. 6. Model ORM - M₂



Rys. 7. Graf funkcjonalny M_2 (rysunek 6)
 Fig. 7. The functional graph M_2 (figure 6)

6 Zakończenie

Zagadnienia opisane w artykule umożliwiają budowę systemu typu CASE do automatycznego wyboru zbioru równoważnych rozwiązań danego problemu na poziomie modelowania. W prowadzonych pracach przedstawiona metoda umożliwiła badanie wpływu zastosowanych różnych elementów struktury informacyjnej np. polimorfizmu i specjalizacji, prowadzących do rozwiązań równoważnych, na jakość modelu reprezentowaną przez złożoność strukturalną [3].

Literatura

1. P.van Bommel, A.H.M. ter Hofstede, Th.P.van der Weide: *Semantics and verification of object-role models*. Information Systems, vol. 16, No.5, str. 471-495, 1991.
2. Hainaut J. L.: *Specification preservation in schema transformations-application to semantics and statistics*. Data & Knowledge Engineering 19 (1996) 99-134..
3. Kruczkiewicz Z.: *Rozprawa doktorska*. Raport serii: Preprinty nr 42/99, Wydawnictwo Politechniki Wrocławskiej, Wrocław 1999.
4. Booch G., Rumbaugh J., Jacobson I. *The Unified Modeling Language- User Guide*. Addison-Wesley 1999.

Streszczenie

W pracy rozważa się równoważność modeli oprogramowania zorientowanego obiektowo. Zdefiniowano zasady równoważności modeli ORM w sensie identycznych zbiorów typów atrybutowych, zbiorów zależności funkcyjnych i niefunkcyjnych oraz populacji spełniającej te zależności. Sformułowano podstawy formalne algorytmu badania równoważności oraz opisano przykład zastosowania opracowanej metody.

THE EQUIVALENCE EXAMINATION METHOD OF THE OBJECT ORIENTED SOFTWARE MODELS

Summary

In the paper, the equivalence of the object oriented software models is considered. The rules of the equivalence of ORM models in the meaning of the identical sets of label types, the identical sets of the functional and non-functional dependences and population which fulfil them are defined. The formal base of the equivalence examination algorithm is formulated and the example of the application of the method is described.

Referat recenzowany

AUTOMATYCZNA GENERALIZACJA LINII W SYSTEMACH TYPU GIS

1 Wstęp

Terminem **GIS** (*Geographical Information System*) określa się system komputerowy służący do zbierania, przechowywania, weryfikacji, integracji, przetwarzania, analizowania i wizualizacji danych odzwierciedlających pozycje jakichś obiektów na powierzchni Ziemi. Zazwyczaj GIS (lub też *Spatial Information System*) używany jest do wykonywania operacji na różnego rodzaju mapach, zaś dane przechowywane są w nim w wielopoziomowych strukturach [8].

Terminem **generalizacja map** określa się proces upraszczania informacji zawartych na mapach, wykonywany głównie przy zmianie (zwiększaniu) ich skali, polegający na zwiększaniu czytelności przy zachowaniu wszystkich istotnych informacji. Generalizacja map jest procesem ściśle zależnym od przyjętej skali powstających map oraz (mniejszych) skal map topograficznych, na bazie których ten proces się odbywa. Ponadto jej przebieg zależy od przyjętych ograniczeń (jak np. przeznaczenia mapy, dokładności danych, itp.). Zazwyczaj z generalizacją wiąże się proces **redukcji detali**, proces **powtórnego (rządszego) próbkowanie** lub proces **redukcji liczby węzłów krzywych (linii)**.

Tradycyjnie proces generalizacji wykonywany był ręcznie przez kartografa, lecz po wprowadzeniu systemów typu GIS udało się go częściowo (a czasem całkowicie) zautomatyzować [8]. W procesie automatycznej generalizacji mapy wyróżnić można następujące etapy [1]: rozpoznanie struktury, rozpoznanie procesu, modelowanie procesu, wykonanie procesu, wizualizacja wyników.

Głównym celem rozpoznania struktury i rozpoznania procesu jest określenie typu struktur i względnej ważności elementów na mapie oraz zidentyfikowanie operatorów generalizacji, które należy przyjąć. Podczas ręcznej generalizacji etapy te wykonuje kartograf. Etapy modelowania i wykonania procesu oraz wizualizacji wiążą się bezpośrednio z implementacją i są uwarunkowane wynikami etapów poprzednich.

Większość metod generalizacji rozwinęła się dla danych w reprezentacji wektorowej. Jednak powstanie technik cyfrowej obróbki obrazów pozwoliło wykorzystać do zadań generalizacji model rastrowy. Obecnie duże zainteresowanie wzbudza możliwość implementacji metod generalizacji operujących na hybrydowych (wektorowych i rastrowych) bazach danych [7].

W niniejszym opracowaniu zawarto opis metod i algorytmów używanych w kartografii w procesie generalizacji (rozdział 2). Szczególną uwagę poświęcono problemowi generalizacji linii i perspektywom zastosowania sieci neuronowych do jego rozwiązania. Przy tej okazji wymieniono tradycyjne metody klasyfikacji krzywych (rozdział 3), które można zastosować w połączeniu z sieciami neuronowymi. W opracowaniu przedstawiono dwa przykłady „naiwnego” użycia sieci neuronowych do problemu generalizacji linii (rozdział 4). W pierwszym przykładzie sieć neuronowa typu *feed-forward* była uczona odwzorowania $X \rightarrow Y$, gdzie X był wektorem współrzędnych punktów linii, Y zaś zero-jedynkową reprezentacją linii zgeneralizowanej. W przykładzie drugim zaś do generalizacji linii użyto sieć rekurencyjną, działającą podobnie do algorytmu VectGen. W podrozdziale ostatnim (rozdział 5) zamieszczono podsumowanie.

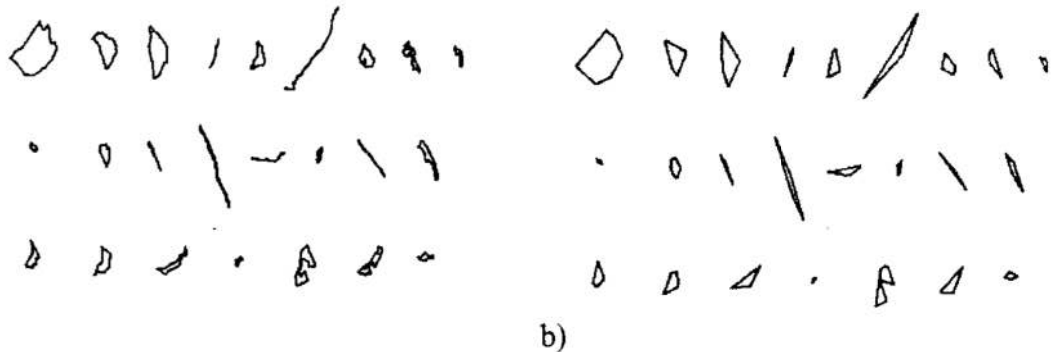
* Instytut Cybernetyki Technicznej Politechniki Wrocławskiej, ul. Janiszewskiego 11/17, Wrocław
e-mail: tkubik@ict.pwr.wroc.pl.

2 Standardowe metody i algorytmy generalizacji linii

Generalizacja linii jest chyba najdokładniej zbadanym zagadnieniem spośród wszystkich problemów generalizacji. Opracowane algorytmy i metody operują na wektorowych i rastrowych reprezentacjach danych. Poniżej wymienione zostaną tylko niektóre z nich.

Algorytmy selekcji punktów znaczących

W algorytmach tych generalizacja polega na usunięciu z krzywej takich punktów, które są mało istotne dla zachowania jej kształtu. Istotność określa się na podstawie jakiegoś kryterium, podając do algorytmu jej wartość w postaci parametru (lub parametrów). Najprostszym z nich jest algorytm pozostawiający co n -ty punkt. Do bardziej skomplikowanych algorytmów (w których obliczane są geometryczne zależności typu odległość od prostej, kąty, powierzchnia efektywna, itp.) należą: algorytm Lang'a, algorytm VectGen, *point relaxation algorithm*, algorytm Reumann'a–Witkam'a [9] oraz algorytm Visvalingham'a i Williamson'a [7], algorytmy selekcji punktów o największej krzywiznie [4] (gdzie używa się pojęcia: *k-cosinus*); algorytm Douglas'a–Peucker'a (*DP*) [2]. Efekty zastosowania algorytmu *DP* do rzeczywistych obiektów (krzywych zamkniętych) przedstawia rysunek 1. W części a) tego rysunku przedstawione są oryginalne kształty, w części b) kształty zgeneralizowane.



Rys. 1. Efekty generalizacji po zastosowaniu algorytmu *DP*: a) obiekty przed generalizacją, b) obiekty po generalizacji
Fig. 1. Generalization results after applying *DP* algorithm: a) original objects, b) generalized objects

Algorytmy tworzące nowe punkty

W algorytmach tych generalizacja polega na znalezieniu nowych punktów reprezentujących krzywą, przy zachowaniu jej istotnych cech. Reprezentantem tej grupy jest algorytm Li i Openshaw'a [5]. Algorytm ten parametryzowany jest rozmiarem najmniejszego, godnego zachowania szczegółu r , nowe punkty zaś definiowane są przez środek odcinka rozpiętego na punktach przecięć krzywej z okręgami o średnicy $4r$.

Algorytmy podziału krzywej na podzbiory

Do tej grupy należą algorytmy klasteryzacji, w których punkty reprezentujące krzywą dzieli się na podzbiory punktów tak, aby dało się je aproksymować liniami prostymi z jak najmniejszym błędem [16]. W algorytmach tych wychodzi się z arbitralnego podziału krzywej na k podzbiorów i w kolejnych krokach, na podstawie odległości od wyznaczonych linii odniesienia, przesuwają się granice tych podzbiorów.

Metody z zastosowaniem transformacji prostych i odwrotnych

Do tej grupy należą metody i algorytmy, w których krzywe transformuje się do nowej postaci, dokonuje pewnych operacji na nowej postaci i przeprowadza się transformację odwrotną. Należą do niej metody analizy częstotliwościowej (jak metody oparte na

transformacji Fouriera) oraz metody wykorzystujące transformację Hough'a (gdzie następuje przejście z przestrzeni kartezjańskiej do przestrzeni parametrów).

Inne algorytmy i metody

Wśród innych metod i algorytmów generalizacji należy wyróżnić metody aproksymacji krzywych wielobokami, metody bazujące na kodzie łańcuchowym Freeman'a, algorytmy wygładzania (do grupy tej należą algorytmy korzystające z funkcji sklepanych, algorytmy dokonujące filtracji krzywej (np. filtrem Gaussa) oraz algorytmy operujące pojęciem wykresu skala-przestrzeń [14]).

3 Metody klasyfikacji krzywych

W metodach tych próbuje się opisać krzywą zestawem parametrów oddającym jej indywidualne cechy i na tej podstawie dokonać jej klasyfikacji [12].

Metoda momentów

Momenty są parametrami, które można interpretować jak momenty bezwładności. Stosuje się je głównie do klasyfikacji krzywych w reprezentacji rastrowej [13]. Analityczna definicja momentu rzędu $p+q$ dwuwymiarowej, ciągłej funkcji dana jest równaniem całkowym. W przypadku obrazów rastrowych czarno-białych wyrażenie na momenty rzędu $p+q$ w dyskretnej wersji upraszcza się do postaci $m_{pq} = \sum_y \sum_x x^p y^q$. Momenty centralne, które są już niezależne od położenia obrazu, obliczane są ze wzoru:

$$\mu_{pq} = \sum_y \sum_x (x - x_s)^p (y - y_s)^q, \text{ gdzie } x_s = m_{10} / m_{00}, \quad y_s = m_{01} / m_{00}.$$

Zazwyczaj momenty centralne obliczane są do rzędu trzeciego i normalizowane. Na podstawie znormalizowanych momentów centralnych obliczanych jest zazwyczaj siedem momentów niezmienniczych, które są niezależne od przesunięcia, obrotu i zmiany skali obiektów. Ze względu na małe wartości oblicza się ich logarytm lub dalej normalizuje.

Wyznaczanie miary

Użycie miar okrągłości i prostoliniowości w połączeniu z innymi cechami daje możliwość klasyfikacji krzywych.

Badanie okrągłości obiektów mają miejsce, kiedy nie tyle zwraca się uwagę na specyficzne cechy kształtu, co bada ich ogólną charakterystykę (np. gładkość lub zgrzebność). Okrągłość c zdefiniowano w następujący sposób: $c = 4 \pi (\text{pole powierzchni}) / (\text{obwód}^2)$.

Pomiar prostoliniowości sd pozwala ocenić jak bardzo prosty lub krzywy jest badany fragment. sd oblicza się sumując kwadraty odległości punktów pośrednich, leżących pomiędzy wyróżnionym punktem początkowym i końcowym, od prostej przechodzącej przez punkt początkowy i końcowy.

Klasyfikacja z wykorzystaniem deskryptorów Fouriera

Posługując się współczynnikami rozwinięcia w szereg Fouriera można dokonywać klasyfikacji krzywych zamkniętych (lub dowolnych krzywych po modyfikacji metody), przy czym do klasyfikacji kształtu wystarczy brać pod uwagę tylko amplitudy. Dokonując wcześniej zmiany reprezentacji [12] uzyskuje się niezależność od położenia, orientacji i skali klasyfikowanych obiektów.

Metody analizy fraktalnej

W analizie fraktalnej bada się właściwości nieregularnych obiektów (fraktali). Ich niezmiennicze, niezależne od skali cechy mogą być wykorzystane do generalizacji [3], jednak najlepszym sposobem wykorzystania tych cech jest wspomaganie procesu generalizacji lub użycie jako metod kompensacji. Cechy fraktali wyznacza się dokonując pomiaru stopnia złożoności ich złożoności. Podczas pomiaru ocenia się, jak szybko wzrasta długość badanego obiektu y (powierzchnia czy też objętość) w miarę wzrostu dokładności pomiaru x

(np. dla długości linii mierzonej cyrklem $x=1/s$, gdzie s – rozstaw cyrkla). Zależność y od x wyraża prawo potęgowe: $y \propto x^d$ (dla linii $0 < d < 1$). Na bazie tego prawa zdefiniowano różne wymiary fraktalne, między innymi: wymiar samopodobieństwa $D_s = \log n / \log (1/s)$, gdzie n – liczba części, na które podzielony jest obiekt, s – współczynnik redukcji; wymiar cyrklowy $D_c = 1 + d$, gdzie d jest nachyleniem wykresu logarytmu długości zmierzonej krzywej w zależności od logarytmu dokładności ($1/s$), s – rozstaw cyrkla; wymiar pudełkowy (oblicza się go umieszczając obiekt na równomiernej siatce, zliczając ilość zajętych przez obiekt oczek przy zmieniającej się ich rozmiarze).

4 Sieci neuronowe w generalizacji

Zastosowaniem sieci neuronowych do celów generalizacji oraz ich użyciem w systemach GIS zajęto się stosunkowo niedawno [6], [11], [15]. Jak dotąd, literatura opisująca zastosowania sieci neuronowych w tym kontekście jest bardzo uboga. Zazwyczaj opisuje się w niej perspektywy zastosowań sieci neuronowych, kładąc szczególny nacisk na sposób reprezentacji danych. Rzeczywiste zastosowania sieci neuronowych ograniczone są do użycia ich jako systemów decyzyjnych bądź też filtrów w obróbce obrazów rastrowych. Szersze możliwości zastosowań sieci neuronowych wiążą się z przyjęciem założenia, że problem generalizację w kartografii można sprowadzić do problemu klasyfikacji. O poprawności tego założenia świadczy fakt, iż dwa najważniejsze spośród wyróżnionych w [1] etapów generalizacji: rozpoznanie struktury i rozpoznanie procesu, są rzeczywiście zadaniami klasyfikacji. W prezentowanych niżej eksperymentach próbowano sprawdzić, czy jednak (dla prostych przypadków) sieci neuronowe nie mogłyby zadziałać jak algorytm selekcji punktów znaczących.

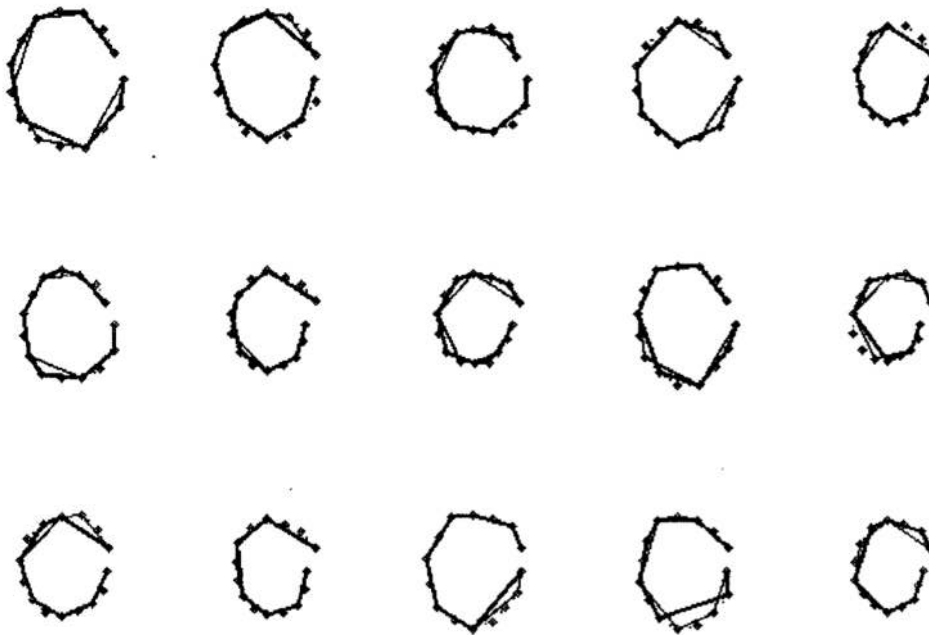
Proste eksperymenty z generalizacją linii

W procesie generalizacji (zalgorytmizowanym lub też wykonywanym ręcznie) z góry nie wiadomo, ile punktów będzie miała generalizowana krzywa, ani też ile ich zostanie po generalizacji. Tak więc sam problem generalizacji charakteryzuje się zmienną liczbą zarówno wejść jak i wyjść. Ponieważ w eksperymentach używane były sieci neuronowe o niezmienną topologią, musiano przyjąć pewne założenia.

I tak: w eksperymencie pierwszym zaprojektowano sieć typu *feed-forward* o 30 wejściach, 3 warstwach ukrytych (liczących po 60, 60 i 75 neuronów) i 15 wyjściach. Funkcją aktywacji była funkcja sigmoidalna. W procesie uczenia na wejście podawany był wektor współrzędnych (x,y) 15 punktów linii wygenerowanych sztucznie („losowych kółek”), na wyjście zaś wektor 15 zer i jedynek (0 oznaczało, że w zgeneralizowanej linii odpowiedni punkt jest nieistotny, 1 – że punkt jest istotny). Zestaw uczący składał się z 300 przykładów linii zgeneralizowanych algorytmem *DP*, zestaw testowy z 30 przykładów. Efekty generalizacji widoczne są na rysunku 2, gdzie linie grube oznaczają odpowiedź sieci na przykład testowy, linie cienkie – wynik generalizacji algorytmem *DP*, izolowane punkty zaś są punktami oryginalnej krzywej

W eksperymencie drugim zasymulowano sieć rekurencyjną o schemacie jak na rysunku 3b. Od sieci tej oczekiwano, że nauczy się generalizować linie w myśl algorytmu VectGen (w algorytmie tym rozważa się trzy kolejne punkty; jeśli odległość punktu środkowego od odcinka łączącego punkt pierwszy z drugim jest większa niż zadany margines, staje się on punktem początkowym następnych trzech punktów, jeśli zaś odległość jest mniejsza, punkt ten jest odrzucany i punkt początkowy z nowymi dwoma punktami brane są następnie pod uwagę). Symulacja polegała na tym, że najpierw sieć uczone rozpoznawać, czy punkt środkowy powinien być odrzucony bez zapiętego sprzężenia (danymi uczącymi były: na wejściu – zestawy współrzędnych trzech punktów; na wyjściu – 0 albo 1), następnie modyfikowano sieć (aby na wyjściu dawała współrzędną pierwszego lub drugiego punktu) i zapinano sprzężenie. Przykładowe efekty generalizacji

(dla zestawu testowego jak w eksperymencie pierwszym) przedstawiono na rysunku 3a. Przedstawione eksperymenty pokazały, że do generalizacji linii nie da się użyć sieci neuronowych wprost. Szersze wnioski płynące z tych wstępnych eksperymentów przedstawia podrozdział następny.

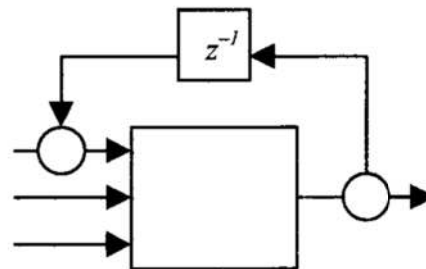


Rys. 2. Przykłady rozwiązań otrzymanych za pomocą generalizującej sieci neuronowej (linia gruba – rozwiązanie z sieci, linia cienka – rozwiązanie z algorytmu DP)

Fig. 2. Examples of results provided by generalizing neural networks (thick line – neural network's output, thin line – results obtained from DP algorithm)



a)



b)

Rys. 3. Generalizacja za pomocą sieci rekurencyjnej a) przykłady wyników (linie grube – wyjście z sieci, linie cienkie – rozwiązania otrzymane algorytmem DP), b) schemat sieci

Fig. 3. Generalization with use of recurrent neural network a) results examples (thick line – neural network's output, thin line – results obtained from DP algorithm), b) neural network's scheme

Charakterystyka sieci neuronowych i potencjalne zastosowania sieci w systemach typu GIS

Ogólnie wiadomo, że nie ma uniwersalnej sieci rozwiązującej wszystkie problemy. Topologię i algorytmy uczenia sieci projektuje się mając na uwadze ich zastosowanie. Sieci neuronowe szczególnie dobrze nadają się do celów klasyfikacji. Klasyfikacja za

pomocą sieci neuronowych: nie wymaga znajomości funkcji klasyfikującej (sieci potrafią się jej nauczyć), jest odporna na zakłócenia (dzięki „uśrednianiu się” danych podczas uczenia), ma cechy generalizacji (sieci mogą znajdować sensowne rozwiązania dla przykładów „leżących pomiędzy” przykładami z ciągu uczącego).

Sieci neuronowe mają holistyczny charakter, tj. rozpoznają cechy całej populacji, nie widzą natomiast cech poszczególnych elementów. Do wad sieci neuronowych, z punktu widzenia ich zastosowań do generalizacji w kartografii, wymienić można: kłopoty z wykonywaniem niektórych operacji (np. dzielenia, co utrudnia obliczanie odległości przy wektorowych reprezentacjach krzywych); niedokładność (brak unikalnych rozwiązań; generalnie sieci neuronowe aproksymują jakieś przekształcenie, a nie oddają je dokładnie; podczas generalizacji w kartografii niedokładności są czasem niedopuszczalne). Występowanie tych wad pokazały przeprowadzone eksperymenty.

Jeśli więc przyjąć założenie, że rola sieci neuronowych podczas generalizacji w kartografii polegać będzie na klasyfikacji, to wtedy jest szansa na uzyskanie pozytywnych rezultatów. Przy tym założeniu wymienione wyżej wady tracą na znaczeniu, a potencjalne możliwości zastosowań sieci neuronowych obejmują zagadnienia:

- klasyfikacji tematycznej (semantycznej) atrybutów cech map,
- rozpoznawania struktury, np. w ocenie ilościowej i klasyfikacji liniowych obiektów na mapach ze względu na proces ich postania,
- wyboru najlepszego rozwiązania,
- wykorzystania holistycznej natury sieci neuronowych.

Tradycyjnie rozpoznawanie struktury odbywa się w oparciu o obliczenia „sygnatury” (reprezentowanej przez różnorodne miary) i użycie statystycznych metod klasyfikacji. Użycie sieci neuronowych (w szczególności sieci o uczeniu z rywalizacją jak sieci Kohonena, sieci ART) może dać lepsze rezultaty, zwłaszcza gdy obliczanie „sygnatur” odbywać się będzie na etapie przetwarzania wstępnego.

5 Podsumowanie

W metodach algorytmicznych bazujący na wektorowym modelu generalizacja polega na zastosowaniu wobec nich pewnych operatorów. Skutków zastosowania tych operatorów, a zwłaszcza ich sekwencji nie można przewidzieć. W reprezentacji rastrowej generalizacja poprzedzana jest operacjami pogrubiania i ścieniania linii konturów. Podobnie jak w przypadku reprezentacji wektorowej również i tutaj trudno przewidzieć i kontrolować efekty generalizacji. Z zastosowaniem sieci neuronowych wiąże się nadzieje, że będą one w sobie wiązać wszystkie cechy wspomnianych operatorów lub że będą wyznaczać ich kolejność wraz z parametrami. Zakłada się, że wygenerowanie odpowiednich ciągów uczących i wytrenowanie nimi sieci neuronowych zwolni człowieka od analizowania osobno każdego przypadku, pozwalając na całkowitą automatyzację procesu generalizacji.

Z zastosowaniem sieci wiąże się kilka zagadnień: problem doboru topologii sieci, problem doboru reprezentacji danych wejściowych, problem doboru algorytmu uczenia sieci, problem interpretacji wyjścia z sieci, problem doboru danych do zbiorów uczących.

Dobór właściwej reprezentacji danych wejściowych lub zbioru uczącego jest dużo ważniejszy niż wybór algorytmu uczenia. Właściwie dobrana reprezentacja danych powinna zapewniać [10]: niezmienniczość (dwie krzywe o tym samym kształcie powinny mieć tę samą reprezentację), jednoznaczność (jeśli dwie krzywe mają różny kształt, powinny one mieć różną reprezentację); stabilność (zmiana reprezentacji powinna zachowywać podobieństwo krzywych); wydajność (reprezentacja powinna ograniczać wymagania czasowe i zasobowe algorytmów); łatwość implementacji; czułość (własności kształtu krzywej powinny być obecne w jej reprezentacji).

Normalnie krzywe przedstawiane są jako uporządkowane zbiory par (x, y) (tj. łamane) lub jako binarne rastry. W tych reprezentacjach każde przekształcenie liniowe (przesu-

nięcie, obrót, jednorodne skalowanie) zmienia reprezentowany obiekt. Znaczy to, że bez zmiany reprezentacji sama sieć musiałaby nauczyć się niezmienniczości. W przykładach przedstawianych w niniejszym opracowaniu tak właśnie postąpiono. Dlatego dużo lepszym rozwiązaniem byłoby dokonanie zmiany reprezentacji obiektów (zobacz rozdział 3) na etapie wstępnym i dopiero potem skorzystanie z sieci neuronowych.

Niniejsze opracowanie powstało w ramach projektu „Automatyzacja procesu generalizacji map cyfrowych – zastosowanie sztucznej inteligencji w kartografii”, 9T12E01115.

Literatura

1. Brassel K.E., Weibel R.: *A review and conceptual framework of automated map generalization*. International Journal of Geographical Information Systems, 2(3): s. 229–244, 1988.
2. Douglas D.H., Peucker T.K.: *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*. Canadian Cartographer, 10(2): s. 112–122, 1973.
3. He Z.: *The research on fractal method of automated generalization of coastlines and rivers*. Proc. 18th ICC, Stockholm, 1997.
4. Kankanhalli M.M.: *An adaptive dominant point detectin algorithm for digital curves*. Pattern Recognition Letters, (14): s. 385–390, 1993.
5. Zhilin Li, Openshaw S.: *A natural principle for the objective generalization of digital maps*. Cartography and geographic information systems, (1): s. 19–29, 1993.
6. Liqiu Meng: *Application of neural network in cartographic pattern recognition*. Proc. 16th Int. Cartographic Conference, (1): s. 192–202, Cologne, 1993.
7. Liqiu Meng: *Automatic generalization of geographic data*. Technical report, VBB Viak, Stockholm, Sweden, 1997.
8. Maguire D.J., Goodchild M.F., Rhind D.W.: *Geographical Information Systems: Principles and applications*. Longman, 1991.
9. MGE Map Generalizer Reference Manual.
10. Mokhtarian F., Mackworth A.: *A theory of multiscale, curvature-based shape representation for planar curves*. IEEE Trans. on Pattern Analysis and Machine Intelligence, 14(8): s. 789–805, 1992.
11. Openshaw S., Wymer C., Cross A.: *Using neural nets to solve some hard analysis problems in GIS*. Proc. EGIS '91, (2): s. 797–807, Bruxelles, 1991.
12. Rauber T.W.: *Two-dimentional shape description*. Technical Report GR UNINOVA-RT-10-94, Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia; Quinta da Torre, Monte de Caparica, Portugal, Lisboa, Portugal, 1994.
13. Reiss T.H.: *The revised fundamental theorem of moment invariants*. IEEE Trans. on Pattern Analysis and Machine Intelligence, 13(8): s. 830–834, August 1991.
14. Rosin Paul L.: *Determining local natural scales of curves*. Pattern Recognition Letters, (19): s. 63–75, 1998.
15. Werschlein T., Weibel R.: *Design considerations for the use of neural networks in line generalization*. Technical report, Internal Report. Department of Geography, University of Zurich, 1994.
16. Peng-Yeng Yin: *Algorithms for stright line fitting using k-means*. Pattern Recognition Letters, (19): s. 31–41, 1998.

Streszczenie

W pracy podjęto się próby pokazania, jakie potencjalne możliwości i niebezpieczeństwa kryją się w zastosowaniu sieci neuronowych do problemu generalizacji w kartografii. Ponieważ poruszone zagadnienie jest słabo opisane w literaturze, w pracy dokonano wprowadzenia w tematykę związaną z systemami typu GIS i generalizacją map. Szczególną uwagę poświęcono problemowi generalizacji linii, tj. standardowym metodom generalizacji, reprezentacji i klasyfikacji linii, jak również perspektywom zastosowania tam sieci neuronowych. Praca zawiera opis dwóch eksperymentów z zastosowaniem sieci neuronowych do generalizacji linii. Eksperymenty te mają charakter wstępny. Ich celem było uwidocznienie, jakie problemy wiążą się z użyciem sieci neuronowych i gdzie należy szukać ich rozwiązań.

AUTOMATED LINE GENERALIZATION IN GEOGRAPHICAL INFORMATION SYSTEMS

Summary

In this paper some problems connected with use of neural networks in line generalization are discussed. Because a subject of possible application of neural networks in cartography is concerned as a new idea, paper contains introductory information about GIS. Some standard methods used in generalization task and line classification are described. In the paper two simple experiments with use of neural network for line generalization task are presented. The aim of these experiments was to highlight, that neural networks needs some preparation for this task and that „naive” use of neural networks is not a good solution.

Referat recenzowany

STEROWANIE ROZMYTE STATKIEM WZDŁUŻ ZADANEJ TRAJEKTORII Z ZADANĄ PRĘDKOŚCIĄ

1 Wprowadzenie

Wzrost rozmiarów i prędkości statków oraz przewożenie przez nie różnych, czasami niebezpiecznych ładunków, wymusza zwiększenie marginesu bezpieczeństwa przy manewrach tych statków, a zwłaszcza na ograniczonych akwenach lub w porcie.

Sterowanie ruchem statku wzdułuż zadanej trajektorii jest zadaniem sterowania programowego. Pożądana trajektoria ruchu statku może być trajektoria bezpieczną, omijającą obszary zagrożenia kolizją lub też trajektoria wynikającą z optymalizacji trasy rejsu. W literaturze proponowane były różne regulatory trajektorii [4], [5], [6], [9]. Inną klasę układów sterowania statkiem stanowią systemy, które umożliwiają zarówno sterowanie statkiem wzdułuż zadanej trajektorii z dowolnym kątem dryfu, jak i operacje dynamicznego pozycjonowania [1], [2], [3].

W ostatnich latach wiele publikacji dotyczy różnorodnych zastosowań teorii sterowania rozmytego. Na bazie tej teorii opracowany został i przedstawiony w referacie algorytm sterowania statkiem wzdułuż zadanej trajektorii ze stabilizacją zadanej prędkości.

2 Sformułowanie problemu

Trajektoria ruchu statku określana jest zazwyczaj jako linia łamana składająca się z odcinków prostych. Zadane prędkości statku dla poszczególnych odcinków trajektorii mogą być różne i określone przez czasy osiągnięcia kolejnych punktów zwrotu. Prędkość statku na prostoliniowym odcinku trajektorii pomiędzy kolejnymi punktami zwrotu jest niezmienna.

Proces sterowania statkiem wzdułuż zadanej trajektorii przedstawia się na płaszczyźnie w następujących układach współrzędnych [6]:

- X_0Y_0 - nieruchomym, prawoskrętnym układzie współrzędnych kartezjańskich z osią X_0 skierowaną w kierunku północy;
- XYZ - prawoskrętnym układzie współrzędnych kartezjańskich związanym z kadłubem statku, z kierunkiem osi X zgodnym z linią diametralną kadłuba i początkiem układu 0 w połowie odległości pomiędzy pionami kadłuba oraz w przecięciu płaszczyzn: diametralnej kadłuba oraz linii wodnej (wyznaczonej powierzchnią wody spokojnej).

Trajektorię zadaną statku, wzdułuż której ma się poruszać statek, wyznacza uporządkowany ciąg elementów $(x_k, y_k) \in X_0Y_0$ opisujących współrzędne kolejnych punktów zwrotu. Kierunek każdego odcinka trajektorii jest określany względem osi X_0 i oznaczony kątem ψ_{zk} . Na każdym odcinku trajektorii zadawana jest również prędkość V_{zk} z jaką ma się poruszać statek. Dodatkowo na każdym z odcinków trajektorii wyznaczana jest droga wyprzedzenia manewru L_{wk} , która określa w jakiej odległości od punktu zwrotu należy rozpocząć manewr zmiany odcinka trajektorii.

* Wyższa Szkoła Morska w Gdyni, Katedra Automatyki Okrętowej, ul. Morska 83, 81-225 Gdynia.
E-Mail: {kskula, lmoraw, tomera}@wsm.gdynia.pl

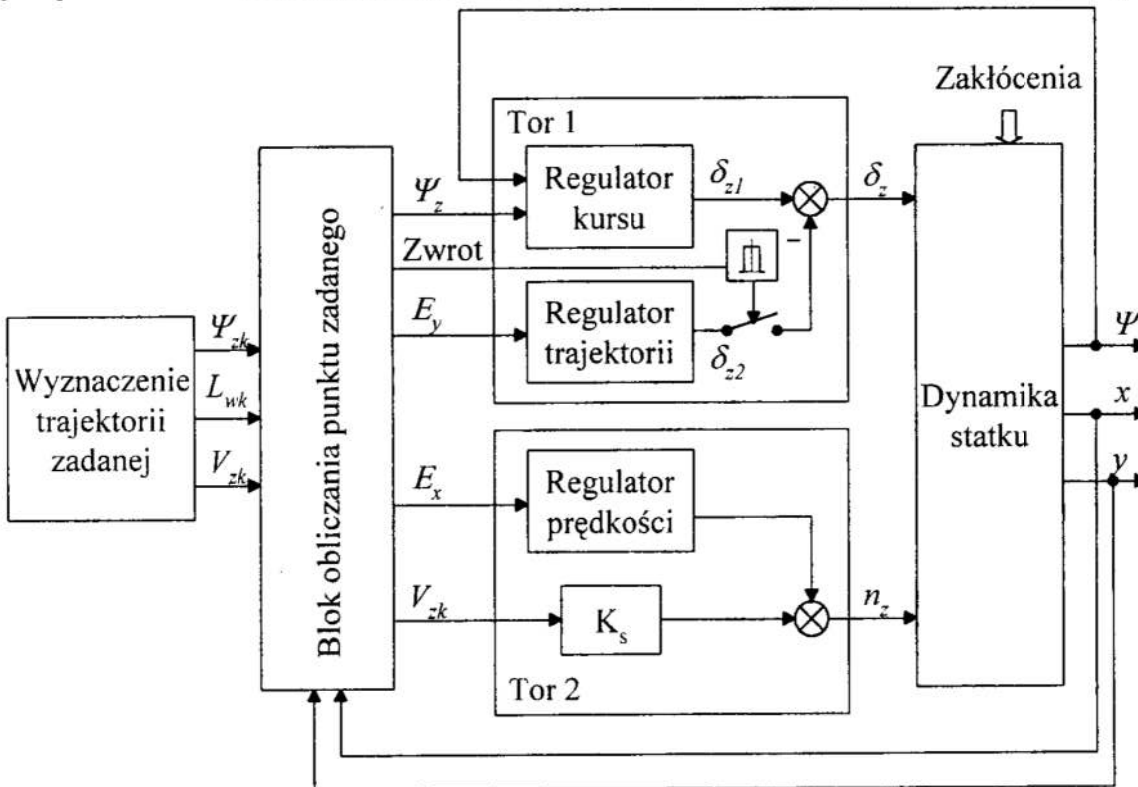
$$E_y = (x_z - x) \cdot \sin \psi_{zk} - (y_z - y) \cdot \cos \psi_{zk} \quad (4)$$

gdzie:

$$L_x = (x_z - x) \cdot \cos \psi_{zk} + (y_z - y) \cdot \sin \psi_{zk} \quad (5)$$

3 Struktura układu sterowania

Układ sterowania realizujący przedstawiony cel sterowania został pokazany na rysunku 2. Składa się on z dwóch rozdzielnych niezależnych torów sterowania. W pierwszym układzie sterowania regulator generuje zadany kąt wychylenia steru, który minimalizuje odchyłkę boczną (poprzeczną) kadłuba E_y od zadanej trajektorii. W drugim układzie za pomocą prędkości obrotowej śruby (siły naporu śruby) zmieniana jest prędkość ruchu statku, która prowadzi do zmniejszenia odchylenia wzdłużnego E_x .



Rys. 2. Układ sterowania statkiem wzdłuż zadanej trajektorii z zadaną prędkością

Fig. 2. The system of ship control along given trajectory at desired speed

W bloku "obliczania punktu zadanego" (Rys.2) wyznaczane są na podstawie wzorów (3) i (4) uchyby: wzdłużny E_x oraz boczny E_y . "Wyznaczenie trajektorii zadanej" odbywa się w bloku nadrzędnym, którym na przykład może być system antykolizyjny określający bezpieczną trajektorię ruchu statku oraz jego prędkości. Trajektorię zadaną może być również ortodroma aproksymowana odcinkami loksodromu. Parametrami wyjściowymi bloku są również wielkości drogi wyprzedzenia manewrów dla każdego odcinka trajektorii, które wyznaczane są na podstawie wcześniej określonych właściwości dynamicznych sterowanego obiektu.

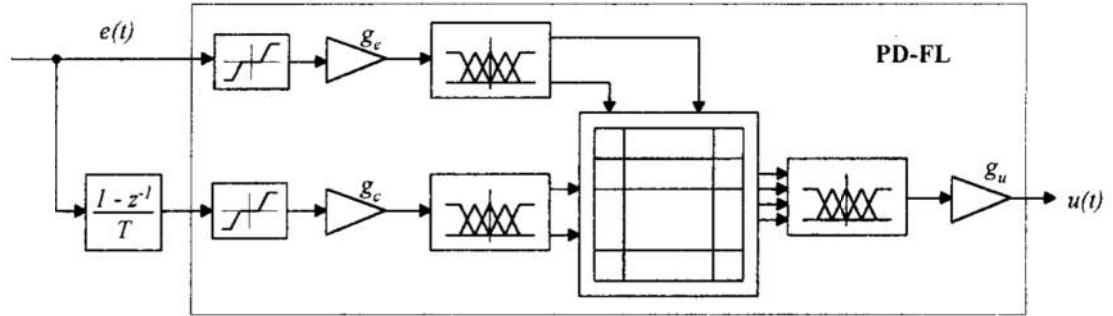
W trakcie wykonywania przez statek manewru zwrotu wyłączany jest regulator trajektorii, natomiast kurs zadany ψ_{zk} jest równy kierunkowi poruszającego się "obiekту odniesienia"

$$\psi_z = \psi_o \quad (6)$$

Po wykonaniu manewru zwrotu kurs zadany ψ_z przyjmuje wartość określoną przez kierunek nowego odcinka trajektorii

$$\psi_z = \psi_{zk} \quad (7)$$

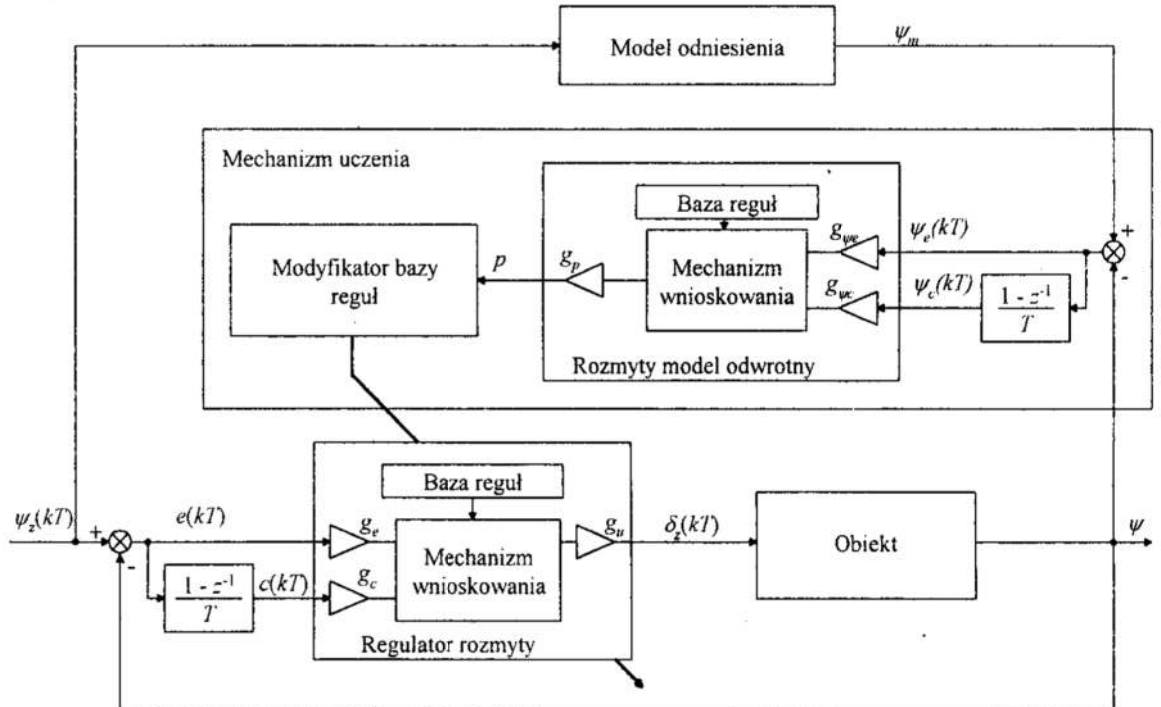
Podstawowym algorytmem sterowania wykorzystywanym w regulatorach rozmytych jest układ pokazany na rysunku 3 z jedną dwuwymiarową bazą reguł, a oznaczany jako PD-FL.



Rys. 3. Algorytm regulatora PD-FL

Fig. 3. PD-FL controller algorithm

Jako regulator kursu zastosowano regulator rozmyty o przestrzajanej dwuwymiarowej bazie ze 121 regułami. Każdy z sygnałów wejściowych rozmywany jest na 11 zbiorów rozmytych o kształcie trójkątnym. Strojenie bazy reguł odbywa się w układzie adaptacyjnym z modelem odniesienia pokazanym na rysunku 4.



Rys. 4. Schemat blokowy regulatora kursu

Fig. 4. Blok diagram of the course controller

Zasada działania regulatora adaptacyjnego (Rys. 4) została szczegółowo opisana w referacie przedstawionym na X symposium SPD-10 w Zakopanem [7]. Potrzeba jego zastosowania wynikała z niestacjonarności sterowanego obiektu, zwłaszcza w zakresie małych prędkości. Przy małych prędkościach w układzie sterowania występowały słabo tłumione oscylacje spowodowane zbyt małym działaniem wyprzedzającym regulatora.

Regulator trajektorii likwiduje odchyłkę boczną statku od trajektorii zadanej. Działanie tego regulatora opiera się na algorytmie sterowania rozmytego typu PDPI-FL [2]. Składa się z dwóch baz reguł bez możliwości automatycznego dostrajania. Regulator ten szczegółowo opisany został w pracy [8].

Zastosowany regulator prędkości posiada taką samą strukturę jak regulator trajektorii. Obydwa regulatory różnią się między sobą innymi wartościami wejściowych i wyjściowych wzmocnień skalujących.

4 Wyniki symulacji

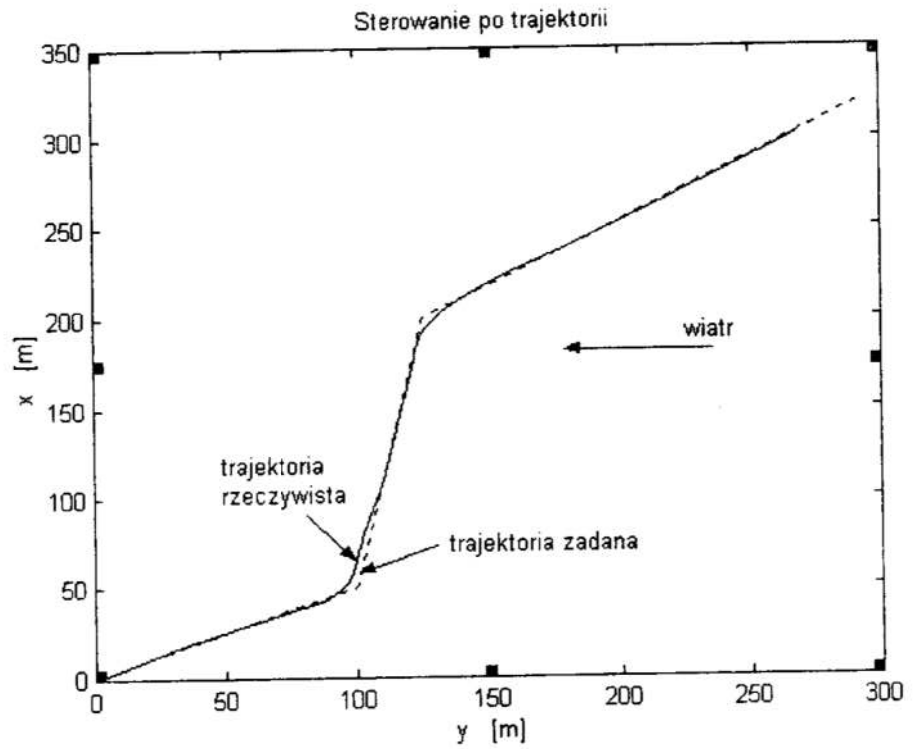
Przedstawiony układ sterowania poddany został badaniom symulacyjnym w środowisku obliczeniowym Matlab/Simulink. Sterowanym obiektem był izomorficzny nieliniowy model matematyczny izomorficznego modelu fizycznego zbiornikowca VLCC klasy E wykorzystywany do szkolenia kapitanów w Ośrodku Manewrowania Fundacji Bezpieczeństwa Żeglugi i Ochrony Środowiska na jeziorze Silm w Iławie-Kamionce. Model statku uwzględnia zmiany parametrów pod wpływem zmian stopnia załadowania, prędkości statku oraz uwzględnia oddziaływanie zakłóceń hydrometeorologicznych. Próby sterowania ruchem zbiornikowca wzdłuż zadanej trajektorii przeprowadzono przy różnych prędkościach, i różnych kształtach trajektorii. Przykład zadanej i otrzymanej rzeczywistej trajektorii ruchu zbiornikowca przedstawiony jest na rysunku 5. Wyniki odpowiadają zakłóceniom wiatrowym o sile 6°B i kierunku wiatru określonym strzałką.

Przebiegi uchybów wzdłużnego E_x i poprzecznego E_y od zadanej trajektorii pokazano na rysunku 6. Na rysunkach 7 i 8 pokazane zostały odpowiednio przebiegi: kursu statku, zadanego wychylenia steru, zadanej i rzeczywistej prędkości statku oraz zadanego kąta ustawienia łopatek śruby napędu głównego.

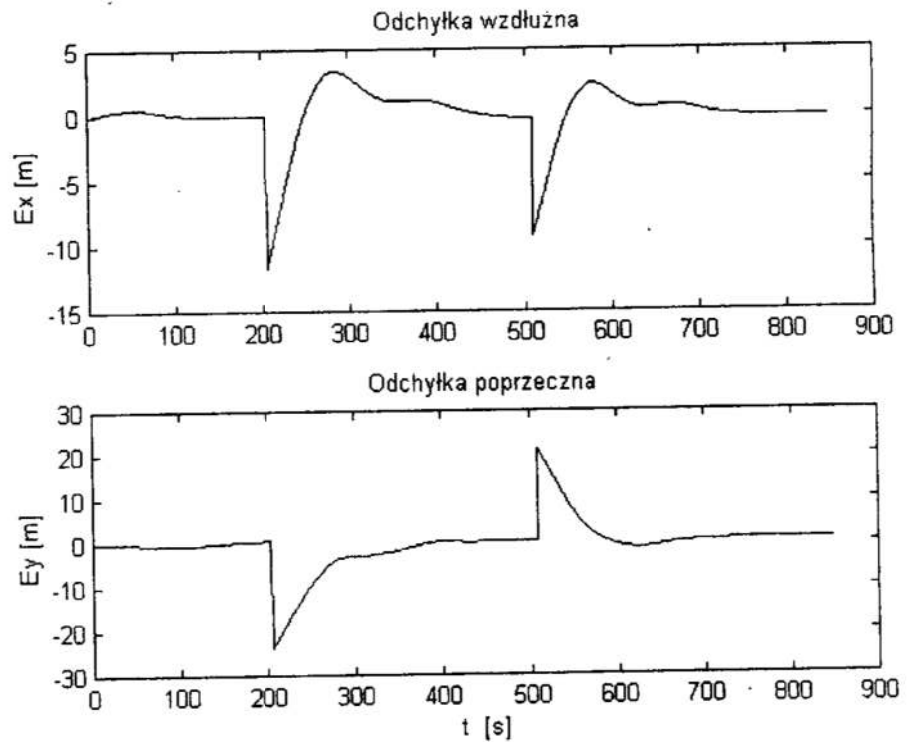
5 Wnioski

Wyniki przeprowadzonych badań symulacyjnych pokazały, że opracowany układ sterowania ruchem statku wzdłuż zadanej trajektorii charakteryzuje mała wrażliwość na zmiany właściwości dynamicznych obiektu sterowania. Wartości ustalone uchybów wzdłużnego i bocznego na prostoliniowych odcinkach trajektorii nie przekraczały odpowiednio 0.2 i 0.4 m. Baza reguł regulatora adaptacyjnego jest efektywnie dostrajana w warunkach znacznego "pobudzenia" układu, co ma miejsce przy zmianach odcinków trajektorii i zmianach prędkości statku. Zastosowany sposób wyłączania regulatora trajektorii podczas zmiany kursu zmniejszył znacząco wzajemne oddziaływanie na siebie obydwu torów sterowania. Dostrajanie bazy reguł regulatora kursu okazało się niezbędne ze względu na znaczne zmiany właściwości dynamicznych statku przy zmianach jego prędkości.

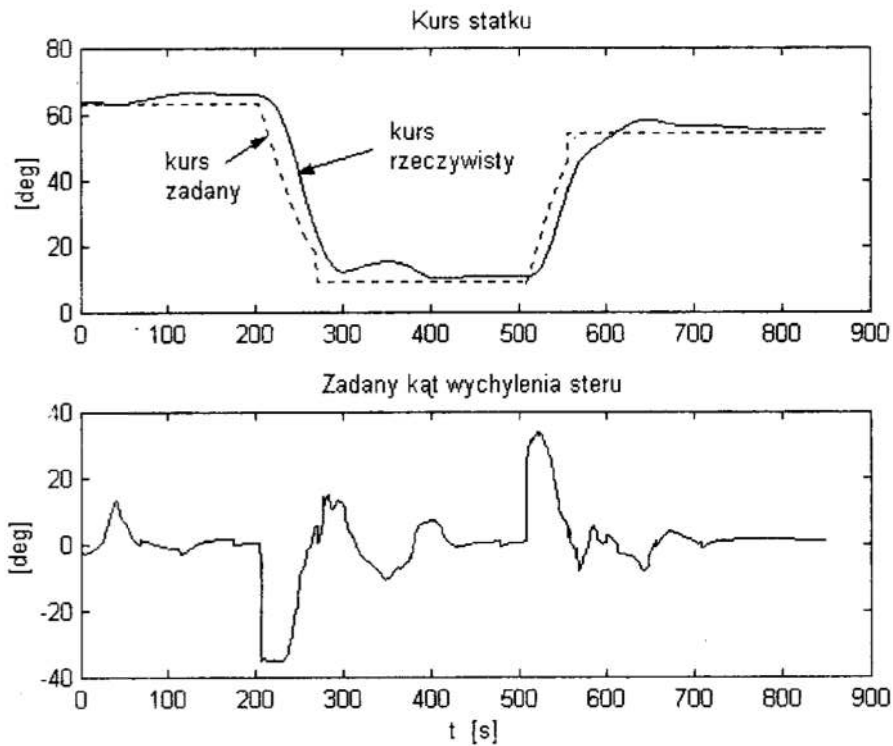
Praca wykonywana była w ramach realizacji grantu nr 8 T11A 015 15, finansowanego przez Komitet Badań Naukowych.



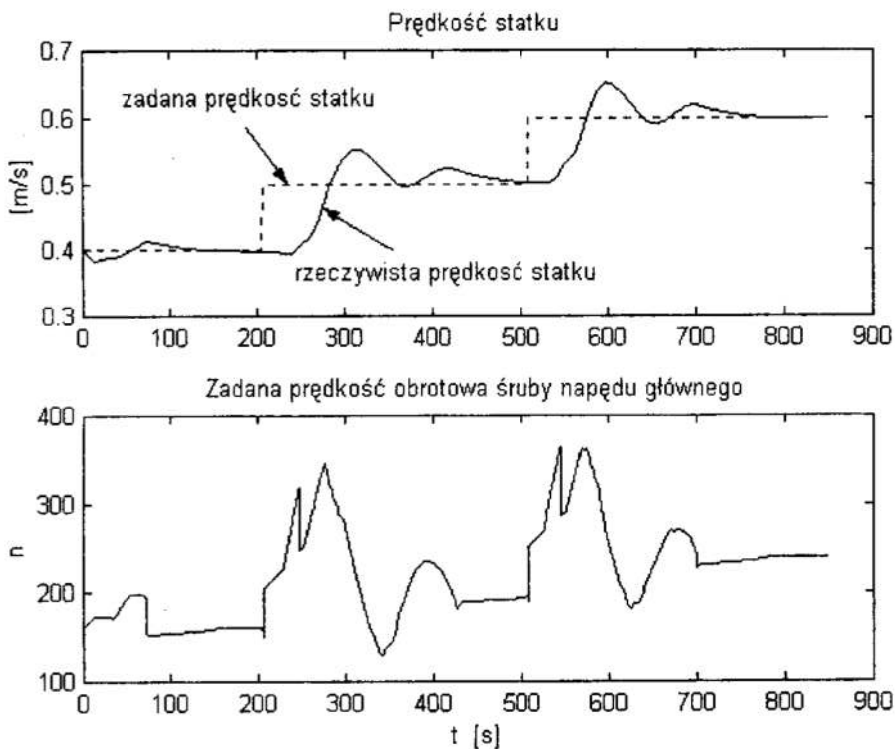
Rys. 5. Sterowanie statkiem wzdłuż zadanej trajektorii
 Fig. 5. Ship trajectory in track-keeping



Rys. 6. Odchyłki od zadanej trajektorii: wzdłużna E_x i poprzeczna E_y
 Fig. 6. Deviation from a given trajectory: longitudinal E_x and transverse E_y



Rys. 7. Kurs statku i zadane wychylenie steru
 Fig. 7. Ship heading and comanded rudder angle



Rys. 8. Prędkość statku i zadany kąt ustawienia łopatek śruby napędu głównego
 Fig. 8. ship velocity and desired revolution of the propeller

Literatura

1. Bertin D.: "Track-keeping controller for a precision manoeuvring autopilot", Proceedings of IFAC Conference Control Application in Marine Systems, CAMS 98, Fukuoka, Japan, pp. 155-160, 1998.
2. Broel-Plater B.: "Fuzzy control for vessel motion", Proceedings of the 5th International Symposium on Methods and Models in Automation and Robotics, Międzyzdroje, Poland, vol. 2, pp. 697-702, 1998.
3. Broel-Plater B.: "Rozmyte sterowanie powolnym ruchem jednostki pływającej", Materiały VII Krajowej Konferencji Naukowo-Dydaktycznej "Automatyzacja i eksploatacja systemów sterowania" AMW, Gdynia, ss. 367-374, 1999.
4. Holzhuter T.: "A high precision track controller for ships", Proceedings of the 11th IFAC World Congress, Tallin, Estonia, pp.118-123, 1990.
5. Kallstrom C.G.: *Identification and adaptive control applied to ship steering*, Ph.D Thesis, Lund Institute of Technology, Sweden, 1979.
6. Morawski L.: *Metody syntezy układów sterowania ruchem statku*, Rozprawa habilitacyjna, Gdynia 1994
7. Morawski L., Tomera M.: "Samouczący autopilot rozmyty", Prace X Sympozjum SPD-10, Zakopane-Kościelisko, pp. 275-281, 1998.
8. Morawski L., Tomera M.: "Samouczący regulator rozmyty z modelem odniesienia", XIII Krajowa Konferencja Automatyki, Opole, vol. 2, pp. 243-246, 1999.
9. Zuidweg J.K.: "Optimal and sub-optimal feedback in automatic track-keeping system" Proceeding of 6th Ship Control Systems Symposium, vol. 3, Ottawa. 1981.

Streszczenie

W pracy zaprezentowano układ regulatora zbudowanego na bazie algorytmów sterowania rozmytego, który obejmuje dwa podukłady sterowania. Pierwszy jest układem regulacji kursu z kątem wychylenia steru jako sygnałem sterującym minimalizującym odchylenia boczne statku od trajektorii. W drugim podukładzie za pomocą obrotów śruby napędu głównego regulowana jest prędkość w celu zminimalizowania odchyłki wzdłużnej statku. Algorytmy sterowania opracowane zostały na bazie stacjonarnych i adaptacyjnych układów sterowania rozmytego.

SHIP TRACK-KEEPING FUZZY CONTROL WITH ASSUMED VELOCITY

Summary

The paper presents a system controlling the ship motion along the given trajectory. The system includes two internal sub-systems. The first of them controls the ship's course using a rudder turning angle as the controlling signal for minimising ship's transverse deviation from the given trajectory. In the second sub-system the ship's speed is controlled with the aid of main propulsion propeller's revolutions in order to minimise ship's longitudinal deviation. Controlling algorithms have been worked out basing on time-invariant and adaptable systems of fuzzy control.

Referat recenzowany

SZEREGOWANIE CZYNNIKÓW Z WYKORZYSTANIEM METODY PORÓWNYWANIA PARAMI W SENSIE AHP

1. Wprowadzenie

Metoda porównywania parami w kategoriach AHP (ang. Analytic Hierarchy Process) wprowadzona przez Saaty'ego [12] może być wykorzystana do szeregowania skończonej liczby czynników w szeregowaniu wieloatrybutowym lub wielokryterialnym. Polega ona na przyporządkowaniu każdej parze rozważanych czynników pewnej liczby z założonej skali. Liczba ta wyraża subiektywną preferencję eksperta w stosunku do jednego z elementów pary w porównaniu z drugim. Przyporządkowanie może być dokonywane również przez grupę ekspertów i wtedy mamy do czynienia z grupowym podejmowaniem decyzji. Często rozważa się sytuację tzw. brakujących danych, kiedy ekspert lub eksperci nie podają oceny danej pary czynników lub pewnej ich liczby. Oceny podane przez ekspertów umieszczane są w tzw. macierzach ocen, i ponieważ zakłada się, że proces oceny czynników przez eksperta uległ zniekształceniu, to utworzona macierz ocen zawiera oceny niedokładne, a zadaniem procesu szeregowania jest znalezienie odpowiadającej jej macierzy z ocenami dokładnymi. Z matematycznego punktu widzenia problem szeregowania sprowadza się do aproksymacji macierzy. Stosuje się tutaj głównie trzy metody: maksymalnej wartości własnej [12], [13], najmniejszych kwadratów [4], [12] oraz logarytmicznych najmniejszych kwadratów [1], [2], [4], [5], [6].

W pracy omawia się krótko podstawowe metody szeregowania dla przypadku jednego eksperta, a następnie przedstawia propozycje rozwiązania zagadnienia dla grupowego podejmowania decyzji z uwzględnieniem problemu brakujących danych. Uwaga skoncentrowana jest głównie na metodach szeregowania. Pomija się zagadnienie agregacji wielopoziomowej w hierarchicznym procesie decyzyjnym. Opis zagadnienia agregacji można znaleźć w książce Saaty'ego [12] lub np. pracy [5].

2. Sformułowanie zagadnienia porównywania parami

Załóżmy, że mamy n czynników F_1, F_2, \dots, F_n , które należy uszeregować. Załóżmy ponadto, że każdej parze czynników (F_i, F_j) , $i, j=1, \dots, n$, ekspert przyporządkowuje liczbę $r_{ij} \in S$, $S = \{1/9, \dots, 1/2, 1, 2, \dots, 9\}$, wyrażającą jego subiektywne preferencje dotyczące pierwszego elementu pary w stosunku do drugiego, zgodnie z zasadami przedstawionymi w Tabelicy 1 [12], [5]. Następnie wyniki umieszczane są w macierzy ocen R :

* Politechnika Gdańska, Wydział Elektrotechniki i Automatyki, G. Narutowicza 11/12, 80-952 Gdańsk
Email: mkwies@ely.pg.gda.pl

** doktorant na Wydziale Elektrotechniki i Automatyki Politechniki Gdańskiej

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}, \quad (1)$$

gdzie:

$$r_{ij} > 0, \quad r_{ij} = 1/r_{ji} \quad \forall i, j = 1, 2, \dots, n. \quad (2)$$

Tablica 1. Oceny odpowiadające preferencjom ekspertów
Table 1. Judgements associated with expert preferences

Ocena	Preferencja
1	Równoważność alternatyw
3	Słaba preferencja jednej z alternatyw
5	Istotna preferencja jednej z alternatyw
7	Wyraźna preferencja jednej z alternatyw
9	Bezwzględna preferencja jednej z alternatyw
2,4,6,8	Wartości pośrednie
Odwrotności powyższych liczb	Relacja odwrotna

Koncepcja Saaty'ego polega na aproksymacji macierzy ocen R za pomocą następującej macierzy ilorazów:

$$P = \begin{pmatrix} p_1/p_1 & p_1/p_2 & \cdots & p_1/p_n \\ p_2/p_1 & p_2/p_2 & \cdots & p_2/p_n \\ \vdots & \vdots & \vdots & \vdots \\ p_n/p_1 & p_n/p_2 & \cdots & p_n/p_n \end{pmatrix}. \quad (3)$$

Macierz R jest macierzą z niezgodnymi ocenami, natomiast macierz P zawiera oceny zgodne, wyrażone w postaci ilorazów $p_{ij} = p_i/p_j, i, j = 1, 2, \dots, n$. Otrzymując macierz P otrzymujemy równocześnie wektor p uszeregowania zgodnego z preferencjami decydenta:

$$p = (p_1, \dots, p_n)^T, \quad (4)$$

który następnie jest poddawany normalizacji arytmetycznej.

W celu znalezienia wektora p stosowane są głównie trzy metody: maksymalnej wartości własnej [12], [13], najmniejszych kwadratów [4], [12], [13] oraz logarytmicznych najmniejszych kwadratów [1], [2], [4], [5], [6].

Jeśli założymy, że w przypadku ogólnym liczby D ekspertów (pewnej skończonej liczby ocen na parę) występuje sytuacja brakujących ocen, wówczas k -tą ocenę pary (i, j) , można oznaczyć jako r_{ijk} , $r_{ijk} \in S$, gdzie $k = 0, 1, 2, \dots, d_{ij} \leq D$, $\forall i, j = 1, 2, \dots, n$, natomiast d_{ij} jest liczbą ocen pary (i, j) .

Zagadnienia z pełnymi macierzami ocen, można rozwiązywać w oparciu o wymienione metody, obliczając znormalizowane uszeregowania dla każdego z ekspertów, a następnie dokonując agregacji uszeregowania względem ekspertów.

W przeciwnym przypadku należy uzupełnić brakujące oceny lub stosować metodę najmniejszych kwadratów albo logarymicznych najmniejszych kwadratów, ponieważ metoda maksymalnej wartości własnej nie dopuszcza przypadków z brakującymi ocenami.

3. Rozwiązanie zagadnienia dla przypadku jednego eksperta bez brakujących danych

Metoda maksymalnej wartości własnej

Koncepcja Saaty'ego [12], [13] opiera się na obserwacji że macierz P z ocenami zgodnymi posiada następującą własność:

$$Pp = np, \quad (5)$$

gdzie p jest wektorem uszeregowania, natomiast P – macierzą ilorazów. Biorąc pod uwagę fakt, że macierz ocen R jest macierzą z niedokładnymi danymi nie można do niej zastosować zależności (5).

Aby sformułować odpowiedni warunek dla macierzy ocen przytoczmy podstawowe fakty z algebry liniowej. Jeśli wektor $\lambda = (\lambda_1, \dots, \lambda_n)^T$ jest wektorem wartości własnych macierzy A : $Ax = \lambda, x, i = 1, \dots, n$ oraz $a_{ii} = 1 \forall i = 1, \dots, n$, to: $\sum_{i=1}^n \lambda_i = n$. Wynika stąd,

że jeśli zachodzi warunek (5) to wszystkie wartości własne oprócz jednej są równe zero. Ta jedna niezerowa jest równa n i jest ona największą wartością własną. Jeśli założymy teraz, że macierz ocen R posiada tylko nieznacznie zaburzone dane w porównaniu z macierzą P , wówczas można oczekiwać, że największa wartość własna macierzy ocen R jest bliska n , a pozostałe jej wartości własne są bliskie zero. Zatem powracając do szeregowania należy znaleźć wektor p spełniający zależność [12], [13]:

$$Rp = \lambda_{\max} P, \quad (6)$$

co sprowadza się w rozważanym przypadku do znalezienia unormowanego do jedności wektora własnego, odpowiadającego maksymalnej wartości własnej macierzy ocen R .

Metoda najmniejszych kwadratów

Obliczenie uszeregowania z wykorzystaniem metody najmniejszych kwadratów [4], [13] opiera się na znalezieniu wektora p minimalizującego wyrażenie:

$$\sum_{i,j=1}^n \left(r_{ij} - \frac{p_i}{p_j} \right)^2, \quad (7)$$

co oznacza znalezienie takiej macierzy ocen zgodnych P , która jest najbliższej położona w stosunku do macierzy ocen R w sensie normy Euklidesowej.

Do aproksymacji macierzy R można wykorzystać metodę aproksymacji poprzez macierz o niższym rzędzie [3], [4], [13], [14], która opiera się na koncepcji rozkładu macierzy na wartości szczególne:

$$A = QSU^T, \quad (8)$$

gdzie macierz A jest dowolną macierzą kwadratową, Q i U są macierzami ortogonalnymi, natomiast macierz S jest macierzą diagonalną, zawierającą wartości szczególne

macierzy A . Macierzą rzędu r ($r < n$) aproksymującą macierz A będziemy nazywali macierz A_r :

$$A_r = Q_r S_r U_r^T, \quad (9)$$

gdzie macierze Q_r oraz U_r^T są częściami odpowiednio macierzy Q oraz U^T , odpowiadającymi pierwszym r kolumnom macierzy S .

Stosując powyższą ideę do aproksymacji macierzy R za pomocą macierzy P otrzymujemy:

$$R = Q S U^T, \quad (10)$$

gdzie Q i U są macierzami ortogonalnymi, natomiast macierz S jest macierzą zawierającą wartości szczególne macierzy R . Biorąc pod uwagę fakt, że macierz ilorazów jest rzędu 1, można ją obliczyć zgodnie z zależnością:

$$P = Q_1 S_1 U_1^T = s_1 q_1 u_1^T, \quad (11)$$

gdzie macierze Q_1 oraz U_1^T są częściami odpowiednio macierzy Q oraz U^T , odpowiadającymi pierwszej kolumnie macierzy S , q_1 oraz u_1 są pierwszymi wektorami ortonormalnymi odpowiednio macierzy Q oraz U , natomiast s_1 jest odpowiadającą im wartością szczególną. Należy wybrać maksymalną wartość szczególną $s_1 = \max(s_i)$, ponieważ wówczas minimalna odległość między macierzami P i R w sensie zdefiniowanej metryki (7) wyrazi się zależnością [3]:

$$\sum_{i,j=1}^n \left(r_{ij} - \frac{p_i}{p_j} \right)^2 = \sum_{i=2}^n s_i^2, \quad (12)$$

gdzie $s_i, 2=1, \dots, n$ są pozostałymi wartościami szczególnymi macierzy R . Rozwiązanie przy zastosowaniu powyższej metody nie zawsze jest rozwiązaniem jedynym.

Metoda logarytmicznych najmniejszych kwadratów

Metoda logarytmicznych najmniejszych kwadratów [1], [2], [4], [6] polega na znalezieniu takiej macierzy ilorazów P , aby uzyskać minimalną wartość wyrażenia:

$$\sum_{i,j>1}^n \left(\ln(r_{ij}) - \ln\left(\frac{p_i}{p_j}\right) \right)^2, \quad (13)$$

czyli minimalną wartość odległości pomiędzy macierzami R i P w oparciu o normę euklidesową w skali logarytmicznej.

Dokonując podstawień $y_{ij} = \ln(r_{ij})$, $x_i = \ln(p_i)$, $\forall i, j = 1, \dots, n$ w (13) otrzymamy zagadnienie optymalizacyjne:

$$\min_{x_i, i=1, \dots, n} \left\{ I = \sum_{i,j=1}^n (y_{ij} - x_i + x_j)^2 \right\}, \quad (14)$$

którego rozwiązanie sprowadza się do przyrównania gradientu I do zera oraz rozwiązaniu odpowiedniego układu równań liniowych [4]:

$$\frac{\partial I}{\partial x_i} = -2 \sum_{j=1}^n (y_{ij} - x_i + x_j) = -2 \left(\sum_{j=1}^n y_{ij} - nx_i + \sum_{j=1}^n x_j \right) = 0, \quad i = 1, \dots, n, \quad (15)$$

przy założeniu [4], [5], [6]:

$$\sum_{i=1}^n x_i = 0, \quad (16)$$

który dla wektora uszeregowan p jest warunkiem normalizacji geometrycznej:

$$\prod_{i=1}^n p_i = 1. \quad (17)$$

Wówczas z zależności (15) otrzymamy:

$$x_i = \frac{1}{n} \sum_{j=1}^n y_{ij}, \quad i = 1, \dots, n. \quad (18)$$

Po powrocie do funkcji wykładniczych zależność (18) przyjmie postać:

$$p_i = \left(\prod_{j=1}^n r_{ij} \right)^{\frac{1}{n}}, \quad i = 1, \dots, n. \quad (19)$$

Warto zwrócić uwagę, że rozwiązanie (19) spełnia wcześniej założony warunek geometrycznej normalizacji (17).

W przypadku wielu ekspertów, kiedy każdy z nich podał ocenę danej pary ($\forall i, j, d_{ij} = D$) wyrażenie (13) przyjmie postać:

$$\sum_{k=1}^D \sum_{i, j > i}^n \left(\ln(r_{ijk}) - \ln \left(\frac{p_i}{p_j} \right) \right)^2, \quad (20)$$

dając rozwiązanie:

$$p_i = \left(\prod_{j=1}^n \prod_{k=1}^D r_{ijk} \right)^{\frac{1}{nD}}, \quad i = 1, \dots, n, \quad (21)$$

dla którego spełniony jest również warunek geometrycznej normalizacji (17).

W oparciu o klasyczny liniowy model regresyjny można wykazać, że przy pewnych założeniach dotyczących rozkładu błędów ocen ekspertów metoda średniej geometrycznej jest zgodna z metodą estymatora o maksymalnej wiarygodności [4].

4. Rozwiązanie zagadnienia dla wielu ekspertów i brakujących danych

Jak już wspomniano, w przypadku brakujących danych bezpośrednio można zastosować metodę najmniejszych lub logarytmicznych najmniejszych kwadratów. Pierwsza z nich nie zawsze daje jedyne rozwiązanie co stanowi jej poważną wadę. Zaletą drugiej jest jej zgodność z podejściem przy pomocy estymatora o największej wiarygodności. Poza tym przy małej liczbie brakujących ocen daje ona jedno rozwiązanie [6].

Znalezienie macierzy ocen P , która znajdowałaby się w najmniejszej odległości od macierzy ocen R w sensie normy euklidesowej z wykorzystaniem skali logarytmicznej [4], [5], [6] sprowadzi się w rozważanym przypadku do minimalizacji:

$$I_{\min} = \min_{p_i, i=1, \dots, n} \left\{ \sum_{i=1, j>i}^n \sum_{k=1}^{d_{ij}} \left(\ln(r_{ijk}) - \ln\left(\frac{p_i}{p_j}\right) \right)^2 \right\} \quad (22)$$

Dokonując podstawień $x_i = \ln(p_i)$ i $y_{ijk} = \ln(r_{ijk})$ $i, j = 1, \dots, n$ $k = 1, \dots, d_{ij}$, zagadnienie minimalizacji (22) sprowadza się do układu równań normalnych [1], [2], [4], [6]:

$$x_i \sum_{i \neq j, j=1}^n d_{ij} - \sum_{i \neq j, j=1}^n d_{ij} x_j = \sum_{i \neq j, j=1}^n \sum_{k=1}^{d_{ij}} y_{ijk}, i = 1, 2, \dots, n, \quad (23)$$

gdzie:

$$d_{ij} \geq 0, \forall i, j, \quad \sum_{i \neq j, j=1}^n d_{ij} > 0, \forall i. \quad (24)$$

Metoda uogólnionej pseudoodwrotności

Układ równań (23) można przedstawić w postaci macierzowej:

$$Ax = b, \quad (25)$$

W pracy [6] pokazano, że jeśli macierz A lewych stron układu równań (25) jest rzędu $n-1$, wówczas rozwiązanie ogólne tego układu równań zawsze istnieje i posiada postać:

$$x = A^+ b + z, \quad (26)$$

gdzie wektor z jest wektorem o stałych współczynnikach, natomiast macierz A^+ jest macierzą pseudoodwrotną do macierzy A w sensie Penrose [9], [10]. Rozwiązanie o minimalnej normie:

$$x = A^+ b \quad (27)$$

spełnia wówczas zależność (16).

Warto zwrócić uwagę, że z uwagi na powrót do funkcji wykładniczych oraz arytmetyczną normalizację wyniku stałe współczynniki wektora z ulegają uproszczeniu. Zatem można tu korzystać z zależności (26), a nie (27). Zależność ta może być stosowana w większości przypadków ponieważ w praktyce macierz A jest rzędu $n-1$. Niższe rzędy zdarzają się dla dużej liczby brakujących ocen, co w praktyce spotykane jest bardzo rzadko.

Metoda alternatywna rozwiązania zagadnienia

Jak wspomniano wcześniej, jeśli rozwiązanie o minimalnej normie układu równań (25) istnieje, to spełnia ono warunek (16). Wykorzystanie tego warunku pozwala na modyfikację podejścia do rozwiązania układu równań normalnych [8]. Polega ona na zastąpieniu dowolnego wiersza układu równań normalnych (25) równaniem (16). Wtedy dla przypadku z rzędem macierzy A równym $n-1$ otrzymamy układ równań z macierzą lewych stron pełnego rzędu:

$$Cx = d. \quad (28)$$

Bez utraty ogólności rozważań można przyjąć, że dokonano wymiany pierwszego wiersza macierzy A . Wówczas macierz C przyjmie następującą strukturę:

$$C = \begin{pmatrix} 1 & 1 & \dots & 1 \\ -d_{21} & \sum_{j=1, j \neq 2}^n d_{2j} & \dots & -d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -d_{n1} & -d_{n2} & \dots & \sum_{j=1, j \neq n}^n d_{nj} \end{pmatrix}, \quad (29)$$

a wektor prawych stron d postać:

$$d = (0 \quad b_2 \quad \dots \quad b_n)^T. \quad (30)$$

Ponieważ sumy elementów wierszy macierzy A są równe zero to łatwo pokazać, że pierwszy wiersz macierzy C jest ortogonalny do wszystkich pozostałych jej wierszy, zatem jeśli macierz A jest rzędu $n-1$ to macierz C będzie rzędu n . Jej wiersze w takim przypadku będą tworzyły zbiór n wektorów liniowo niezależnych.

Przedstawiona modyfikacja metody rozwiązania zagadnienia znacznie upraszcza obliczenia. Warto podkreślić, że rozwiązanie:

$$x = C^{-1}d \quad (31)$$

jest rozwiązaniem o minimalnej normie i jest równe rozwiązaniu równania (27).

5. Uwagi i wnioski

W pracy przedstawiono trzy metody szeregowania czynników w oparciu o porównywanie parami w kategoriach AHP. Dwie z nich, a mianowicie największej wartości własnej oraz najmniejszych kwadratów nie mogą być stosowane do przypadków z brakującymi ocenami. Zaproponowano, zatem wykorzystanie w takich przypadkach metody logarytmicznej regresji, którą sprowadzono do rozwiązania układu równań liniowych z wykorzystaniem pseudoodwrotności. Warto podkreślić, że dla przypadku małej liczby brakujących ocen, czyli kompetentnego zespołu ekspertów, znormalizowane rozwiązanie zagadnienia uszeregowania z wykorzystaniem tej metody jest rozwiązaniem jedynym. Przedstawiono modyfikację metody rozwiązania układu równań normalnych, sprowadzając go do układu równań o pełnym rzędzie.

Zaproponowana metodyka podejścia do zagadnienia porównywania parami z brakującymi ocenami może być łatwo rozwinięta dla przypadku z ocenami rozmytymi [6], [9] i zastosowana do analizy rozwiązania zagadnienia rozmytego.

Z punktu widzenia przydatności przedstawionych metod aproksymacji macierzy celowe byłoby zbadanie zgodności ocen ekspertów i jej wpływu na otrzymane uszeregowanie.

Literatura

1. Boender C.G.E., de Graan J.G. and Lootsma F.A.: *Pairwise-comparison methods using fuzzy numbers on ratio scales with geometric progression*. Rep. of Department of Mathematics and Informatics 85-24, Delft University of Technology, 1985.
2. Boender C.G.E., de Graan J.G. and Lootsma F.A.: *Multi-criteria decision analysis with fuzzy pairwise comparisons*. Fuzzy Sets and Systems, 29:133-143, 1989.

3. Eckart C., Young G.: *The approximation of one matrix by another matrix of lower rank*. Psychometrika, 1 (3): 211-218, 1936.
4. Grawford G. and Williams C.: *A note on the analysis of subjective judgement matrices*. Journal of Mathematical Psychology, 29: 387-405, 1985.
5. Kwiesielewicz M.: *Grupowe wielokryterialne podejmowanie decyzji w oparciu o metodę porównywania parami*. Materiały XII Ogólnopolskiej Konferencji Polioptymalizacja i Komputerowe Wspomaganie Projektowania: 91-98, 1994.
6. Kwiesielewicz M.: *The logarithmic least squares and generalised pseudoinverse in estimating ratios*. Europ. J. Operational Research, 93:611-619, 1996.
7. Kwiesielewicz M.: *A note on the fuzzy extension of Saaty's priority theory*. Fuzzy Sets and Systems, 95: 161-172, 1998.
8. Kwiesielewicz M., van Uden E.: *A note on the pseudoinverse approach to estimating ratios*, 1999 (Submitted to Journal of Mathematical Psychology).
9. Laarhoven van P.J.M. and Pedrycz W.: *A fuzzy extension of Saaty priority theory*. Fuzzy Sets and Systems, 11:229-241,1983.
10. Penrose R.: *A generalized inverse of matrices*. Proc. Cambridge Philos. Soc., 51: 406-413, 1955.
11. Pyle L.D.: *The generalised inverse in linear programming. Basic structure*. SIAM J. Appl. Math., 22:335-355, 1972.
12. Saaty T.L. *The Analytic Hierarchy Process*. Mc-Grew Hill New York 1980.
13. Saaty T.L and Vargas L.G.: *Comparison of eigenvalue, logarithmic least squares and least squares method in estimating ratios*. Math. Modelling,5:309-324, 1984.

Streszczenie

W pracy przedstawia się główne metody wykorzystane do aproksymacji macierzy ocen otrzymywanych w wyniku porównywania parami czynników w kategoriach AHP. Metodę logarytmicznej regresji rozwija się do przypadku grupowego podejmowania decyzji i sprowadza do tzw. układu równań normalnych, który potem rozwiązywany jest z wykorzystaniem pseudoodwrotności. Wprowadza się modyfikację metody rozwiązania układu równań normalnych poprzez sprowadzenie go do układu o pełnym rzędzie

RANKING FACTORS USING PAIRWISE COMPARISON METHOD IN SENSE OF AHP

Summary

In the paper main methods of estimation of judgement matrices obtained as a result of the pairwise comparison method of factors ranking in sense of AHP are considered. The logarithmic regression is developed for cases of group decision making with lacking judgements. This method leads to solution of so called normal equation set, which is solved using a pseudoinverse approach. A modification of the solution method is introduced.

Referat recenzowany

KONCEPCJA TWORZENIA REGUŁ HEURYSTYCZNYCH DLA PROJEKTOWANIA ERGONOMICZNEGO

1 Wstęp

W pracy przedstawiono koncepcję tworzenia reguł heurystycznych dla komputerowego wspomaganie projektowania ergonomicznego, w oparciu o strukturę procesu projektowania ergonomicznego i o ergonomiczne kryteria projektowe zaproponowane przez E. Tytyka. Tworzone reguły heurystyczne powinny uwzględniać:

- etapy procesu projektowania,
- wagę podejmowanych decyzji (strategiczne, operacyjne),
- specyfikę ergonomicznych kryteriów projektowych,
- logikę przebiegu procesu projektowania ergonomicznego.

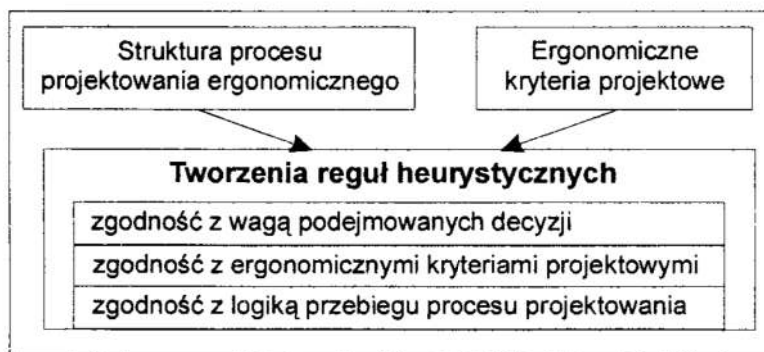
Wymieniona specyfikacja zawęży obszar poszukiwań rozwiązań oraz ukierunkowuje pracę na analizę metodologii projektowania ergonomicznego.

Wyznaczono zasady tworzenia heurystyk w celu zapewnienia zgodności i spójności reguł heurystycznych z procesem projektowania ergonomicznego.

2 Koncepcja tworzenia reguł heurystycznych

Analiza zagadnienia pozwoliła na wyróżnienie następujących zasad tworzenia reguł heurystycznych (Rys. 1):

- zasada zgodności z wagą podejmowanych decyzji,
- zasada zgodności z ergonomicznymi kryteriami projektowymi,
- zasada zgodności z logiką przebiegu procesu projektowego.



Rys. 1. Tworzenia reguł heurystycznych

Fig. 1. Creating the heuristic rules

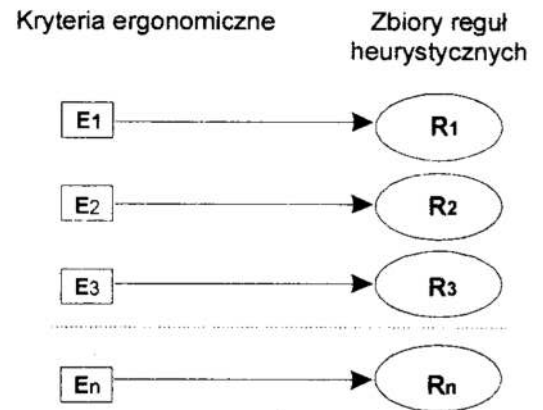
Zasada zgodności z wagą podejmowanych decyzji wiąże się z zasadą zgodności

* Wyższa Szkoła Pedagogiczna, Instytut Techniki, al. Wojska Polskiego 69, 65-077 Zielona Góra

e-mail: A.Lasota@wsp.zgora.pl WWW: <http://asia.aw.wsp.zgora.pl/~lasota/>

Pracę wykonano w ramach projektu badawczego Nr 7 T07C 019 18 finansowanego przez Komitet Badań Naukowych w 2000 roku

z ergonomicznymi kryteriami projektowymi. Ergonomiczne kryteria projektowe podzielone są na grupy, w których uwzględnione są etapy procesu projektowania (formułowanie założeń ergonomiczno–techniczno–ekonomicznych, projektowania wstępnego, itd.) oraz kolejność zagadnień rozwiązywanych podczas projektowania systemu człowiek–obiekt techniczny [4]. Wśród kryteriów występują kryteria o znaczeniu strategicznym i operacyjnym. Aby uzyskać reguły o odpowiednich cechach, kryteria ergonomiczne zostały potraktowane jako „źródło generujące” specyficzne zbiory reguł heurystycznych. Pozwala to na utworzenie dla każdego ergonomicznego kryterium projektowego określonego zbioru reguł heurystycznych (Rys. 2).

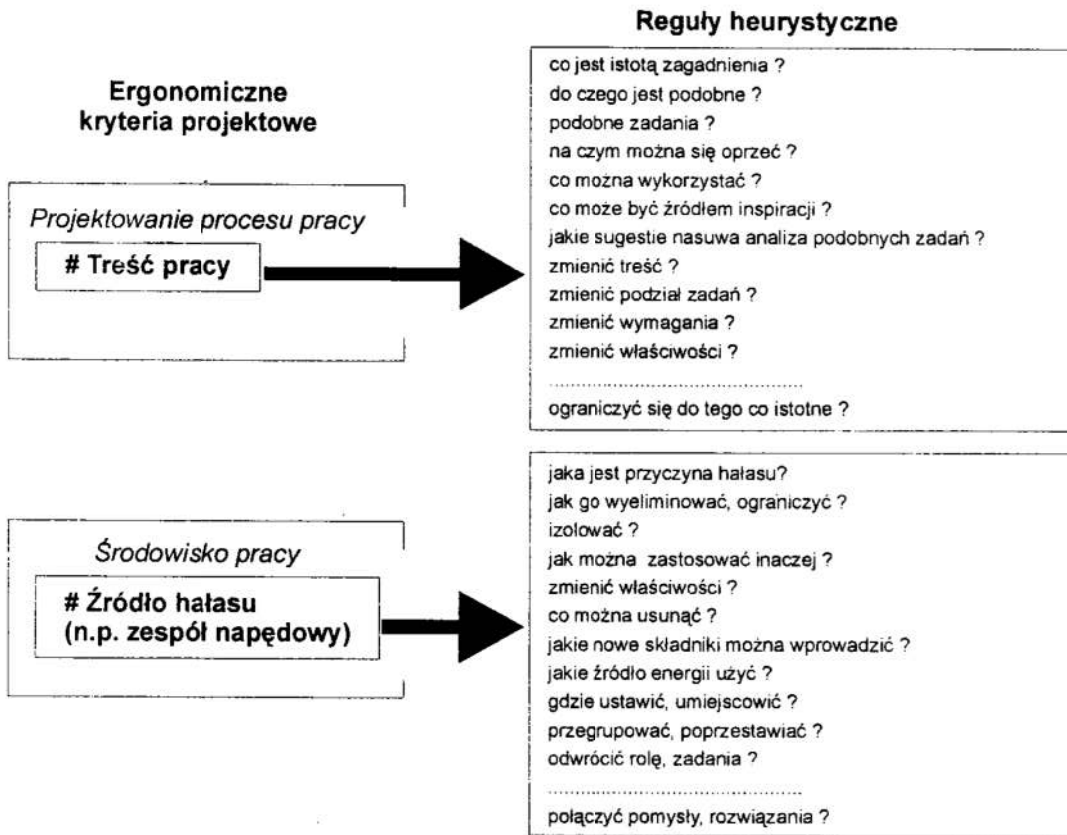


Rys. 2. Zbiory reguł heurystycznych odpowiadające określonym ergonomicznym kryteriom projektowym
 Fig 2. Sets of heuristic rules corresponding with a certain criteria

Przykład (rys. 3) ilustruje tworzenie zbiorów reguł heurystycznych dla dwóch kryteriów ergonomicznych. Kryterium „Treść pracy” dotyczy faz: formułowania założeń do projektu (ZETE) oraz projektowania wstępnego. Należy do kryteriów o znaczeniu strategicznym. Utworzone reguły posiadają postać ogólną i wspierają osiągnięcie celów strategicznych. Kryterium „Źródło hałasu” dotyczy fazy projektowania szczegółowego i należy do kryteriów o znaczeniu operacyjnym (ze względu na stopień szczegółowości rozpatrywanych zagadnień). Projektant musi pamiętać, że projektując np. zespół napędowy, przekładnię, sposób obróbki materiału przy użyciu narzędzia z ostrzem, itd. projektuje „źródła” zakłóceń środowiskowych w sposób najczęściej niezamierzony i nieświadomie. Jako produkt uboczny i wielce niepożądany może, w efekcie działań projektowych, powstać hałas, drgania, zanieczyszczenia, itp. Bardzo istotnym jest zagadnienie poziomu wiedzy i świadomości projektanta, gdyż jego rozwiązania mogą tworzyć źródła zakłóceń środowiskowych i określone skutki dla środowiska pracy. Problem ten musi być doceniony przed przystąpieniem do realizacji konkretnych zadań szczegółowych. W procesie projektowania projektant musi wyeliminować, unikać lub ograniczać skutki uboczne w postaci niepożądanych czynników zakłócających parametry otoczenia (emisja hałasu, ciepła, szkodliwej energii, substancji lotnych, itp.). Utworzone reguły ukierunkowują i wspomagają poszukiwanie rozwiązań zapewniających wyeliminowanie lub ograniczenie poziomu hałasu do niezbędnego minimum.

Zasada zgodności z logiką przebiegu procesu projektowego – struktura procesu projektowania ergonomicznego [3] narzuca określoną logikę postępowania oraz przebieg procesu projektowania. W określonych fazach i krokach procesu projektowego muszą być uwzględnione odpowiednie ergonomiczne kryteria projektowe, po których następuje możliwość doprojektowania do człowieka określonych podzespołów technicznych. Pozwala to, w następnych etapach, na zaprojektowanie prototypu systemu człowiek–obiekt techniczny. Wykorzystanie ergonomicznych kryteriów projektowych do budowy reguł, cel zbiorów reguł heurystycznych powoduje, że zbiór reguł należy

umieścić przed określonym ergonomicznym kryterium projektowym. Projektant powinien posłużyć się regułami w celu znalezienia satysfakcjonującego rozwiązania spełniającego wymogi zawarte w kryteriach ergonomicznych.



Rys. 3. Przykład tworzenia reguł heurystycznych
Fig. 3. The sample of creating heuristic rules

3 Uwagi końcowe

Tworzenie reguł heurystycznych dla komputerowego wspomaganie ergonomicznego projektowania wynikają przede wszystkim z metody, która opiera się na strukturze projektowania ergonomicznego oraz na ergonomicznych kryteriach projektowych. Zasadniczy wpływ ma nadrzędny cel metody: wspomaganie heurystycznych procesów decyzyjnych w ergonomicznym projektowaniu maszyn. Dekomponując ten cel można wyznaczyć cele:

- bezpośredni – zadaniem jest wspomaganie twórczego wysiłku projektanta, wspomaganie procesów myślowych, ukierunkowywanie działań w procesie podejmowania decyzji w celu spełnienia wymagań zawartych w kryterium projektowym w sposób poprawny z punktu widzenia ergonomii;
- pośredni – zadaniem jest wspomaganie heurystycznych procesów decyzyjnych zachodzących w procesie projektowania ergonomicznego poprzez ułatwienie osiągnięcia wymogów zawartych w kryteriach ergonomicznych w celu zapewnienia wymaganej jakości ergonomicznej projektowanego obiektu technicznego.

Wyznaczone cele posiadają istotny wpływ na zasady tworzenia reguł, ukierunkowują oraz wyznaczają przestrzeń rozwiązań.

Utworzone reguły w oparciu o przedstawione zasady znajdą zastosowanie w komputerowym programie wspomagania heurystycznych procesów decyzyjnych w ergonomicznym projektowaniu maszyn[1].

Literatura

1. Lasota A.: *Koncepcja programu komputerowego wspomagania procesów decyzyjnych w projektowaniu maszyn ukierunkowana na wykorzystanie wiedzy o człowieku*. XVII Ogólnopolska Konferencja Polioptymalizacja i Komputerowe Wspomaganie Projektowania Mielno'99, Zeszyty Naukowe Wydziału Mechanicznego Politechniki Koszalińskiej nr 26, s. 162–169, Koszalin 1999.
2. Tarnowski W. (red): *Wspomaganie komputerowe CAD,CAM. Podstawy projektowania technicznego*. WNT Warszawa 1997.
3. Tytyk E.: *Metodologia projektowania ergonomicznego w budowie maszyn*. Wydawnictwo Politechniki Poznańskiej rozprawy nr 252, Poznań 1991.
4. Tytyk E.: *Projektowanie ergonomiczne – metoda nowoczesnego projektowania*. Zeszyty Naukowe Politechniki Poznańskiej, Organizacja i Zarządzanie nr 15, Poznań 1995.

Streszczenie

W pracy przedstawiono tworzenie reguł heurystycznych dla komputerowego wspomaganie projektowania ergonomicznego w oparciu o strukturę procesu projektowania ergonomicznego oraz ergonomiczne kryteria projektowe. Utworzenie reguł w oparciu o przedstawione zasady pozwoli na wykorzystanie ich w programie komputerowego wspomaganie heurystycznych procesów decyzyjnych w ergonomicznym projektowaniu maszyn.

THE CONCEPT OF CREATING HEURISTIC RULES FOR THE ERGONOMIC DESIGN

Summary

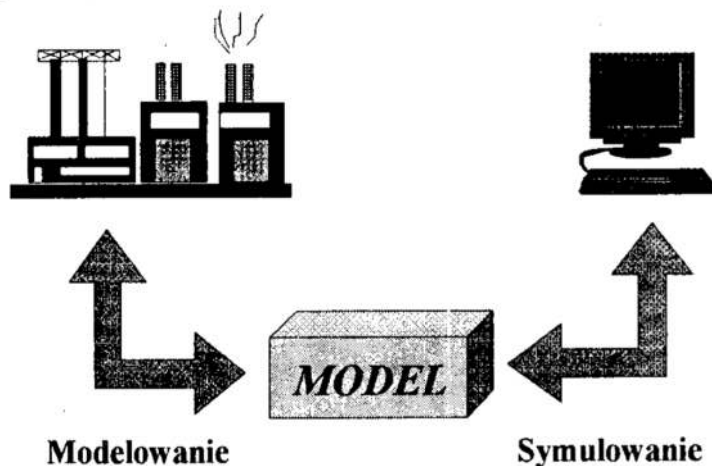
In this article there was presented creating of heuristic rules for computer supporting the ergonomic design on a base of the structure of ergonomic design process and ergonomic design criteria. Creating the rules on a base of presented rules will enable to use them in the programme of computer supporting the heuristic decision processes in the ergonomic machines design.

Referat recenzowany

WYKORZYSTANIE MODELOWANIA I SYMULACJI DO ANALIZY ODLEWNICZYCH SYSTEMÓW WYTWARZANIA

1. Wprowadzenie

Wzrost wymagań co do jakości i efektywności procesów wytwarzania rodzi konieczność poszukiwania nowych rozwiązań w zakresie organizacji i sterowania produkcją. Przedsiębiorstwa produkcyjne zmuszane są do wytwarzania wyrobów atrakcyjnych rynkowo, o wysokiej jakości, w krótkim czasie, przy jak najmniejszych kosztach własnych, eliminując jednocześnie warunki pracy szkodliwe dla zdrowia człowieka. Do rozwiązywania tych problemów stosuje się coraz częściej metody modelowania i symulacji [2].



Rys. 1. Wzajemne powiązania między rzeczywistym systemem a modelem
Fig. 1. Relations between real system and simulation model

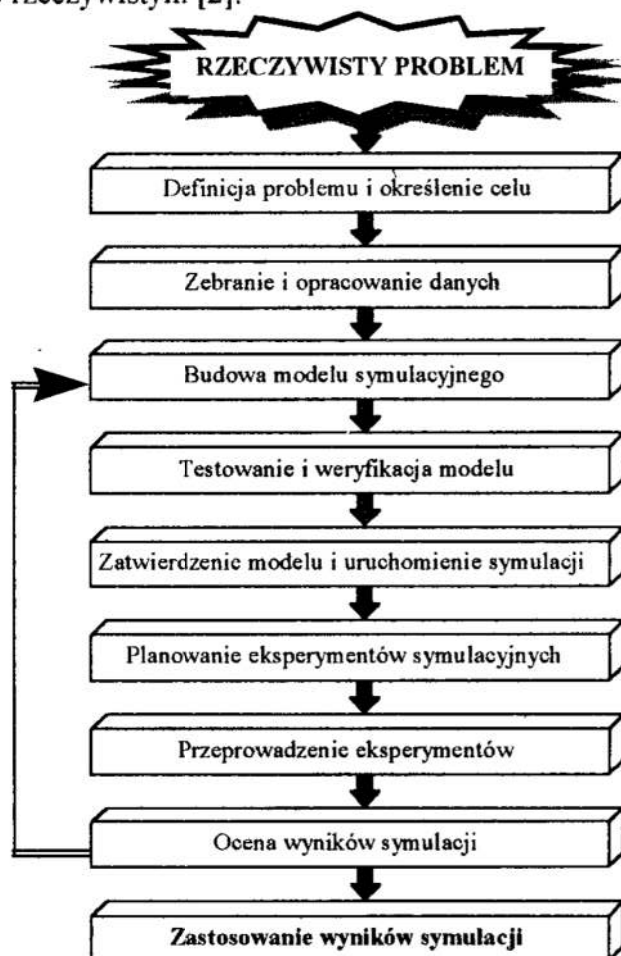
Modelowanie i symulacja jest techniką umożliwiającą szybkie przeanalizowanie różnych wariantów rozwiązań produkcyjnych na komputerowo zamodelowanym systemie. W tym celu tworzy się komputerowy model systemu rzeczywistego, na którym przeprowadza się szereg eksperymentów symulacyjnych. W wyniku przeprowadzonych symulacji uzyskuje się zestaw raportów pozwalających na dokonanie wyboru najlepszego rozwiązania przy projektowaniu nowego systemu produkcyjnego lub na opracowanie programu zmian, jakie należałoby wprowadzić w istniejącym już systemie, aby osiągnąć założony cel [3].

* Józef Matuszek - OBR MiR „REDOR” w Bielsku-Białej, ul. Paderewskiego 11a, 43-300 Bielsko-Biała, E-Mail: matuszek@kip.pb.bielsko.pl

** Sławomir Kukla - Katedra Inżynierii Produkcji, Politechnika Łódzka Filia w Bielsku-Białej, ul. Willowa 2, 43-309 Bielsko-Biała, E-Mail: kukla@kip.pb.bielsko.pl

2. Zasady wykonywania projektów symulacyjnych

Pierwszym krokiem projektu symulacyjnego jest identyfikacja problemu. Należy określić cel projektu oraz kryteria pozwalające na porównywanie różnych wariantów rozwiązań. Kolejnym krokiem projektu jest zebranie i przygotowanie niezbędnych danych. Do typowych danych należą: plan produkcji, schemat przepływu materiałów, przepływ informacji, ilość pracowników, liczba dostępnych maszyn i urządzeń, czasy produkcji (t_j , t_{pz}), itp. Po opracowaniu danych i zbudowaniu modelu poddaje się go testowaniu i weryfikacji. Następnie przeprowadza się eksperymenty symulacyjne pozwalające na ocenę modelowanego systemu. Jeśli uzyskane wyniki nie są zadowalające (zależnie od zdefiniowanego problemu), to należy zmienić konfigurację parametrów wejściowych lub też zmienić model symulacyjny. W przypadku uzyskania satysfakcjonujących rezultatów należy wyciągnąć odpowiednie wnioski i wprowadzić zmiany w systemie rzeczywistym [2].



Rys. 2. Etapy projektu symulacyjnego

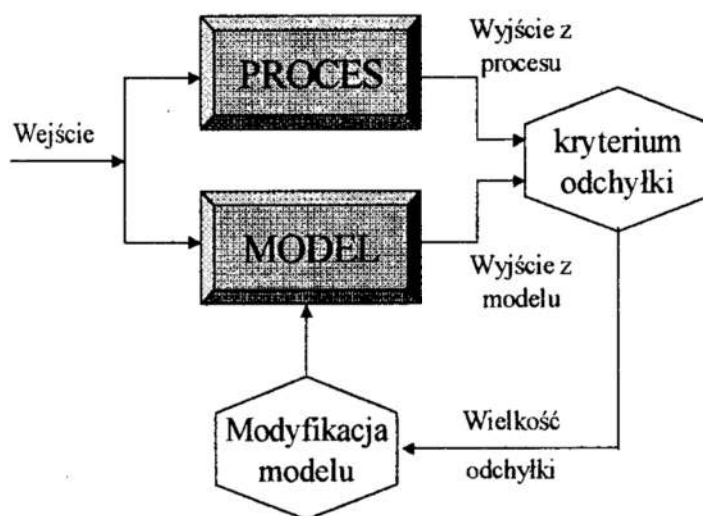
Fig. 2. Main steps in simulation project

Podstawowe założenia symulacji [4]:

- Symulacja nie jest metodą optymalizacyjną, ani inną metodą automatycznie wyszukującą gotowe rozwiązania problemu. Jest przede wszystkim metodą eksperymentalną pomagającą projektantowi znaleźć odpowiedź na pytanie: „Co się stanie w przypadku, gdy...”.

- Przy symulacji wymagana jest szczegółowa znajomość modelowanego systemu oraz posiadanie umiejętności matematycznych, analitycznych, projektowych i programowania komputerowego.
- Model symulacyjny powinien być tak szczegółowy, jak to tylko jest konieczne i tak uproszczony, jak to tylko jest możliwe.
- Zastosowanie modelu, który w pożądanym stopniu nie odzwierciedla systemu rzeczywistego i wykorzystuje nieściśle dane nie ma sensu.
- Rezultaty symulacji muszą być umiejętnie interpretowane. Należy brać pod uwagę uproszczenia w budowie modelu oraz to, że często opiera się on o statystycznie opracowane wielkości.

Wartości parametrów modelu należy dobrać w ten sposób, aby zgodność systemu rzeczywistego z komputerowym była maksymalna. Należy przeprowadzić optymalizację parametrów modelowanego systemu (rys. 3).



Rys. 3. Schemat optymalizacji parametrów modelu

Fig. 3. The scheme of the optimization of model parameters

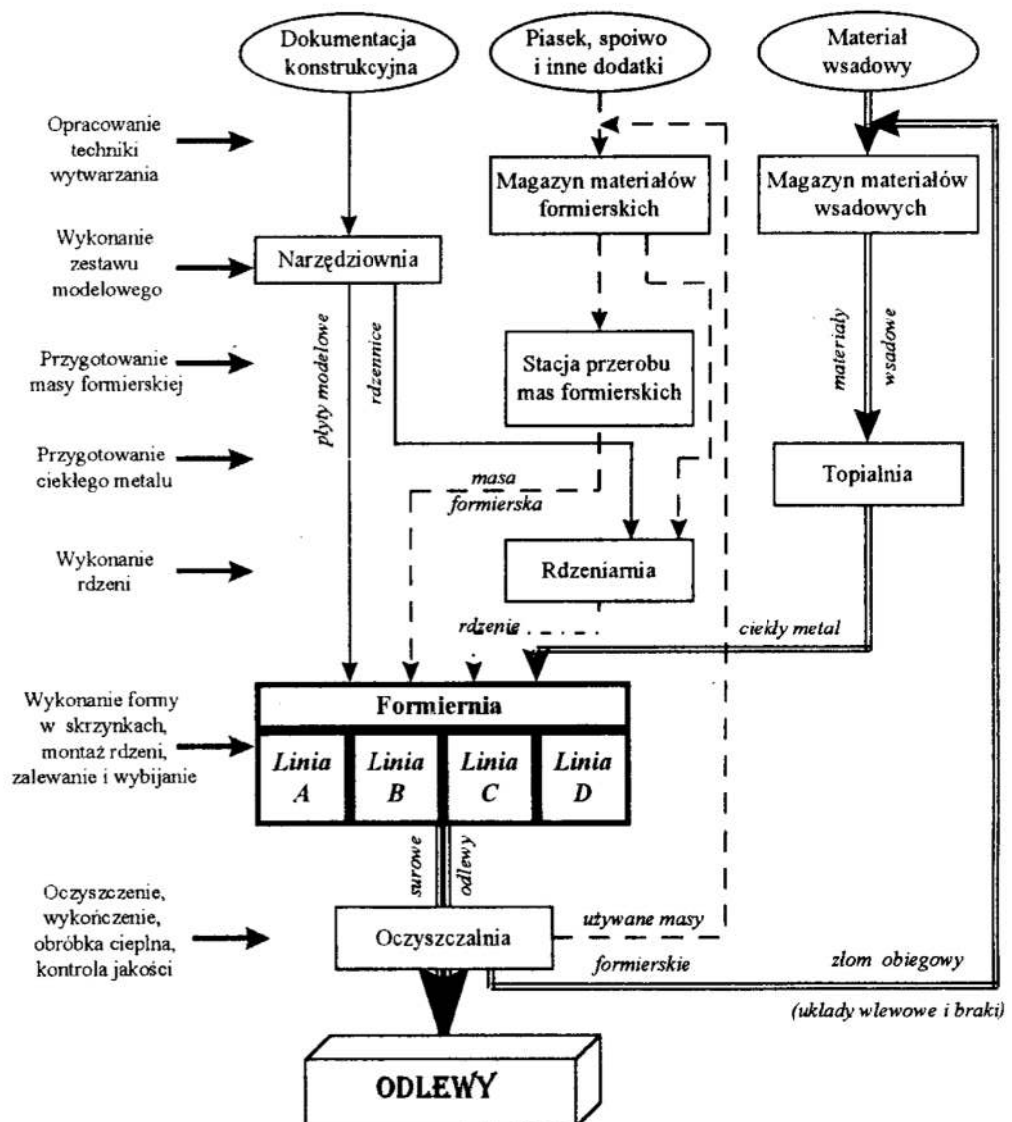
Dzięki zastosowaniu modelowania i symulacji można przeprowadzać eksperymenty na złożonym systemie przed rozpoczęciem realizacji projektu w praktyce. W ciągu kilku minut można prześledzić przebieg złożonych procesów produkcyjnych, które w rzeczywistości trwają kilka dni, tygodni czy nawet miesięcy.

Modelowanie i symulacja złożonych decyzyjnych sytuacji w procesach produkcji odlewów może przyczynić się do:

- Wzrostu produktywności.
- Obniżenia globalnych kosztów zarządzania spowodowanych awariami, uruchomieniem pieców oraz uruchomieniem i przebrojeniem linii odlewniczych, utrzymaniem ruchu, zużyciem energii i materiałów, brakami oraz koniecznością ich obróbki.
- Obniżenia kosztu jednostkowego wyrobu dzięki lepszemu wykorzystaniu stanowisk produkcyjnych.
- Wzrostu realnej zdolności produkcyjnej urządzeń.
- Lepszej organizacji produkcji.

3. Modelowanie i symulacja procesów wytwarzania odlewów

Przykładem wykorzystania metody modelowania i symulacji może być model symulacyjny wytwarzania odlewów w Odlewni Żeliwa Teksid Poland S.A. w Skoczowie, która jest nowoczesnym zakładem zajmującym się wytwarzaniem odlewów z żeliwa szarego i sferoidalnego przeznaczonych przede wszystkim dla potrzeb przemysłu samochodowego. Z uwagi na pracochłonność robót oraz szkodliwe warunki pracy w odlewni, zautomatyzowanie produkcji ma bardzo duże znaczenie techniczne, ekonomiczne oraz społeczne. Automatyzacja procesu formowania i zalewania została zrealizowana przez wprowadzenie do odlewni automatycznych linii odlewniczych. Linie te składają się z zespołu maszyn i urządzeń, połączonych urządzeniami transportowymi, wykonujących operacje związane z wytworzeniem odlewu takie jak: formowanie, zakładanie rdzeni, składanie form, zalewanie ciekłym metalem, chłodzenie, wybijanie odlewów z form oraz ich czyszczenie.



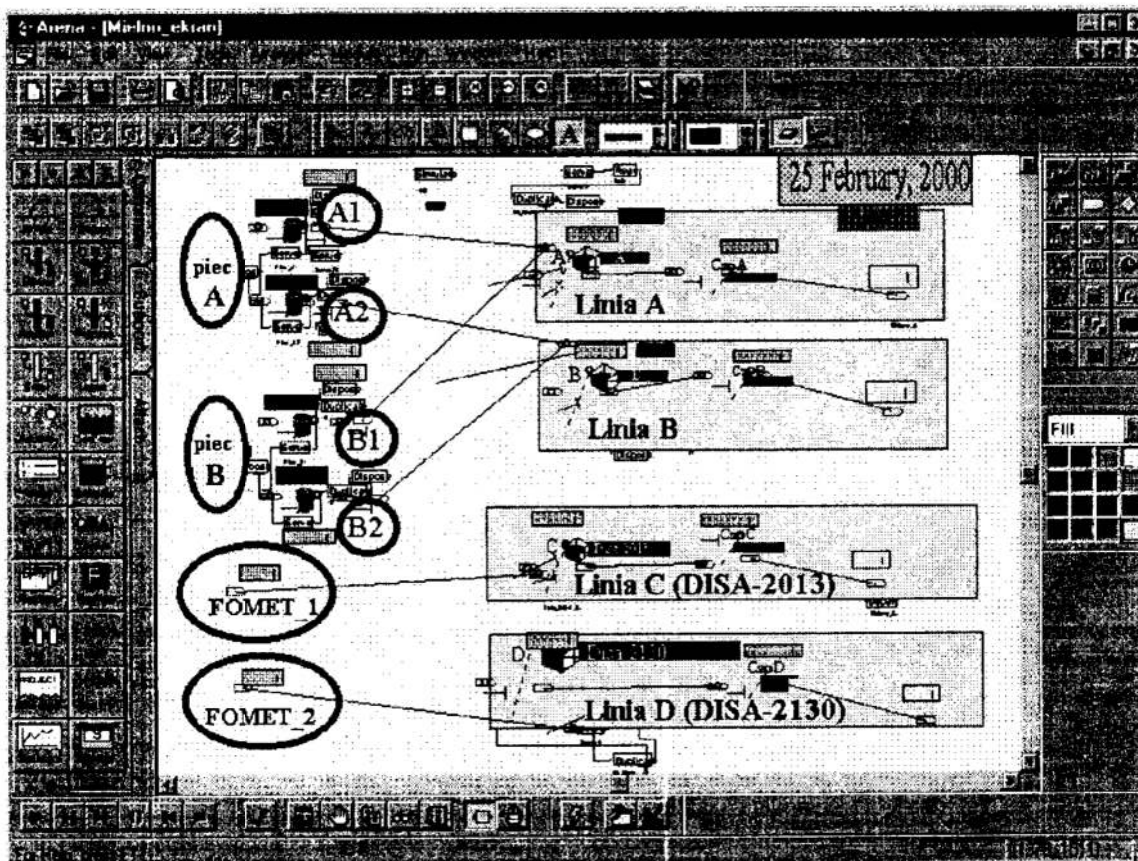
Rys. 4. Schemat systemu produkcyjnego

Fig. 4. Diagram of the manufacturing system

Problem przedstawiono na przykładzie odlewniczego systemu produkcyjnego wyposażonego w piece topiące i przetrzymujące ciekły metal oraz w cztery automatyczne linie odlewnicze pracujące w sposób równoległy, niezależny:

- *A* - linia formowania skrzynkowego, poziomego,
- *B* - linia formowania skrzynkowego, poziomego,
- *C* - linia formowania bezskrzynkowego, pionowego - DISAMATIC 2013,
- *D* - linia formowania bezskrzynkowego, pionowego - DISAMATIC 2130.

Do budowy modelu wykorzystano pakiet komputerowy do modelowania i symulacji procesów produkcyjnych ARENA firmy System Modeling Corporation (USA). ARENA jest systemem pracującym w środowisku WINDOWS. Tworzenie modelu symulacyjnego w tym systemie polega na wyborze i umiejscowieniu elementów składowych w zdefiniowanej przestrzeni symulowanego systemu i określeniu połączeń logicznych między obiektami oraz podstawowych parametrów niezbędnych do przeprowadzenia projektu symulacyjnego. Dzięki kompatybilności systemu ARENA z Windows 95 i Windows NT istnieje możliwość wykorzystania techniki ODBC. Poprzez ODBC można importować dane z tabel arkuszy kalkulacyjnych i baz danych (np. Microsoft EXCEL, Microsoft ACCES czy dBase). Możliwy jest także eksport danych z opracowanego modelu do tabel tych programów [3].



Rys. 5. Model systemu – pakiet do modelowania i symulacji ARENA
 Fig. 5. System model – software to modelling and simulation ARENA

Przy opracowaniu modelu wykorzystano następujące dane:

- Informacje na temat przebiegu procesu produkcyjnego w odlewni realizowanego na automatycznych liniach odlewania,
- Informacje dotyczące dostępnych zasobów produkcyjnych - pieców topialnych i przetrzymawczych, linii odlewniczych oraz środków transportowych,
- Informacje związane z planowaniem produkcji.

Zaprojektowany w ten sposób model symulacyjny, po sprawdzeniu i weryfikacji, nadaje się do przeprowadzenia szeregu eksperymentów symulacyjnych pozwalających na ocenę danego systemu. Dla przeprowadzenia symulacji sporządzono bazę danych testowych w arkuszu kalkulacyjnym EXCEL zawierającą wartości zmiennych zadeklarowanych w pakiecie ARENA.

Przyjęto średnie prędkości cyklu pracy automatycznych linii wynoszące:

- 160 form/godz. – dla linii A
- 130 form/godz. – dla linii B
- 250 form/godz. – dla linii C
- 220 form/godz. – dla linii D

Masa metalu wlewanego do formy mieści się w granicach 22 do 86 kg/formę. Symulację przeprowadzono dla jednego dnia roboczego (24 godziny), uwzględniając czas potrzebny na wykonanie zabiegów niezbędnych dla prawidłowego przebiegu procesu produkcyjnego związanych z obsługą linii odlewniczych (przezbieranie linii, czyszczenie płyt modelowych i leja) oraz czas związany z uruchomieniem linii.

Przeprowadzono następujące eksperymenty:

- Eksperyment 1 - przeprowadzony dla założonego planu produkcji, przy zasilaniu linii A z pieca FOMET_1, linii B z pieca FOMET_2, linii C z pieców A1 i B1 oraz linii D z pieców A2 i B2. Wyniki zamieszczono w tabeli 1.
- Eksperyment 2 - przeprowadzony dla identycznego planu produkcji, jak w eksperymencie 1, lecz przy zasilaniu linii A naprzemiennie z pieca indukcyjnego A1 i B1, linii B z pieca A2 oraz B2, linii C z pieca FOMET_1 i linii D z pieca FOMET_2. Wyniki zamieszczono w tabeli 2.
- Eksperyment 3 - przeprowadzony dla innego wariantu planu produkcji przy zasilaniu linii ciekłym metalem jak w eksperymencie 2. Wyniki zamieszczono w tabeli 3.
- Eksperyment 4 - przeprowadzony dla planu, jak w eksperymencie 2 i 3 po wprowadzeniu do systemu dodatkowej linii E. Wyniki zamieszczono w tabeli 4.

W wyniku przeprowadzonej symulacji uzyskano następujące wyniki:

Tabela 1. Wyniki eksperymentu 1

Table 1. Results of experiment 1

	Linia A	Linia B	Linia C	Linia D
Liczba form zalanych	2780	2895	5142	4199
Liczba form wykonanych na gotowo	2580	2748	5059	4063
Dostępność linii [%]	96,2	96,9	87,6	88,3
Wykorzystanie linii [%]	72,4	94,5	85,7	80,7

Tabela 2. Wyniki eksperymentu 2

Table 2. Results of experiment 2

	Linia A	Linia B	Linia C	Linia D
Liczba form zalanych	3586	2917	5226	3334
Liczba form wykonanych na gotowo	3385	2770	5148	3242
Dostępność linii [%]	96,2	96,9	87,6	88,3
Wykorzystanie linii [%]	93,4	95,2	87,1	63,8

Tabela 3. Wyniki eksperymentu 3

Table 3. Results of experiment 3

	Linia A	Linia B	Linia C	Linia D
Liczba form zalanych	3572	2917	5226	3456
Liczba form wykonanych na gotowo	3371	2768	5145	3368
Dostępność linii [%]	96,2	96,9	87,5	88,3
Wykorzystanie linii [%]	93,0	95,2	87,1	66,2

Z analizy wyników przeprowadzonych eksperymentów, biorąc pod uwagę produktywność całego systemu, można wyciągnąć wniosek, że sposób zasilania linii ciekłym metalem zastosowany w eksperymencie 1 jest mniej korzystny, pomimo większej wydajności linii D. Dodatkowym problemem tego rozwiązania jest krzyżowanie się dróg transportowych. Analizując wyniki wcześniej przytoczonych eksperymentów oraz wielu innych nie zamieszczonych w pracy można zauważyć, że możliwości produkcyjne pieców topiących i przetrzymujących metal są większe niż zapotrzebowanie linii. W związku z tym należałoby zastanowić się nad inwestycją w zakładzie w postaci kolejnej automatycznej linii odlewniczej. Propozycję tego rozwiązania wykonano wprowadzając do modelu symulacyjnego linię DISAMATIC „E”, która jest identyczna jak linia C. Linia E zasilana jest z pieca FOMET_1. Wyniki symulacji zamieszczono w tabeli 4. Potwierdzają one celowość tej inwestycji.

Tabela 4. Wyniki eksperymentu 4

Table 4. Results of experiment 4

	Linia A	Linia B	Linia C	Linia D	Linia E
Ilość form zalanych	3586	2917	4879	3334	3260
Ilość form wykonanych na gotowo	3385	2770	4797	3242	3218
Dostępność linii [%]	96,2	96,9	87,6	88,3	87,7
Wykorzystanie linii [%]	93,4	95,2	87,1	63,8	54,3

Istotne znaczenie dla określenia poprawności działania całego systemu produkcyjnego ma problem niezawodności pracy linii odlewniczej w określonych warunkach eksploatacyjnych w wymaganym przedziale czasu. Uszkodzenie jednego z elementów automatycznej linii odlewania może spowodować postój całej linii wywierając przez to wpływ na ilość produkowanych wyrobów i efekt ekonomiczny przedsiębiorstwa. Jednym z możliwych rozwiązań problemu wyznaczenia optymalnych okresów obsługi zespołów linii z uwzględnieniem awarii, remontów planowych, czy też napraw uprzedzających jest zastosowanie techniki modelowania i symulacji [1].

4. Podsumowanie

Modelowanie i symulacja jest techniką stosowaną z powodzeniem na całym świecie. Pozwala na zmniejszenie kosztów związanych z nieefektywnym wykorzystaniem zasobów produkcyjnych, daje możliwość porównywania różnych rozwiązań produkcyjnych i alternatywnych procesów wytwarzania, umożliwia planowanie realizacji zleceń produkcyjnych. Projekt symulacyjny nie stwarza możliwości znalezienia optymalnego rozwiązania, ale jest narzędziem wspomagającym pracę projektanta umożliwiającym szybkie przetestowanie na komputerowo zamodelowanym systemie różnych rozwiązań i na wybranie wariantu najlepiej spełniającego założone kryteria. Stosując odpowiedni program badań symulacyjnych przy różnych parametrach pracy można określić optymalne wartości realizowanego procesu symulując różne sytuacje decyzyjne, np. awarie linii czy awarie pieców topiących żeliwo. Można symulować skutki zakłóceń, określać podejmowane decyzje związane z zakłóconym procesem.

Modelowanie i symulacja jest metodą dającą się skutecznie zastosować w warunkach dyskretniej produkcji rytmicznej, a taką produkcją charakteryzuje się Odlewna Żeliwa Teksid Poland S.A. w Skoczowie.

Literatura

1. Cybulski J.: *Poliptymalizacja harmonogramów wymiany elementów automatycznej linii odlewniczej*. Praca doktorska, Łódź 1982.
2. Gregor M., Halušková M., Hromada J., Košturiak J., Matuszek J.: *Simulation of Manufacturing System*. Wydawnictwo Politechniki Łódzkiej Filii w Bielsku-Białej, Bielsko-Biała 1998.
3. Kelton W., Sadowski R. Sadowski D.: *Simulation with Arena*. WCB/McGraw-Hill, Sewickley 1998.
4. Skalík P.: *Počítačová simulace procesu v oblasti oprav a údržby strojních zařízení*. Disertační práce, Ostrava 1998.

Streszczenie

W artykule przedstawiono pojęcie modelowania i symulacji procesów produkcyjnych. Omówiono możliwość zastosowania tej techniki do projektowania odlewniczych procesów wytwarzania. Podano przykład analizy systemu składającego się z automatycznych linii odlewniczych oraz pieców topiących i przetrzymujących żeliwo. Zagadnienie przedstawiono wykorzystując do tego celu pakiet do modelowania i symulacji – ARENA.

USING OF MODELLING AND SIMULATION IN ANALYSIS OF FOUNDRY MANUFACTURING SYSTEMS

Summary

The paper introduces a concept of the modelling and simulation production process. This article presents possibility application the technology to design and analysis of the foundry manufacturing process. The article gives example of the model production process realized on automatic casting lines and melting furnaces of cast iron. This problem was presented trough application of the computer program – ARENA.

Referat recenzowany

KOSZTY PRODUKCJI W MODELOWANIU I SYMULACJI PROCESÓW PRODUKCYJNYCH

1 Wprowadzenie

Sprostanie wymaganiom rynku możliwe jest jedynie poprzez wprowadzenie efektywnej organizacji, umożliwiającej skrócenie przebiegów produkcyjnych, obniżkę kosztów wytwarzania oraz wysoką jakość produktów. Konieczne jest także podejmowanie coraz bardziej skomplikowanych i złożonych procesów produkcyjnych, co znacznie komplikuje i zwiększa liczbę zadań przedsiębiorstwa. Synchronizacja narastającej ilości czy nników ma szczególne znaczenie w odniesieniu do procesów planowania i zarządzania produkcją.

Na rozwój planowania i sterowania produkcją znaczący wpływ wywierają nowe tendencje z zakresu zarządzania wytwarzaniem, które mają również wpływ na systemy komputerowe. Do najważniejszych z nich zaliczyć można: Just in Time (JIT), Kanban, Optimized Production Technology (OPT), Material Requirements Planning (MRP I), Manufacturing Resource Planning (MRP II), Enterprise Requirements Planning (ERP lub MRP III).

Ponadto coraz istotniejszą metodą wspomagającą zarządzanie produkcją staje się modelowanie i symulacja. Symulacja pozwala na poprawę funkcjonowania systemów produkcyjnych dzięki możliwości rozwiązania złożonych problemów w bardzo krótkim czasie. Przeprowadzenie badań na opracowanym modelu jest szczególnie wskazane, gdy rozwiązanie problemu metodami analitycznymi nie jest możliwe. Kolejną zaletą jest możliwość weryfikacji dokonanych założeń jeszcze przed ich zastosowaniem w praktyce. Podstawą symulacji jest stworzenie modelu, na którym przeprowadza się symulacje. Na ich podstawie dokonywany jest wybór rozwiązania.

Dodatkowo uwzględnienie kosztów w modelowaniu i symulacji zwiększa efektywność przeprowadzanych badań.

2 Koszty produkcji

Od dawna poszukuje się sposobu zwiększenia przychodów, przy równoczesnym utrzymaniu na możliwie minimalnym poziomie nakładów, niezbędnych do prowadzenia działalności gospodarczej. Inżynierowie mają możliwość wpływania na wielkość kosztów przez kształtowanie wyrobu, procesu wykonywania, jak również organizacji produkcji w przedsiębiorstwie.

Podstawą analizy procesu produkcyjnego jest, przy zachowaniu warunków jakościowych, określenie czasu i kosztów własnych produkcji wyrobu i jego elementów

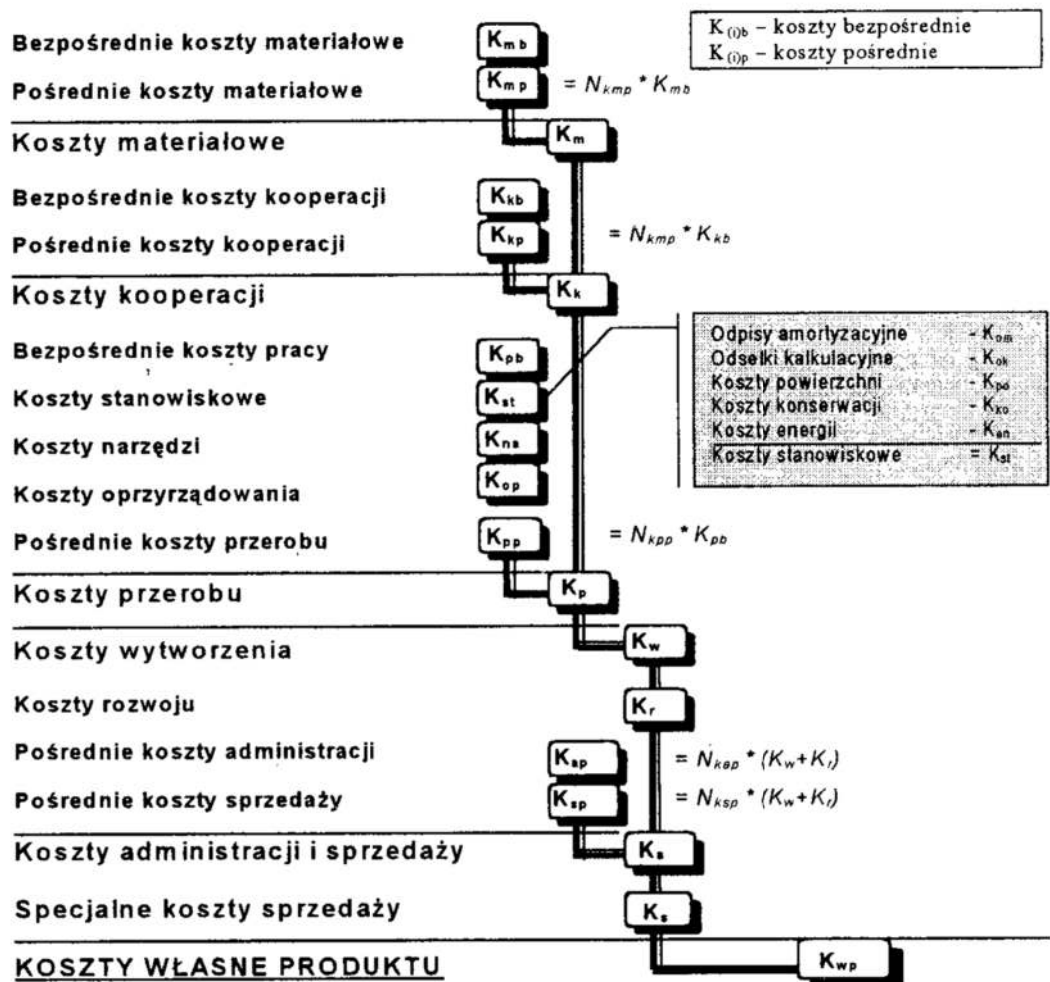
* *Józef Matuszek - OBR MiR REDOR, ul. Paderewskiego 11a, 43-300 Bielsko-Biała., E-mail: matuszek@kip.pb.bielsko.pl*

** *Dariusz Plinta - Katedra Inżynierii Produkcji w Filii Politechniki Łódzkiej w Bielsku-Białej; ul. Willowa 2, 43-309 Bielsko-Biała, E-mail: plinta@kip.pb.bielsko.pl*

składowych. Określenie czasów realizacji procesów produkcyjnych można otrzymać rozwiązując problem harmonogramowania postawionych zadań np. w oparciu o oprogramowanie do modelowania i symulacji.

Kolejne metody wyznaczania kosztów rozwijały się w miarę wzrostu złożoności wytwarzania, narastających więzi kooperacyjnych, różnorodności wyrobów finalnych, rozwoju systemów informatycznych i coraz większych wymagań co do dokładności obliczeń.

Do najważniejszych algorytmów określania kosztów produkcji można zaliczyć: kalkulację podziałową prostą (suma kosztów rodzajowych przez ilość wyprodukowanych wyrobów w rozpatrywanym okresie czasu), oraz kalkulację doliczeniową (koszty rodzajowe są stopniowo rozliczane na wyroby za pomocą narzutów lub bezpośrednio za pomocą wskaźników wyodrębnionych działań) [1, 4].

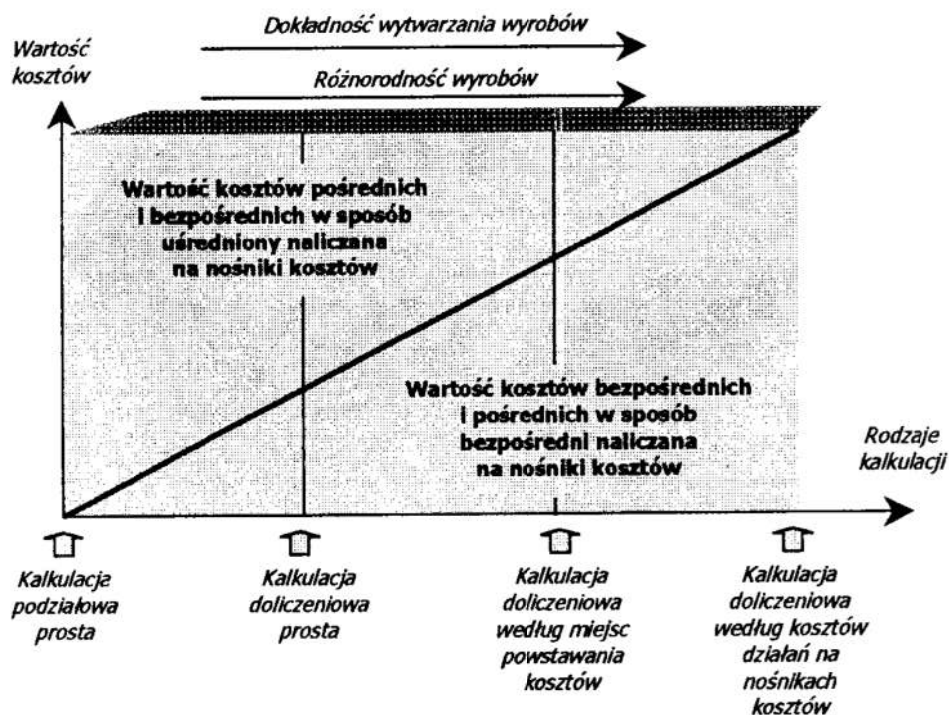


Rys. 1. Przebieg kalkulacji doliczeniowej [1]

Fig. 1. Run of the cost accounting [1]

W kalkulacji doliczeniowej pośrednie koszty są rozliczane na wyroby według miejsc powstawania kosztów - za pomocą wskaźników narzutów, lub bezpośrednio poprzez wskaźniki wyodrębnionych działań ($N_{kmp}, N_{kpp}, \dots = 0$).

Wraz z zastosowaniem kolejnych metod doliczeniowych (kalkulacja prosta, kalkulacja według miejsc powstawania kosztów i kalkulacja według kosztów działań na nośnikach kosztów) maleje udział kosztów pośrednich w koszcie własnym wyrobu określanych za pomocą narzutów – rysunek 2.



Rys. 2. Wartość poszczególnych składników kosztowych w wybranych algorytmach obliczeniowych kosztów

Fig. 2. The value of particular costs in following costs accounting algorithms

Niektóre operacje mogą być wykonane na różnych obrabiarkach. Inne są wykonywane w kooperacji. Stosowanie w algorytmach obliczeń wskaźników narzutów może zniekształcić wyniki obliczeń. Wyroby wykonane na obrabiarkach prostych, o małych kosztach eksploatacji często wykonane są za drogo, a wyroby obrabiane na drogich stanowiskach pracy mają ustalone koszty produkcji zbyt nisko. Występuje tu niebezpieczeństwo ustalenia niewłaściwych kosztów wyrobu, co często powoduje straty. Stąd dążenie do minimalizacji wskaźników narzutów.

Sprowadzenie wartości narzutów do zera, przejście na kalkulację doliczeniową według metody kosztów działań – rysunek 2 - też nie zawsze musi okazać się opłacalne. Koszt zebrania wszystkich niezbędnych informacji może przewyższyć efekty wynikające z dokładności wyznaczonych kosztów.

W warunkach jednostkowej i małoseryjnej produkcji, w warunkach szerokiego asortymentu produkcji, ciągle zmieniającego się programu produkcyjnego, duży wpływ na optymalny układ harmonogramu wykonywanych działań ma dokładność określania kosztów, które w takich warunkach są trudne do wyznaczenia. Problem należy rozwiązać przez wyodrębnienie grupy kosztów działań związanych np. z pracami technicznego przygotowania dokumentacji konstrukcyjnej, zamawianymi pomocami

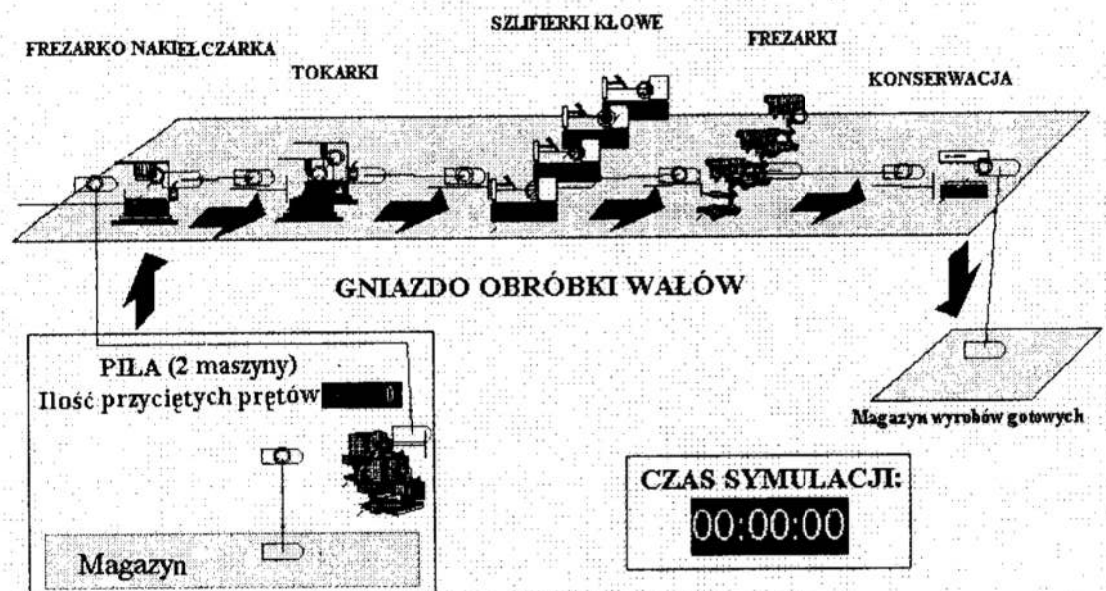
warsztatowymi, z reklamą. Pozostałe koszty pośrednie należy uzyskać przy pomocy zredukowanych do minimum narzutów resztkowych [1, 2, 3, 4].

3 Przykład gniazda obróbki wałów

Rozwiązanie problemu przedstawiono na przykładzie gniazda obróbki wałów.

Tablica 1. Przykład gniazda obróbki wałów - proces wytwarzania, maszyny
Table 1. Example of the shaft working seat - manufacture process, machines

Op.	Nazwa operacji	Stanowisko (liczba maszyn)
1	Cięcie na długość	Piła taśmowa (2)
2	Planowanie na długość i nakiełkowanie	Frezarko nakiełczarka (1)
3	Toczenie	Tokarka DFS 450 NC (2)
4	Szlifowanie	Szlifierka kłowa (4)
5	Frezowanie rowka	Frezarka sekwencyjna (3)
6	Konserwacja smarem	Stanowisko ręczne



Rys. 3. Model symulacyjny gniazda obróbki wałów
Fig. 3. Simulation model of the shaft working seat

Po zamodelowaniu gniazda produkcyjnego wałów i symulacji miesięcznej produkcji (21 dni roboczych, praca na jedną zmianę) uzyskano informacje o obciążeniu stanowisk, i czasie realizacji zleceń.

Wnioski z pierwszej symulacji:

- Najbardziej obciążonymi stanowiskami są dwie tokarki i frezarko-nakiełczarka; wąskim gardłem są tokarki, obciążone prawie w 100%. W przypadku planowania nowych inwestycji związanych z zakupem nowych, czy też modernizacją

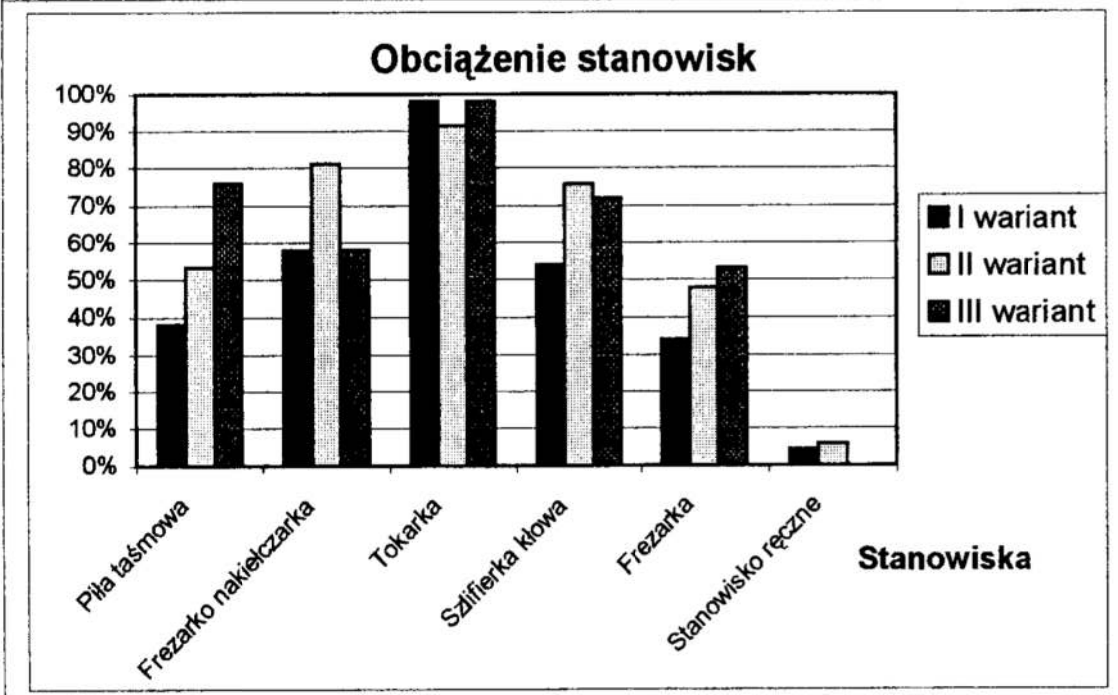
istniejących maszyn w pierwszej kolejności należałoby wziąć pod uwagę właśnie te maszyny.

- Obciążenie stanowiska do konserwacji obrobnionych wałów wynosi 4%; operacja ta może być wykonywana bezpośrednio na stanowiskach frezarskich po operacji frezowania rowków.
- Wśród stanowisk maszynowych najmniej obciążone są dwie piły i trzy frezarki do rowków; do zrealizowania założonej wielkości produkcji wystarczy jedna piła i dwie frezarki; można również zmniejszyć ilość szlifierek do trzech.

Tablica 2. Wyniki symulacji

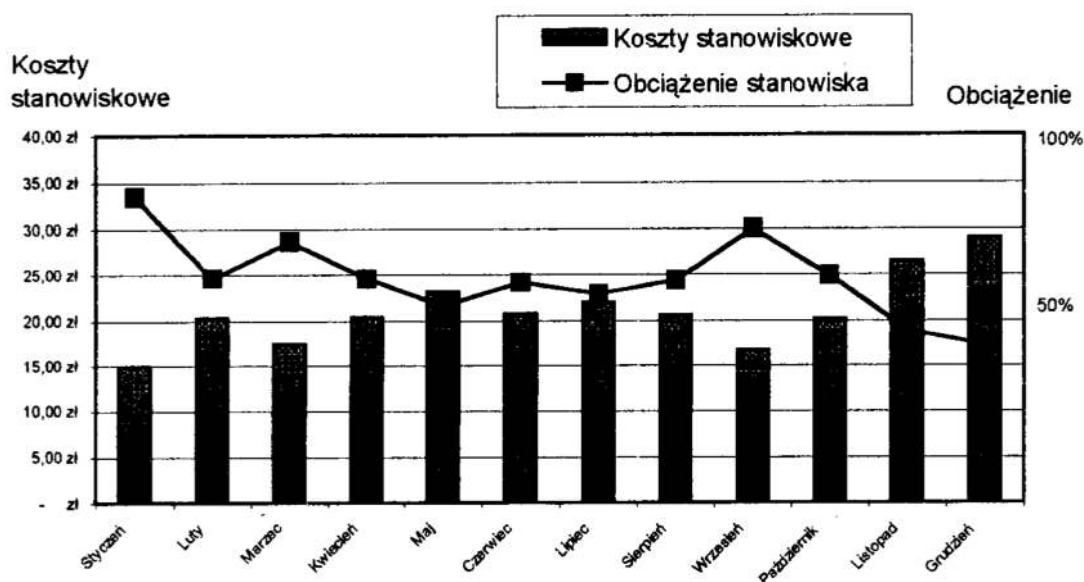
Table 2. Simulation results

Stanowisko	I wariant		II wariant		III wariant	
	Il. st.	Obciążenie	Il. st.	Obciążenie	Il. st.	Obciążenie
Piła taśmowa	2	38%	2	53%	1	76%
Frezarko nakielczarka	1	58%	1	81%	1	58%
Tokarka	2	98%	3	91%	2	98%
Szlifierka kłowa	4	54%	4	76%	3	72%
Frezarka	3	34%	3	48%	2	53%
Stanowisko ręczne	1	4%	1	6%		
I wariant	obecny stan, czas symulacji (realizacji zaplanowanej produkcji) 168 godz.					
II wariant	dodanie jednej tokarki i skrócenie czasu symulacji do 150 godz.					
III wariant	zmniejszenie ilość pił, szlifierek i frezarek, operacja konserwacji realizowana na stanowisku frezarskim (168 godz.)					



Aby uwzględnić koszty produkcji należy dodatkowo przeprowadzić analizę kosztochłonności stanowisk. Konieczne jest wyznaczenie kosztów godziny eksploatacji dla każdego stanowiska gniazda (iloraz kosztów stanowiskowych przez rzeczywisty czas obciążenia maszyny, wyznaczony np. z symulacji).

Koszt godziny eksploatacji stanowiska pracy nie jest kosztem stałym (rys.4.). Zmienia się dynamicznie zależnie od rzeczywistego obciążenia stanowiska. Im bardziej maszyna jest obciążona tym mniejszy koszt godziny, a co za tym idzie mniejszy koszt wytworzenia elementu.



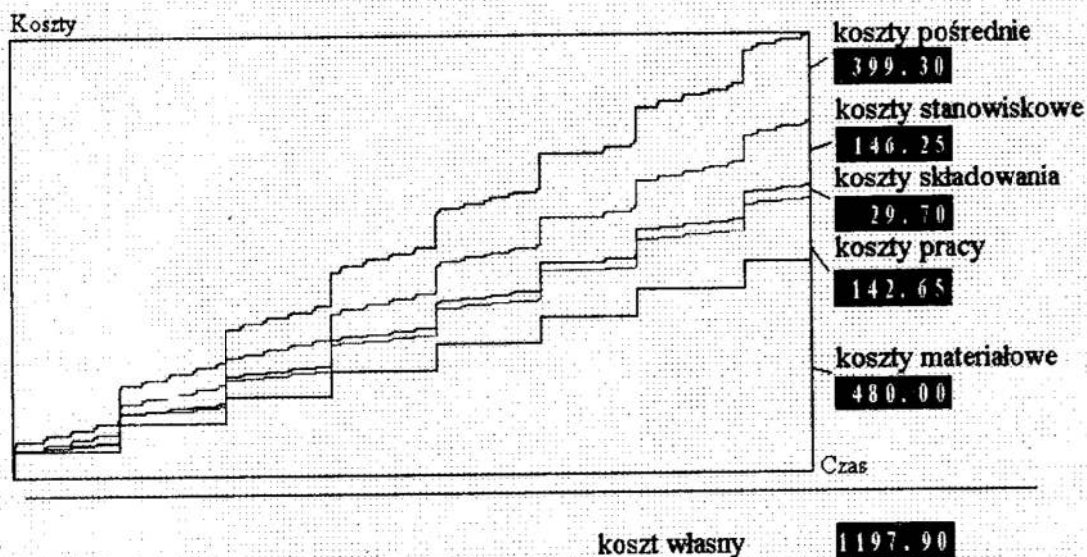
Rys. 4. Przykładowe koszty stanowiskowe pily taśmowej
 Fig. 4. Example of workplaces costs for band saw

Koszt stanowiskogodziny przemnożony przez czas obróbki elementu na stanowisku, wraz z kosztami materiałowymi i kosztami pracy stanowią koszty wytworzenia.

Koszt materiału definiuje się na wejściu zamodelowanego systemu. W czasie symulacji każda nowa część wprowadzona do systemu produkcyjnego ma automatycznie przypisany koszt materiału. Koszt pracy i koszt maszynogodziny stanowią parametr stanowiska produkcyjnego. Koszty te należy określić dla każdego stanowiska. Podczas symulacji program automatycznie może naliczać koszty dla każdej przetwarzanej części. Rysunek 5 przedstawia narastające koszty powstałe w analizowanym gnieździe. Suma tych kosztów podzielona przez liczbę wykonanych wałów daje koszty poniesione na wytworzenie jednego wyrobu.

Ze względu na dużą liczbę różnych typów wałów warto je podzielić na grupy wg kryterium podobieństwa konstrukcyjnego. Dla każdej grupy należy dodatkowo określić przedstawiciela, dla którego zostanie przeprowadzona symulacja. Wyodrębnienie takich grup upraszcza model symulacyjny, ponieważ nie musimy wyznaczać kosztów dla każdego wyrobu, a tylko dla przedstawicieli grup. Odpowiednie zdefiniowanie grup sprawia, że różnice kosztów wytwarzania wałów należących do jednej grupy są na tyle

małe w stosunku do kosztów rzeczywistych, że nie ma niebezpieczeństwa ustalenia niewłaściwych kosztów.



Rys. 5. Symulacja kształtowania się kosztów produkcji
Fig. 5. Simulation of the production costs forming

Tablica 2. Wyniki z symulacji produkcji miesięcznej
Table 2. Simulation results of the monthly production

RAPORT Z SYMULACJI						
Obciążenie stanowisk (czas symulacji: 168 godz.)						
Stanowisko (ilość maszyn)	Czas pracy [godz]		Obciążenie [%]			
Piła taśmowa (2)	63.84		38			
Frezarka nakielczarka (1)	97.44		58			
Tokarka (2)	164.64		98			
Szlifierka kłowa (4)	90.64		54			
Frezarka (3)	57.12		34			
Stanowisko ręczne	6.77		4			
Zestawienie kosztów / szt.						
yrób (wał)	Koszty Materiałowe	Koszty pracy	Koszty składowania	Koszty stanowiskowe	Koszty Pośrednie	Suma kosztów wytworzenia
A	16,84 zł	2,25 zł	2,16 zł	12,95 zł	17,10 zł	51,30 zł
B	20,15 zł	3,31 zł	2,48 zł	14,85 zł	20,39 zł	61,18 zł
C	35,88 zł	3,96 zł	2,91 zł	17,45 zł	30,10 zł	90,30 zł
D	49,13 zł	4,56 zł	4,78 zł	28,70 zł	43,59 zł	130,76 zł
E	60,44 zł	4,88 zł	5,21 zł	31,25 zł	50,89 zł	152,67 zł
F	67,62 zł	6,65 zł	6,03 zł	36,20 zł	58,25 zł	174,76 zł
G	74,38 zł	6,07 zł	5,74 zł	34,45 zł	60,32 zł	180,96 zł
H	95,83 zł	6,48 zł	5,95 zł	35,70 zł	71,98 zł	215,94 zł
I	107,58 zł	15,30 zł	12,41 zł	74,44 zł	104,86 zł	314,59 zł
J	159,65 zł	10,78 zł	9,23 zł	55,40 zł	117,53 zł	352,60 zł

4 Podsumowanie

Wzrost złożoności algorytmów wyznaczania kosztów pociąga za sobą zwiększenie nakładu pracy na zebranie wszystkich niezbędnych informacji dla modelu kosztów. Dokładne określenie kosztów jest wręcz niemożliwe bez zastosowania odpowiedniego systemu informatycznego. Obecny poziom techniki komputerowej sprawia, że wymienione powyżej metody określania kosztów mogą znaleźć zastosowanie niemalże w każdym przedsiębiorstwie.

Ponadto, zastosowanie metody modelowania i symulacji z uwzględnieniem kosztów, zwiększa efektywność prowadzonych badań - projektowania lub analizy funkcjonowania istniejących systemów. Daje nam możliwość zaprojektowania systemu pracy zbliżonego do optymalnego pod względem kosztów i czasu realizacji zadań.

Przedstawiona metoda postępowania poszerza zakres zastosowania programów do modelowania i symulacji. Programy te, jak np. ARENA, często nie posiadają opcji dokładnych obliczeń kosztowych. Przytoczone procedury obliczeń eliminują ten brak.

Literatura

1. Matuszek J.: *Inżynieria kosztów produkcji*; Materiały konferencyjne „Inżynieria Produkcji '99”: Informatyka, organizacja i zarządzanie; Zeszyt nr 6; Politechnika Łódzka Filia w Bielsku-Białej, Bielsko-Biała 1999
2. Sołtys D.: *Rachunek kosztów - elementy rachunkowości zarządczej*; Wydawnictwo Akademii Ekonomicznej, Wrocław 1999
3. Wärnecke H.J., Bullinger H.-J., Hichert R., Voegele A.: *Rachunek kosztów dla inżynierów*; WNT Warszawa 1993
4. Więcek Dorota: *Rachunek kosztów działań*; Materiały konferencyjne „Inżynieria Produkcji '99”: Informatyka, organizacja i zarządzanie; Zeszyt nr 6; Politechnika Łódzka Filia w Bielsku-Białej, Bielsko-Biała 1999

Streszczenie

W artykule przedstawiono algorytmy określania kosztów stosowane w współczesnych przedsiębiorstwach produkcyjnych. Omówiono metodologię określania kosztów przy modelowaniu i symulacji systemów produkcyjnych. Artykuł zilustrowano przykładem z praktyki produkcyjnej, związany z optymalizacją przebiegu procesu wytwarzania w gnieździe obróbki wałów do silników elektrycznych

PRODUCTION COSTS IN MODELLING AND SIMULATION OF PRODUCTION PROCESSES

Summary

Algorithms of costs determining used in present production enterprises have been presented in the paper. There have been presented the methodology of costs calculation in the modelling and simulation of production systems. This paper presents the example from production practice connected with optimisation of the manufacturing process run in the working seat of shafts for electric motors.

Referat recenzowany

STEROWANIE STANEM W SYSTEMIE UTRZYMANIA MASZYN KRYTYCZNYCH PRZEMYSŁU SPOŻYWCZEGO

1. Wprowadzenie

Ze względu na stale rosnące ceny energii, materiałów, środków smarowniczych, problem racjonalnego utrzymania maszyn w ruchu ma zasadnicze znaczenie dla rentowności zakładu produkcji spożywczej.

Racjonalny system utrzymania maszyn w ruchu to oszczędności 0,5–5% produkcji czystej przedsiębiorstwa, które są wynikiem zmniejszenia kosztów napraw o 35%, wzrost wydajności pracy nawet o 65% w tym zwiększenie dyspozycyjności maszyn o 15–20%, zmniejszenie kosztów obsługi maszyn o 20%, zmniejszenie zużycia części wymiennych do 20%, zwiększenie o 1,3–2,5 razy czasu poprawnej pracy do uszkodzenia [3].

Z analizy procesu użytkowania maszyn i urządzeń w przemyśle spożywczym wynika mała sprawność, dyspozycyjność i wysokie koszty ich eksploatacji, których przyczynami są [3]:

- stosowanie obsługi i napraw w zależności od czasu użytkowania lub w zależności od uszkodzenia,
- stosowanie przestarzałych metod oceny stanu technicznego maszyn opartych na weryfikacji części i ich demontażu,
- stosowanie mało skutecznych metod regeneracji części i niskiej jakości napraw,
- braku stosowania metod kontroli stanu technicznego z wykorzystaniem diagnostyki technicznej, opartych na obiektywnych pomiarach sygnałów diagnostycznych.

Zadaniem służby technicznej w zakładach przemysłu spożywczej jest takie obsługiwanie maszyn i urządzeń, aby użytkować je jak najdłużej w toku eksploatacji, z optymalną wydajnością, minimalnymi nakładami na naprawę i dużą niezawodnością eksploatacyjną. Cele te są z reguły możliwe do osiągnięcia tylko poprzez systematyczną obsługę techniczną maszyn polegającą na:

- ♦ utrzymaniu kapitału inwestycyjnego,
- ♦ uniknięciu nieprzewidzianych przerw,
- ♦ zwiększeniu bezpieczeństwa ich użytkowania.

Szczególnie ważna jest realizacja tych celów w przypadku maszyn krytycznych tzn. takich, od których poprawnego funkcjonowania zależne jest utrzymanie ciągłości produkcji. Przyjęcie kryteriów ważności maszyn i opracowanie odpowiednich metod diagnostycznych jest podstawą wdrożenia systemu utrzymania maszyn w zakładzie, co jest związane z ważną rolą ekonomiczną. Odbija się to wyraźnie na nakładach, które ponosi się na okresowe naprawy, modernizacje i wymianę linii technologicznych w poszczególnych działach produkcji. Dzięki zastosowaniu nowoczesnych metod obsługi technicznej maszyn i urządzeń uwarunkowanej ich stanem staje się możliwe znaczne zaoszczędzenie kosztów na utrzymanie przy jednoczesnym zwiększeniu niezawodności eksploatacyjnej.

* Uniwersytet Warmińsko-Mazurski, Wydział Nauk Technicznych, Instytut Inżynierii Mechanicznej,
ul. M. Oczapowskiego 11, 10-736 Olsztyn

2. System obsługi w ujęciu sterowania stanem maszyn

System obsługi maszyn przemysłu spożywczego można opisać następująco [4]:

$$S_O = \langle I_{oo}, D_{oo}, P_{on}, P_{op}, P_{og}, P_{or}, P_{ol}, D_g \rangle$$

gdzie:

I_{oo} , D_{oo} , P_{on} , P_{op} , P_{og} , P_{or} , P_{ol} , D_g – odpowiednio podsystemy: informacyjny, decyzyjny, napraw bieżących (obsługiwanie bieżącego), przechowywania, konserwacji, napraw głównych, napraw zespołów, likwidacji, diagnostyczny. Na rysunku 1 przedstawiono schemat systemu obsługi maszyn.

Utrzymanie obiektów w stanie zdadności może być realizowane w praktyce tylko w przypadku właściwie zorganizowanego podsystemu informacyjnego I_{oo} . Podsystem informacyjny jest powiązany z innymi podsystemami od których uzyskuje informacje o stanie obiektów technicznych, poziomie kosztów obsługi, czasie pracy efektywnej obiektów itp. Spełnia on rolę łącznika zapewniającego funkcjonowanie systemu obsługi.

W podsystemie decyzyjnym D_{oo} są podejmowane wszystkie decyzje związane z racjonalnym obsługiwaniem obiektów technicznych, wśród których można wyróżnić dwa podstawowe poziomy problemów decyzyjnych: wyższy i niższy.

Poziom wyższy obejmuje problemy decyzyjne związane z:

- wyborem strategii utrzymania obiektów technicznych w stanie zdadności, np. przy wykorzystaniu metod diagnostyki technicznej;
- dopasowaniem podsystemu obsługi do potrzeb podsystemu użytkowania.

Na niższym poziomie znajdują się problemy decyzyjne dotyczące między innymi:

- planowania i obsługi;
- organizacji i funkcjonowania podsystemu obsługi;
- określenia optymalnych terminów diagnozowania obiektów;
- wyboru metod i urządzeń diagnostycznych;
- programowania asortymentu części wymiennych;
- inne.

W systemie napraw bieżących P_{on} (obsługiwanie bieżącego) są wykonywane ustalone rodzaje obsługi i naprawy bieżące. Dowolny rodzaj obsługi jest realizowany w układzie:

- ◆ zbiór sprawdzeń wynikający z ustalonych algorytmów diagnozowania;
- ◆ operacje, które należy wykonać w wyniku postawienia diagnozy (np. regulacje, naprawy bieżące);
- ◆ operacje, które należy wykonać w miarę potrzeb (np. sprawdzenie i ustalenie poziomów płynów eksploatacyjnych).

W podsystemie przechowywania P_{ol} są przechowywane obiekty użytkowania bieżącego lub oczekujące na naprawę.

Z podsystemu przechowywania wydzielono podsystem konserwacji P_{ok} obiektów, ponieważ nie są one użytkowane i podlegają szczególnym operacjom diagnostyczno – obsługowym. Prawidłowe zakonserwowanie obiektów jest decydującym czynnikiem warunkującym uzyskanie i utrzymanie wysokiej wartości współczynnika gotowości technicznej.

W podsystemie napraw głównych P_{og} wykonuje się naprawy główne. Należy zaznaczyć, że w małych przedsiębiorstwach podsystem ten z reguły nie występuje i obiekty techniczne są przekazywane do specjalistycznych zakładów naprawczych.

W podsystemie napraw zespołów P_{oz} wykonywane są naprawy specjalistyczne zespołów, podzespołów, a także może być wykonana regeneracja części, podsystem ten również nie występuje w małych przedsiębiorstwach.

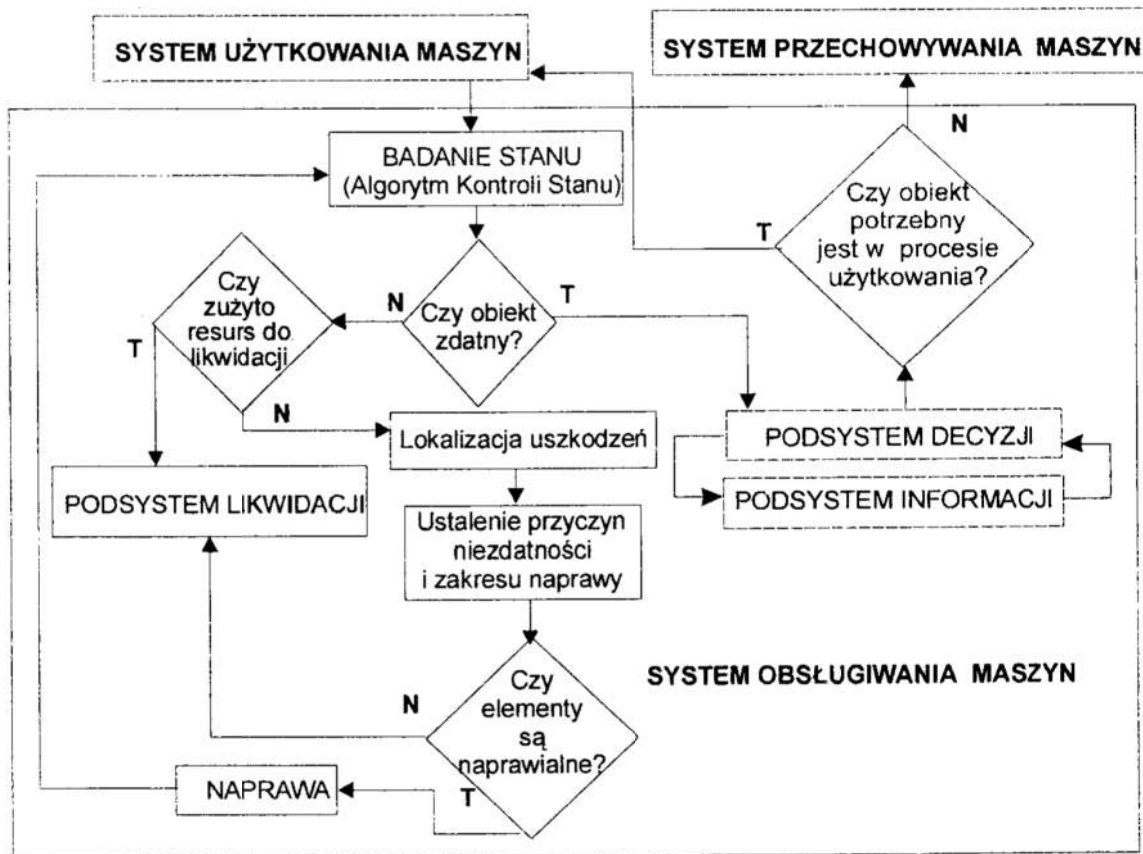
Zadaniem likwidacji P_{ol} jest wykorzystanie i wycofanie zużytych obiektów technicznych, zespołów, podzespołów, mechanizmów, a nawet części z eksploatacji.

Podsystem diagnostyczny jest podstawą budowy i funkcjonowania dynamicznego systemu eksploatacji obiektów technicznych.

Podsystem diagnostyczny to zespół diagnostów, zbiór metod i środków uzyskiwania, przetwarzania, przedstawiania i gromadzenia informacji oraz zbiór obiektów technicznych, ich modeli diagnostycznych i algorytmów diagnozowania, prognozowania i generowania stanów, a także relacji między tymi elementami, przeznaczony do podejmowania wiarygodnych decyzji o przynależności badanego obiektu do określonej klasy stanów. Podsystem ten charakteryzują trzy zasadnicze czynniki: cel, obserwowalność i proces diagnozy.

Celem tak zbudowanego podsystemu jest uzyskanie informacji o właściwościach przedmiotu diagnozy, a więc o jego stanie, zmianie tego stanu i przyczynach zaistniałego stanu.

Obserwowalność oznacza możliwość ustalenia stanu przedmiotu diagnozy na podstawie obserwacji jego wielkości wyjściowych (np. pomiaru wartości parametrów sygnału drganiowego).



Rys. 1. Schemat sterowania utrzymaniem maszyn w stanie zdatości
 Fig. 1. Scheme of maintenance control of machines in operational state

Proces diagnozy to ciąg operacji zmierzających do uzyskania informacji o stanie badanego obiektu. Ta informacja diagnostyczna powinna dotyczyć nie tylko obiektu diagnozy, ale

również szeroko rozumianego otoczenia. Oznacza to, że postać informacji diagnostycznej warunkują właściwości przedmiotu diagnozy, jego otoczenia i właściwości użytkownika (decydenta).

Przyjmując, że sterowanie jest to celowe oddziaływanie na obiekt sterowania, mające doprowadzić do uzyskania założonego celu [1], można stwierdzić, iż w przypadku utrzymania maszyn sterowanie odbywa się na podstawie kompleksowych działań diagnostycznych, których celem jest utrzymanie ich w stanie zdatności przy możliwie najmniejszych kosztach eksploatacji. Na rysunku 1 przedstawiono uproszczony algorytm sterowania zmianą stanu, rozumianego jako działania obsługowe w systemie utrzymania ruchu. Jednym z zasadniczych działań w tym algorytmie jest diagnozowanie stanu, od którego wyniku zależą podejmowane decyzje eksploatacyjne. Poza tym opracowanie odpowiedniej metody diagnozowania jest warunkiem wdrożenia metody obsługiwanie według stanu, która powinna szczególnie znaleźć zastosowanie w przypadku maszyn krytycznych.

3. Metoda oceny stanu dynamicznego maszyny na przykładzie wirówki do mleka

Źródła procesów drganiowych

Diagnostyka drganiowa oparta jest na obserwacji jednego z procesów resztkowych generowanych poprzez ruch elementów roboczych maszyny. Obserwacja procesów drganiowych daje bezpośrednią informację o zaawansowaniu procesów ich zużycia i degradacji, czyli o bieżącym stanie technicznym. Z wielowymiarowości informacyjnej procesów dynamicznych maszyn wynika możliwość oceny stopnia zużycia łożysk, wirnika, przekładni zębatych, osiowości itp. [2].

Do podstawowych źródeł drgań maszyn wirnikowych można zaliczyć:

- niewyrównoważenie: statyczne – powodujące równe reakcje (drgania) na obu łożyskach, momentowe – powodujące drgania łożysk równe co do wartości ale przeciwne w fazie itp., obserwowane w postaci silnych impulsów z częstotliwością f_n (f_n – częstotliwość odpowiadająca prędkości obrotowej),
- nieosiowość wałów i sprzęgieł: powoduje drgania nieharmoniczne występujące przede wszystkim w kierunku osiowym, których amplituda jest proporcjonalna do rozosiowania promieniowego i kontowego, w widmie drgań widoczna dla częstotliwości $2f_n$,
- zużycie łożysk tocznych: generuje drgania maszyn o szerokim obrazie widma drgań w paśmie nisko-, średnio- i wysoko częstotliwościowym, zależnym od rodzaju i wielkości zużycia,
- zużycie przekładni zębatych: generuje drgania o intensywności zależnej od rodzaju zazębienia, obciążenia itp.. W widmie drgań dominują harmoniczne częstotliwości zazębienia zależne od liczby zębów (n) i częstotliwości obrotowej (f_n) poszczególnych wałów.

Wobec powyższego należy stwierdzić, że analiza obrazu widma drgań elementów maszyn wirnikowych w funkcji częstotliwości pozwala na ocenę stanu poszczególnych elementów maszyn.

Metodyka pomiarów i charakterystyka obiektu badań

Na podstawie przeprowadzonych badań teoretyczno – eksperymentalnych systemu utrzymania wirówek do mleka typu LWG 47, przyjęto następującą procedurę sterowania ich stanem z wykorzystaniem diagnostyki drganiowej:

- diagnostyka ogólna stanu maszyny (na podstawie wartości RMS drgań dla odbioru szerokopasmowego),
- diagnostyka silnika (na podstawie analizy widmowej),

- diagnostyka układu: hamulec – sprzęgło (na podstawie analizy widmowej),
- próba wodna i ocena stanu dynamicznego maszyny.

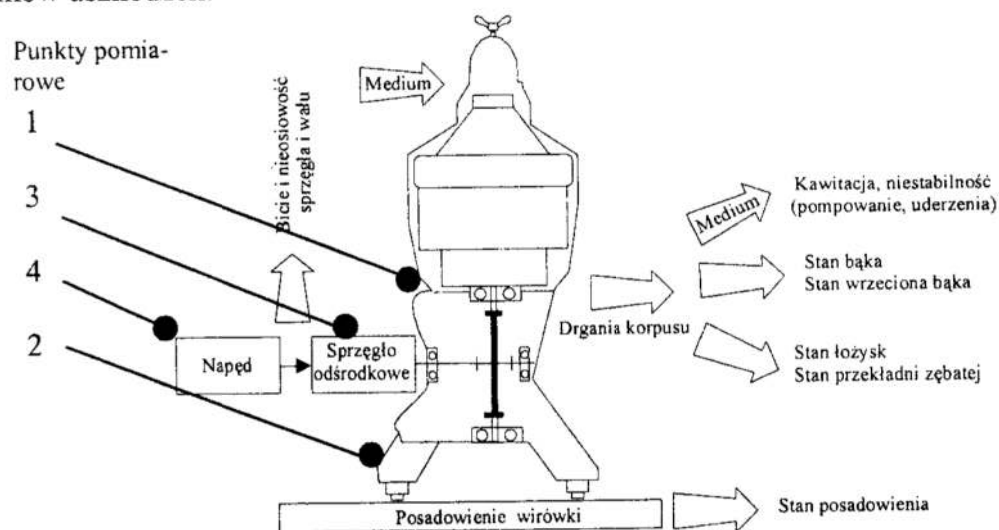
W przypadku stwierdzenia konieczności wyważania poszczególnych zespołów wirówki, zespoły te kierowano na specjalnie zaprojektowane i wykonane stanowiska do wyważania (bąka, silnika i układu napędowego).

Do pomiaru drgań i ich analizy wykorzystano analizator drgań śledzący ATR-2, w którego skład wchodzi: moduł pomiarowy, dwa czujniki indukcyjne do pomiaru drgań i czujnik foto do pomiaru fazy drgań.

Kryteria oceny stanu dynamicznego maszyny wirnikowej przyjęto na podstawie normy PN-90/N-01358 [5] i analizy kinematycznej maszyny oraz mocy i sposobu posadowienia maszyny.

Wirówka napędzana jest silnikiem elektrycznym o mocy 11 kW i prędkości obrotowej 1500 obr/min. (25 Hz). Napęd z silnika przekazywany jest poprzez sprzęgło odśrodkowe i przekładnię zębatą hipoidalną (o przełożeniu 1:4) na bąk obracający się z prędkością 6000 obr/min (100 Hz). Wirówka usadowiona i wypoziomowana jest na gumowych poduszkach osadzonych w posadzce. Moc silnika oraz sposób usadowienia wirówki pozwala zaliczyć ją do II grupy maszyn wg normy PN-90/N-01358.

Wirówkę można podzielić umownie na dwa zespoły: napęd (układ poziomy) i zespół roboczy – (bąk, wrzeciono). Podział taki pozwala na szybszą i z głębszą diagnozę oddziaływań siłowych. Napęd oprócz przekazywania ruchu obrotowego może jednocześnie być przyczyną występowania drgań związanych z nieosiowością tego układu. Bąk maszyny może być niewyważony statycznie i dynamicznie, co prowadzi do uszkodzenia ułożyskowania i wrzeciona bąka. Na rysunku 2 przedstawiono poglądowo mechanizm oddziaływania dynamicznego w wirówce do mleka i możliwe miejsca odbioru symptomów uszkodzeń.



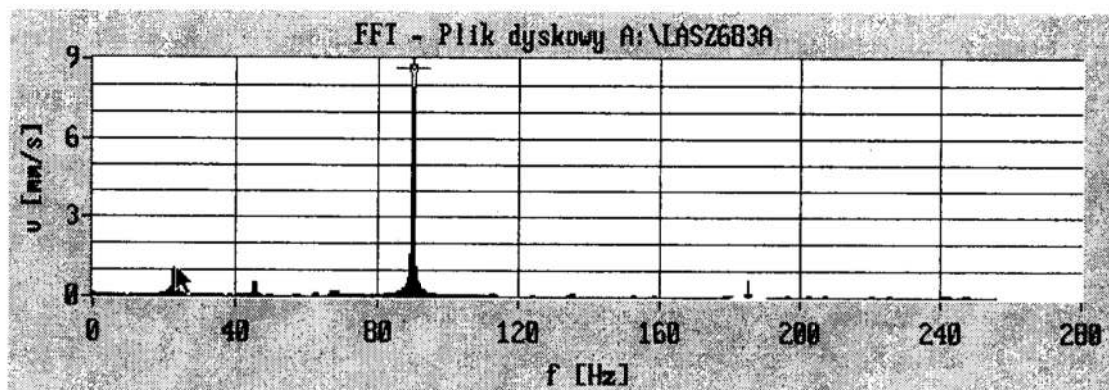
Rys. 2. Schemat oddziaływań siłowych w wirówce do mleka
Fig. 2. Scheme of forces action in cream separator

Na podstawie analizy schematu oddziaływań siłowych dokonano wyboru punktów pomiarowych. Przykładowo do ceny niewyważenia bąka, drgania mierzono w punktach 1, 2, a w celu określenia stanu układu przeniesienia napędu w punktach 3 i 4.

Przykładowe wyniki oceny stanu dynamicznego wirówki

Przedstawione przykłady wyników pomiarów pochodzą z przeprowadzonych badań wirówki LWG 47 o numerze 298 4683, produkcji Spomasz Gniezno (na licencji Alfa-Laval), rok budowy 1981.

Na rysunku 3 podano widmo zarejestrowanych drgań na obudowie kadłuba wirówki (punkt pomiarowy 1 – patrz rysunek 2).

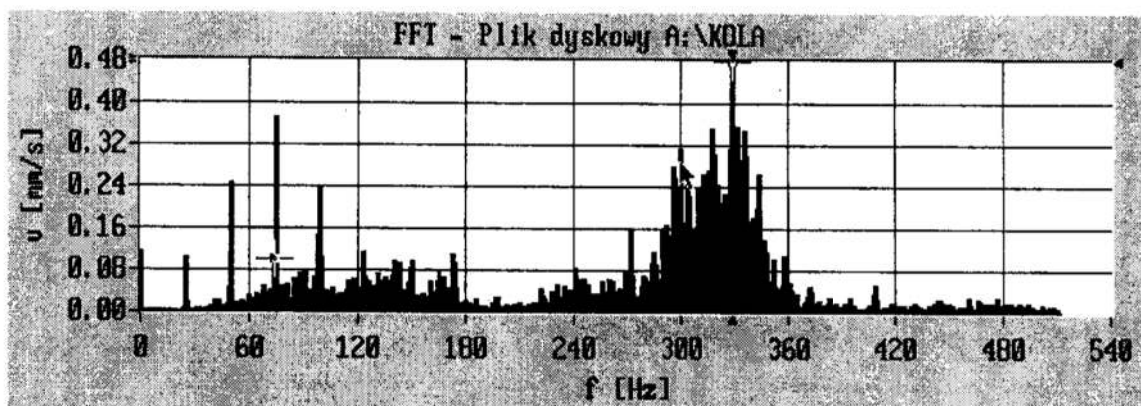


Rys. 3. Widmo drgań wirówki typu LWG 47 przed wyważeniem bąka
Fig. 3. Vibration spectrum of scream separator LWG 47 before balancing of the top

Z analizy widma drgań (rys. 3) wynika, iż maksymalne drgania występujące na częstotliwości 98 Hz wynoszące 8,9 [mm/s], są spowodowane niewyważeniem układu pionowego bąka wirówki. Pik odpowiadający 25 Hz o wartości 1,2 [mm/s], jest spowodowany niewyważeniem zespołu koła–sprzęgła–hamulca–silnika, jednakże wartość tych drgań mieści się w granicach dopuszczalnych dla tego typu układu.

Z porównania wartości pomiaru V_{RMS} z kryteriami oceny stanu maszyn wg [4] stwierdzić można iż wirówka znajduje się w stanie niedopuszczalnym.

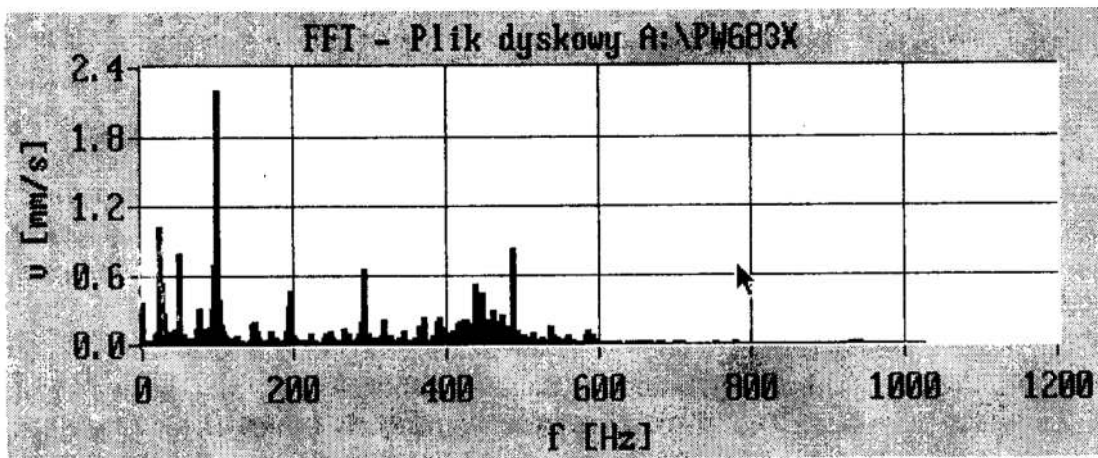
Następnie po zdemontowaniu bąka dokonano pomiaru drgań w punkcie pomiarowym 3 i przeprowadzono analizę widma drgań w paśmie częstotliwości do 1000 Hz (rys. 4).



Rys. 4. Widmo drgań przekładni zębatej wirówki
Fig. 4. Vibration spectrum of the gear box of cream separator

Występujące wstęgi boczne wokół 320 Hz, widoczne na rysunku 4 świadczą o nadmiernym zużyciu przekładni zębatej (koła napędowego).

Po wyważeniu bąka, zgodnie z procedurą wyważania dwupłaszczyznowego na odpowiednio zaprojektowanym stanowisku do wyważania i wymianie kół zębatach przeprowadzono próbę wodną i wykonano pomiary kontrolne. Na rysunku 5 przedstawiono widmo drgań zarejestrowane podczas próby wodnej.



Rys. 5. Widmo drgań po wyważeniu i naprawie wirówki
 Fig. 5. Vibration spectrum after balancing and repair of cream separator

Z rysunku 5 i innych wyników pomiarów kontrolnych wynika, że po wyważeniu układu pionowego wirówki i wymianie kół zębatach stan wirówki można uznać za dostateczny.

4. Podsumowanie

Na podstawie rozważań teoretyczno – eksperymentalnych przedstawiono problem utrzymania maszyn krytycznych w przemyśle spożywczym z punktu widzenia strategii ich obsługiwanie według stanu technicznego. Opracowano metodę i algorytm sterowania zmianą stanu technicznego maszyn w systemie utrzymania ruchu. W metodzie tej zastosowano diagnostykę drganiową do kontroli stanu technicznego maszyny krytycznej na przykładzie wirówki typu LWG47.

Badania weryfikacyjne metody i algorytmu sterowania stanem wirówki w systemie napraw przeprowadzono w P.P.H.U. „PROHATUR” w Nidzicy. Wyniki tych badań zostały pozytywnie ocenione i wdrożono do systemu napraw wirówek algorytm sterowania ich stanem technicznym.

Zatem poprawnie opracowane procedury diagnostyczne nadają całemu systemowi utrzymania maszyn charakter dynamiczny i integracyjny poszczególnych podsystemów, w aspekcie sterowania stanem. Stwarza to podstawę do skutecznego i ekonomicznego funkcjonowania systemów utrzymania, co w szczególności jest istotne dla maszyn krytycznych.

Literatura

1. Baborski A. i inni: *Elementy cybernetyki ekonomicznej*. PWE, Warszawa, 1983.
2. Cempel Cz.: *Podstawy wibroakustycznej diagnostyki maszyn*. WNT, Warszawa, 1982.
3. Michalski R. Mikołajczak P.: *Zastosowanie systemu ekspertowego utrzymania w ruchu maszyn przemysłu spożywczego*. *Zagadnienia Eksploatacji Maszyn*. Z.1 (109) vol. 32, 1997.
4. Michalski R. Niziński S.: *Podstawy eksploatacji obiektów technicznych*. Wydawnictwo ART w Olsztynie nr 35, 1997.
5. PN-90 N-01358 – *Metody pomiarów i oceny drgań maszyn*.

Streszczenie

W pracy przedstawiono problemy utrzymania maszyn krytycznych w przemyśle spożywczym. Opisano system obsługiwanania oraz metodę oceny stanu dynamicznego maszyn. Zaproponowano procedurę postępowania diagnostycznego w czasie eksploatacji i naprawy na przykładzie wirówek do mleka.

CONTROLLING STATE IN THE MAINTENANCE SYSTEM OF CRITICAL MACHINES IN FOOD INDUSTRY

Summary

In the paper there were presented issues of maintenance of the critical machines in food industry. The system of maintenance and the method of examining of the dynamic state of the machines were described. There was presented the procedure of the diagnostic actions during the exploitation and repair on an example of the cream separator.

Referat recenzowany

FILTRACJA ZBIORU ROZWIĄZAŃ OPTYMALNYCH W SENSIE PARETO METODĄ PRZEDZIAŁÓW NIEROZRÓŻNIALNOŚCI

1 Wprowadzenie

Wprowadzenie nowoczesnych komputerowych technik obliczeniowych w procesie projektowania złożonych obiektów technicznych z uwzględnieniem wielu przeciwstawnych kryteriów, pozwala na uzyskania zbiorów rozwiązań dopuszczalnych o dużej liczebności. Wyznaczenie podzbioru rozwiązań niezdominowanych prowadzi tylko do częściowej redukcji liczebności analizowanego zbioru. W tej sytuacji stosowanie metod [1], [2], [5], [10], [11] uważanych za klasyczne, a polegające na tworzeniu rankingu rozwiązań niezdominowanych nie daje właściwego przeglądu analizowanego zbioru. W przeświadczeniu autora należy dążyć do redukcji liczebności zbioru rozwiązań niezdominowanych [7], [8], [9], a następnie wyznaczać podzbiór rozwiązań reprezentatywnych [6], lub ewentualnie stosować metody porządkujące rozwiązania w ranking.

2 Optymalność w sensie przedziału nierozróżnialności

Optimum w sensie Pareto (OSP). Rozwiązania niezdominowane, często nazywane rozwiązaniami optymalnymi w sensie Pareto, są jednoznacznie określone matematycznie. Rozwiązanie jest optymalne w sensie Pareto, jeśli żadne z kryteriów $F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_n(\mathbf{x})$ nie może być poprawione bez jednoczesnego pogorszenia przynajmniej jednego z nich. Element $\mathbf{x}^* \in X$ jest nazywany rozwiązaniem optymalnym w sensie Pareto wtedy i tylko wtedy, gdy w zbiorze X nie istnieje taki element \mathbf{x} , że dla każdego $n \in N$:

$$F_n(\mathbf{x}) \leq F_n(\mathbf{x}^*) \quad \text{oraz istnieje } j \in N, \text{ takie że} \quad F_j(\mathbf{x}) < F_j(\mathbf{x}^*) \quad (1)$$

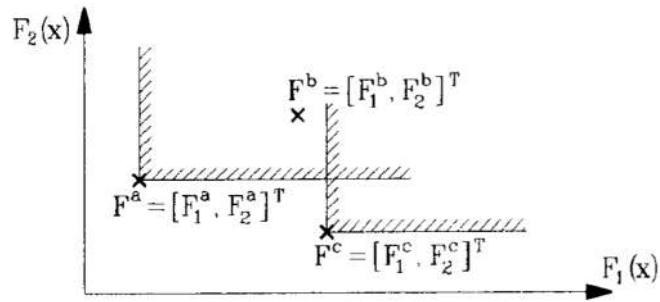
gdzie: X - niepusty zbiór rozwiązań dopuszczalnych.

Do wyznaczenia rozwiązań optymalnych w sensie Pareto stosujemy relacje porządku częściowego określone przez stożek. Stożek dodatni o wierzchołku w zerze - C_0 oraz w punkcie $F^s = [F_1^s, \dots, F_m^s]^T$ - C_s , należący do m -wymiarowej euklidesowej przestrzeni definiujemy następująco [12]:

$$\begin{aligned} C_0 &= \{F = [F_1, \dots, F_m]^T : F_i \geq 0 \ (i=1, \dots, m)\} \\ C_s &= \{F = [F_1, \dots, F_m]^T : (F_i - F_i^s) \geq 0 \ (i=1, \dots, m)\} \end{aligned} \quad (2)$$

Według relacji porządku częściowego ($^c <$) określonego przez stożek dodatni, element $F^a = [F_1^a, \dots, F_m^a]^T$ jest mniejszy od elementu $F^b = [F_1^b, \dots, F_m^b]^T$, jeżeli element F^b należy do stożka dodatniego o wierzchołku w punkcie F^a : $F^a \ ^c < F^b$, jeśli $F^b \in C_{F^a}$. Rysunek 1 przedstawia niezdominowane elementy - F^a i F^c oraz element zdominowany - F^b .

* Politechnika Lubelska, Katedra Podstaw Techniki, 20-618 Lublin, ul Nadbystrzycka 38a,
e-mail: montus@antenor.pol.lublin.pl



Rys.1. Relacja mniejszości określona przez stożek dodatni
Fig.1. The minority relation determined by the positive cone

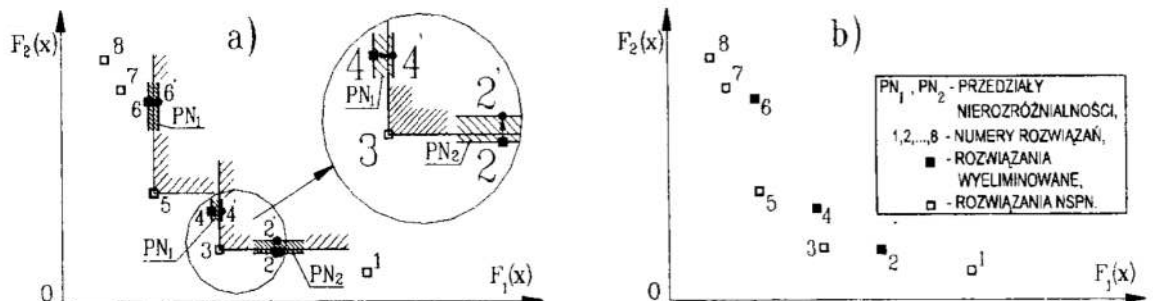
Optymalność w sensie przedziału nierozróżnialności (OSP). Prezentowana OSPN bazuje na wprowadzeniu pojęcia przedziału nierozróżnialności (PN), który określany jest przez projektanta w procentach wartości analizowanego kryterium [5]. Analiza wielokryterialna rozwiązań niezdominowanych przeprowadzana jest w przestrzeni kryterialnej i dąży do ustalenia, czy rozwiązanie „pogorszone” przez przyjęty PN pozostaje nadal rozwiązaniem niezdominowanym i dołączane jest do aktualnie tworzonego podzbioru rozwiązań niezdominowanych. Dla przypadku minimalizacji kryteriów element $\mathbf{x}^{\wedge} \in \Omega$ będzie optymalny w sensie przedziału nierozróżnialności (niezdominowany w sensie przedziału nierozróżnialności NSPN) wtedy i tylko wtedy, gdy w zbiorze Ω nie ma takiego elementu \mathbf{x}^+ , że dla każdego $l \in N$, dla którego:

$$\text{gdy } F_l(\mathbf{x}^{\wedge}) \geq 0 : F_l(\mathbf{x}^{\wedge}) < F_l(\mathbf{x}^+) \text{ zachodzi } (1 + \frac{PN_l}{100}) F_l(\mathbf{x}^{\wedge}) > F_l(\mathbf{x}^+) \quad (3)$$

$$\text{gdy } F_l(\mathbf{x}^{\wedge}) < 0 : F_l(\mathbf{x}^{\wedge}) < F_l(\mathbf{x}^+) \text{ zachodzi } (1 - \frac{PN_l}{100}) F_l(\mathbf{x}^{\wedge}) > F_l(\mathbf{x}^+)$$

gdzie: Ω - niepusty zbiór rozwiązań optymalnych w sensie Pareto,

Rysunek 2 prezentuje ideę wyznaczania rozwiązań niezdominowanych w sensie przedziału nierozróżnialności (RNSPN) opisanych warunkiem (3), według relacji porządku częściowego ($\overset{\circ}{<}$) określonego przez stożek dodatni, z ośmioelementowego zbioru rozwiązań optymalnych w sensie Pareto (ROSP) przy minimalizacji dwóch kryteriów. Rysunek 2a przedstawia działanie przedziału nierozróżnialności PN_1 – dla kryterium $F_1(x)$, w odniesieniu do rozwiązań nr 4 i 6 oraz PN_2 – dla kryterium $F_2(x)$, w odniesieniu do rozwiązania nr 2. Rozważmy szczegółowo rozwiązania nr 3 i nr 4.

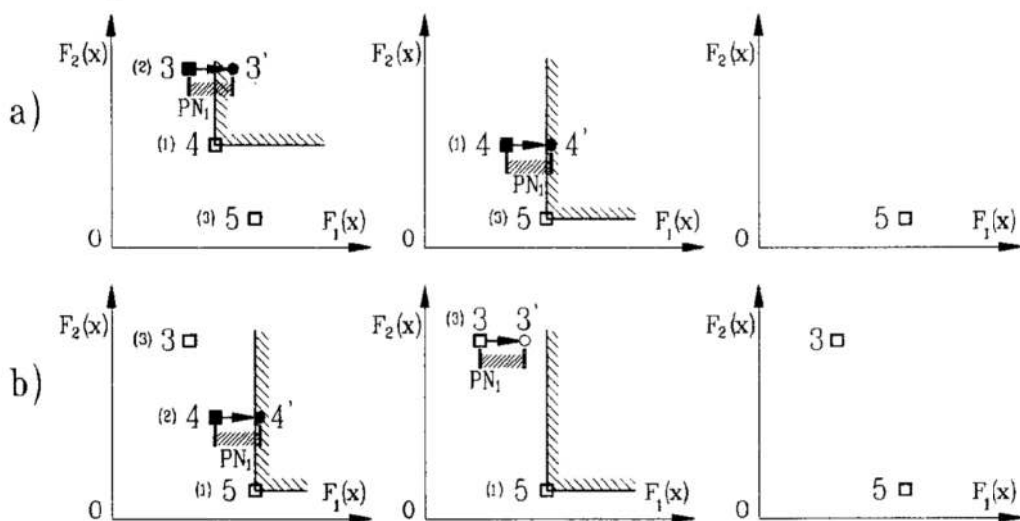


Rys. 2. Wyznaczanie rozwiązań NSPN przy dwukryterialnej minimalizacji
Fig. 2. Determining solutions in the sense of undifferentiation interval with bicriterial minimisation

Rozwiązanie 4 ma większą wartość kryterium $F_2(x)$ i mniejszą wartość kryterium $F_1(x)$, niż rozwiązanie 3 i dlatego jest optymalne w sensie Pareto. Stosując metodę przedziału nierozróżnialności „pogarszamy” chwilowo wartość składowej F_1 rozwiązania 4, o przyjęty PN_1 , przemieszczając rozwiązanie 4 wzdłuż kierunku F_1 do położenia 4'. W zaistniałej sytuacji rozwiązanie 4' należy do stożka dodatniego o wierzchołku w punkcie 3. Wynika z tego, że rozwiązanie 4 nie jest NSPN i zostaje wyeliminowane. W wyniku porównania rozwiązania 2 z rozwiązaniem 3 następuje eliminacja rozwiązania 2. W tym przypadku zdecydował PN_2 odnoszący się do kryterium $F_2(x)$. Rozwiązanie 2 przemieszczone wzdłuż kierunku F_2 do położenia 2' należy również do stożka dodatniego o wierzchołku w punkcie 3 i zostało wyeliminowane. Zgodnie z tą procedurą rozwiązanie nr 5 eliminuje rozwiązanie nr 6. Rysunek 2b prezentuje otrzymany podzbiór rozwiązań NSPN (RNSPN). Należy podkreślić, że rozwiązania 2 i 4 nie mogą (przy przyjętych wartościach PN_1 i PN_2) w żaden sposób wyeliminować rozwiązania 3.

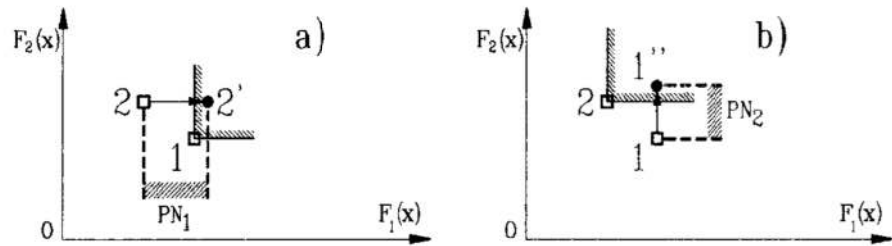
W rzeczywistości położenie elementów zbioru ROESP może być takie, że wyznaczanie rozwiązań NSPN w przestrzeni wielowymiarowej będzie nastroczać pewne kłopoty. Poniżej przedstawiono szczegółowo dwie sytuacje.

Zjawisko łańcucha eliminacji (ZŁE) pojawia się wtedy, gdy zbiór ROESP jest analizowany przy różnych porządkach wprowadzania danych. Przy porządku pobierania danych ze zbioru ROESP: 4-3-5; rysunek 4a, rozwiązanie 3 zostało wyeliminowane przez 4, a następnie rozwiązanie 5 wyeliminowało rozwiązanie 4. Tak więc podzbiór RNSPN składa się z tylko z elementu 5. Wystąpiło ZŁE poprzez chwilowe wykorzystanie rozwiązania 4 do eliminacji rozwiązania 3. Przy zmianie porządku pobierania danych na: 5-4-3 - rysunek 4b, rozwiązanie 4 zostaje w pierwszej kolejności wyeliminowane przez rozwiązanie 5. Podzbiór RNSPN składa się więc z rozwiązań 5 i 3.



Rys. 3. Zjawisko łańcucha eliminacji (ZŁE) przy dwukryterialnej minimalizacji
 Fig. 3. The phenomenon of elimination chain (PEC) with bicriterial minimisation

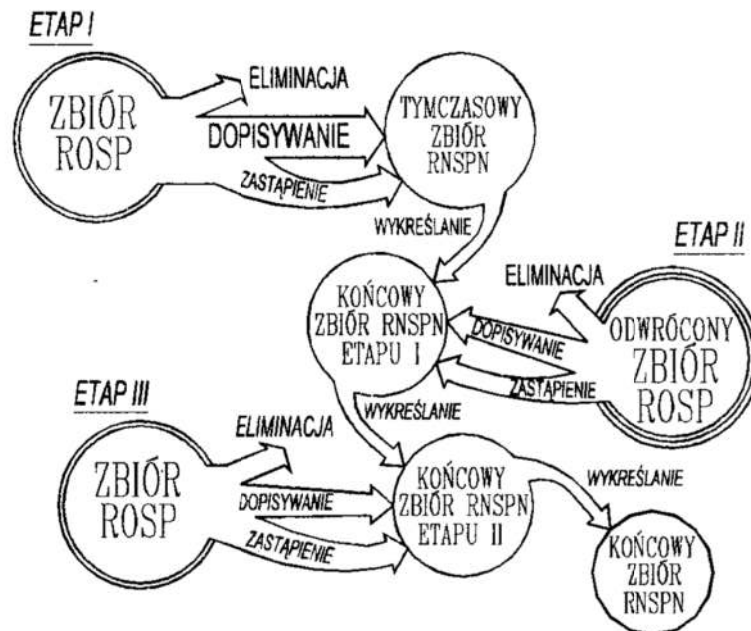
Zjawisko wzajemnego wykluczenia (ZWW) występuje, gdy oba porównywane rozwiązania eliminują się nawzajem przy analizie poszczególnych kryteriów. Oznacza to, że należą odpowiednio do stożków dodatnich o wierzchołku w 1 (Rys. 4a) i 2 (Rys. 4b). W takim przypadku oba porównywane rozwiązania są dopisywane do tworzonego podzbioru RNSPN.



Rys. 4. Zjawisko wzajemnego wykluczania (ZWW) przy dwukryterialnej minimalizacji
 Fig. 4. The phenomenon of mutual exclusion (PME) with bicriterial minimisation

3 Algorytm metody przedziału nierozróżnialności (MPN)

Prezentowany algorytm bada obszar przestrzeni kryterialnej, przy czym warunek (3) sprawdzany jest w obie strony, zarówno w odniesieniu do rozwiązania pobieranego ze zbioru RO SP, jak i rozwiązań już znajdujących się w wyznaczanym podzbiorze RNSPN.



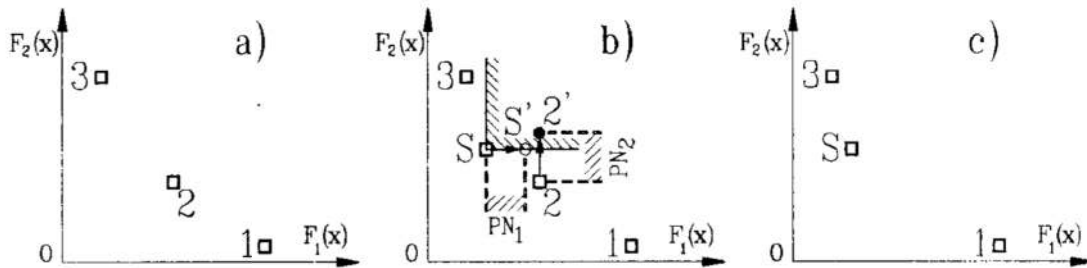
Rys. 5. Idea działania algorytmu MPN

Fig. 5. An idea of the functioning of the undifferentiated interval method (UIM) algorithm

W każdym przebiegu algorytmu następuje wykonanie następujących działań:

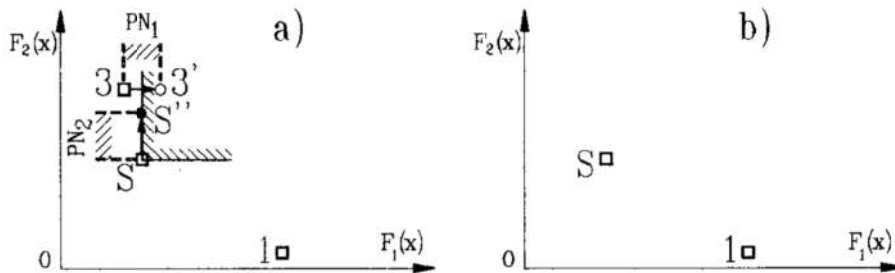
- dopisanie do tworzonego podzbioru RNSPN aktualnie analizowanego rozwiązania – warunek (3) nie jest spełniony dla rozwiązania pobieranego ze zbioru RO SP,
- eliminowanie aktualnie analizowanego rozwiązania przez rozwiązanie należące do podzbioru RNSPN – warunek (3) jest spełniony dla rozwiązania pobranego ze zbioru RO SP,
- zastąpienie istniejącego elementu podzbioru RNSPN analizowanym rozwiązaniem – warunek (3) jest spełniony dla rozwiązania pobranego ze zbioru RNSPN,
- wykreślanie elementów podzbioru RNSPN przez rozwiązanie, które wcześniej zastąpiło istniejący element podzbioru – warunek (3) jest spełniony dla kolejnego rozwiązania pobranego ze zbioru RNSPN.

Poniżej pokazano działanie *zastąpienia*. Zbiór RNSPN składa się z trzech elementów: 1, 2, 3 (Rys. 6a). Po wprowadzeniu rozwiązania S ze zbioru ROSP (Rys. 6b) okazuje się, że przy przyjętym PN_2 rozwiązanie 2 należy do stożka dodatniego o wierzchołku w punkcie S. Warunek (3) jest spełniony i następuje zastąpienie rozwiązania 2 przez rozwiązanie S (Rys. 6c).



Rys. 6. Zastąpienie
Fig. 6. Replacement

Na rysunku 7 przedstawiono graficznie działanie *wykreślenia*. Z dalszego porównywania rozwiązania S z następnym elementem zbioru RNSPN, w tym przypadku z rozwiązaniem 3, wynika, że rozwiązanie 3' należy do stożka dodatniego o wierzchołku w punkcie S (warunek 3 jest spełniony) i dlatego rozwiązanie 3 zostaje wyeliminowane ze zbioru RNSPN.

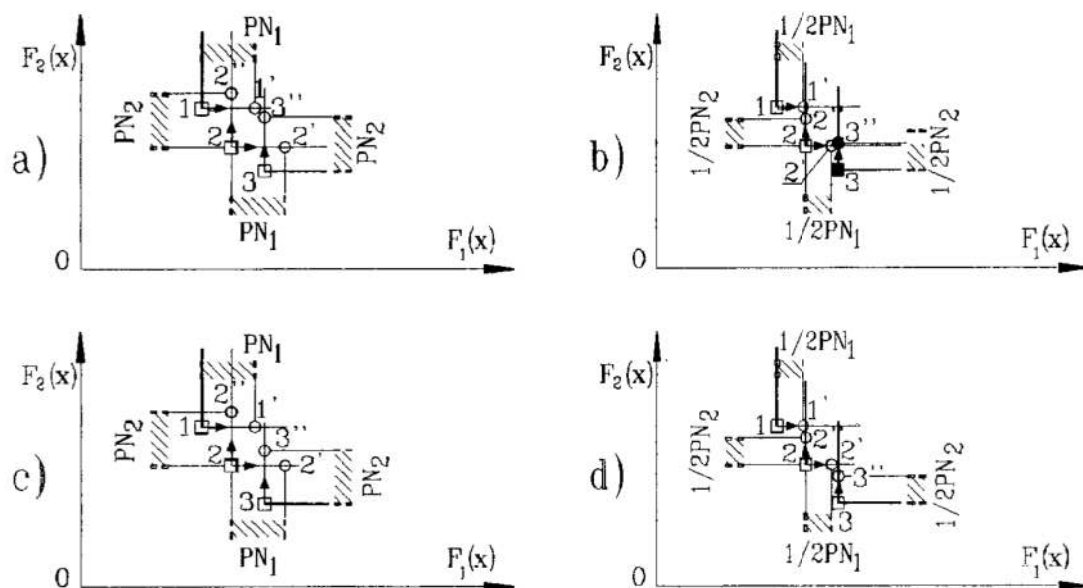


Rys. 7. Wykreślenie
Fig. 7. Cancellation

W prezentowanym algorytmie występowanie ZWW częściowo zlikwidowano przez zastosowanie korekcji przedziału nierozróżnialności (KPN), która polega na dodatkowej analizie tych samych rozwiązań (Rys. 8) i sprawdzeniu warunku (3) przy zmniejszonej wartości PN. Rysunek 8a prezentuje takie położenie elementów zbioru RNSPN: 1 i 3 oraz rozwiązania ze zbioru ROSP - 2, że powtórna analiza przy zmniejszeniu wartości PN_1 i PN_2 o połowę powoduje eliminację rozwiązania 3 (Rys. 8b).

Położenie elementów zbioru RNSPN: 4 i 6 oraz rozwiązania 5 ze zbioru ROSP, prezentowanych na rysunku 8c jest takie, że powtórna analiza przy dwukrotnym zmniejszeniu wartości PN_1 i PN_2 nie przyczynia się do eliminacji żadnego rozwiązania (Rys. 8d).

Występowanie ZŁE częściowo wyrugowano poprzez prowadzenie powtórnej analizy zbioru ROSP przy wprowadzaniu danych w odwrotnej kolejności (etap II i etap III - Rys. 5).



Rys. 8. Korekcja przedziału nierozróżnialności (KPN) podczas ZWW
 Fig. 8. Corection of the undifferentiation interval (CUI) during the PME

4 Przykład zastosowania MPN

Przykład obliczeniowy przedstawiony w pracy [3], a dotyczący optymalizacji impulsowej obróbki elektrochemicznej (PECM) oraz w pracy [4] poświęconej tworzeniu rankingu wariantów parametrów tej obróbki, posłużył do testowania algorytmu realizującego MPN. Analizowany zbiór liczył 30 elementów, które ponumerowano od 1 do 30 zgodnie z pozycją [3]. Ranking najlepszych wariantów obróbki, który był generowany metodą wagowo-korelacyjną w zbiorze rozwiązań dopuszczalnych przedstawia się następująco: 23, 5, 20, 11, 17, 27, 26, 4, 1, 10, 16, 13, 24, 19, 8, 22, 29, 9, 28, 30, 2, 12, 21, 15, 7, 18, 14, 6, 3. Rozwiązanie nr 25 nie było klasyfikowane.

Tablica 1. Wyniki obliczeń Metodą Przedziałów Nierozróżnialności (MPN)
 Table 1. Calculation results according to the UIM

1	2	3	4	5	6	7
Lp.	PN [%]	Liczba elementów		Elementy zbioru RNSPN		Rozwiązania zawsze eliminowane
		Algorytm KPN	Algorytm KPN z powtórna analizą	Zawsze występujące	Czasami występujące	
1.	1,5	18	18	1,3,4,5,8,10,11,12,13,16,17,19,20,22,23,24,26,28	-	2,30
2.	4,5	13	13	4,5,10,11,16,17,19,20,22,23,24,26,28	-	1,2,3,8,12,13,30
3.	6,0	14	14	4,5,10,11,13,16,17,19,20,22,23,24,26,28	-	1,2,3,8,12,30
4.	9,0	12 (4x) 11 (3x)	12 (5x) 11 (2x)	4,5,10,11,16,17,19,23,24,26,28	22	1,2,3,8,12,13,20,30

W cytowanych pracach wyodrębniono cztery kryteria optymalizacji: kryterium 1. – średni wzrost grubości szczeliny międzyelektrodowej podczas obróbki, kryterium 2. – średnia chropowatość powierzchni obrobionej, kryterium 3. – kąt nachylenia obrobionej powierzchni, kryterium 4. – błąd prostoliniowości obrobionej powierzchni. Pierwsze kryterium było maksymalizowane, zaś trzy pozostałe minimalizowano. Zbiór ROSP zawierał 20 elementów o następujących numerach: 1, 2, 3, 4, 5, 8, 10, 11, 12, 13, 16, 17, 19, 20, 22, 23, 24, 26, 28, 30. Wyznaczanie podzbioru RNSPN przeprowadzono dla siedmiu różnych kolejności wprowadzania danych, a wyniki analizy przeprowadzonej MPN pokazano w tabeli 1.

5 Podsumowanie

1. Zastosowanie MPN umożliwia rozdział zbioru ROSP na trzy podzbiory:
 - a) podzbiór RNSPN – jego liczebność jest uzależniona od rozpiętości wartości elementów zbioru ROSP oraz wartości PN (kolumna 5),
 - b) podzbiór rozwiązań, które są zawsze eliminowane (kolumna 7),
 - c) podzbiór RNSPN, który pojawia się na skutek trudności wyrugowania ZŁE. Liczebność tego podzbioru jest wielokrotnie mniejsza od podzbioru RNSPN, a czasami wynosi 0 (kolumna 6).
2. Wydzielenie podzbioru RNSPN przyczynia się do znacznej redukcji liczebności zbioru ROSP (dla $PN=1,5\%$ redukcja o 10%; dla $PN=4,5\%$ redukcja o 35%), co umożliwia łatwiejsze przeglądanie wyników i podjęcie właściwej decyzji.
3. Algorytm z KPN daje stabilne wyniki, o czym świadczy zmniejszająca się liczba elementów zbioru RNSPN przy zwiększaniu wartości PN. Odstępstwem są jedynie wyniki dla $PN=6\%$. Nastąpił nieznaczny wzrost liczebności zbioru RNSPN na skutek niezadziałania korekcji (Rys. 8c i 8d).
4. Zastosowanie algorytmu z powtórными analizami (PA) przy odwróconych kolejnościach wprowadzania danych, przyczyniło się do otrzymywania identycznych lub zbliżonych liczebności zbioru RNSPN. Działanie PA można zaobserwować przy $PN=9\%$. Wynik 12 elementów uzyskano dla 5 porządków wprowadzania danych (kolumna 4), wcześniej wartość tą otrzymano tylko dla 4 porządków (kolumna 3).
5. Porównując ranking rozwiązań z pracy [4], otrzymany metodą wagowo-korelacyjną z podzbiórami RNSPN otrzymanymi dla różnych wartości PN widać, że MPN nie eliminuje rozwiązań, które zajmowały najwyższe pozycje w rankingu. Z jedenastu pierwszych miejsc rankingu 9 elementów wchodzi w skład wszystkich wygenerowanych podzbiorów RNSPN (przy $PN=1,5$ nawet 10 elementów).
6. Przedstawiona w pracy MPN może zostać z powodzeniem włączona do systemów wspomaganie decyzji wielokryterialnych oraz do eliminacji ROSP wygenerowanych na różnych etapach procesu obliczeniowego, np.: w metodach losowych czy metodach z wykorzystaniem algorytmów genetycznych.

Literatura

1. Brans J.P., Vincke P., Mareschal B.: *How to select and how to rank project: The PROMETHEE method*. European Journal of Operational Research (EJOR) vol. 24, 1986, pp. 228-238

2. Keeney R., Raiffa H.: *Decisions with multiple objectives: preferences and value trade-offs*. John Wiley & Sons, Chichester 1991
3. Łubkowski K., Zawora J.: *Optymalizacja impulsowej obróbki elektrochemicznej (PECM)*, ZN Wydz. Mech. nr 19, WSIŃz. w Koszalinie, 1995, ss. 127-134
4. Łubkowski K., Pogorzelski W.: *Ranking wariantów impulsowej obróbki elektrochemicznej*, ZN Wydz. Mech. nr 19, WSIŃz. w Koszalinie, 1995, ss. 135-138
5. Matarazzo B.: *Preference ranking global frequencies in multicriterion analysis (PRAGMA)*. EJOR vol. 10, 1982, pp. 51-55
6. Montusiewicz J.: *Method of searching for compromise solutions*. w: *Rozwój techniki a problemy kształcenia technicznego*, LTN, Lublin 1996, pp. 50-57
7. Montusiewicz J.: *Reducing the Pareto optimal set by means of the undifferentiation interval method*. Proc. of the Second World Congress of Structural and Multidisciplinary Optimization vol. 1, Zakopane 1997, pp. 97-102
8. Montusiewicz J.: *Algorytmy eliminacji rozwiązań optymalnych w sensie Pareto metodą przedziałów nierozróżnialności*. ZN Wydz. Mech., nr 26, Pol. Koszalińska, Koszalin-Mielno, 1999, ss. 184-191
9. Rosenman M. A., Gero J. S.: *Reducing the Pareto Optimal Set in Mulicriteria Optimization (with Applications to Pareto Optimal Dynamic Programming)*, Eng. Opt., vol. 8, 1985, pp. 189-206
10. Roy B.: *Wielokryterialne wspomaganie decyzji*. WNT Warszawa 1990
11. Vincke P.: *Multicriteria decision-aid*. John Wiley & Sons, Chichester 1992
12. Yu P.L.: *Cone convexity, cone extrema points and nondominated solutions in decision problems with multiobjectives*. Journal of Optimization Theory and Applications vol. 14, No. 3, 1974, pp. 319-377

Streszczenie

W pracy przedstawiono metodę wyznaczania podzbiorów rozwiązań niezdominowanych w sensie przedziału nierozróżnialności (NSPN) ze skończonego zbioru rozwiązań optymalnych w sensie Pareto. Podano warunki, jakie musi spełniać rozwiązanie, aby zostało dopisane do wyznaczanego podzbioru rozwiązań NSPN. Przedstawiono ideę działania algorytmu realizującego metodę przedziałów nierozróżnialności (MPN) oraz zamieszczono przykład obliczeniowy bazujący na danych z literatury.

FILTERING OF THE SET OF OPTIMAL SOLUTIONS IN PARETO SENSE BY MEANS OF THE UNDIFFERENTIATION INTERVAL METHOD

Summary

The paper shows a method of determining subsets of nondominated solutions in the sense of an undifferentiation interval (NSUI) from a finite set. The conditions are provided which a solution must fulfil in order to be included in the NSUI solution subset currently determined. An idea of the operation of the algorithm capable of realising the undifferentiation interval method (UIM) is then introduced. A calculation example is also included, based on data from literature.

Referat recenzowany

FILTRACJA ROZWIĄZAŃ PARETO OPTYMALNYCH W OPTYMALIZACJI WIELOKRYTERIALNEJ Z WYKORZYSTANIEM ALGORYTMÓW GENETYCZNYCH

1. Wprowadzenie

Metody optymalizacji wielokryterialnej oparte na algorytmach genetycznych mają stosunkowo długą historię ze względu na swoje zalety. Metody te nie stawiają żadnych ograniczeń co do modelu optymalizacyjnego i pozwalają na generowanie pełnego zbioru rozwiązań Pareto optymalnych w czasie jednego przebiegu programu optymalizacyjnego. Ma to ogromne znaczenie w problemach podejmowania decyzji przy konfliktowych celach, ponieważ decydent ma pełny obraz możliwych rozwiązań kompromisowych. Zaleta ta w pewnych przypadkach może stać się wadą zmniejszającą efektywność działania algorytmu genetycznego, szczególnie ze względu na czas obliczeń. Podczas realizacji algorytmu genetycznego dla problemu optymalizacji wielokryterialnej ze zmiennymi decyzyjnymi ciągłymi, zbiór rozwiązań Pareto optymalnych rośnie wraz z liczbą generacji. Dla pewnych zadań końcowym wynikiem wykonania programu jest zbiór Pareto, który zawiera setki a nawet tysiące rozwiązań i niektóre z nich mogą być bardzo zbliżone co do wartości funkcji celu. Dla podejmującego decyzję analizowanie wszystkich tych rozwiązań jest trudne i męczące. Ponadto, w większości metod w każdej generacji nowo wygenerowane rozwiązanie jest porównywane ze wszystkimi rozwiązaniami w zbiorze Pareto. Jeżeli zbiór rozwiązań Pareto optymalnych powiększa się to czas obliczeń zaczyna znacząco wzrastać. Problem ten został zauważony przez Osyczkę i Montusiewicza (1994) w czasie stosowania prostej metody losowego przeszukiwania. W celu eliminacji tego zjawiska zaproponowali oni pewne metody wyboru reprezentatywnego podzbioru rozwiązań Pareto optymalnych. Montusiewicz, (1998, 1999) rozwinął dalej jedną z tych metod. Dla metod optymalizacji wielokryterialnej opartych na algorytmach genetycznych podejście takie może być bardzo przydatne zarówno ze względu na czas obliczeń jak i na problem podejmowania decyzji. W pracy została opisana metoda poszukiwania reprezentatywnego zbioru rozwiązań Pareto optymalnych wykorzystująca tzw. przedział nierozróżnialności a następnie zostało przedstawione jej wykorzystanie do zwiększenia efektywności algorytmów genetycznych.

* Politechnika Krakowska, Wydział Mechaniczny, Al. Jana Pawła II 37, 31-864 Kraków
email: osyczka@mech.pk.edu.pl, krenich@mech.pk.edu.pl

** Politechnika Lubelska, Katedra Podstaw Techniki, ul. Nadbystrzycka 38a, 20-618 Lublin
email: montus@antenor.pol.lublin.pl

2. Sformułowanie ogólnego problemu optymalizacji wielokryterialnej

Problem optymalizacji wielokryterialnej dla programowania nieliniowego można sformułować następująco:

$$\min_{\mathbf{x} \in R^n} f(\mathbf{x}) \quad (1)$$

przy spełnieniu ograniczeń nierównościowych:

$$g_k(\mathbf{x}) \geq 0 \quad k=1, 2, \dots, K \quad (2)$$

i ograniczeń równościowych:

$$h_m(\mathbf{x}) = 0 \quad m=1, 2, \dots, M. \quad (3)$$

gdzie:

$\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ – jest wektorem zmiennych decyzyjnych,

$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_l(\mathbf{x})]^T$ – jest wektorem funkcji celu.

Rozwiązaniem zadania optymalizacji wielokryterialnej jest znalezienie zbioru rozwiązań niezdominowanych czyli rozwiązań Pareto optymalnych.

3. Metoda przedziału nierozróżnialności

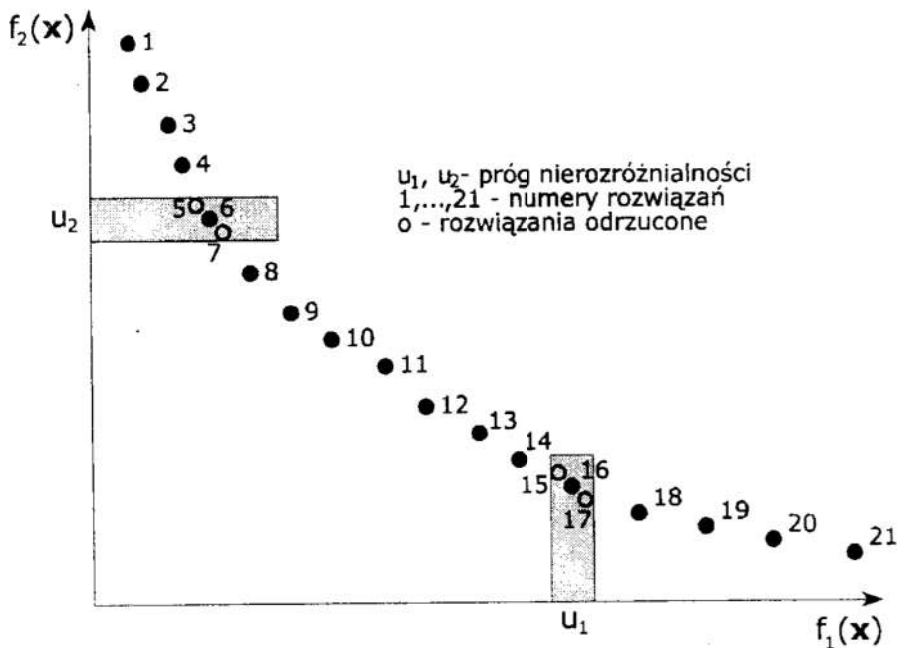
Główna idea tej metody polega na odrzucaniu ze zbioru rozwiązań Pareto optymalnych, tych rozwiązań, które różnią się nieznacznie co do wartości funkcji celu. Metoda ta wprowadza pojęcie przedziału nierozróżnialności u_i dla i -tej funkcji celu. Przedział ten jest podawany w procentach i określa różnicę pomiędzy wartościami funkcji celu, która kwalifikuje rozwiązanie do odrzucenia ze zbioru Pareto. Innymi słowy, jeżeli jest więcej niż jedno rozwiązanie w przedziale u_i , wówczas jedno rozwiązanie pozostaje w zbiorze Pareto a pozostałe są odrzucane. Metoda opiera się na następującej definicji: *Rozwiązanie $\mathbf{x}^{\wedge} \in X$ jest rozwiązaniem niezdominowanym w sensie przedziału nierozróżnialności wtedy i tylko wtedy gdy nie istnieje rozwiązanie $\mathbf{x}^+ \in X$ takie że:*

$$\text{dla } f_i(\mathbf{x}^{\wedge}) \geq 0 : f_i(\mathbf{x}^{\wedge}) < f_i(\mathbf{x}^+) \quad \text{i} \quad f_i(\mathbf{x}^{\wedge}) \cdot \left(1 + \frac{u_i}{100}\right) > f_i(\mathbf{x}^+)$$

$$\text{dla } f_i(\mathbf{x}^{\wedge}) < 0 : f_i(\mathbf{x}^{\wedge}) < f_i(\mathbf{x}^+) \quad \text{i} \quad f_i(\mathbf{x}^{\wedge}) \cdot \left(1 - \frac{u_i}{100}\right) > f_i(\mathbf{x}^+)$$

Definicja ta została graficznie zilustrowana na rysunku 1, na którym zacieniowane obszary przedstawiają przedziały nierozróżnialności u_1 i u_2 . W tym przykładzie dla pierwszej funkcji celu rozwiązanie 16 eliminuje rozwiązanie 15 i 17. Dla drugiej funkcji celu rozwiązanie 6 eliminuje rozwiązanie 5 i 7. Szczegółowy opis tej metody znajduje się w pracach Montusiewicza (1998, 1999).

Zauważmy, że metoda wyboru reprezentatywnego podzbioru rozwiązań Pareto optymalnych może być stosowana również po zrealizowaniu obliczeń z wykorzystaniem dowolnego algorytmu genetycznego dla optymalizacji wielokryterialnej. W tym przypadku zamiast podejmować decyzję na podstawie setek rozwiązań Pareto optymalnych wygenerowanych przez algorytm genetyczny, podejmujący decyzję ma do dyspozycji kilka lub kilkanaście rozwiązań, które są równomiernie rozłożone w zbiorze rozwiązań Pareto optymalnych.



Rys. 1 Graficzna ilustracja metody progów nierozróżnialności
 Fig. 1 Graphical illustration of the undiscernibility interval method

4. Metoda algorytmu genetycznego z wyborem reprezentatywnego podzbioru rozwiązań Pareto

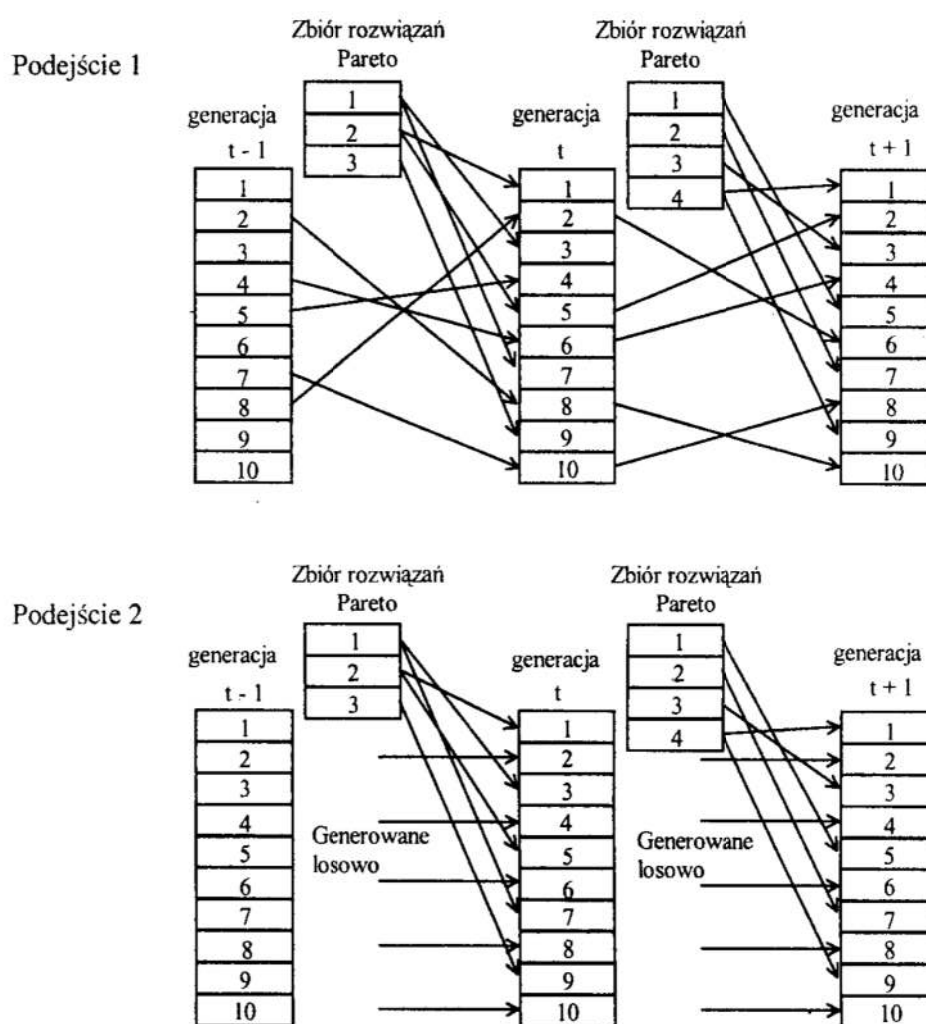
Idea tej metody polega na zredukowaniu zbioru rozwiązań Pareto optymalnych po realizacji pewnej liczby generacji przy użyciu opisanej powyżej metody wyboru reprezentatywnego podzbioru rozwiązań Pareto.

Kroki tej metody są następujące:

1. Podstaw $t=1$, gdzie t jest numerem aktualnie realizowanej generacji.
2. Wygeneruj zbiór rozwiązań Pareto optymalnych używając dowolnego algorytmu genetycznego.
3. Sprawdź czy zbiór rozwiązań Pareto optymalnych zawiera więcej rozwiązań niż ustalona liczba P , gdzie P jest przyjmowanym przez użytkownika progiem uruchomienia filtracji. Jeżeli tak, to wybierz reprezentatywny podzbiór rozwiązań Pareto używając metody przedziału nierozróżnialności i przejdź do kroku 4. W przeciwnym przypadku przejdź bezpośrednio do kroku 4.
4. Podstaw $t = t + 1$ i jeśli $t \leq T$, gdzie T jest z góry ustaloną liczbą generacji, to przejdź do kroku 2. W przeciwnym przypadku zakończ obliczenia.

Wybór liczby P powinien być dokonany z dużą ostrożnością. Jeżeli P jest za małe to liczba rozwiązań Pareto może być zbyt mało reprezentatywna dla rozpatrywanego problemu i algorytm genetyczny może nie osiągnąć rzeczywistego zbioru rozwiązań Pareto optymalnych. Jeżeli P jest za duże to tracimy efekt zmniejszenia czasu obliczeń. Bardzo ważny jest również wybór przedziału nierozróżnialności u_i . Jeżeli przedział ten jest za mały to za mała jest również liczba odrzuconych rozwiązań i wtedy nie ma efektu zmniejszenia zbioru rozwiązań Pareto optymalnych. Natomiast zbyt duża wartość u_i może powodować, że otrzymany podzbiór rozwiązań jest za mały aby być reprezentatywnym.

Do generowania zbiorów rozwiązań Pareto optymalnych w poszczególnych generacjach została wykorzystana metoda rozprowadzania rozwiązań Pareto optymalnych. Metoda ta została zaproponowana przez Osyczkę i Tamurę (1996) a jej główna idea polega na utrzymywaniu między kolejnymi generacjami zbioru rozwiązań Pareto optymalnych, które są używane do tworzenia populacji w następnej generacji. W każdej generacji zbiór rozwiązań Pareto jest tworzony na podstawie istniejącego zbioru Pareto otrzymanego w poprzedniej generacji i rozwiązań wygenerowanych przez algorytm genetyczny w rozpatrywanej generacji. Sposoby tworzenia nowej populacji zostały przedstawione na rysunku 2. Łatwo można zauważyć, że zbiór rozwiązań Pareto optymalnych będzie rósł wraz ze wzrostem liczby generacji. Liczba rozwiązań w zbiorze Pareto nie zależy od rozmiaru populacji. Może się więc zdarzyć, że po przebiegu pewnej liczby generacji liczba rozwiązań Pareto optymalnych będzie większa niż rozmiar populacji. Problem ten jest szczególnie widoczny dla ciągłych problemów programowania matematycznego gdzie liczba wygenerowanych rozwiązań Pareto



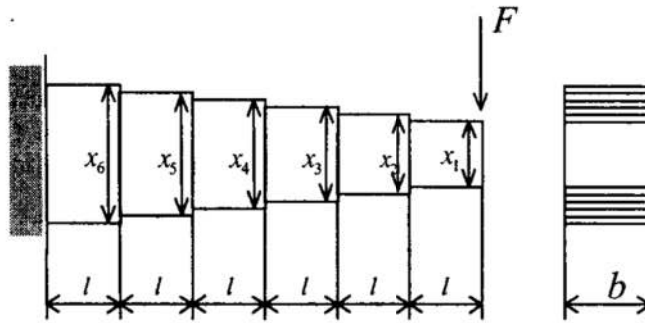
Rys. 2 Graficzna ilustracja metody rozprowadzania zbioru rozwiązań Pareto optymalnych

Fig. 2 Graphical illustration of the Pareto Set Distribution Method

optymalnych może być tak duża, że przekroczy zadeklarowane tablice w programie optymalizacyjnym. W tym przypadku konieczne wydaje się stosowanie opisywanej metody.

5. Przykład numeryczny

Rozpatrzmy przykład optymalizacji wielokryterialnej belki o zmiennych decyzyjnych ciągłych przedstawionej na rysunku Rys. 3.



Rys. 3 Schemat belki
Fig. 3 Scheme of the beam

Problem optymalizacji jest sformułowany następująco:

Wektor zmiennych decyzyjnych:

$$\mathbf{x} = [x_1, x_2, \dots, x_N]^T$$

gdzie: x_n jest szerokością n -tej części belki.

Funkcje celu są następujące:

- Objętość belki:

$$f_1(\mathbf{x}) = bl \sum_{n=1}^N x_n$$

- Przemieszczenie spowodowane działaniem siły F :

$$f_2(\mathbf{x}) = \frac{Fl^3}{2E} \left(\frac{1}{I_1} + \sum_{n=2}^N \frac{i^3 - (i-1)^3}{I_i} \right)$$

gdzie: $I_n = \frac{bx_n^3}{12}$ dla $n = 1, 2, \dots, N$

Ograniczenia są następujące:

$$\frac{6F \times nl}{bx_n^2} \leq \sigma_g \quad \text{dla } n = 1, 2, \dots, N$$

$$0 \leq x_1, x_n \leq x_{n+1}, x_N \leq d \quad \text{dla } n = 1, 2, \dots, N-1$$

Dla rozwiązania tego problemu przyjęto następujące wartości parametrów:

$N = 6$, $l = 50$ [mm], $b = 50$ [mm], $F = 10000$ [N], $E = 2.06 \times 10^5$ [N/mm²],
 $\sigma_g = 360$ [N/mm²], $d = 32$ [mm]

W wyniku optymalizacji z wykorzystaniem algorytmu genetycznego bez stosowania metody przedziału nierozróżnialności, otrzymano bardzo liczny zbiór rozwiązań Pareto optymalnych sięgający nawet kilku tysięcy rozwiązań. Na przykład dla rozmiaru populacji równej 200 osobników i ilości generacji równej 400 otrzymano 1284 rozwiązań Pareto optymalnych. Znaczna część tych rozwiązań jest do siebie bardzo podobna i nie wnosi żadnych nowych informacji do procesu podejmowania optymalnej decyzji, natomiast bardzo spowalnia działanie algorytmu genetycznego. Przeprowadzone eksperymenty pokazały, że zupełnie bez szkody dla końcowego zbioru rozwiązań Pareto optymalnych można w trakcie działania algorytmu genetycznego stosować filtrację rozwiązań metodą przedziałów nierozróżnialności. Zastosowanie tej metody znacznie przyspiesza działanie algorytmu genetycznego co zostało przedstawione poniżej. Eksperyment przeprowadzono używając metodę rozprowadzania zbioru Pareto dla kilkunastu przebiegów algorytmu genetycznego przy różnych parametrach generatora liczb losowych bez filtracji oraz przy zastosowaniu filtracji. Zastosowano następujące parametry algorytmu genetycznego:

- wielkość populacji 200, ilość generacji 200,
- prawdopodobieństwo krzyżowania 0.6, prawdopodobieństwo mutacji 0.08,
- współczynnik kary $r = 10^4$,
- próg nierozróżnialności dla każdego kryterium 10 %,
- próg uruchomienia filtracji: 100 rozwiązań w zbiorze Pareto.

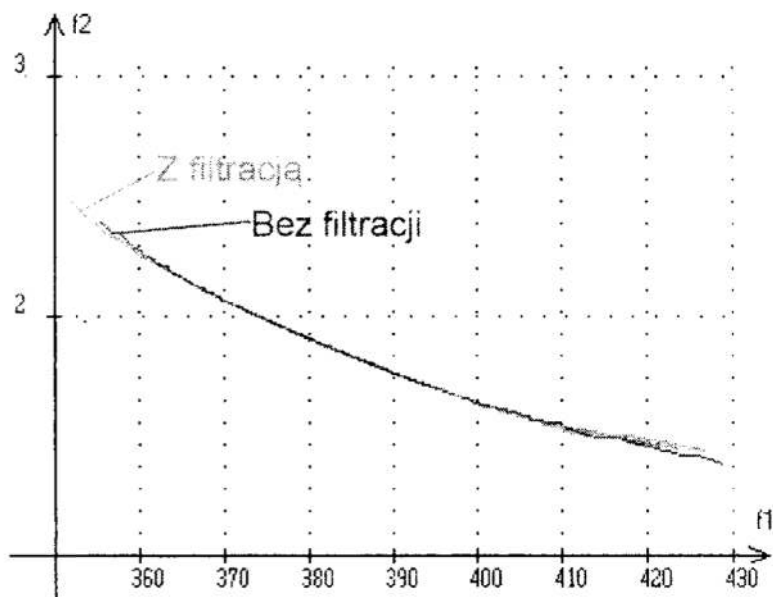
Uzyskane wyniki dla dwóch przykładowych przebiegów algorytmu przedstawiono w tabelach Tab.1 i Tab.2 oraz zilustrowano graficznie na rysunkach Rys.4 i Rys.5. Analizując uzyskane wyniki można stwierdzić, że proponowana metoda daje kilkakrotne skrócenie czasu obliczeń przy pełnej reprezentatywności zbioru rozwiązań Pareto optymalnych uzyskanych na końcu przebiegu algorytmu genetycznego. Podobne rezultaty uzyskano dla innych przebiegów algorytmu genetycznego dla tego zadania.

Tab. 1 Wyniki eksperymentu I dla belki stopniowanej
Tab. 1 Results of the experiment I for the beam.

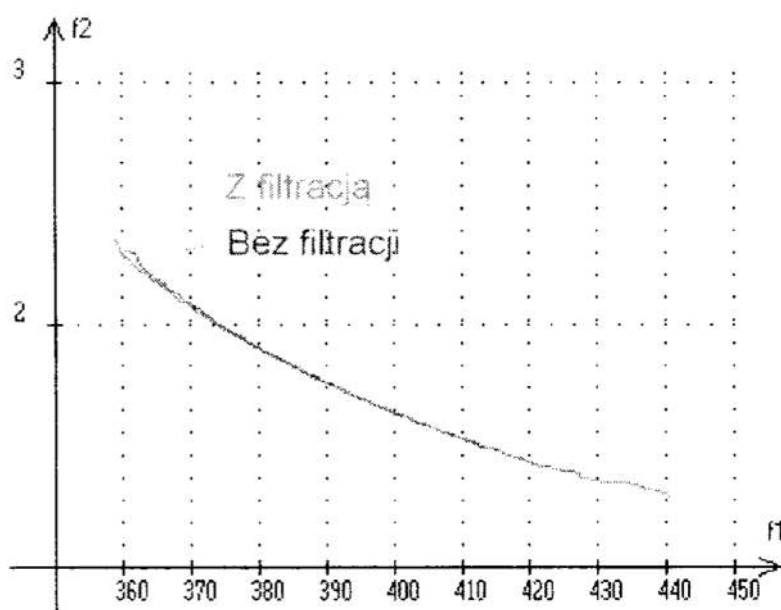
Metoda generacji zbioru Pareto optymalnego	Czas generacji zbioru Pareto [s]	Liczebność zbioru Pareto optymalnego
z filtracją	28	97
bez filtracji	146	782

Tab. 2 Wyniki eksperymentu II dla belki stopniowanej.
Tab. 2 Results of the experiment II for the beam.

Metoda generacji zbioru Pareto optymalnego	Czas generacji zbioru Pareto [s]	Liczebność zbioru Pareto optymalnego
z filtracją	29	77
bez filtracji	173	810



Rys. 4 Zbiory rozwiązań Pareto optymalnych - eksperyment I
 Fig. 4 Sets of Pareto optimal solutions - the experiment I



Rys. 5 Zbiory rozwiązań Pareto optymalnych - eksperyment II
 Fig. 5 Sets of Pareto optimal solutions - the experiment II

6. Wnioski

Przeprowadzone eksperymenty pokazują, że zupełnie bez szkody dla końcowego zbioru rozwiązań Pareto optymalnych można w trakcie działania algorytmu genetycznego stosować filtrację rozwiązań metodą przedziałów nierozróżnialności. Należy przy tym zwrócić uwagę, że na efektywne wykorzystanie metody ogromny wpływ ma odpowiedni dobór progu filtracji P oraz przedziałów nierozróżnialności u_i . Dla metod

optymalizacji wielokryterialnej opartych na algorytmach genetycznych zastosowanie tej metody może być bardzo przydatne zarówno ze względu na czas obliczeń jak i na problem podejmowania decyzji, co zostało potwierdzone także dla innych przykładów optymalizacji wielokryterialnej ze zmiennymi decyzyjnymi ciągłymi.

Praca została wykonana w ramach grantu KBN PB-1429/T07/98/14

Literatura

1. Montusiewicz J.: *Pewna metoda eliminacji rozwiązań optymalnych w sensie Pareto*. Zeszyty Naukowe Wydziału Mechanicznego, z.23, pp. 222-229, Politechnika Koszalińska, Koszalin, 1998.
2. Montusiewicz J.: *Division of the Set of Nondominated Solutions by Means of the Undifferentiation Interval Method*. [in:] The Technological Information Systems, pp. 65-72, Societas Scientiarum Lublinensis, Lublin, 1999.
3. Osyczka A., Montusiewicz J.: *A Random Search Approach to Multicriterion Discrete Optimization*. IUTAM Symposium on Discrete Structural Optimization, Springer-Verlag, pp. 71-79, 1994.
4. Osyczka A., Tamura H.: *Pareto Set Distribution Method for Multicriteria Optimization Using Genetic Algorithm*. In: Proceedings of Genetic Algorithms '96, (MENDEL-96)-International Conference, Brno, Czech Republic, pp.135-143, 1996
5. Osyczka A., Tamura H., Saito Y.: *Pareto Set Distribution Method for Multicriteria Design Optimization Using Genetic Algorithm*. In: Proceedings of Engineering Design and Automation Conference, Bangkok, Thailand, pp. 228-231, 1997.
6. Steuer R.E.: *Multiple Criteria Optimization – Theory, Computation and Application*. Wiley, New York, 1986.

Streszczenie

W pracy została opisana metoda poszukiwania reprezentatywnego zbioru rozwiązań Pareto optymalnych wykorzystująca tzw. przedział nierozróżnialności a następnie zostało przedstawione jej wykorzystanie do zwiększenia efektywności działania algorytmów genetycznych. Eksperyment przeprowadzono używając metodę rozprowadzania zbioru Pareto dla przykładowego ciągłego problemu optymalizacji belki z wykorzystaniem metody oraz bez jej użycia. Uzyskane wyniki wskazują na znaczne zwiększenie efektywności działania algorytmu genetycznego.

GENETIC ALGORITHM BASED METHOD FOR MULTICRITERIA OPTIMIZATION WITH FILTRATION PARETO SET

Summary

In the paper the method of selection of a representative subset of Pareto optimal solutions is presented. The method uses indiscernibility intervals to remove some solutions from the whole Pareto set. Application of this method to genetic algorithm techniques makes the search of a Pareto frontier much more effective. An example presented in the paper shows this effectiveness.

Referat recenzowany

DYDAKTYKA EKSPERYMENTALNA Z WYKORZYSTANIEM PROGRAMOWALNYCH KLOCKÓW LEGO MINDSTORMS™

1 Wstęp

Czy można optymalizować dydaktykę? Można, a nawet trzeba! Wraz z szybkim rozwojem nauki i techniki, dotychczasowy paradygmat scholastyki bazujący głównie na wykładaniu wiedzy, staje się coraz mniej adekwatny do wyzwań stawianych współczesnej uczelni akademickiej. Problem ten dobrze ilustrują słowa:

„Tell me, and I forget. Teach me, and I may remember.
Involve me, and I learn” - Benjamin Franklin

O tym, że problem jest istotny świadczy fakt, że większość uczelni zagranicznych od kilku lat prowadzi intensywne prace badawcze nad nowymi formami nauczania określanymi jako: *learning by doing*, *hands-on learning* lub *problem-based learning* [1,2,3]. Tematyka ta jest na tyle aktualna, że na tegorocznej międzynarodowej konferencji IEEE System, Man, and Cybernetics 2000, organizowana jest specjalna sesja tematyczna *Robotic Construction: An Integrated Systems Approach to Teaching* [7].

W bieżącym roku akademickim, również na Politechnice Wrocławskiej została zainicjowana nowa forma zajęć dydaktycznych: „Seminarium eksperymentalnego konstruowania i programowania robotów” dla studentów różnych specjalności elektroniki i informatyki. Narzędziem, jakie wykorzystano podczas seminarium, były programowalne klocki (z wbudowanym mikroprocesorem) opracowane przez zespół naukowców z MediaLab na uczelni MIT [5], a produkowane od roku przez firmę LEGO w postaci zestawów *LEGO MindStorms - Robotic Invention System* [6].

Intencją tego referatu jest podzielenie się doświadczeniem i efektami jakie uzyskano podczas tego seminarium oraz rozpropagowanie idei między-uczelnianych konkursów zaawansowanych projektów studenckich [4].

Ponieważ trudno jest znaleźć adekwatny polski odpowiednik w/w haseł *learning by doing*, *hands-on* ... itd., w dalszej części referatu przyjmujemy roboczą nazwę „dydaktyka eksperymentalna” (w skrócie DE) podkreślającą dużą rolę eksperymentów (rozwiązywania rzeczywistych problemów) w takim stylu nauczania.

2 Dydaktyka eksperymentalna

Dydaktyka eksperymentalna to taka strategia nabywania wiedzy i umiejętności, która promuje aktywne formy uczenia się poprzez **rozwiązywanie problemów**, zamiast klasycznego **zapamiętywania faktów**. Podstawowe cechy charakterystyczne dla DE:

- kluczowym elementem jest aktywna postawa studenta,
- nauka odbywa się w małych grupach studenckich,
- nauczyciel jest pomocnikiem lub przewodnikiem (ale nie liderem grupy),
- rozwiązywanie odpowiednio wybranego problemu inicjuje i ukierunkowuje proces zdobywania wiedzy teoretycznej w bezpośrednim połączeniu z nabywaniem praktycznych umiejętności,
- nowa wiedza nabywana jest w procesie samo-uczenia, którego przebieg w dużym stopniu zależy od samych studentów.

* Instytut Cybernetyki Technicznej, Politechnika Wroclawska, E-mail: piasecki@ict.pwr.wroc.pl

** Instytut Cybernetyki Technicznej, Politechnika Wroclawska, E-mail: pawel@ict.pwr.wroc.pl

Standardowy cykl dydaktyczny:

wykład → ćwiczenia → laboratorium → seminarium → projekt / eksperyment

wymagane zaangażowanie studentów

podstawowy cel: UTRWALANIE WZORCÓW
podstawowe pytanie: JAKIE JEST ROZWIĄZANIE ?

Dydaktyka eksperymentalna:

eksperyment / projekt → seminarium $\left\{ \begin{array}{l} \text{samodzielne studia literaturowe} \\ \text{wykłady / ćwiczenia / laboratoria} \end{array} \right.$

wymagane zaangażowanie studentów

podstawowy cel: POSZUKIWANIE WZORCÓW
podstawowe pytanie: JAKA JEST METODA ROZWIĄZUJĄCA ?

Rys. 1. Ilustracja podstawowych cech dydaktyki eksperymentalnej

Fig. 1. Diagram illustrating basic features of Problem-Based learning

Taka forma dydaktyki jest próbą odpowiedzi na podstawowy problem współczesnych uczelni: pomimo, iż studenci otrzymują dużą ilość wiedzy, to nie potrafią jej zastosować w praktyce. Nie potrafią pracować zespołowo. Nie mają wprawy w poszukiwaniu i wykorzystywaniu nowych zasobów technicznych i naukowych.

Niestety, istniejące zajęcia laboratoryjne są najczęściej jedynie bierną ilustracją wykładów, a samodzielne projekty pojawiają się dopiero na ostatnich latach studiów. O tym, że można inaczej, świadczy przykład amerykańskich uczelni (np. University of Colorado at Boulder) realizujących program *Integrated Teaching and Learning* - ITL, gdzie studenci realizują zaawansowane projekty już od pierwszego roku studiów [10].

Dydaktyka eksperymentalna bardzo często wykorzystuje interdyscyplinarne problemy projektowo-badawcze. Na uczelniach technicznych najbardziej popularne stało się wykorzystywanie robotyki jako dziedziny łączącej: elektronikę, elektrotechnikę, mechanikę, inżynierię komputerową, sztuczną inteligencję, bioinżynierię, itd. Popularność ta znacznie wzrosła wraz z pojawieniem się na rynku „względnie tanich” zestawów robotycznych, umożliwiających swobodne eksperymentowanie i prototypowanie, bez wymagania zaawansowanej wiedzy studenta we wszystkich w/w dziedzinach jednocześnie.

3 Programowalne klocki LEGO MindStorms

Jednym z takich produktów jest rodzina programowalnych klocków *LEGO MindStorms*. W pierwotnym zamiarze klocki te miały służyć głównie starszym dzieciom i uczniom szkół gimnazjalnych (wiek 12+). Wkrótce okazało się jednak, że po opracowaniu bardziej zaawansowanych narzędzi programistycznych oraz wykorzystywaniu kilku programowalnych klocków jednocześnie, zestawy te stały się idealnym narzędziem dydaktyki eksperymentalnej na uczelniach wyższych.

Zestawy *MindStorms* składają się z kompletu standardowych klocków *LEGO* (np. elementy nośne konstrukcji, łączniki, ośki, koła), klocków *LEGO Technic* (np. silniki prądu stałego, koła zębate, przekładnie pasowe, ślimakowe, różnicowe, czujniki) oraz klocka-mikrokomputera w plastikowej obudowie umożliwiającej bezpośrednie łączenie z pozostałymi klockami. Taki zestaw umożliwia szybkie budowanie różnorodnych konstrukcji mechanicznych, które mogą być sterowane za

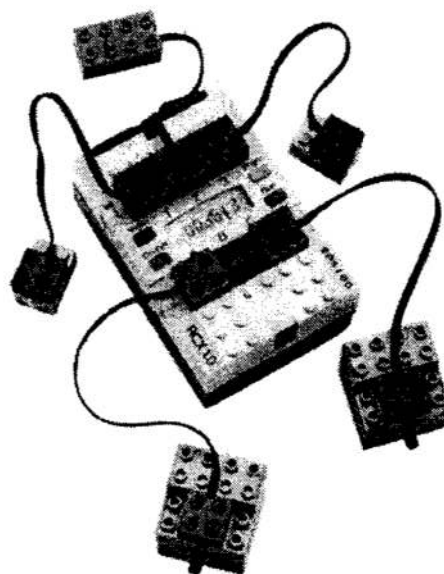
pomocą wbudowanego mikrokontrolera „pokładowego” lub sterowane zdalnie z zewnętrznego komputera o dużej mocy obliczeniowej. W tabeli 1 zestawiono rodzaje aktualnie dostępnych programowalnych klocków LEGO.

Tabela 1. Zestawienie programowalnych klocków LEGO
Table 1. Examples of programmable LEGO bricks

	Scout	Cybermaster	RCX
Zestaw:	Robotics Discovery Set	CyberMaster	Robotic Invention System
Komunikacja:	sterownie „pilotem” (Scout’s Command Center)	kanał radiowy do komputera PC	kanał na podczerwień do komputera PC i innych klocków RCX
Firmware:	wbudowany	wbudowany	ładowny
Wejścia	2 uniwersalne + wbudowany czujnik światła	3 uniwersalne	3 uniwersalne
Wyjścia	2 uniwersalne	1 uniwersalne + 2 wbudowane silniki	3 uniwersalne

Do celów edukacyjnych najbardziej użyteczny jest klocek RCX, który wchodzi w skład podstawowego zestawu RIS (*Robotic Invention System*). Klocek RCX posiada wbudowane:

- mikrokontroler Hitachi H8 taktowany zegarem 16 MHz,
- 32 KB pamięci RAM, do której ładowany jest *firmware* oraz programy sterujące,
- 16 KB pamięci ROM,
- kanał transmisji na podczerwień do komunikacji z komputerem PC lub innym klockiem RCX,
- 3 uniwersalne kanały wejściowe (10 bitowe przetworniki A/D), do których mogą być dołączane różne sensory,
- 3 kanały wyjściowe do prądowego sterowania silników lub innych elementów wykonawczych,
- wyświetlacz LCD,
- głośniczek.



Rys. 2. Klocek RXC
Fig. 2 RCX brick

Podstawowy zestaw RIS zawiera m.in. 1 klocek RCX, 2 silniki, 2 sensory dotykowe, 1 punktowy czujnik oświetlenia, komunikacyjny transponder podczerwieni dołączany do komputera PC oraz zestaw 700 standardowych klocków umożliwiających zbudowanie różnorodnych konstrukcji mechanicznych (np. kołowe lub gąsienicowe roboty mobilne, model kopiarki itp.). Uzupełnieniem zestawu RIS są zestawy rozszerzające, które zawierają dodatkowe silniki i sensory (np. *RoboSports* - do konstrukcji robotów grających w piłkę, *ExplorationMars* - do konstrukcji pojazdów nawigujących w nieznanym otoczeniu). Uniwersalna konstrukcja wejść i wyjść klocka RCX umożliwia również dołączanie innych niestandardowych układów wykonawczych i sensorycznych zaprojektowanych samodzielnie przez studentów.

Do tworzenia programów dla klocka RCX można wykorzystać środowisko programistyczne dostarczane wraz z zestawem RIS:

- „RCX Code” - graficzny edytor do tworzenia programów z „klocków - instrukcji” (Windows 95/98/NT),
- spirit.ocx - biblioteka do programowania w środowisku Visual Basic, Visual C, Delphi lub C++ Builder (Windows 95/98/NT),

lub alternatywne narzędzia programistyczne dostępne w sieci Internet:

- NQC (*Not Quite C*) - język o składni wzorowanej na C (Windows 95/98/NT, Unix),
- RcxCC (*RCX Command Center*) - zintegrowane środowisko do programowania i sterowania klockiem RCX lub CyberMaster (Windows 95/98/NT),
- LegoOS - „system operacyjny” dla klocków LEGO (Unix)
- emulegOS - emulator klocka RCX dla LegOS (Unix).

Środowisko graficzne „RCX Code” dostarczane wraz z zestawem RIS umożliwia budowanie wyłącznie prostych programów reaktywnych, w których sekwencje instrukcji sterujących są inicjowane przez sygnały z sensorów lub z wbudowanego zegara.

Bardziej uniwersalne programy mogą być pisane w języku NQC o składni zbliżonej do języka C. W języku NQC można wykorzystać do 9 procesów (zadań), które mogą być wykonywane współbieżnie. Programy NQC mogą zawierać dowolnie zagnieżdżone instrukcje warunkowe i iteracyjne. Niestety, dostępne są tylko 32 zmienne całkowite. Język NQC nadaje się do tworzenia programów behawioralnych, w których procesy - zadania implementują złożone „zachowania” budowanej konstrukcji. Kompilatory języka NQC mogą być uruchamiane zarówno w środowisku Windows, jak i Unix. W środowisku Windows kompilator może być wywoływany bezpośrednio z programu RcxCC, który umożliwia edycję, kompilację, uruchamianie programów oraz testowanie stanu RCX (podgląd zmiennych, stan sensorów, mapa pamięci itp.). Kod wynikowy programu jest ładowany do pamięci RCX i jest wykonywany pod kontrolą LEGO *firmware*.

Najbardziej zaawansowane możliwości programowania RCX zapewnia środowisko LegOS. Jest to pakiet procedur bibliotecznych, które umożliwiają tworzenie i uruchamianie programów pisanych w języku C++. Do kompilacji programów wykorzystywany jest standardowy kompilator GNU gcc dostępny w systemie Unix. Kod wynikowy programu jest ładowany do pamięci RCX i zastępuje LEGO *firmware*. To pozwala na wykorzystanie wszystkich możliwości wbudowanego mikrokontrolera oraz całej pamięci RAM na kod programu i zmienne. Tworzone programy mogą stosować złożone algorytmy sterowania wykorzystujące model globalny otoczenia, które działają zgodnie ze schematem: *percepcja otoczenia* → *modelowanie* → *planowanie* → *działanie*.

Uruchamianie i testowanie programów tworzonych w systemie LegoOS ułatwia pakiet procedur emulegOS, który po konsolidacji z programem wynikowym tworzy aplikację emulującą działanie mikrokontrolera RCX w systemie Unix.

4 Seminarium konstruowania i programowania robotów

Obserwując prace naukowe i wdrożeniowe nad DE, prowadzone na wielu zachodnich uczelniach, podjęliśmy decyzję o przygotowaniu podobnej formy zajęć dla studentów kierunku informatyka wydziału elektroniki. Pierwotnie zajęcia te miały się rozpocząć dopiero po formalnym włączeniu do obowiązujących siatek zajęć (w ramach modyfikacji programów pojawi się przedmiot „Informatyka systemów autonomicznych” prowadzony na zasadach DE). Jednakże, aby uniknąć ponad rocznego opóźnienia oraz zdobyć jakieś doświadczenie, zaproponowaliśmy studentom formułę otwartego seminarium odbywającego się nadobowiązkowo (na zasadach koła naukowego).

Seminarium trwało 9 tygodni i odbywało się w miesiącach listopad'99 ÷ styczeń'00. Rozpoczęło się wstępną prezentacją 26.10.99, na którą przyszło około 40 studentów.

Chęć aktywnego uczestnictwa zgłosiło 18 studentów (większość informatyków, kilku elektroników, 1 robotyk), których podzielono na 5 zespołów.

Przebieg seminarium

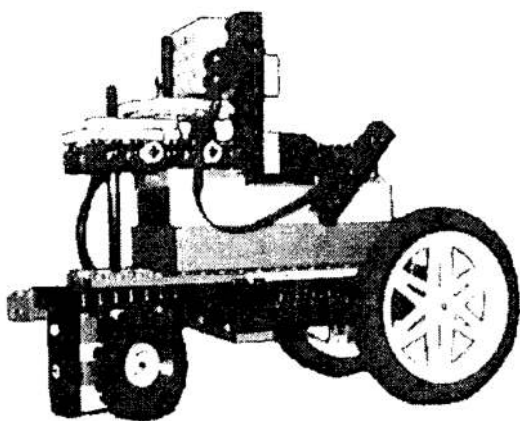
Spotkania zespołów odbywały się raz w tygodniu w godzinach popołudniowych oraz w wolne soboty. Typowe spotkanie trwało 2÷3 godziny zegarowe. Każda z grup otrzymała identyczne zadanie: w ciągu 2 miesięcy należy zbudować robota mobilnego, który będzie potrafił poruszać się w labiryncie przeszkód, rozpoznawać graficzne znaki terenowe i odnajdywać nadajnik promieniowania podczerwonego. Ograniczeniem było wykorzystanie tylko elementów z jednego zestawu LEGO *Robotic Invention System*. Opis szczegółowych zasad tego zadania umieszczony jest w sieci Internet [4].

Pierwsze spotkanie prowadzone było przez pracowników ICT, na następnych rola prowadzących sprowadzała się do kilkuminutowych dyskusji / konsultacji pozwalających ocenić i ukierunkować postęp prac projektowo-konstrukcyjnych. Tematyka, program pracy i przebieg kolejnych spotkań zależały głównie od studentów. W niektórych grupach występował bardzo wyraźny podział kompetencji na osoby zajmujące się oprogramowaniem i osoby zajmujące się mechaniką. Były też zespoły, które wszystko robiły wspólnie. Część zespołów prowadziła równoległe prace nad konstrukcją i oprogramowywaniem. Dwa zespoły najpierw zajęły się głównie konstruowaniem mechaniki, a dopiero pod koniec seminarium podjęły prace programistyczne (co przyczyniło się do tego, że zabrakło czasu na dopracowanie algorytmów).

Konkurs finałowy

Wzorem innych uczelni [8,9] seminarium zakończyło się publicznym konkursem, który odbył się 5 stycznia 2000 w jednej z sal seminaryjnych PWr. Poligonem testowania opracowanych robotów był stół o rozmiarach 2×2m., na którym znajdowały się przeszkody o rozmiarach 24×6cm oraz obudowany transponder nadający i odbierający sygnały wykorzystujący promieniowanie podczerwone. Zadaniem robota było:

- kierując się czujnikami oraz znakami graficznymi umieszczonymi na podłożu i na ścianach przeszkód omijać przeszkody znajdujące się stole,
- odnaleźć transponder promieniowania podczerwonego,
- usunąć mechaniczną zasłonę blokującą odbieranie sygnałów,
- nadać sygnał kodujący numer zespołu do programu „sędziego”, który nadzorował przebieg konkursu.



Rys. 3. Jedna z konstrukcji opracowanych na seminarium oraz przygotowania do konkursu
Fig. 3. One of the robots constructed by students and getting ready for the contest

Konkurencja odbywała się w dwóch etapach. Najpierw dla każdego zespołu osobno (tylko jeden robot na scenie), a na końcu dwie najlepsze konstrukcje jednocześnie (dwa roboty na scenie). Przed startem każdy z zespołów miał 5-cio minutową „konferencyjną” prezentację opracowanej konstrukcji oraz wykorzystywanych algorytmów. Osiągnięcia zespołów były oceniane w czterech kategoriach:

- praktyczna skuteczność opracowanego rozwiązania: czas zużyty na odnalezienie i odblokowanie transmitera,
- oryginalność stosowanych algorytmów programowych (pomysłowe wykorzystanie ograniczonych zasobów obliczeniowych klocka RCX, obliczeń współbieżnych, agregację danych pomiarowych w czasie i przestrzeni, itp),
- oryginalność konstrukcji robota (pomysłowe wykorzystanie klocków, silników, sensorów),
- jakość 5-minutowej publicznej prezentacji osiągnięć zespołu.

Szczegółową klasyfikację osiągnięć zespołów oraz galerię zdjęć dokumentującą przebieg konkursu można znaleźć na stronie WWW [4].

Osiągnięte cele dydaktyczne

Z merytorycznego punktu widzenia: uczestnicy seminarium zapoznali się z elementami różnych technik programowania systemów autonomicznych. Dyskutowano i testowano różne paradygmaty poczynając od tradycyjnego podejścia wykorzystującego modelowanie otoczenia oraz wnioskowanie wieloetapowe, poprzez programowanie behawioralne polegające na składaniu podstawowych zachowań robota zależnych od aktualnego stanu, po programowanie reaktywne polegające na wykonywaniu prostych odruchów pobudzanych przez sygnały pochodzące z sensorów.

Doświadczyli również inżynierskich problemów związanych z tworzeniem produktów działających w świecie rzeczywistym. Działanie robotów było w dużym stopniu uzależnione od różnych zjawisk naturalnych: od pogody (oświetlenia), od stanu baterii zasilających, od poślizgów układu jezdnego itp. Wykorzystywane algorytmy musiały się zatem adaptować do aktualnego stanu otoczenia. Dla części studentów nowością było wykorzystywanie programowania współbieżnego.

Zupełnie nowym doświadczeniem była samodzielna organizacja pracy zespołu (jak w małej firmie), planowanie harmonogramu czasowego prac oraz rozwiązywanie typowych problemów przemysłowych. Np. przeprowadzenie eksperymentów wymagało wcześniejszego zadbania o naładowanie akumulatorów a rozstrzygnięcie sporów o wzajemne „podkradanie” rozwiązań realizowano poprzez zgłaszanie patentów.

Popelnione błędy

W trakcie prowadzenia seminarium posiadaliśmy jedynie dwa zestawy RIS na pięć zespołów studenckich. W rezultacie konieczne było każdorazowe budowanie od nowa mechanicznej części robota (co jednej osobie zajmowało około 45 minut). Efektem tego była niepotrzebna strata czasu, unikanie bardziej złożonych konstrukcji mechanicznych oraz kilkukrotne problemy z odtworzeniem robota. Pewnym plusem była konieczność stosowania różnych form dokumentacji technicznej.

Zbyt późno (na dwa tygodnie przed konkursem) udostępniliśmy studentom szczegółowe parametry zadania konkursowego (rozmiary sceny, przeszkód, kolory, próbki materiałów, konstrukcję zasłony nadajnika). Również sam stół, na którym przeprowadzono konkurencje, został zmontowany dopiero tuż przed konkursem. W rezultacie nie było czasu na iteracyjne testowanie i udoskonalanie konstrukcji.

Ze względu na bardzo krótki czas trwania seminarium oraz i tak duże nadobowiązkowe zaangażowanie studentów, brak było wspólnych spotkań seminaryjnych. W efekcie prowadzący tracili czas na wielokrotne powtarzanie „wprowadzeń”, natomiast studenci nie angażowali się w przeglądanie oraz prezentacje literatury technicznej i naukowej.

Seminarium i konkurs '2001

Zachęcenii sukcesem i popularnością pierwszego seminarium będziemy kontynuować takie zajęcia w następnym roku. Tym razem seminarium będzie trwało przez cały zimowy semestr akademicki 2000/2001. W odróżnieniu do poprzedniego seminarium, oprócz cotygodniowych roboczych spotkań samych zespołów planujemy również miesięczne spotkania seminaryjne z krótkimi wykładami oraz prezentacją ciekawych materiałów wybranych przez studentów. Seminarium zakończy się publicznym konkursem w lutym 2001. Tym razem tematem projektu będzie skonstruowanie robota grającego w „piłkę siatkową”.

Podobnie jak w konkursie organizowanym przez Lawrence Technical University [8] podstawowym ograniczeniem będzie wykorzystanie klocka RCX jako jedynej jednostki obliczeniowej oraz 6 akumulatorów 1.5V jako jedynej źródła zasilania. Chcemy jednak rozszerzyć formułę zeszłorocznego seminarium o możliwość opracowywania i wykorzystywania własnych czujników oraz elementów wykonawczych.

5 Podsumowanie

Przeprowadzone seminarium w pełni potwierdziło skuteczność i atrakcyjność dydaktyki eksperymentalnej. Udało się pobudzić zaangażowanie studentów w realizację zadania konkursowego oraz aktywność w studiowaniu różnych technik informatycznych. Większość uczestników zadeklarowała swoje uczestnictwo w kolejnej edycji konkursu.

Pozytywnie oceniamy przydatność programowalnych klocków RCX do dydaktyki eksperymentalnej oraz różnorodnych laboratoriów z zakresu: informatyki, automatyki i robotyki. Do realizacji bardziej zaawansowanych projektów potrzebne jest jednak opracowanie klocka o większej mocy obliczeniowej lub wykorzystywanie kilku klocków RCX jednocześnie. Inną potrzebą jest rozszerzenie listy czujników i elementów wykonawczych, jakie można podłączyć do klocka RCX (np. matryca czujników dotykowych, czujniki zbliżeniowe, dalmierze, liczniki obrotów i przesunięć, itp.).

Proponujemy również zainicjowanie konkursu o zakresie międzyuczelnianym, co pozwoli uzupełnić zadanie projektowe o elementy „konkurencji i wolnego rynku”. Do konkursu '2001 zapraszamy zespoły z innych uczelni. Oferujemy pomoc i wymianę doświadczeń z ośrodkami akademickimi, które chciałyby prowadzić formy nauczania o charakterze dydaktyki eksperymentalnej lub uczestniczyć w kolejnej edycji konkursu. Zgłoszenia prosimy nadsyłać na adres naszego instytutu lub pocztą elektroniczną na adres: mindstorms@ict.pwr.wroc.pl.

Literatura

1. Cyr M., Miragila V. Nocera T. Rogers C.: *A Low-Cost, Innovative Methodology for Teaching Engineering Through Experimentation* Journal of Engineering Education, Vol. 86, No. 2
2. Miglino O., Lund H.H., Cardaci M.: *Robotics as an Educational Tool*. Journal of Interactive Learning Research, 10(1), pp. 25-47, 1999
3. Resnick M., Martin F., Sargent R., Silverman B.: *Programmable Bricks: Toys to Think With*. IBM Systems Journal, vol. 35, no. 3-4, pp. 443-452., 1996
4. <http://www.ict.pwr.wroc.pl/mindstorms> : strona domowa *Seminarium eksperymentalnego konstruowania i programowania robotów* na Politechnice Wrocławskiej
5. <http://el.www.media.mit.edu/projects/programmable-brick/> : strona domowa: *The MIT Programmable Brick Project*
6. <http://www.legomindstorms.com/> : strona domowa firmy LEGO oraz produktów z rodziny LEGO MindStorms

7. <http://mot.vuse.vanderbilt.edu/smc2000/> : strona domowa konferencji *IEEE System, Man, and Cybernetics 2000*.
8. <http://www.robofest.net/> : strona domowa: *Lawrence Robotics Festival 2000*
9. <http://www.media.mit.edu/mindfest/> : strona domowa *MindFest™, A Festival of Designing, Inventing, and Learning, by MIT Media Lab*
10. <http://itll.colorado.edu> : strona domowa *Integrated Teaching and Learning Laboratory*.

Streszczenie

Tematem referatu jest prezentacja doświadczeń z dydaktyką eksperymentalną na uczelni technicznej. W pierwszej części przedstawiono podstawowe idee i cechy takiej formy dydaktyki. Następnie omówiono programowalne klocki LEGO MindStorms jako potencjalne narzędzie do realizacji dydaktyki eksperymentalnej na kierunkach o charakterze elektroniczno-informatycznym. W dalszej części referatu opisano przebieg oraz problemy i wnioski z seminarium „Eksperymentalnego konstruowania i programowania robotów” jakie odbyło się na Politechnice Wrocławskiej.

LEARNING BY DOING WITH PROGRAMMABLE BRICKS LEGO MINDSTORMS™

Summary

The paper discusses the basic ideas and experience with problem-based learning at a technical university. The authors analyse potential use of programmable bricks produced by LEGO, in students projects covering the area of electronics and informatics. The discussion is illustrated by educational report from the seminar entitled „Experimental constructing and programming of autonomous robots” which took part at the Wrocław University of Technology.

Referat recenzowany

KOMPUTEROWA ANALIZA DZIAŁANIA ROBOTÓW W ZAWODACH RoboCup

1 Wstęp

Przeglądając literaturę naukową z ostatnich lat, można wyraźnie zaobserwować wzrost intensywności prac nad systemami wieloagentowymi. Zjawisko to spowodowane jest szybkim rozwojem sieci komputerowych (agentami są programy działające na stacjach lub serwerach), próbami komputerowo wspomaganego kierowania zespołami ludzkimi oraz przewidywaniami dalszego rozwoju automatyzacji i robotyki (agentami mogą być inteligentne budynki i systemy sterujące różnymi sprzętami gospodarstwa domowego). W pracach tych rozpatruje się zarówno podstawowy przypadek koegzystencji, jak i bardziej złożone problemy: koordynacji, kooperacji, konkurencji. Wszystkie te zagadnienia stają się szczególnie trudne w przypadku niepełności i/lub niepewności wykorzystywanej informacji oraz ograniczeń komunikacyjnych pomiędzy agentami, gdy nie jest możliwe opracowanie jednego centralnego systemu (algorytmu) sterowania.

Bardzo dobrym przykładem systemu wieloagentowego z elementami konkurencji i kooperacji są różnego rodzaju zespołowe gry sportowe [7]. Nic zatem dziwnego, że jako jedno z zadań testowych w pracach nad dziedziną systemów wieloagentowych wybrano grę w piłkę nożną, realizowaną za pomocą zespołu skomputeryzowanych zawodników [8]. Aby umożliwić porównanie rezultatów badań prowadzonych na różnych uczelniach, od pięciu lat odbywają się różnego rodzaju mistrzostwa piłki nożnej komputerów lub robotów określane najczęściej jako *RoboCup*. A od dwóch lat (jako odpowiedź na tragiczne trzęsienia ziemi) pojawiła się również „dyscyplina” *RoboCup Rescue* dotycząca analizy i planowania działań zespołów ratowniczych działających w sytuacjach klęsk żywiołowych i katastrof.

O tym jaka jest skala złożoności problemu *RoboCup* świadczy pobieżne wyliczenie różnego rodzaju decyzji, jakie w każdym momencie meczu musi podejmować zarówno rzeczywisty, jak i skomputeryzowany zawodnik [1]:

- **gdzie patrzeć:** czy na piłkę?, czy na boisko? (bramkę, linie boiska, flagi), czy śledzić ruch innych zawodników, a jeżeli tak, to których?
- **jak się poruszać:** czy stać w miejscu (oszczędzać energię), czy biegać?, jaki powinien być kierunek i tor ruchu?, czy iść wolno?, czy biec z maksymalną prędkością?
- **jak wpływać na ruch piłki:** czy kopnąć? (w którym momencie?, w jakim kierunku?, z jaką siłą?), czy zatrzymać?, czy dryblować?, czy podać do innego zawodnika? (którego?), czy strzelać w kierunku bramki?
- **jaką zająć pozycję** względem piłki i innych zawodników: czy śledzić ruch?, czy omijać? (z prawej czy z lewej?), czy zasłaniać / blokować?, czy od strony piłki, czy od strony innego zawodnika?

Większość dotychczasowych prac badawczych nad *RoboCup* dotyczyła głównie poszukiwania różnych form lokalnych algorytmów sterowania agentami, które zapewniłyby pracę całego systemu w czasie rzeczywistym. Szczególnie dobre rezultaty otrzymano po zastosowaniu różnego rodzaju sieci neuronowych oraz drzew decyzyjnych [2,6]. Cechą charakterystyczną takich rozwiązań jest wykorzystywanie architektur reaktywnych lub behawioralnych oraz założenie, że decyzje zawodników mogą być

* Instytut Cybernetyki Technicznej, Politechnika Wroclawska. E-mail: piasecki@ict.pwr.wroc.pl

** Instytut Cybernetyki Technicznej, Politechnika Wroclawska. E-mail: pawel@ict.pwr.wroc.pl

podejmowane jednokroково (na podstawie stanu boiska z ostatniego momentu gry podejmowany jest wybór jednej akcji wykonywanej w kolejnym kroku gry).

W odróżnieniu od tych prac, dziedziną naszych zainteresowań naukowych jest zbudowanie systemu, który wspomagałby budowanie wysokopoziomowych strategii gry pojedynczego zawodnika lub całego zespołu. Jako dodatkowe założenie przyjęliśmy, że system taki powinien pozwalać na budowanie algorytmów działania w przestrzeni pojęć wykorzystywanych przez piłkarzy-ludzi. W takim przypadku, proponowany system można by wykorzystać przynajmniej na cztery sposoby:

- a) ewaluacja: czy reguły dotyczące piłki nożnej ludzi sprawdzają się w przypadku piłki nożnej robotów,
- b) komputerowe wspomaganie ewaluacji i analizy stylów gry różnych zespołów piłki nożnej robotów lub ludzi,
- c) wstępna inicjacja algorytmu zawodnika RoboCup poprzez wykorzystanie wybranych i sprawdzonych poprzez (a) zasad gry podanych przez ludzi (piłkarzy, trenerów),
- d) automatyczne generowanie reguł gry RoboCup w postaci przyswajalnej dla człowieka (co umożliwi bezpośrednie sprawdzenie zgodności lub sprzeczności z intuicją).

W naszych pracach założyliśmy również, że w przypadku symulacyjnej ligi RoboCup, wiedzę taką da się wyekstrahować z zarejestrowanych nagrań dokumentujących przebieg wszystkich meczy, jakie były rozgrywane na wszystkich dotychczasowych zawodach (na podobnej zasadzie jak szachiści analizują przebieg różnych partii, a trenerzy piłkarscy wykorzystują nagrania video z różnych meczy przeciwników) [4,5].

Zanim przejdziemy do omówienia metod analizy plików rejestrowych, potrzebne jest ustalenie podstawowych pojęć opisu stanu boiska i działania zawodników.

2 Definicje przestrzeni konfiguracji i ruchów

W symulacyjnej lidze RoboCup czas meczu jest zdyskretyzowany [1]. Kolejne chwile t są wyznaczone przez kolejne kroki symulacji realizowanej przez serwer. Opis stanu gry w wybranym momencie t powinien uwzględniać stan wszystkich obiektów na boisku:

- parametry boiska p_b – stałe (lub quasi-statyczne) składowe wpływające na rozgrywany mecz (np. rozmiary boiska i bramek, maksymalne przyśpieszenia zawodników, nazwy drużyn, itp.),
- konfiguracja piłki $k_p(t)$ – położenie, prędkość i kierunek ruchu,
- konfiguracja zawodnika $k_z^i(t)$ – położenie, prędkość i kierunek ruchu, zmęczenie,
- konfiguracje innych obiektów – sędziego $k_s(t)$ (np. czas do końca meczu, liczniki strzelonych goli), siła i kierunek wiatru $k_w(t)$.

Globalna konfiguracja boiska (w skrócie konfiguracja $k(t)$) jest to wektor w/w konfiguracji częściowych:

$$k(t) = (p_b, k_p(t), (k_z^i(t) \ i=1,2,\dots,22), k_s(t), k_w(t)).$$

Zbiór wszystkich konfiguracji, które mogą zaistnieć w czasie dowolnego meczu tworzy globalną przestrzeń konfiguracji K .

Aktywnymi obiektami, które wpływają na zmiany konfiguracji boiska są: zawodnicy, sędzia oraz serwer (modelujący losowe zakłócenia ruchu piłki i zawodników). Można wyróżnić następujące typy takich oddziaływań:

- $r_i(t)$ – elementarne ruchy i -tego zawodnika odpowiadające bezpośrednio podstawowym komendom ruchu: *dash*, *turn*, *kick*, *catch*, *turn_neck*,
- $s(t)$ – decyzje sędziego (które są ogłaszane jako komendy: *kick_off*, *play_on*, itp.),
- $u(t)$ – wektor zakłóceń ruchu piłki i zawodników modelujący: wiatr, poślizgi zawodników, zderzenia obiektów, nierówności terenu itp.

Aby uprościć zapis, oznaczmy przez $r(t)$ wektor elementarnych posunięć wszystkich zawodników $r(t) = (r_1(t), r_2(t), \dots, r_{22}(t))$. Zbiór wszystkich możliwych wektorów $r(t)$ tworzy globalną przestrzeń ruchów R .

Zmiana konfiguracji $k(t)$ po wykonaniu ruchu $r(t)$ jest określona pewną zależnością funkcyjną, która jest realizowana przez serwer:

$$k(t+1) = f(k(t), r(t), u(t)).$$

Każdy rozegrany mecz jest ciągiem kolejnych konfiguracji $k(t)$, $t = 0, 1, 2, \dots, N$, które będziemy nazywać trajektorią konfiguracji. Wycinek trajektorii konfiguracji od chwili t do chwili $t+\Delta t$ będziemy nazywać śladem i oznaczać $\gamma\{t, \Delta t\}$. Zbiór wszystkich śladów tworzy przestrzeń śladów Γ .

Podobnie ciąg kolejnych ruchów zawodników $r(t)$ $t=0, 1, 2, \dots, N$ będziemy nazywać trajektorią ruchów. Wycinek trajektorii ruchów od chwili t do chwili $t+\Delta t$ będziemy nazywać akcją $a\{t, \Delta t\}$. Zbiór wszystkich akcji tworzy przestrzeń akcji A .

Nową konfigurację $k(t+\Delta t)$, powstającą po wykonaniu akcji $a\{t, \Delta t\}$ w konfiguracji $k(t)$, można wyznaczyć przez Δt -krotne złożenie funkcji serwera f :

$$k(t+dt) = f(\dots f(f(k(t), r(t), u(t)), r(t+1), u(t+1)) \dots, r(t+dt), u(t+dt)),$$

gdzie $r(t)$, $r(t+1)$, $r(t+2)$... $r(t+\Delta t)$ są kolejnymi ruchami akcji $a\{t, \Delta t\}$. Takie złożenie będziemy w skrócie oznaczać $f(k(t), a\{t, \Delta t\})$.

3 Podstawowe klasy algorytmów sterowania zawodnikami

Algorytmy sterowania zawodników definiują odwzorowanie globalnej przestrzeni konfiguracji K w przestrzeń ruchów R . Odwzorowania takie mogą być tworzone różnymi metodami.

1. Proste jednoetapowe algorytmy sterowania zakładają jednoznaczne odwzorowanie bieżącej konfiguracji $k(t)$ w ruch $r(t)$, które można zapisać w postaci funkcji $r(t) = g(k(t))$. W najprostszym przypadku definiuje się zestaw reguł postaci: jeżeli x to y , gdzie x jest warunkiem sprawdzającym przynależność $k(t)$ do pewnego podzbioru K , natomiast y jest punktem przestrzeni R . Odwzorowania g można również budować za pomocą odpowiednio wyuczonej sieci neuronowej lub drzewa decyzyjnego [6].
2. Bardziej złożone algorytmy sterowania zakładają istnienie kryterium oceny „jakości” konfiguracji $O(k)$. Planują one sekwencję optymalnych ruchów, czyli akcję $a^*\{t, \Delta t\}$, prowadzącą do konfiguracji $k^*(t+\Delta t)$, która maksymalizuje kryterium O . Metody te zakładają również znajomość funkcji prognozującej $f'(k(t), a\{t, \Delta t\})$, pozwalającej estymować konfigurację boiska $k'(t+\Delta t)$ spełniającą warunek:

$$k'(t+\Delta t) \approx k(t+\Delta t) \text{ gdzie } k(t+\Delta t) = f(k(t), a\{t, \Delta t\}).$$

Możliwe są dwie metody wyznaczenia optymalnej akcji $a^*\{t, \Delta t\}$:

- Pierwsza metoda polega na bezpośrednim przeszukiwaniu przestrzeni akcji A celem znalezienia rozwiązania $a^*\{t, \Delta t\} \in A$ takiego, że:

$$O(f'(k(t), a^*\{t, \Delta t\})) = \max_{a \in A} O(f'(k(t), a\{t, \Delta t\})).$$

- Druga metoda zakłada istnienie miary odległości $diff(k_1, k_2)$, pomiędzy dowolnymi konfiguracjami $k_1, k_2 \in K$, pozwalającej oszacować ilość ruchów potrzebnych do przejścia od k_1 do k_2 . Metoda ta polega na przeszukiwaniu fragmentu przestrzeni konfiguracji $K^o(k(t), \Delta t) \subset K$, zawierającego konfiguracje $k(t+\Delta t)$ z bezpośredniego otoczenia $k(t)$ spełniające warunek odległości:

$$diff(k(t), k(t+\Delta t)) \leq \Delta t.$$

Celem tego przeszukiwania jest znalezienie konfiguracji optymalnej $k^*(t+\Delta t)$ maksymalizującej kryterium jakości O :

$$O(k^*(t+\Delta t)) = \max_{k(t+\Delta t) \in K^o(k(t), \Delta t)} O(k(t+\Delta t)).$$

Następnie w przestrzeni A wyszukuje się taką akcję $a^*\{t, \Delta t\}$, która realizuje przejście od $k(t)$ do $k^*(t+\Delta t)$ w minimalnej ilości kroków.

Ze względu na rozmiary przestrzeni konfiguracji oraz ruchów planowanie jest trudne do realizacji w czasie rzeczywistym. Ponadto wymaga skonstruowania metody oceny konfiguracji O , metody predykcji kolejnych konfiguracji f' oraz metody oceniającej podobieństwa konfiguracji $diff$.

3. Zupełnie odmiennym sposobem syntezy algorytmu zawodnika jest próba przeniesienia odwzorowań $K \rightarrow R$ w odpowiednie przestrzenie abstrakcyjnych pojęć stosowanych przez piłkarzy-ludzi. Wówczas cały algorytm działania zawodnika mógłby być wyrażony za pomocą reguł lingwistycznych możliwych do bezpośredniego zinterpretowania przez człowieka. Załóżmy, że konfiguracja boiska będzie opisywana jako wektor cech $m(t) = (m_1, m_2, \dots, m_n)$ zdefiniowanych przez ludzi (np. odległość piłki od bramki, odległość najbliższego zawodnika od piłki, liczbę zawodników w określonym sektorze boiska, rozmiar wolnej przestrzeni wokół zawodnika itp.). Zbiór wszystkich możliwych wektorów $m(t)$ tworzy przestrzeń modelu M .

Algorytm działania zawodnika buduje się jako odwzorowanie $h: M \rightarrow A$ przestrzeni modelu M w określoną przestrzeń A . Odwzorowanie h może być tworzone w postaci bazy reguł W kreowanej przez trenera (aktualnie człowieka).

W tym przypadku ruch $r(t)$ jest generowany w następujących krokach:

- transformacja konfiguracji $k(t)$ w wektor cech $m(t)$,
- wybór akcji $a\{t, d\Delta t\}$ na podstawie bazy reguł W ,
- rozkład akcji $a\{t, d\Delta t\}$ na ciąg posunięć zawodników $r(t), r(t+1), \dots, r(t+\Delta t)$,
- realizacja kolejnych sekwencji ruchów r w całości lub częściowo. W skrajnym przypadku realizowany jest tylko pierwszy element akcji $r(t)$, a cykl kreacji nowego sterowania jest powtarzany w każdej kolejnej chwili t .

Podstawowym problemem przy budowie takiego algorytmu sterowania jest wybór cech tworzących przestrzeń modelu M oraz budowa odwzorowania h . Obecnie wybór ten jest dokonywany arbitralnie przez ludzi (twórców drużyny RoboCup) w oparciu o znajomość reguł gry, doświadczenie nabyte w czasie dotychczasowych prac oraz intuicje (analogie do gry ludzi w piłkę). Nasze eksperymenty pokazują, że równie cennym źródłem informacji może być analiza rozegranych meczy.

4 Zawartość plików rejestrowych meczy RoboCup

Aby umożliwić późniejszą analizę rozgrywanych meczy program serwera dokumentuje wybrane elementy konfiguracji boiska k w pliku rejestrowym meczu (*ang. logfile*). W dalszej części tego referatu pliki te będziemy oznaczać jako LOG. Typowy plik LOG składa się z 3000 rekordów l dokumentujących stany boiska w kolejnych krokach symulacji jednej połowy meczu (czas trwania połowy meczu = 3000 kroków). Rekordy l zawierają następujące dane:

- *time* – czas jaki upłynął od początku meczu (nr kroku symulacji),
- *pmode* – tryb gry ustalany przez sędziego: *kick_off_right*, *play_on*, *goal_left*, itd.,
- *name* – nazwy „lewego” i „prawego” zespołu,
- *score* – ilości bramek zdobytych przez „lewy” i „prawy” zespół,
- (x, y) – kartezyjskie pozycje i kąty orientacji (*angle*) 22 graczy oraz piłki,
- dla każdego zawodników dodatkowo przechowywana jest przynależność do zespołu (*side*) oraz numer „koszulki” zawodnika (*unum*).

Niestety, rekordy I nie zawierają wszystkich cech wektora konfiguracji boiska k . W szczególności nie zawierają maksymalnych oraz aktualnych prędkości i przyspieszeń ruchu zawodników lub piłki. Wielkości te trzeba rekonstruować poprzez różnicowanie wartości pozycji $(x, y, angle)$ z kolejnych następujących po sobie rekordów I .

Innym ograniczeniem jest brak w rekordzie I informacji o ruchach lub akcjach wykonywanych przez zawodników oraz o wielkości zakłóceń wprowadzanych przez serwer. Efektem tego jest niemożność dokładnej rekonstrukcji wektorów ruchu r , jakim poddawane były obiekty piłki i zawodników. Np. nie można dokładnie zidentyfikować, co jest przyczyną zakłóceń regularności ruchu piłki: czy jest to efekt zakłóceń ruchu modelowanych przez serwer, czy też efekt kopnięcia przez zawodnika. Szczególnie trudna sytuacja występuje w przypadku gdy, w pobliżu piłki znajduje się kilku zawodników. Nie znając wektora r nie można również zidentyfikować: kto kopnął piłkę, w jakim kierunku i z jaką mocą. A przecież te wartości są podstawowymi decyzjami podejmowanymi przez zawodników.

Podsumowując, aby umożliwić wykorzystanie plików LOG do konstruowania algorytmów gry konieczne jest opracowanie metody rekonstrukcji wektorów konfiguracji k i ruchu r na podstawie sekwencji rekordów I .

Warto w tym miejscu zauważyć, że analogiczne pliki LOG mogą być generowane przez systemy wizyjne wykorzystywane w ligach małych robotów, gdzie kamera zawieszona bezpośrednio nad boiskiem dostarcza obraz umożliwiający taką dwuwymiarową rekonstrukcję stanu boiska. Możliwe jest także zbudowanie podobnego systemu wizyjnego do nagrywania w plikach LOG przebiegu meczy rozgrywanych przez ludzi.

5 Statystyczna analiza plików rejestrowych LOG

Statystyczna analiza plików LOG musi być przeprowadzona w kilku etapach. W pierwszym etapie konieczne jest wyznaczenie wektora stałych parametrów boiska, piłki i zawodników oznaczonego wcześniej jako p_b :

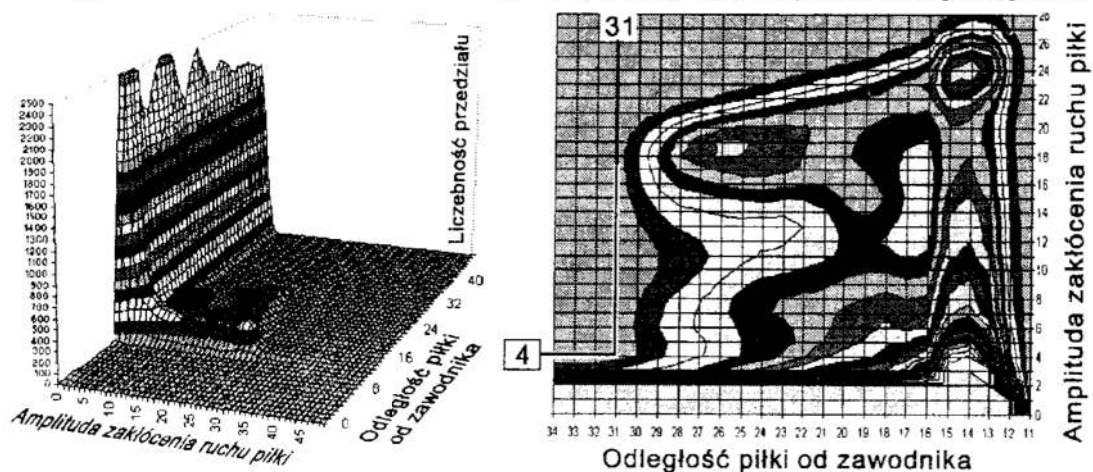
1. Wyznaczenie maksymalnych prędkości i przyspieszenia (liniowych i kątowych) ruchu zawodników i piłki.

W obu przypadkach wartości prędkości i przyspieszeń rekonstruowano na podstawie równań różnicowych 1 i 2 rzędu wykorzystujących sekwencje pozycji $(x, y, angle)$ z kolejnych rekordów I . Pierwotnie jako wartości ekstremalne przyjmowano maksimum z całego meczu. Okazało się jednak, że w trakcie meczu występują szczególne sytuacje, gdy zawodnicy są nagle „przestawiani” przez sędziego na pozycje zgodne z przepisami. Efektem było generowanie bardzo dużych wielkości parametrów. Niektóre z tych przypadków (np. powrót zawodników po strzeleniu bramki „na swoją połowę”) można zidentyfikować na podstawie trybu gry $pmode$ zadawanego przez sędziego. Żeby uniknąć pozostałych zakłóceń wprowadzono dyskretyzację estymowanych parametrów oraz zliczanie liczebności wystąpień w poszczególnych przedziałach (histogram). Jako wartość ekstremalną parametru przyjmowano środek ostatniego przedziału dyskretyzacji, którego liczebność przekraczała próg 1% liczebności wszystkich przedziałów.

2. Wyznaczenie maksymalnej wielkości (amplitudy) zakłócenia ruchu piłki spowodowanego szumem modelowanym przez serwer.

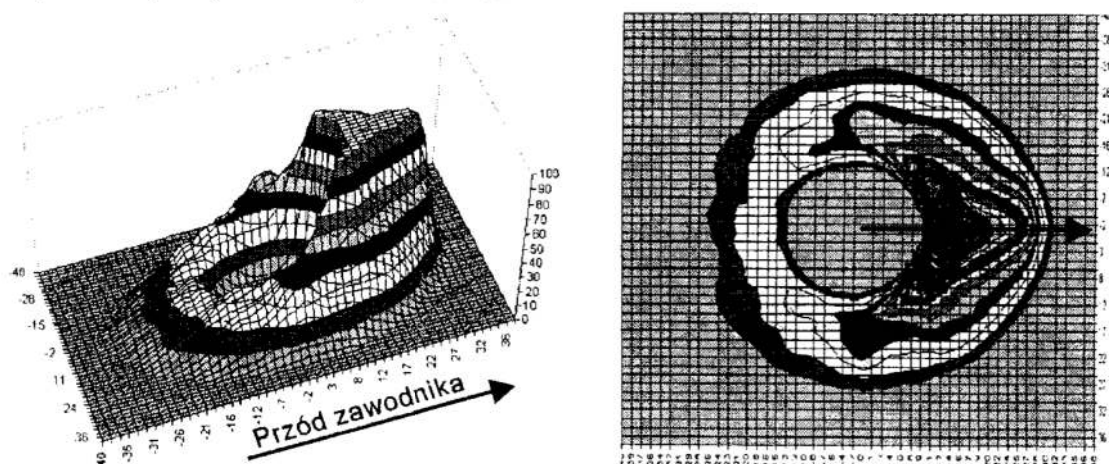
Wykorzystując parametry wyznaczone w kroku (1) można zbudować równania modelujące jednostajnie opóźniony samodzielny ruch piłki po boisku. Stosując takie równanie można prognozować pozycję piłki po jednym cyklu symulacji. Różnica pomiędzy pozycją prognozowaną a rzeczywistą (odczytaną z kolejnego rekordu I) stanowi zakłócenie regularności ruchu piłki. To zakłócenie najczęściej jest powodowane przez serwer modelujący nierówności boiska. Oprócz wpływu serwera, ruch piłki zakłócany jest „kopnięciem piłki” przez zawodnika oraz incydentalnymi przestawieniami piłki przez sędziego. Ponieważ rekordy I nie zawierają informacji

o takich ruchach, konieczne było połączenie zbierania danych o amplitudzie zakłócenia z odległością od najbliższego zawodnika. Aby odfiltrować incydentalny wpływ sędziego, podobnie jak w etapie (1), posłużono się dyskretyzacją estymowanych parametrów oraz zliczaniem liczebności przedziałów. Dwuwymiarowy histogram ilustrujący liczebności przedziałów przedstawiono na rysunku 1. Po odrzuceniu przedziałów, których liczebność nie przekraczała 5% maksymalnej liczebności, oszacowano maksymalną amplitudę zakłóceń generowanych przez serwer jako 4. Z tego samego histogramu oszacowano również maksymalny zasięg wpływu piłkarza na piłkę jako 31.



Rys. 1. Histogram zależności zakłócenia ruchu piłki względem odległości od zawodnika
 Fig. 1. Histogram: occurrence of ball disturbance relative to ball-player distance

- Wyznaczenie wielkości piłki i zawodnika. Na histogramie z rysunku 1, dla odległości piłki od zawodnika o wartościach mniejszych od 11 liczebności przedziałów są równe zero. Przyczyną tego jest niemożność nachodzenia się obszarów piłki i zawodnika (kontrolowana przez serwer). Po zbudowaniu podobnego histogramu dla minimalnej odległości pomiędzy dwoma dowolnymi zawodnikami można wyznaczyć promień obszaru zajmowanego przez zawodnika (oszacowano jako 8), a następnie oszacować promień piłki jako różnicę: $11 - 8 = 3$.



Rys. 2. Histogram wystąpień zakłóceń ruchu piłki względem zawodnika (pozycje kopnięć)
 Fig. 2. Histogram: positions of ball disturbance relative to centre of player

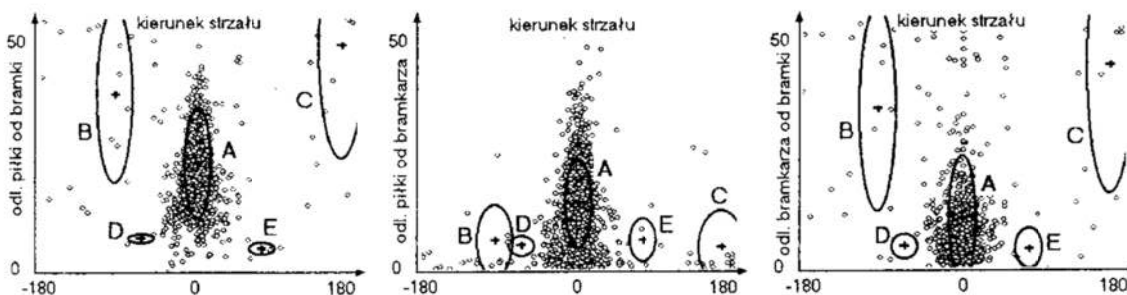
- Wyznaczenie obszaru wpływu zawodnika na piłkę.
 Z histogramu na rysunku 1 oszacowano, że maksymalny zakres wpływu zawodnika na ruch piłki wynosi 31. Interesującym parametrem jest również kształt obszaru wpływu

zawodnika na ruch piłki. Na rysunku 2 przedstawiono histogram ilustrujący liczebność wystąpień zakłócenia ruchu piłki w zależności od pozycji piłki względem najbliższego zawodnika. Żeby zwiększyć czytelność rysunku, z histogramu po lewej stronie usunięto bardzo wysoki wierzchołek o szczycie (liczebność 650) w punkcie (14,0). Wierzchołek ten wyraźnie pokazuje, że zawodnicy najczęściej kopią piłkę, która znajduje się tuż przed nimi, natomiast najrzadziej gdy piłka znajduje się z tyłu. W przypadku symulacyjnej ligi robotów (gdzie zawodnicy są „symetryczni”) wyniki takiej analizy są oczywiste. Natomiast bardzo ciekawe wyniki można by uzyskać stosując podobną technikę do analizowania gry piłkarzy-ludzi. W takim przypadku, niesymetryczności histogramu pozwolą zidentyfikować, którą nogą najczęściej kopie zawodnik, jakie są różnice w maksymalnym zasięgu obu nóg i jakie są siły uderzenia. Wiedzę taką można potem wykorzystać podczas tworzenia algorytmów działania w „pojedynekach”.

6 Kategoryzacja zagrań zawodników

Statystyczna rekonstrukcja podstawowych parametrów boiska oraz zakresu oddziaływania zawodników na piłkę pozwala na odtworzenie wektorów ruchu $r(t)$. Analiza trajektorii ruchów i odpowiadającej jej trajektorii konfiguracji pozwala na identyfikację podstawowych zagrań (np. podanie, strzał, obrona itp.). Sposób identyfikacji zostanie przedstawiony na przykładzie analizy gry strzelca i bramkarza w „sytuacjach podbramkowych”. Strzał poprzedzający zdobycie bramki jest identyfikowany jako akcja $a(t, \Delta t)$ i odpowiadająca jej trajektorii konfiguracji $k(t), \dots, k(t+\Delta t)$ taka, że pierwszy wektor ruchu $r(t)$ zawiera składową „kick”, pozostałe wektory $r(t+1), \dots, r(t+\Delta t)$ nie zawierają składowej „kick”, natomiast w ostatniej konfiguracji $k(t+\Delta t)$ nastąpiła zmiana wyniku meczu. Podobnie można zidentyfikować strzały, po których nastąpiła udana interwencja bramkarza.

Po przeanalizowaniu wszystkich plików LOG zawierających zapis meczy rozegranych na mistrzostwach świata RoboCup '98, wydzielono 322 skuteczne strzały na bramkę oraz 559 udanych interwencji bramkarza. Tak uzyskane sytuacje można kategoryzować za pomocą algorytmów analizy skupień [3]. W tym celu konfiguracje $k(t)$ opisujące stan boiska w chwili strzału przetransformowano do przestrzeni modelu M . Jako wektor charakterystycznych cech przyjęto następujące parametry: kierunek strzału, odległość piłki od bramki, orientacja bramkarza względem bramki, odległość bramkarza od bramki, odległość bramkarza od piłki, orientacja bramkarza względem piłki, skuteczność strzału. Następnie dokonano grupowania w/w wektorów cech za pomocą centroidalnej metody aglomeracji skupień [3]. Jako kryterium odległości przyjęto ważoną odległość euklidesową, w której cechą dominującą był kierunek strzału. W wyniku grupowania wydzielono 5 grup, które można zinterpretować jako: A – strzały na wprost, B (C) – strzały z lewego (prawego) skrzydła, E (F) – strzały z pobliża bramki po lewej (prawej) stronie. Liczebność poszczególnych grup była następująca (w nawiasach podano liczbę strzelonych bramek): A–835 (318), B–24 (0), C–16 (0), D–2(2), E–4 (4).



Rys. 3. Ilustracja wyników grupowania wektorów cech dla sytuacji podbramkowych
 Fig. 3. Illustration of clustering results for situations close to goal area

Punkty na rysunku 3 ilustrują wybrane współrzędne wektora cech, natomiast elipsoidy ilustrują położenie centroidu grupy i wariancję cech wewnątrz grupy.

7 Podsumowanie

Zaproponowana metoda kategoryzacji pozwala na wydzielanie typowych „zagrań zawodników” opisanych w kategoriach pojęć stosowanych przez ludzi i kojarzenie ich z wektorami cech tworzących przestrzeń modelu. Uzyskane w ten sposób wyniki mogą być wykorzystywane przez trenera (twórcę drużyny) podczas tworzenia bazy reguł W definiujących algorytm sterowania zawodnika. Jeżeli znane będą oceny rezultatów tych typowych zagrań, to baza reguł W może posłużyć również jako podstawa skonstruowania funkcji ewaluacji O oceniającej „jakość” dowolnej konfiguracji boiska. Istnienie takiej funkcji warunkuje stosowanie złożonych algorytmów planowania działań (patrz rozdział 3). Zachęteni wstępnymi rezultatami, planujemy dalsze prace nad automatyzacją analizy plików rejestrowych meczy pod kątem oceny gry zespołowej, a nie tylko zagrań pojedynczych zawodników.

Literatura

1. Hiroaki K. Et all: *The RoboCup Synthetic Agent Challenge 97*. Conference proceedings of IJCAI-97, 1997.
2. Hiroaki K.: *RoboCup-97: Robot Soccer World Cup I*. Springer-Verlag, June 1998.
3. Kucharczyk J.: *Algorytmy analizy skupień w języku ALGOL 60*. PWN, W-wa 1982.
4. Letia I.A., et all: *State/Action Behavioral CClassifiers for Simulated Soccer Players*. Proceed. of the second RoboCup Workshop, 1998
5. Takahashi T., Naruse T.: *From Play Recognition to Good Plays Detection*, Proceed. Procc. of the second RoboCup Workshop, 1998
6. Veloso M., Stone P.: *A layered Approach to Learning Client Behaviours in the RoboCup Soccer Server*. Applied Artificial Intelligence (AAI), Volume 12, 1998.
7. Veloso M., Stone P., Han K.: *The CMUnited-97 Robotic Soccer Team: Perception and Multiagent Control*. Proc. of Second International Conference on Autonomous Agents (Agents '98) Minneapolis/St. Paul, 1998.
8. <http://www.robocup.org>

Streszczenie

Przedmiotem referatu jest poszukiwanie metod pozyskiwania wiedzy niezbędnej do planowania i realizacji działań zawodników RoboCup. Jako podstawową metodę zdobywania takiej wiedzy zaproponowano analizę przebiegu wcześniej rozgrywanych meczy. Uzyskane w ten sposób dane wykorzystano następnie do identyfikacji i wydzielenia elementarnych zachowań zawodników. Zaproponowano również metodę grupowania i oceny wydzielonych zachowań w oparciu o algorytmy analizy skupień.

COMPUTER ANALYSIS OF ROBOT'S BEHAVIOUR IN RoboCup COMPETITION

Summary

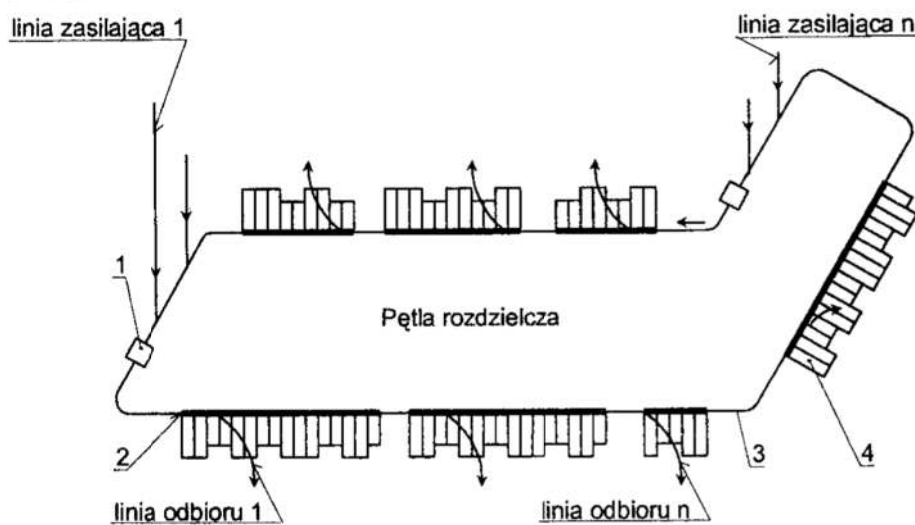
The paper presents analysis of RoboCup logfiles as a basic method for acquiring domain specific knowledge about simulated soccer. Simple statistical analysis was used to estimate basic models of player-ball interactions. Cluster analysis method was proposed for extracting typical player actions.

Referat recenzowany

DYNAMICZNY MODEL PROCESU ZGARNIANIA JAKO OBIEKT OPTYMALIZACJI ROZDZIAŁU STRUMIENIA ŁADUNKÓW JEDNOSTKOWYCH

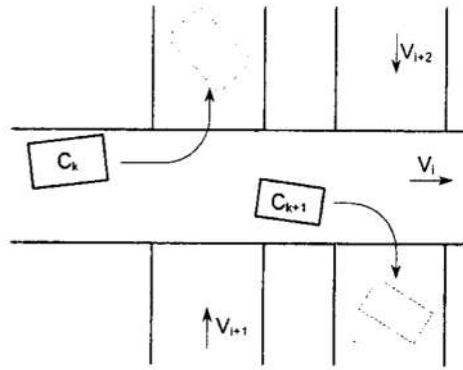
1 Identyfikacja potrzeby

Problem selekcji i podziału towarów w formie ładunków jednostkowych (paczek) istnieje tam, gdzie jest ich duża koncentracja, np.: w węzłowych urządzeniach pocztowych, portach lotniczych, domach sprzedaży wysyłkowej, w zakładach przemysłowych wielu branż. Zautomatyzowany rozdział strumienia ładunków na kierunki zgodnie z przeznaczeniem realizowany jest za pośrednictwem rozdzielaczy – elementów wykonawczych maszyn rozdzielczych wplecionych w system przenośników dosyłający – odbiorczych (rys.1). Podział potoku ładunków odbywa się na podstawie identyfikacji cech charakterystycznych wyróżniających poszczególne ładunki (rys.2). Każdemu rozpoznanemu przez system detekcji ładunkowi przypisywany jest odpowiedni kierunek dalszego transportowania. Na podstawie tych cech można np. wydzielić z ogólnej masy towarów ładunki o niepożądanych zawartościach (zawierające broń, materiały wybuchowe, materiały radioaktywne, narkotyki), dokonać selekcji towarów (wg masy, wymiarów zewnętrznych), podzielić je wg kodów adresowych (pocztowych, towarowych, graficznych, magnetycznych).



Rys.1. Linia automatycznego sortowania ładunków: 1 – stanowisko odczytu cech ładunków, 2 – maszyna rozdzielcza, 3 – przenośnik, 4 – ześlizgi

Fig.1. Automatic sorting line: 1 – reading stand of the load feature, 2 – separating machine, 3 – conveyor, 4 – skids



Rys. 2. Rozdział strumienia ładunków: C_k – cecha charakterystyczna ładunku, V_i – prędkość liniowa

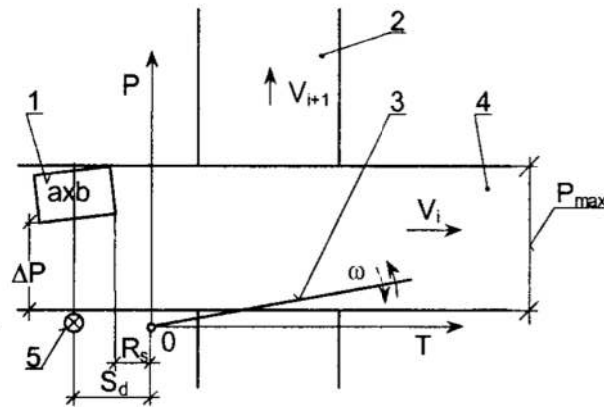
Fig. 2. Separation of the load stream: C_k – characteristic features of the load, V_i – linear speed

2 Możliwości realizacji procesu sortowania

O niezawodności całego systemu decyduje niezawodność jego najsłabszego elementu. Ogniwem najbardziej skomplikowanym i wykonującym najtrudniejsze zadanie w procesie sortowania ładunków jest maszyna rozdzielcza, a zwłaszcza jej elementy wykonawcze – rozdzielacze. Ich praca w zdecydowany sposób wpływa na ogólne zachowanie się systemu. Rozdzielanie ładunków może być prowadzone na wiele sposobów. Wszystkie łączy wspólna cecha. W celu rozsortowania potoku ładunków na kierunki konieczne jest przekazanie poszczególnym ładunkom impulsu siły (momentu) we właściwy sposób i w odpowiednim momencie, powodując zmianę ich kierunku przemieszczania. Ten impuls siły może być uzyskiwany różnymi metodami. Jedną z nich jest transportowanie ładunków na wózkach prowadzonych w torze, których powierzchnie są uchylnymi tacami albo poprzecznymi przenośnikami taśmowymi. Innym sposobem rozdzielania jest rozmieszczenie wzdłuż przenośnika taśmowego, transportującego rozdzielane ładunki, urządzeń zgarniających. Urządzenia te mogą być wyposażone w ramiona, które są wprawiane w ruch obrotowy, wahadłowy lub postępowy poprzeczny albo skośny. Zmiana kierunku przemieszczania ładunku może być także powodowana w skutek oddziaływania sprężonego powietrza.

3 Model fizyczny procesu zgarniania

Głównym elementem maszyny (pracującej na zasadzie zgarniania, rys.3) jest taśmociąg wzdłuż którego rozmieszczone są przemiennie po obu stronach urządzenia zgarniające. Zgarnięcie paczki 1 następuje w momencie jej wejścia do strefy pracy odpowiedniego urządzenia zgarniającego, tzn. po przecięciu bariery świetlnej 5. Podczas otwierania się ramienia zgarniaka 3, paczka zostaje zgarnięta do ześlizgu 2 umieszczonego po przeciwnej stronie przenośnika rozdzielczego 4 względem zamocowania zgarniaka.



Rys.3. Model fizyczny procesu zgarniania zgarniakiem wychyłowym: 1 – zgarniany ładunek (o wymiarach axb), 2 i 4 – przenośniki, 3 – ramię zgarniaka, 5 – bariera świetlna, R_s – położenie ładunku w chwili zadziałania zgarniaka, S_d – położenie bariery świetlnej, ΔP – odległość paczki od brzegu przenośnika, V_i – prędkość liniowa

Fig.3. Physical model of the scraping process with a deflection scraper: 1 – scraped load (dimension axb), 2 i 4 – conveyors, 3 – scraper's arm, 5 – light barrier, R_s – position of the load at the moment when scraper's arm begin working, S_d – position of the light barrier, ΔP – distance of the load from border of the conveyor, V_i – linear speed

Rozwiązanie zagadnienia optymalizacji procesu zgarniania podjęto w oparciu o analizę teoretyczną, budując odpowiedni model numeryczny i przeprowadzając symulację działania tego modelu.

Model fizyczny procesu zgarniania uwzględnia występowanie następujących podstawowych zjawisk fizycznych: uderzenie ukośne ciał szorstkich, tarcie ślizgowe suche, ruch płaski paczki i ruch obrotowy ramienia zgarniaka. W modelowaniu przyjęto uproszczenia, zakładając, że nie będą one miały istotnego wpływu na wnioskowanie z wyników symulacji modelu numerycznego:

- paczka traktowana jest jako płaskie ciało sztywne o równomiernie rozłożonej masie,
- powierzchnie zetknięcia paczki z elementami maszyny rozdzielczej są także nieodkształcalne, tzn. siła normalna nacisku paczki ani towarzysząca jej siła tarcia nie wywołują odkształcenia sprężystego lub odkształceń plastycznych w okolicy zetknięcia,
- założono jednorodne właściwości cierne na całej powierzchni taśmy przenośnika i zgarnianej paczki,
- pominięto zjawisko precesji,
- pominięto oddziaływanie oporu powietrza na ruch paczki.

W ciągłym procesie zgarniania paczek na maszynie rozdzielczej można wyodrębnić dyskretne stany paczki o cechach charakterystycznych. Wynikają one z wzajemnego położenia i oddziaływania zgarniaka, paczki i przenośnika maszyny rozdzielczej. Wyróżnionymi dyskretnymi stanami fizycznego modelu procesu zgarniania są:

- F1 uderzenie ukośne narożem paczki o zgarniak – pojawienie się największych oddziaływań dynamicznych paczki w całym procesie zgarniania,
- F2 ruch płaski paczki przy odskoku po udarze – poślizg paczki względem taśmy przenośnika rozdzielczego, paczka poza taśmą przenośnika nie jest w kontakcie z żadnym innym elementem maszyny rozdzielczej,

- F3 powtórne ukośne uderzenie o zgarniak – bez odskoku paczki od zgarniaka, oddziaływanie dynamiczne mniejsze niż przy pierwszym uderzeniu,
- F4 obrót paczki wokół naroża ocierającego się o ramię zgarniaka – paczka porusza się na taśmie przenośnika rozdzielczego ruchem płaskim, jednocześnie jeden z narożników paczki jest w ciągłym kontakcie z ramieniem zgarniaka,
- F5 ruch paczki wzdłuż ramienia zgarniaka – paczka bokiem ociera się o ramię zgarniaka i przemieszcza po nim na przeciwległą stronę taśmociągu względem zamocowania urządzenia zgarniającego,
- F6 zgarnianie paczki ułożonej najbliżej osi obrotu zgarniaka, nie mieszczącej się na ramieniu – paczka na taśmie przenośnika rozdzielczego wykonuje ruch płaski, jednocześnie jednym narożnikiem przemieszcza się po burcie taśmociągu, a drugim – po ramieniu zgarniaka.

4 Warunki optymalizacji

Optymalizacja dotyczy przede wszystkim poszukiwania nastawów procesu zgarniania (prędkości taśmy V , czasu cyklu pracy zgarniaka t_c , położenia czoła paczki w chwili zadziałania zgarniaka R_s , położenia bariery świetlnej S_d). Należy podkreślić, że uzyskane rezultaty w procesie optymalizacji mogą być wykorzystane dla formułowania założeń konstrukcyjnych przy podejmowaniu nowych zadań projektowo – konstrukcyjnych związanych z urządzeniami procesów segregacji i rozdziału. Obrane zmienne decyzyjne (dostępne dla obsługi podczas eksploatacji) mają bezpośredni wpływ na wydajność i niezawodność procesu zgarniania. Za funkcję celu optymalizowanego zadania przyjęto maksymalizację wydajności procesu zgarnia:

$$\max W(t_c, V, R_s, S_d) = \frac{3600}{t_c + \frac{R_s - S_d}{V}} \quad (1)$$

Najkorzystniej dla maksymalizacji funkcji celu jest, gdy $R_s = S_d$ (w rzeczywistości $R_s \geq S_d$). Wówczas funkcja celu przybiera postać:

$$\max W(t_c) = \frac{3600}{t_c} \quad (2)$$

Ograniczeniami zadania optymalizacji są wymogi, których spełnienia bezwzględnie oczekuje się od procesu zgarniania. Należą do nich:

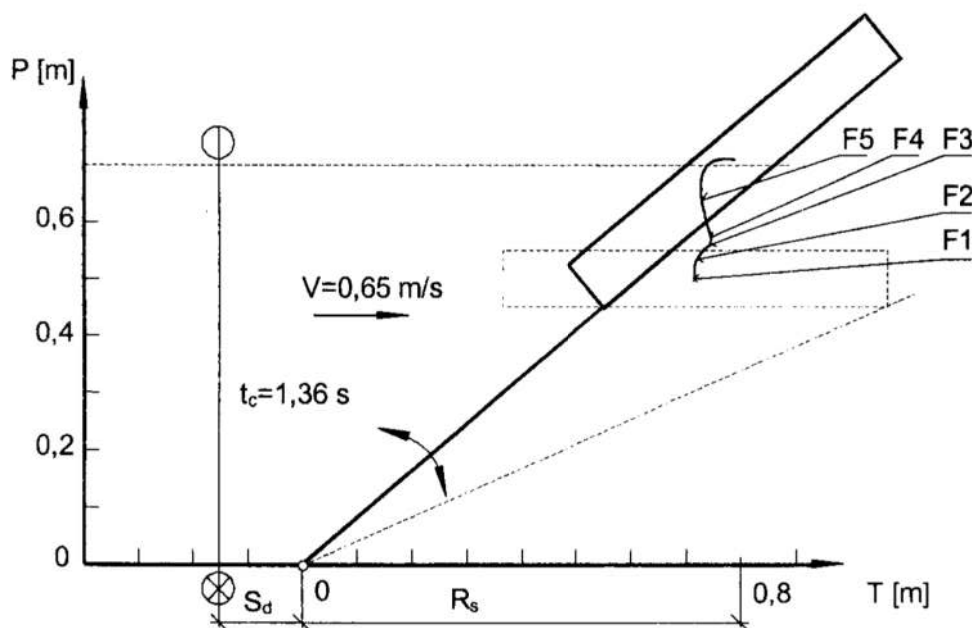
- nieuszkodzenie zgarnianych paczek:
 - $S_{dop} - S(t_c, V) \geq 0$ – warunek nieprzekraczania przeciążeń dopuszczalnych przez paczkę podczas zgarniania dla najbardziej niekorzystnych warunków zgarniania, tj. gdy uderzenie zachodzi w chwili osiągnięcia największej prędkości kątowej ramienia zgarniaka, w największej odległości od jego osi obrotu (S_{dop} – dopuszczalna wartość impulsu siły),
- niezawodne zgarnięcie wszystkich wprowadzonych do pętli rozdzielczej paczek:
 - $P(t_c, V, R_s) - P_{max} \geq 0$ – warunek przekroczenia przez środek ciężkości paczki podczas zgarniania przeciwległej burty taśmociągu ($P(t_c, V, R_s)$ – przemieszczenie środka ciężkości paczki w kierunku poprzecznym do osi przenośnika),

- $T_g(t_c, V, R_s) \geq 0$ – warunek zaistnienia kontaktu paczki ze zgarniakiem (sprawdzenie, czy paczka nie „ucieknie” zgarniakowi).

Kryterium optymalizacji (równanie 2) zależy monotonicznie od zmiennej decyzyjnej t_c sugerując, że optymalizacja jest trywialna. Jednakże występujące nieliniowe ograniczenia sprawiają, że zadanie optymalizacji jest problemem skomplikowanym. Szczególnie złożonym jest ograniczenie, które dotyczy konieczności pokrywania przez tor ruchu środka ciężkości paczki całej szerokości taśmociągu (czyli spełnienie warunku niezawodnego zgarnięcia). Sprawdzenie tego ograniczenia wymaga użycia modelu procesu zgarniania z występującymi w nim złożonymi zależnościami kinematyczno – dynamicznymi (wraz z tarciem suchym o nieliniowym charakterze).

5 Analiza modelu

Na podstawie modelu fizycznego zbudowano model matematyczny, a następnie model numeryczny procesu [2]. W oparciu o ten ostatni przeprowadzono symulację procesu zgarniania. Model weryfikowano w oparciu o istniejącą maszynę rozdzielczą. W trakcie badań np. okazało się, że przy obecnie stosowanych parametrach zgarniania, zalecanych przez dokumentację techniczną maszyny [3], nie można zgarnąć paczki o granicznych wymiarach 700x100x100 mm (dopuszczonej do automatycznego dzielenia), gdyby przed zgarnięciem znalazła się przy wewnętrznym skraju lub nawet w osi taśmociągu. Zgarnięcie tej paczki staje się możliwe, gdy znajdzie się ona w odległości $\Delta P=0,45$ m (rys.4).



Rys.4. Tor ruchu paczki podczas zgarniania zgarniakiem wychyłowym, parametry zgarniania: $t_c=1,36$ s, $V=0,65$ m/s, $R_s=0,8$ m, $\Delta P=0,45$ m; F1, F2, F3, F4, F5 – dyskretne stany procesu zgarniania

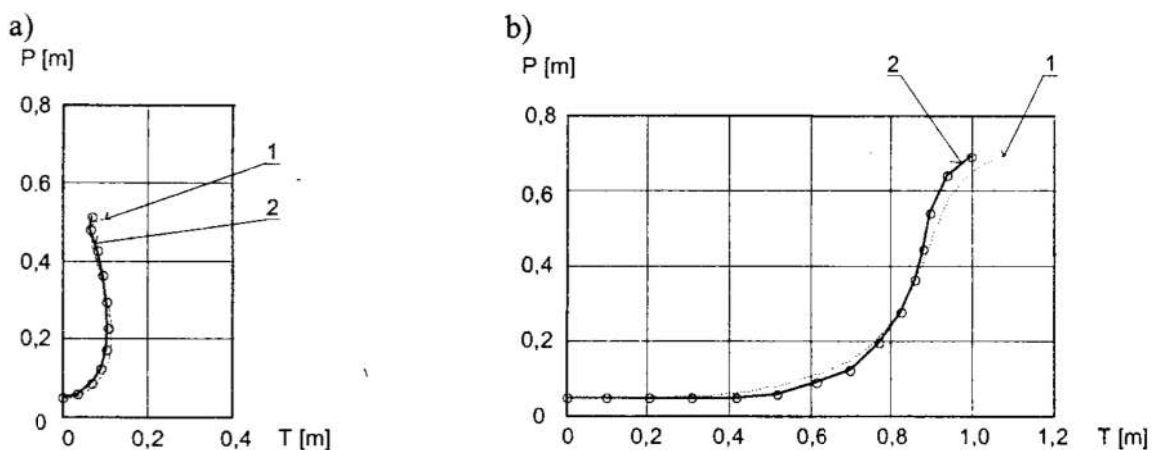
Fig.4. The path of package, scraping data: $t_c=1,36$ s, $V=0,65$ m/s, $R_s=0,8$ m, $\Delta P=0,45$ m; F1, F2, F3, F4, F5 – discreet states of the scraping process

Po zastosowaniu parametrów wyznaczonych drogą optymalizacji numerycznej (metodą gradientową z zewnętrzną funkcją kary i metodą genetyczną) procesu zgarniania zgarnięcie

takiej paczki do ześlizgu nie przysparza szczególnych problemów [2]. Optymalizacja procesu zgarniania metodą genetyczną wykazała ponadto istnienie bardzo wąskiego zakresu zmiennych decyzyjnych, które nie naruszałyby ograniczeń procesu zgarniania. Oznacza to, że przypadkowy dobór parametrów zgarniania przez obsługę jest praktycznie niemożliwy.

6 Badanie doświadczalne

Wyznaczanie rzeczywistego toru ruchu geometrycznego środka paczki (wykonanej zgodnie z PN [4]) przeprowadzono rejestrując proces jej zgarniania kamerą wideo. Na rysunku 5 wykreślono obok siebie toru ruchu paczek względem ramy przenośnika rozdzielczego - uzyskane drogą symulacji komputerowej (1) i zarejestrowane kamerą video (2). Rysunek 5 przedstawia tor ruchu środka ciężkości paczki o wymiarach 700x100x100 mm, gdy przed zgarnięciem znajdowała się przy wewnętrznym skraju taśmociągu. Przy niskiej prędkości taśmociągu V uzyskano wysoką zgodność torów (teoretycznego i rzeczywistego - rys.5a), nieco mniejszą zgodność przy wyższej prędkości (rys.5b). Porównując przebiegi krzywych doświadczalnych (2) i teoretycznych (1), można stwierdzić, że przyjęte założenia w modelowaniu analitycznym pozwalają określić, czy dane parametry zgarniania zapewnią zgarnięcie paczki do ześlizgu. Rozbieżność torów wynika z odmiennego przemieszczenia się

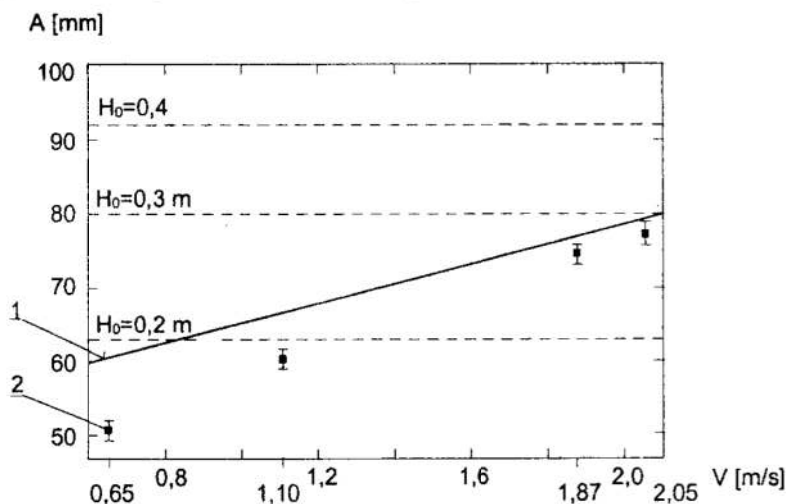


Rys.5. Tor ruchu środka ciężkości paczki (700x100x100 mm, $\Delta P=0$) względem ramy przenośnika rozdzielczego: 1 – wynik symulacji modelu, 2 – zarejestrowany kamerą video; parametry zgarniania: a) $t_c=1,36$ s, $V=0,65$ m/s, $R_s=0,8$ m, b) $t_c=1,36$ s, $V=2$ m/s, $R_s=0$ m

Fig.5. The path of package (700x100x100 mm, $\Delta P=0$): 1 – the model testing results, 2 – recorded by camera; scraping data: a) $t_c=1,36$ s, $V=0,65$ m/s, $R_s=0,8$ m, b) $t_c=1,36$ s, $V=2$ m/s, $R_s=0$ m

geometrycznego środka paczki wzdłuż osi taśmociągu, w kierunku ruchu taśmy. Rozbieżności te mogą być powodowane m.in. trudnym do zamodelowania zmiennym charakterem współczynnika tarcia. Wpływ mają także właściwości sprężysto-tłumiące paczki, które w modelowaniu opisano jedynie przy pomocy współczynnika restytucji. Podobne rezultaty uzyskano podczas zgarniania paczek o różnych wymiarach i dowolnie ułożonych na przenośniku taśmowym. Wyższą zgodność uzyskiwano dla ładunków o zwartej budowie, mniejszą o smukłej (podatniejszych na odkształcenia). Dla oszacowania przeciążeń jakim poddawana jest paczka w czasie jej zgarniania, czy podczas swobodnego spadku (operacje związane z sortowaniem paczki nie powinny

powodować przeciążeń dynamicznych większych niż odpowiadający im spadek z wysokości 0,3 m [1]), zamontowano w paczce specjalnie skonstruowany przetwornik sejsmiczny wraz z układem rejestrującym. Na rysunku 6 przedstawiono wyniki badań doświadczalnych przeciążeń paczki podczas zgarniania na maszynie rozdzielczej (pasma rozrzutu 2) i przy swobodnym spadku (poziome linie przerywane). Umieszczono na nim także krzywą (1) obrazującą wyniki obliczeń teoretycznych oddziaływania dynamicznego zgarniaka na paczkę. Naniesiono ją tak, że na rysunku 6 osiąga wartości przeciążeń odpowiadających wysokości swobodnego spadku $H_0=0,2$ m i $H_0=0,3$ m dla takich samych prędkości taśmy V jak uzyskanych drogą obliczeń matematycznych [2]. Analizując rysunek 6 okazuje się, że podczas badań rejestrowano przeciążenia paczki mniejsze niż te, które uzyskiwano z obliczeń matematycznych.



Rys. 6. Przeciążenia osiągnięte przy swobodnym spadku i przy zgarnianiu na maszynie sortującej dla cyklu pracy zgarniaka $t_c=1,36$ s: 1 – modelowanie teoretyczne, 2 – badania doświadczalne; A – amplituda wychylenia masy sejsmicznej
 Fig. 6. Overload reached during free fall of the load and during scraping on the machine, scraping cycle $t_c=1,36$ s: 1 – theoretical modelling, 2 – experimental research; A – amplitude of seismic mass swing

Dla niższych prędkości taśmy rzeczywiste oddziaływanie dynamiczne zgarniaka na paczkę bardziej odbiegało od teoretycznego, a dla prędkości dochodzącej do 2 m/s osiągnęto wyższą zgodność wyników. Wykorzystując model matematyczny do oceny powstałych przeciążeń paczki w zakresie przebadanych prędkości taśmy nie istnieje groźba, że zostaną przekroczone rzeczywiste przeciążenia dopuszczalne (osiągane przy spadku z wysokości $H_0=0,3$ m). Rozbieżność wyników znajduje się w obszarze bezpiecznym dla paczki niedopuszczając do jej uszkodzenia.

7 Wnioski

1. Wyniki badań modelu są poznawcze z punktu widzenia możliwości wpływania na konstrukcję i eksploatację maszyny rozdzielczej.
2. Proponowany analityczny model procesu zgarniania w odniesieniu do ograniczeń tego procesu spełnia postawione przed nim oczekiwania:
 - doświadczalne badanie toru ruchu paczki wykazało, że model matematyczny procesu zgarniania umożliwia prawidłową ocenę przemieszczenia paczki

- w kierunku poprzecznym do osi taśmy, tj. właściwie określa cechę, która służy do szacowania stopnia spełnienia warunku poprawnego jej zgarnięcia,
- doświadczalna weryfikacja przeciążeń paczki, jakie pojawiają się podczas zgarniania na maszynie rozdzielczej i przy swobodnym spadku, daje podstawę sądzić, iż teoretyczny sposób określania oddziaływań dynamicznych paczki przy jej zderzeniu ze zgarniakiem jest trafny w odniesieniu do przeciążeń dopuszczalnych.
3. Biorąc pod uwagę ograniczenia przyjętego do rozważań modelu rozdziału poprzez dynamiczne zgarnianie, uznano za celowe przeprowadzenie dalszej szczegółowej analizy możliwie szerokiego spektrum fizycznych modeli procesu sortowania.

Literatura

1. *Informacja o dopuszczeniu do obrotu pocztowego paczek pocztowych w typowych opakowaniach i kartonach*. Technika i Eksploatacja Poczty 4'93.
2. Piątkowski T.: *Analiza procesu zgarniania na przykładzie paczkowej maszyny rozdzielczej*. Praca doktorska, ATR Bydgoszcz 1999.
3. Wiatr R., Rawłuszko J.: *Analiza pracy pocztowej maszyny rozdzielczej pod kątem zwiększenia wydajności i / lub niezawodności sortowania paczek*. Zeszyty Naukowe ATR Nr 169, Bydgoszcz 1990.
4. PN-O-79100:1986, Opakowania transportowe, wymagania i badania.

Streszczenie

Referat dotyczy zagadnień związanych z procesem automatycznego dzielenia strumienia ładunków jednostkowych (paczek). Przedstawiono propozycję modelowania procesu zgarniania paczki, wyniki symulacji modelu oraz weryfikację eksperymentalną modelu teoretycznego. Wykonane badania potwierdziły wartość modelu teoretycznego.

DYNAMIC MODEL OF SCRAPING PROCESS AS AN OBJECT OF DISTRIBUTION PROCESS OPTIMIZATION OF UNITARY LOAD STREAM

Summary

The paper deals with the problem of automatic separation of the unitary loads (parcels). The report presents a proposal for a model of the scraping process, the model testing results and experimental verification of the theoretical model. Magnitude of the theoretical model was confirmed by realised experiments.

Referat recenzowany

PROBLEM EFEKTYWNOŚCI DOSTĘPU DO INFORMACJI SKŁADOWANYCH W RELACYJNEJ BAZIE DANYCH – ANALIZA OPERACJI ŻAŁAŃ

1. Wstęp

Tematem opracowania jest zagadnienie optymalizacji struktury bazy danych multimedialnych. Charakter tego rodzaju aplikacji jest złożony i możliwe jest wykorzystanie wielu kryteriów optymalizacji. Fundamentalnym krokiem w optymalizacji struktury bazy danych jest określenie głównych czynników, decydujących o poprawności funkcjonalnej i sprawności działania. W niniejszym opracowaniu omówiono zagadnienie określenia takich kryteriów oraz sposobów ich optymalizacji.

Efektywność systemu zarządzania relacyjną bazą danych zależy przede wszystkim od dwóch czynników. Pierwszy z nich to reprezentacja relacji w postaci takiej struktury danych, która pozwala na szybki dostęp na wiele różnych sposobów do poszczególnych relacji. Drugi z czynników dotyczy stosowania technik optymalizacji przy interpretacji języka manipulowania danymi.

W opracowaniu wykorzystano fragment bazy danych przeznaczonej do zarządzania elementami składowymi multimedialnych aplikacji dydaktycznych. Dane, na jakich przeprowadzone będą rozważania, to informacje określające parametry danych wizualnych w postaci obrazów (rozmiary, informacje o rodzaju palety obrazu oraz dane składowe obrazu) oraz informacje o lekcjach i pakietach lekcyjnych, których są składnikami.

Obszar optymalizacji obejmuje analizę struktury drzew żądań pod kątem efektywnego wyboru ścieżek dostępu, ułatwiających wykonanie operacji na danych. Relacja OBRAZY rozważanej bazy danych zawiera informacje o obrazach i lekcjach, których są elementami składowymi. Lekcje są elementami pakietów lekcyjnych obejmujących pewien kurs. Relacje zawierające opis danych multimedialnych i sposób organizacji lekcji, w których skład wchodziły zostały przedstawione na rysunku 1.

OBRAZY	HASŁA	PAKIETY_LEKCyjne	LEKCJA_OBRAZY	ROZKŁAD_LEKCJI
ID_OBRAZU WYSOKOŚĆ SZEROKOŚĆ PALETA BOOL ID_HASŁA	ID_HASŁA HASŁO	ID_PAKIETU ID_LEKCJI ST_ZAAW	ID_LEKCJI ID_OBRAZU	ID_LEKCJI ID_PAKIETU NR_LEKCJI W_PAKIECIE

Rys. 1. Zbiór relacji reprezentujący informacje związane z zarządzaniem elementami składowymi aplikacji wspomagających nauczanie – fragment bazy

Fig. 1. Fragment of database – a set of relations representing information associated with basic elements of Computer Aided Learning application

* Politechnika Zielonogórska, ul. Podgórna 50 65–246 Zielona Góra, A.Mowna@iie.pz.zgora.pl

Modelowana baza przechowuje następujące informacje opisujące analizowane elementy: numer obrazu (ID_OBRAZU), jego wysokość i szerokość podana w pikselach (odpowiednio $WYSOKOSC$, $SZEROKOSC$), rodzaj palety ($PALETA = „mono”$ lub $„kolor”$) oraz znacznik poprawności transmisji ($BOOL$). Ostatni atrybut określa hasło kluczowe, jakim został opisany obraz. Ponadto przechowywane są informacje określające przynależność obrazu do określonej części aplikacji dydaktycznej: numer lekcji (ID_LEKCJI), stopień zaawansowania informacji zawartych w lekcji (ST_ZAAW), numer pakietu z lekcjami obejmującymi jeden kurs ($ID_PAKIETU$) oraz numery lekcji znajdujących się w pakiecie ($NR_LEKCJI_W_PAKIECIE$).

2. Problem efektywności dostępu do danych składowanych w bazie

Uwzględnienie fazy optymalizacji zapytań w relacyjnych systemach zarządzania bazą danych stanowi jedną z najistotniejszych cech systemu. W dalszej części opracowania przedstawiona zostanie metoda zwiększenia efektywności dostępu do danych poprzez zapytania wyrażone w języku algebraicznym na przykładzie omawianej bazy danych. Następnie omówiona zostanie metoda do optymalnego wyboru ścieżki dostępu wśród istniejących ścieżek dostępu do danych oraz podany zostanie przykład.

Zdefiniowanie problemu

W przykładowej bazie danych zdefiniowano następujące relacje:

- 1) $ID_OBRAZU, WYSOKOSC, SZEROKOSC, PALETA, BOOL \in OBRAZ$
- 2) $ID_PAKIETU, ID_LEKCJI, ST_ZAAW \in PAKIETY_LEKCYJNE$
- 3) $ID_LEKCJI, ID_O \in LEKCJA_OBRAZY$
- 4) $ID_LEKCJI, ID_PAKIETU, NR_LEKCJI_W_PAKIECIE \in ROZKLAD_LEKCJI$
- 5) $ID_HASLA, HASLO \in HASLA$

gdzie $ID_OBRAZU, WYSOKOSC, SZEROKOSC, PALETA, BOOL$ oznaczają odpowiednio numer, wysokość, szerokość i paletę obrazu oraz znacznik poprawności transmisji; atrybuty $ID_HASLA, HASLO, ID_LEKCJI, ID_PAKIETU, ST_ZAAW$ to: numer hasła, hasło tematyczne, numery lekcji, pakietu i stopień zaawansowania informacji zawartych w lekcji; natomiast $ID_PAKIETU, NR_LEKCJI_W_PAKIECIE$ oznaczają odpowiednio numery pakietu i lekcji w pakiecie.

Zapytanie typu: „podaj listę numerów obrazów, które wchodzą w skład lekcji o poziomie zaawansowania wyższym niż 1 (podstawowy) i zawarte są pierwszej lekcji pakietu” można zapisać w języku algebraicznym na dwa różne sposoby. W notacji symbole „*”, „:”, „[]” oznaczają odpowiednio operację złączenia, selekcji i projekcji.

$$((PAKIETY_LEKCYJNE * LEKCJA_OBRAZY) : NR_LEKCJI_W_PAKIECIE = 1) [ID_OBRAZU] \quad (1)$$

$$((PAKIETY_LEKCYJNE : NR_LEKCJI_W_PAKIECIE = 1) [ID_LEKCJI]) * LEKCJA_OBRAZY [ID_OBRAZU] \quad (2)$$

W przypadkach (1) i (2) uzyskamy inny czas wykonania tego samego zapytania.

Efektywność systemu zarządzania relacyjną bazą danych zależy przede wszystkim od dwóch czynników. Pierwszy z nich to reprezentacja relacji w postaci takiej struktury

danych, która pozwala na szybki dostęp na wiele różnych sposobów do poszczególnych relacji. Drugi z czynników dotyczy stosowania technik optymalizacji przy interpretacji języka manipulowania danymi.

Metoda optymalizacji zadań dla języków algebraicznych

Algorytm. Przekształcenie wyrażeń algebraicznych

Dane wejściowe: drzewo wyrażenia algebraicznego.

Dane wyjściowe: drzewo zoptymalizowane.

Etap 1: zastąpić wszystkie operatory wyrażenia za pomocą operatorów elementarnych – patrz Tabela 1.

Etap 2: zastosować przekształcenie (patrz Tabela 2) - rozbicie operacji selekcji na ciąg operacji selekcji (przekształcenie P1).

Etap 3: dla każdego z operatorów selekcji zastosować przekształcenia (patrz Tabela 2): inwersja projekcja–selekcja, inwersja selekcja–iloczyn kartezjański, inwersja selekcja–suma, inwersja selekcja–różnica (odpowiednio przekształcenia P3, P4, P5, P6) tak, aby operator ten sprowadzić na jak najniższy poziom drzewa żądania.

Etap 4: dla każdego operatora projekcji zastosować przekształcenia (patrz Tabela 2): połączenie ciągu dwóch operacji w jedną (przekształcenie P2), inwersja projekcja–iloczyn kartezjański (P7), inwersja projekcja–suma (P8), aby operator ten przesunąć na jak najniższy poziom drzewa.

Etap 5: pogrupować operacje selekcji korzystając z przekształcenia rozbicia selekcji na ciąg operacji selekcji. Zastosować przekształcenie (P2) połączenia ciągu dwóch operacji w jedną w celu wyeliminowania zbędnych operacji projekcji. Jeśli jest to możliwe, przekształcić sekwencję iloczyn kartezjański–selekcja do postaci złączenia lub θ -złączenia. (patrz Tabela 2).

Algorytm ten wyraża następującą, podstawową zasadę: operacje projekcji ($\|R(X, Y)\|X\|$) i selekcji ($R : W$, gdzie W stanowi warunek wyboru) powinny się znaleźć na jak najniższym poziomie drzewa wyrażenia. W wyniku jego zastosowania nie zostaje określony taki porządek wykonania poszczególnych operacji, który minimalizuje strumienie danych odpowiadające łukom drzewa wyrażenia algebraicznego. Algorytm nie określa sposobu wykonania operacji selekcji, projekcji, złączenia, itd.

Tabela 1. Reguły zastępowania dowolnych operatorów operatorami podstawowymi [1]

Table 1. The rules of replacing any operators with a basic operator else [1]

R1	$R + S$	$\neg((\neg R) * (\neg S))$, potem R7 i R2
R2	$R(X_1, X_2) * S(Y_1, Y_2)$	$((R \times S) : X_1 = Y_1)\ X_1, X_2, Y_2\ $
R3	$R \cap S$	$\neg((\neg R) \cup (\neg S))$, potem R7
R4	$R(X_1, X_2) * (X_1 \theta Y_1) S(Y_1, Y_2)$	$(R \times S) : X_1 \theta Y_1$
R5	$R(X, Y) \div S(Y)$	$R[X] - ((R[X] \times S[Y]) - R(X, Y))\ X\ $
R6	$R(X, Y)Y\ $	$R(X, Y) \div R(X)$, potem R5
R7	$\neg R(X_1, X_2, \dots, X_n)$	$R[X_1] \times R[X_2] \times \dots \times R[X_n] - R(X_1, X_2, \dots, X_n)$

Tabela 2. Tablica przekształceń [1]
Table 2. The table of transformations [1]

P1	$R: F_1 \wedge F_2 \wedge \dots \wedge F_k$	$((R: F_1): F_2) \dots : F_k$
P2	$R(X, Y)[Z]$	$R[Z]$, jeśli $Z \subseteq Y$
P3	$(R[Y]): F(X)$ $(R: F(X))[Y]$	$(R: F(X))[Y]$, jeśli $X \subseteq Y$ $((R[Y \cup X]: F(X)))[Y]$, jeśli $X \not\subseteq Y$
P4	$(R(X) \times S(Y)): F$ $(R(X) \times S(Y)): F_1 \wedge F_2$	$(R(X): F) \times S(Y)$, jeśli atrybuty z F są zawarte w X $(R(X): F_1) \times (S(Y): F_2)$, jeśli atrybuty z F_1 i F_2 są zawarte odpowiednio w X i Y
P5	$(R \cup S): F$	$(R: F) \cup (S: F)$
P6	$(R - S): F$	$(R: F) - S$
P7	$(R(X) \times S(Y))[Z]$	$R[X \cap Z] \times S[Y \cap Z]$
P8	$(R \cup S)[Z]$	$R[Z] \cup S[Z]$

Zdefiniowano na podstawie relacji istniejących w bazie danych obraz. Nosi on nazwę PAKIETY. Stanowi on relację o dwóch atrybutach: ID_PAKIETU i ID_LEKCJI. Określa on pakiety lekcyjne o wyższym stopniu zaawansowania niż podany. Definicję obrazu PAKIETY, można zapisać:

```
PAKIETY_LEKCYJNE1 (ID_PAKIETU1, ID_LEKCJI1, ST_ZAAW1) ← PAKIETY_LEKCYJNE
PAKIETY (ID_PAKIETU, ID_LEKCJI) ← ( (PAKIETY_LEKCYJNE* (ID_PAKIETU=ID_PAKIETU1)
PAKIETY_LEKCYJNE1) : ST_ZAAW1 < ST_ZAAW) [ID_PAKIETU, ID_LEKCJI]
```

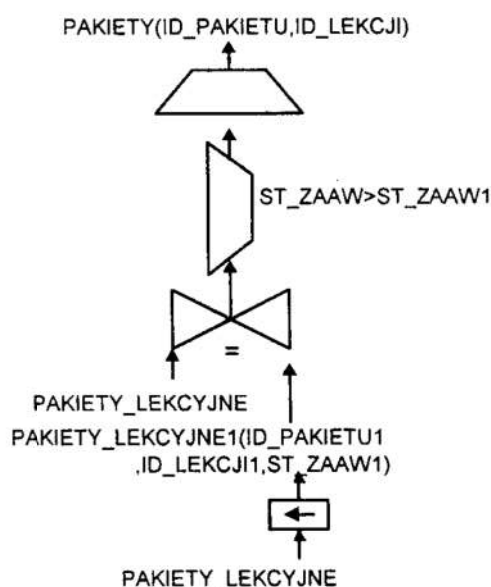
Drzewo powyższego zapytania zostało przedstawione na rysunku 2. Operacja selekcji oznaczona jest symbolem \triangleright , projekcję opisano \triangleleft , operacjom przypisania i złączenia odpowiadają odpowiednio \bowtie i \leftarrow .

Drugi ze skonstruowanych obrazów zawiera informacje o dotyczące rozkładu lekcji w pakietach lekcyjnych. Obrazowi temu odpowiada relacja PAKIETY_LEKCJE stanowiąca złączenie relacji PAKIETY i ROZKŁAD_LEKCJI, więc:

```
PAKIETY_LEKCJE ←
PAKIETY * ROZKŁAD_LEKCJI
```

Atrybuty relacji PAKIETY_LEKCJE są określone jednoznacznie przez atrybuty relacji PAKIETY i ROZKŁAD_LEKCJI.

Zgodnie z dokonanymi wyżej założeniami zapytanie postaci: „podaj listę numerów obrazów, które wchodzi w skład lekcji o poziomie zaawansowania wyższym niż 1 i zawarte są pierwszej lekcji pakietu”. Postać wyrażenia algebraicznego:



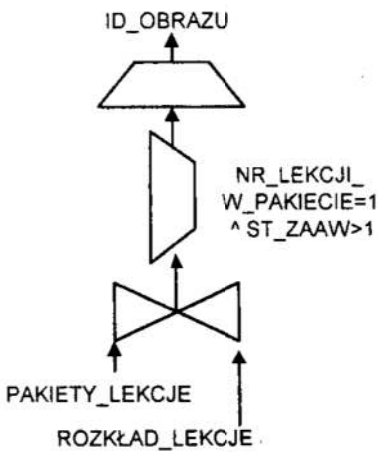
Rys. 2. Drzewo odpowiadające definicji obrazu PAKIETY
Fig. 2. A tree of PAKIETY definition

$$\left((PAKIETY_LEKCJE * LEKCJA_OBRAZY) : NR_LEKCJI_W_PAKIECIE = 1 \wedge ST_ZAAW > 1 \right) [ID_OBRAZU] \quad (3)$$

Drzewo powyższego żądania przedstawia rysunku 3.

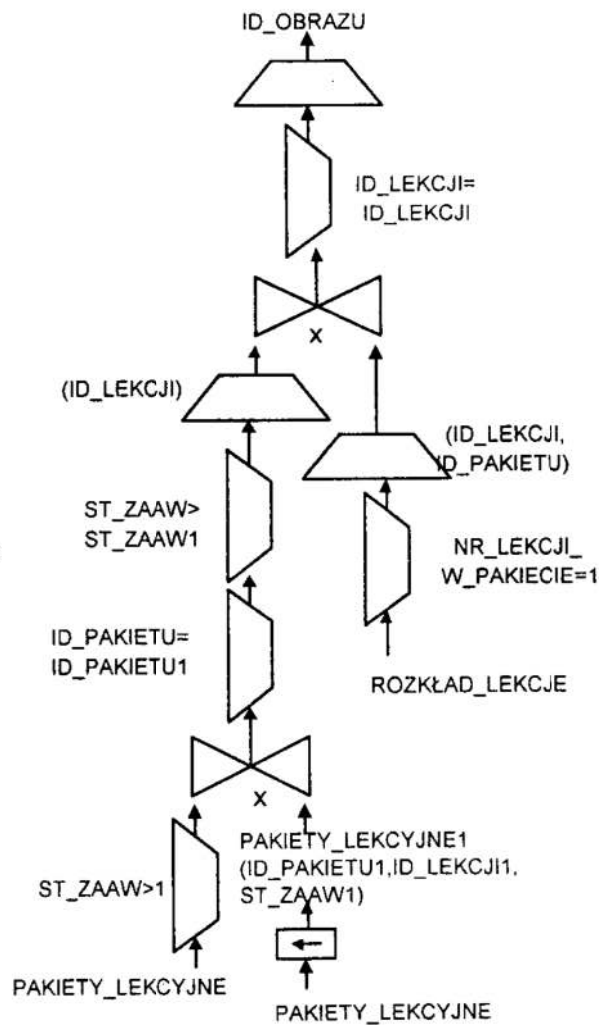
Zgodnie z drugim etapem algorytmu warunek wyboru:

$NR_LEKCJI_W_PAKIECIE = 1 \wedge ST_ZAAW > 1$ należy rozbić na dwa warunki elementarne. Operację selekcji $ST_ZAAW > 1$ można „przesunąć” w dół do poziomu relacji PAKIETY_LEKCYJNE, w wyniku zastosowania przekształcenia inwersja selekcja–suma (P4) oraz w wyniku skorzystania z własności przemienności operacji selekcji. Analogicznie operację selekcji $NR_LEKCJI_W_PAKIECIE = 1$ można przesunąć w dół do poziomu relacji ROZKŁAD_LEKCJI. Czwarty etap algorytmu polega na analizie możliwości przekształceń związanych z operacją projekcji. W przypadku projekcji [ID_OBRAZU] zastosowano drugą regułę przekształcenia inwersja selekcja–projekcja. W wyniku otrzymano drzewo jak z rysunku 4.



Rys. 3. Drzewo żądania korzystające z obrazu PAKIETY_LEKCJE

Fig. 3. PAKIETY_LEKCJE – a request tree



Rys. 4. Drzewo żądania po zastosowaniu omawianego algorytmu
Fig. 4. A result of apply described algorithm – a request tree

3. Podsumowanie

Przeprowadzone badania wstępne nad możliwością zastosowania omawianego algorytmu do rozwiązania problemu optymalizacji żądań prowadzą do wniosku, iż przedstawiona metoda zapewnia jedynie „dobre” rezultaty, a nie optymalne. Zgodnie z ogólnymi zasadami optymalizacji, operacje selekcji i powinny stanowić wierzchołki położone jak najbliżej liści drzewa żądania. Przekształcenia oparte na odpowiednich własnościach operatorów algebraicznych zastosowane w algorytmie umożliwiają takie właśnie przekształcenie drzewa żądań. Przy zastosowaniu tego algorytmu w omawianym przypadku zwiększona została efektywność dostępu do rekordów w bazie danych.

4. Literatura

1. Delobel C., Adiba M.: *Relacyjne bazy danych*. WNT Warszawa 1989.
2. Hwaryszkiewicz J.T.: *Database Analysis and Design*. MacMillan 1991.
3. Wirth N.: *Algorytmy+struktury danych=programy*. WNT Warszawa 1980.
4. Pławiak-Mowna A.: *Optymalne systemy baz danych do składowania danych multimedialnych*. I Ogólnopolskie Warsztaty Doktoranckie OWD'99 Istebna Zaolzie 17–21.10.99.
5. Pławiak-Mowna A., Zając W.: *Zagadnienie zastosowania techniki genetycznej w rozwiązaniu problemu optymalizacji operacji żądań w bazie danych oraz jakości składowanych w niej danych multimedialnych*. Konferencja Metody Sztucznej Inteligencji w Automatyzacji Procesów, Rabka, 13–15.04.2000.

Streszczenie

W artykule przedstawiono sposoby poprawy efektywności dostępu do danych składowanych w bazie przy wykorzystaniu operacji żądań [1]. Przeanalizowano algorytm optymalizacji drzew żądań dostępu do informacji w bazach danych. Wyniki badań zaprezentowano w artykule i zilustrowano przykładem, za który posłużył fragment bazy danych przeznaczonej do zarządzania elementami składowymi multimedialnych aplikacji dydaktycznych.

RDB ACCESS EFFECTIVITY PROBLEM ANALYSIS – REQUEST OPTIMISATION METHOD

Summary

In the paper a method of optimisation of the operation of data request is presented. A set of algorithms to make a tree of data request more effective is described. Results are presented and illustrated by a practical example. As an example the author used the database – a part of Computer Aided Learning application.

Referat recenzowany

OPTYMALIZACJA PROCESU DOKSZTAŁCANIA Z WYKORZYSTANIEM TECHNOLOGII NAUCZANIA NA ODLEGŁOŚĆ

1 Wstęp

W obecnych czasach wiedza staje się dobrem coraz mocniej cenionym. Pracodawcy starają się zatrudniać w swoich firmach osoby z możliwie dużym zasobem wiedzy. Nieustanny rozwój technologiczny, ciągłe zmiany metod pracy powodują, że pracodawcy dążą do podnoszenia kwalifikacji kadry pracowniczej. Obecnie potrzeba ciągłego uczenia się i doksztalcania zawodowego jest coraz bardziej oczywista, staje się wręcz niezbędna dla utrzymania się na coraz bardziej konkurencyjnym rynku. Człowiek uczy się już od pierwszych dni swego życia, naukę kontynuuje w szkole, później na uczelni. Należy zdawać sobie jednak sprawę z tego, że jest to dopiero początek naszej edukacji, wstęp pozwalający na zdobycie informacji niezbędnych do rozpoczęcia pracy zawodowej. Doksztalcanie się jest na stałe wpisane w rozwój zawodowy – pracodawca wymaga bowiem od pracowników, aby zdobywali oni nowe umiejętności i wiadomości, znali na bieżąco wszystkie zmiany jakie zachodzą w branży. Również pracobiorca decydując się na pracę w danej firmie, uzależnia swą decyzję o przyjęciu bądź odrzuceniu oferty pracy od możliwości własnego rozwoju w firmie. Nasuwa się więc dość oczywisty wniosek – zauważamy, że bez nieprzerwanego doksztalcania się nie tylko zatrzymamy się, ale wręcz zaczniemy cofać się w rozwoju.

Wydawać więc by się mogło, że zarówno pracodawca jak i pracownik mają wspólny cel, do którego starają się za wszelką cenę dążyć. Rzeczywistość jest jednak bardziej skomplikowana, często zdarza się bowiem, że rachunek ekonomiczny nie pozwala na organizowanie szkoleń w zakresie oczekiwanym przez obie zainteresowane strony. Koszty związane ze szkoleniem, dojazdami oraz noclegami mogą stanowić znaczne obciążenie finansowe, decydujące o rezygnacji ze szkolenia. W takiej sytuacji wartość pracowników, ich zaangażowanie zawodowe oraz efektywność może znacznie spaść, to zaś odbija się niekorzystnie na funkcjonowaniu całego przedsiębiorstwa.

Pewnym sposobem na wyeliminowanie opisanych powyżej zagrożeń może być wykorzystanie komputera i Internetu w procesie doskonalenia zawodowego. Dynamiczny rozwój technik informatycznych i telekomunikacyjnych znosi bariery, które jeszcze niedawno wydawały się nie do pokonania. Komputery posiadają bogate możliwości gromadzenia, przetwarzania, wyszukiwania i udostępniania informacji. Technika multimedialna udostępniająca tekst, dźwięk, grafikę i animacje jest bardzo dobrym narzędziem służącym do prezentacji wiedzy, powoduje, że edukacja staje się żywa i barwna [1]. Wykorzystanie Internetu daje możliwość przekazywania informacji z jednego miejsca w drugie z bardzo dużą elastycznością. Przesyłanie danych jest łatwe i tanie w porównaniu z tradycyjnymi mediami.

Połączenie możliwości multimedialnych komputera z dynamicznie rozwijającą się *Globalną Siecią Komputerową* (ang. *World Wide Web*), zaowocowało powstaniem nowej jakości w technologii zdalnego nauczania określanego jako nauczanie (kształcenie) na odległość. Działanie sieci Internet oparte jest o protokół TCP/IP [2]. Dzięki temu protokołowi, komputery o skrajnie różnych architekturach i możliwościach mogą ze sobą współpracować i przesyłać w sposób poprawny informacje.

* Politechnika Zielonogórska, Instytut Informatyki i Elektroniki, 65-246 Zielona Góra, ul. Podgórna 50, e-mail: A.Poplawski@iie.pz.zgora.pl

2 Nauczanie na odległość

O nauczaniu na odległość (*ang. distance learning*) możemy mówić wówczas, gdy proces dydaktyczny odbywa się w następujących warunkach:

- nauczycieli i uczniów dzieli pewna (nawet bardzo duża) odległość,
- do przekazywania informacji pomiędzy uczniem a nauczycielem wykorzystywane są współczesne media (radio, telewizja, Internet),
- nie jest wymagana obecność w tym samym czasie i miejscu grup uczących się, chociaż pomiędzy nauczycielem i uczniem istnieje dwustronna komunikacja,
- nad całym procesem nauczania czuwa instytucja edukacyjna (szkoła, uczelnia, firma itp.).

Nauczanie na odległość może odbywać się w dwóch trybach, asynchronicznym lub synchronicznym.

Nauczanie asynchroniczne

Nauczanie na odległość przyjmuje formę nauczania asynchronicznego wtedy, gdy uczestnicy kursu (szkolenia) uczestniczą w nim w dowolnie wybranym przez siebie czasie [3]. Spotkanie uczestników o umówionej porze celem przeprowadzenia kursu nie jest wymagane. Mogą się zdarzyć sytuacje kiedy w danej chwili więcej niż jedna osoba korzysta z zasobów systemu nauczania, jednak nie ma to żadnego wpływu na przebieg szkolenia. Uczestnicy kursu mogą w każdej chwili zwrócić się do prowadzącego szkolenie lub do innych uczestników kursu z pytaniami bądź też z wątpliwościami. Odpowiedź nie jest otrzymywana natychmiast, ale każdy uczestnik kursu może liczyć, że po kilku godzinach ją otrzyma.

Nauczanie synchroniczne

Szkolenie odbywające się w trybie synchronicznym oznacza, że użytkownicy kursu muszą w tym samym czasie uczestniczyć w szkoleniu [3]. Nauczanie w tym trybie ma formę zajęć grupowych, zaś jego przebieg jest na bieżąco monitorowany przez instruktora. W skład tej formy edukacji wchodzi wykłady, ćwiczenia, dyskusje i przerwy, jest to więc forma najbardziej zbliżona do tradycyjnych metod szkolenia. Instruktor dysponuje odpowiednim zestawem narzędzi służących kontroli całego kursu, możliwości zadawania pytań, udzielania głosu oraz monitorowania kto w danej chwili uczestniczy w zajęciach.

Nauczanie synchroniczne wymaga od osób uczestniczących w kursie większej sprawności umysłowej i skupienia uwagi. Jest procesem znacznie bardziej absorbującym zarówno dla ucznia jak i nauczyciela. Proces nauczania asynchronicznego daje jego użytkownikowi znacznie więcej czasu na przemyślenie tematyki jakiegoś zagadnienia, umożliwia znacznie większą indywidualizację procesu nauczania.

Nauczanie asynchroniczne jest zdecydowanie lepsze dla wszelkiego rodzaju ćwiczeń, nie wymaga natychmiastowego zaangażowania się w problematykę danego zadania. System nauczania w trybie asynchronicznym oferuje możliwości:

- uczenia się tego, czego chcę się uczyć (dowolna kolejność przerabianego materiału),
- uczenia się tam, gdzie chcę się uczyć (praca, dom, podróż),
- uczenia się wtedy, kiedy chcę się uczyć (dowolna pora dnia).

Ogólny schemat nauczania na odległość polega na uruchomieniu komputera, podłączeniu się do sieci oraz zalogowaniu się do systemu informatycznego udostępniającego kurs na określony temat. System dokonuje identyfikacji użytkownika oraz rejestruje jego poczynania. Administrator systemu nauczania na odległość (pracodawca) może być na bieżąco informowany o postępach poszczególnych osób (pracowników) w zdobywaniu wiedzy.

3 Możliwości optymalizacji procesu edukacyjnego

Firmy komercyjne mogą korzystać z nauczania na odległość, aby zmniejszyć koszty związane z organizacją szkoleń. Dzięki zastosowaniu tej technologii nauczania w skład kosztów szkolenia nie wchodzi już dojazd i noclegi. Większość przedsiębiorstw posiada wystarczającą liczbę komputerów, umożliwiającą przeprowadzenie szkoleń. Nawet jeżeli dana firma nie dysponuje dostępem do komputerów podłączonych do sieci, z pewnością znacznie taniej jest wynająć salę komputerową, niż posyłać pracowników na szkolenia.

Należy zdawać sobie sprawę z tego, że w każdej organizacji funkcjonują określone zasady BHP, reguły bezpieczeństwa, ściśle zdefiniowane zasady i procedury postępowania. Pracodawca oczekując przestrzegania tych reguł, powinien je w pierwszej kolejności przyswoić pracownikom, co oznacza krótsze lub dłuższe szkolenia. Kursy tego typu można z powodzeniem przeprowadzić przy wykorzystaniu technologii nauczania na odległość. Zagadnienia trudne do nauki bądź też nużące, można stosunkowo łatwo uatrakcyjnić wykorzystując bogate możliwości multimedialne komputera. Każdy nowo przyjmowany pracownik może już w pierwszym dniu pracy rozpocząć poznawanie firmy, zbioru zasad i reguł w niej panujących. Tym samym proces jego adaptacji do stawianych mu zadań zostaje znacznie skrócony, szybciej więc zacznie efektywnie pracować dla przedsiębiorstwa.

Szczególnego wymiaru nabiera technologia nauczania na odległość wśród osób takich profesji jak: prawnicy, sędziowie, prokuratorzy, urzędnicy państwowi, księgowi, itp. Przepisy prawne, rozporządzenia, ustawy, akty wykonawcze i inne przepisy, poddawane są stosunkowo częściej modyfikacji. W interesie pracodawcy leży dostarczenie podwładnym rzetelnej i fachowej informacji o wszelkich zmianach w przepisach, oraz ich interpretacji.

Wykorzystanie komputera do przeprowadzenia szkoleń znacznie skraca czas kursu, umożliwia wręcz natychmiastowy dostęp do nowych informacji. Firmy mające swoje przedstawicielstwa w różnych miastach, a nawet w różnych krajach mogą stosunkowo łatwo ujednoczyć system szkolenia panujący w przedsiębiorstwie. Wprowadzając nowe rozwiązania i technologie, można jednym ruchem przyuczyć pracowników zatrudnionych we wszystkich placówkach firmy. Możliwe jest także poddawanie pracobiorców okresowym sprawdzianom wiadomości, eliminując najslabsze jednostki lub kierując je na dodatkowe szkolenia.

Użycie techniki multimedialnej, żywego i barwnego obrazu przekazywanej wiedzy skojarzonej z dźwiękiem, symulacje omawianych zjawisk, znacznie zwiększają atrakcyjność nauki za pośrednictwem komputera. Wykorzystanie Internetu umożliwia w zależności od potrzeb przesłanie na kilka sposobów następujących informacji [4]:

- zwykłego tekstu (litery, cyfry, znaki specjalne),
- prostej grafiki (linie, koła, prostokąty),
- obrazu fotograficznego,
- dźwięku (głos, muzyka, dowolne efekty specjalne),
- obrazu wideo (ruchome obrazy, animacje),
- obrazu telewizyjnego w czasie rzeczywistym.

Należy jednak z całą stanowczością powiedzieć, że nie jest to system idealny, prosty w realizacji oraz możliwy do zastosowania w każdych warunkach. Stosowność tego rozwiązania ogranicza możliwość przekazu praktycznej wiedzy, rozwoju zdolności manualnych, przeprowadzenia pewnych ćwiczeń i eksperymentów. Jego zastosowanie dla nauk technicznych jest ograniczone tylko do pewnego rodzaju zagadnień. Pracownikom firm nie mających styczności z komputerem, szkolenie przy udziale nowych technik informatycznych, może stanowić dodatkową barierę nie do pokonania.

4 Podsumowanie i wnioski

Przedstawiony w artykule system edukacji na odległość oparty jest na wykorzystaniu komputerów i sieci Internet. Wszędzie tam, gdzie zapewniony jest dostęp do wyżej wymienionych środków technicznych, możliwe jest znaczne zredukowanie kosztów kształcenia i doskonalenia zawodowego pracowników. Pracodawca nie musi już opłacać kosztów przejazdu, zakwaterowania bądź wynajmu i utrzymania sal wykładowych. Opracowanie i wdrożenie odpowiednich kursów w technologii nauczania na odległość można zrealizować we własnym zakresie, bądź też skorzystać z pomocy wykwalifikowanej firmy. Wykorzystanie komputera w procesie doksztalcania niesie ze sobą szereg zalet, zaliczyć do nich można m.in.:

- stosunkowo dużą atrakcyjność nauki,
- możliwość nauki w pracy, domu i podróży,
- jednoczesne oddziaływanie na zmysł wzroku i słuchu,
- możliwość dostosowania tempa nauki do indywidualnych predyspozycji,
- możliwość ćwiczenia tego samego materiału kilkakrotnie.

Od dyrekcji przedsiębiorstw zależy wprowadzenie opisanej technologii w życie. Często brak zaufania do nowoczesnych metod dydaktycznych, strach przed wykorzystaniem komputera i nowoczesnych technik audiowizualnych może przesądzić kwestię wyboru. Całkiem realna wydaje się jednak taka perspektywa, w której ten kto odkryje potencjał nauczania na odległość jako pierwszy i zredukuje tym samym koszty własnej działalności, wyprzedzi konkurencję i pozostawi ją w tyle.

Literatura

1. Popławski A., Zając W.: *Multimedialny proces dydaktyczny*. Materiały Ogólnopolskiej Konferencji Studenckich Kół Naukowych i Młodych Pracowników Nauki, Białystok 1999.
2. Glistler P.: *Internet – przewodnik użytkownika*. WNT, Warszawa 1999
3. PC World Komputer Online, <http://www.pcworld.pl/artykuly/>
4. Wydawnictwa Szkolne i Pedagogiczne, <http://www.wsip.com.pl/serwisy/naucz/>

Streszczenie

W artykule omówiono możliwości optymalizacji procesu doksztalcania zawodowego pracowników przy wykorzystaniu nowoczesnych technik informatycznych. Omówiono metodę nauczania na odległość przy wykorzystaniu komputera i sieci Internet. Opisano różnice w nauczaniu dla trybu synchronicznego i asynchronicznego. Przedstawiono zalety jakie niesie ze sobą proces edukacji na odległość, zwrócono uwagę na wady prezentowanego rozwiązania. Zaprezentowano potencjalne możliwości zastosowań nowego medium edukacyjnego.

OPTIMISATION OF EDUCATION PROCESS BY USING A DISTANCE LEARNING TECHNOLOGY

Summary

In the paper possibilities of optimisation of education and re-education process are presented. A method of learning by using computer, Internet and distance learning technology is described and differences between synchronous and non synchronous education are presented. The potential advantages and disadvantages of the proposed solutions are shown and possibilities of new education medium usage are presented.

Referat recenzowany

PEWNA METODA OPTIMALIZACJI STRUKTUR DANYCH PRZECHOWYWANYCH W PAMIĘCI SYSTEMU KOMPUTEROWEGO

1 Wstęp

Rozwój sprzętu komputerowego jaki dokonał się w ostatnich latach jest bardzo duży. Stale powiększają się częstotliwości taktowania procesorów, dyski twarde umożliwiają zapamiętanie coraz większej ilości informacji w coraz krótszym czasie. W ślad za rozwojem sprzętu idzie rozwój oprogramowania. Jeszcze nie tak dawno, kiedy komputery dysponowały pamięcią RAM rzędu kilkudziesięciu kilobajtów, zaś jako pamięci stałej używano głównie dyskietek, dużą uwagę poświęcano oszczędności miejsca jakie zajmowały dane przechowywane w pamięci (czego dowodem jest tzw. problem roku 2000). Współcześnie dyskietek używa się coraz rzadziej na rzecz dysków twardech, płyt CD-ROM, napędów typu ZIP itp. Urządzenia te charakteryzuje duża pojemność, dostępność, oraz stosunkowo niska cena. Czynniki te wpływają na to, że coraz częściej zdarza się, iż pamięć dyskowa jest wykorzystanie nieoptymalnie, to znaczy w danych występuje nadmiarowość, część przestrzeni dyskowej jest zarezerwowana, lecz nigdy się z niej nie korzysta. Zachodzą jednak sytuacje, kiedy wielkość przechowywanej informacji ma kluczowe znaczenie, kiedy dysk musi być optymalnie wykorzystany, tak aby rozmiar pliku z danymi był możliwie mały.

W pracy rozważa się problem optymalnego wykorzystania przestrzeni dyskowej przy przechowywaniu danych. Zadanie polegało na modyfikacji istniejącego oprogramowania w taki sposób, aby obsługa aplikacji pozostała niezmienną, zaś dane programu zajmowały mniejszą przestrzeń dyskową.

2 Opis problemu

Najważniejszą kwestią przy realizacji zagadnienia było zagwarantowanie zgodności przechowywanych danych z pierwowzorem oraz ich spójności. W pracy nie zajmowano się problemem optymalnego wykorzystania pamięci komputera, dlatego też struktura danych przechowywanych w pamięci RAM pozostała niezmienną. W tym kierunku prowadzone będą dalsze prace. Na obecnym etapie realizacji tematu skupiono się na poszukiwaniu rozwiązań, pozwalających na odmienne interpretowanie danych zapisanych na dysku. Głównym zamierzeniem zaimplementowanych algorytmów było więc odtworzenie z pamięci dyskowej do pamięci operacyjnej komputera danych zgodnie z pierwotnym ich kształtem, oraz późniejsze ich przeniesienie z powrotem do pamięci dyskowej.

Struktura danych przechowywanych na dysku pierwotnie została określona w następujący sposób:

```
type osoba=record
  Dzień, Miesiąc, Rok:word;
  Zonaty, Plec:boolean;
  NIP:string[13];
  Pesel:string[11];
  Kod_pocztowy:string[6];
  Imię:string[20];
  Nazwisko:string[30];
end;
```

* Politechnika Zielonogórska, Instytut Informatyki i Elektroniki, 65-246 Zielona Góra, ul. Podgórna 50,
e-mail: A.Poplawski@iie.pz.zgora.pl

Taka konstrukcja struktury danych powoduje, że zapis jednego rekordu na dysku zajmuje 93 bajty. Należy zwrócić uwagę, że zapamiętanie m.in. takich pól jak: *Imię* czy *Nazwisko* jest w większości przypadków nadmiarowe, nieczęsto bowiem się zdarza, aby imię składało się z 20 liter, zaś nazwisko zawierało 30 znaków. Także inne pola rekordu zadeklarowano w sposób nieoptymalny. Poniżej przedstawiono mechanizmy pozwalające w sposób bardziej ekonomiczny wykorzystać przestrzeń dyskową.

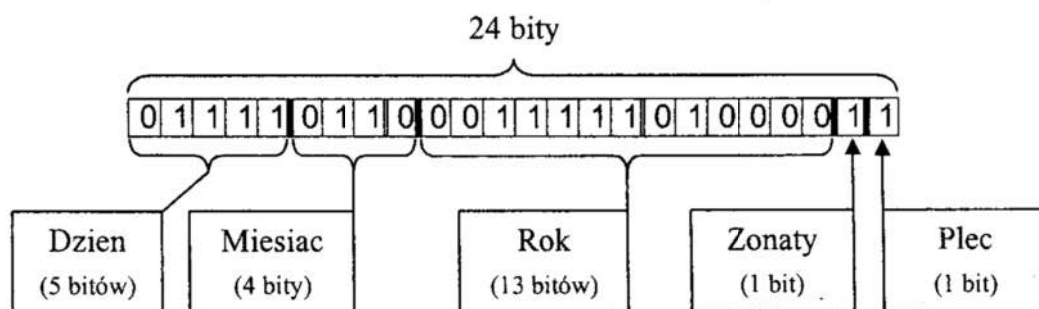
3 Struktura przyjętego rozwiązania

W wyniku przeprowadzonych prac powstał zbiór kilkunastu procedur i funkcji, służących głównie do odpowiedniego przygotowania danych, ich konwersji do innego formatu, zapisu i odczytu informacji na dysku. W prezentowanym rozwiązaniu uwagę zwracano nie tylko na dopasowanie długości zapamiętanych danych do informacji podanych przez użytkownika, ale także na możliwie pełne wykorzystanie wszystkich wolnych bitów.

Zoptymalizowany zapis daty

Pierwotnie dla zapamiętania daty użyto trzech zmiennych typu word, co w sumie powodowało rezerwację 48 bitów danych. Znaczna część zarezerwowanej tu pamięci nie była nigdy wykorzystywana. Dla przechowania dnia miesiąca wystarczy bowiem wykorzystać 5 bitów ($2^5=32$), zaś w celu zapamiętania numeru miesiąca wystarczą 4 bity ($2^4=16$) [1,2]. Przy rozpatrywaniu możliwości zachowania roku należy zwrócić uwagę, że wartość ta nie znajduje się w ściśle określonym przedziale, ale nieustannie rośnie. Należało więc zaproponować rozwiązanie, które będzie aktualne przez możliwie długi, lecz skończony i sensowny z punktu widzenia zajętości pamięci czas. Zdecydowano ograniczyć z 16 do 13 liczbę bitów wykorzystaną dla zachowania tej informacji. Takie rozwiązanie umożliwia podanie liczby od 0 do 8191 dla zachowania roku. Reasumując – chcąc przechować datę wystarczy użyć 22 bitów (21 jeśli ograniczymy rok do 4095), co daje ponad dwukrotną oszczędność pamięci, tj. 26 bitów w stosunku do pierwotnego kształtu aplikacji.

W rekordzie występują dwie zmienne typu logicznego. Przeprowadzone testy dowiodły, że ich zapisanie na dysku powoduje rezerwację dwóch bajtów, podczas gdy teoretycznie wystarczą tylko dwa bity. Przyjęto więc rozwiązanie, w którym zapis zmiennych *Zonaty* oraz *Plec* zajmuje $\frac{1}{4}$ bajta. Oszczędność pamięci w tym przypadku wyniosła więc 800%, co wynika w głównej mierze z nieoptymalnego zarządzania pamięcią przez kompilator języka Pascal [3]. Na rysunku 1 przedstawiono reprezentację zapisu daty oraz dwóch zmiennych logicznych w pamięci komputera.



Rys. 1. Reprezentacja zapisu daty w pamięci komputera

Fig. 1. Representation of a date information in a computer memory

Tak określone rozwiązanie zapisu daty i zmiennych logicznych pozwoliło na zaoszczędzenie 5 bajtów w stosunku do oryginalnej wersji programu. Procedura służąca

do zamiany reprezentacji danych zgodnie z podanym algorytmem `Koduj_Date` została zdefiniowana następująco:

```
Procedure Koduj_Date(Dzien,Miesiac,Rok:word; var d,m,r:byte);
Begin
  d:=0; m:=0; r:=0;
  d:=(Dzien shl 3) or (Miesiac shr 1);
  m:=(Miesiac shl 7) or (Rok shr 6);
  r:=Rok shl 2;
End;
```

Działanie procedury polega na odpowiednim przesunięciu bitowym zmiennych wejściowych *Dzien*, *Miesiac* i *Rok*, oraz umieszczeniu zakodowanego wyniku w trzech zmiennych wyjściowych *d*, *m*, *r*.

Odpowiednio procedurę służącą do przywrócenia pierwotnego kształtu danych `Odkoduj_Date` zdefiniowano następująco:

```
Procedure Odkoduj_Date(d, m, r:byte;var Dzien,Miesiac,Rok:word);
Begin
  Dzien:=0;Miesiac:=0;Rok:=0;
  Dzien:=d shr 3;
  d:=d shl 5; d:=d shr 4;
  Miesiac:=d or (m shr 7);
  m:=m shl 1;
  Rok:=(m shl 5) or r shr 2;
End;
```

Jej algorytm działania zbliżony jest do procedury kodującej datę. Danymi wejściowymi są zmienne *d*, *m*, *r*, zaś wyjściowymi *Dzien*, *Miesiac* oraz *Rok*.

Następna procedura `Koduj_Zonaty_Plec` służy do zapamiętania zgodnie z opisanym wyżej algorytmem dwóch zmiennych logicznych. Jej działanie polega na ustawieniu dwóch ostatnich bitów zmiennej *d* (domyślnie bity są wyzerowane). Definicja procedury przedstawia się następująco:

```
Procedure Koduj_Zonaty_Plec(Zonaty,Plec:boolean; var d:byte);
Begin
  If Zonaty=True Then d:=d+2;
  If Plec=True Then d:=d+1;
End;
```

Odpowiednio procedura `Odkoduj_Zonaty_Plec` została zdefiniowana następująco:

```
Procedure Odkoduj_Zonaty_Plec(d:byte;var Zonaty,Plec:boolean);
Begin
  If ((d and 00000010)=2) Then Zonaty:=True else Zonaty:=False;
  If ((d and 00000001)=1) Then Plec:=True else Plec:=False;
End;
```

Zoptymalizowany zapis danych numerycznych

Następnym w kolejności polem rekordu, które poddano optymalizacji to numer *NIP*. Wartość ta charakteryzuje się tym, że składa się wyłącznie ciągu cyfr od 0 do 9 oraz znaku myślnika, służącemu rozdzieleniu kolejnych trójek lub dwójek cyfr. Ponieważ znaki „-” występują w stałym miejscu, zdecydowano o ich wyrugowaniu z pliku. W zamian, myślniki będą automatycznie wstawiane przy odczycie numeru *NIP* z dysku.

Standardowo zdefiniowany zestaw znaków w kodach ASCII do przechowania informacji o jednej cyfrze potrzebuje 7 bitów, umożliwia to zapamiętanie 128 znaków (w praktyce wykorzystuje się 8 bitów) [4]. Ponieważ w rozważanym przypadku wystarczy zapamiętać tylko 10 różnych znaków, stworzono własne kodowanie znaków – kodowanie 4 bitowe. Daje to możliwość zapamiętania wprawdzie tylko 16 znaków, lecz w rozważanym przypadku jest w zupełności wystarczające. W zamian osiągnięto 50% oszczędność miejsca w pamięci, ponieważ przy kodowaniu 4 bitowym, na 8 bitach można zachować dwa znaki zamiast jednego. Procedurę temu służącą Koduj_Kod4 zdefiniowano w następujący sposób:

```

Procedure Koduj_Kod4(s:string; poz:byte; var tab:array of byte);
const
  Kod4: array[0..15] of Char = ' ()+-0123456789_';
var i,j,k:byte;
Begin
  j:=poz-1;
  For i:=1 to length(s) do
    Begin
      If i mod 2=1 Then
        Begin
          Inc(j);
          tab[j]:=0;
        End;
        For k:=0 to 15 do If Kod4[k]=s[i] Then
          Begin
            tab[j]:=tab[j]+k;
            break;
          End;
        If i mod 2=1 Then tab[j]:=tab[j] shl 4;
      End;
    End;
  End;

```

Zmienna wejściowa s jest przetwarzana zgodnie z przyjętym algorytmem przy udziale tablicy kodowej Kod4, zaś wynik jest zapisywany do tablicy tab, począwszy od pozycji poz. Odpowiednio przywrócenie informacji do stanu początkowego dokonuje się przy udziale procedury Odkoduj_Kod4 zdefiniowanej następująco:

```

Procedure Odkoduj_Kod4(tab:array of byte; dl,poz:byte;
  var s:string);
var i,j,z:byte;
Begin
  j:=poz-1;
  For i:=1 to dl do
    Begin
      If i mod 2=1 Then
        Begin
          Inc(j);
          z:=tab[j] shr 4;
        End Else
        Begin
          z:=tab[j] shl 4;
          z:=z shr 4;
        End;
      s[i]:=Kod4[z];
    End;
  s[0]:=Chr(dl);
End;

```

Dwie powyższe procedury wykorzystano także do zapamiętania numeru *Pesel*. Jest to wartość liczbową składająca się z 11 cyfr. Ponieważ początkowych 6 cyfr jest ściśle związanych z datą urodzenia osoby, która jest już przechowywana w innych polach rekordu, na dysku zapisywanych jest jedynie 5 ostatnich cyfr tej wartości. Funkcja służąca do przywrócenia pierwotnej, 11 cyfrowej wartości przedstawia się następująco:

```
Function Przekształc_Pesel (Dzien, Miesiac, Rok:word;
                           s:string):string;
var i:byte; Pesel:string;
Begin
  Rok:=Rok mod 100;
  Pesel[1]:=Chr((Rok div 10)+48);
  Pesel[2]:=Chr((Rok mod 10)+48);
  Pesel[3]:=Chr((Miesiac div 10)+48);
  Pesel[4]:=Chr((Miesiac mod 10)+48);
  Pesel[5]:=Chr((Dzien div 10)+48);
  Pesel[6]:=Chr((Dzien mod 10)+48);
  For i:=1 to 5 do Pesel[i+6]:=s[i];
  Pesel[0]:=Chr(11);
  Przekształc_Pesel:=Pesel;
End;
```

Zapamiętanie opisywanej wartości wymagało zarezerwowania 3 bajtów, co jest 4 krotną oszczędnością w stosunku do 12 bajtów w oryginalnej wersji programu.

Optymalizując pole *Kod_pocztowy* osiągnięto ponad dwukrotną oszczędność pamięci. Wykorzystano wcześniej zdefiniowane procedury do kodowania liczb w formacie 4-bitowym, co pozwoliło na przechowanie tej wartości na 3 bajtach (poprzednio 7).

Ostatnimi elementami rekordu poddanymi optymalizacji są pola *Imie* i *Nazwisko*. W pierwotnie przyjętym rozwiązaniu, rezerwowano 52 bajty pamięci dla zachowania tych zmiennych, niezależnie od liczby znaków na które składa się dana wartość. Przeprowadzone badania dowiodły, że w przeważającej większości przypadków zarezerwowana pamięć wykorzystywana była w niespełna 50%. Należało więc poszukać innej konstrukcji rekordu, eliminującej „puste” miejsca w pamięci. Zaproponowano rozwiązanie polegające na zastosowaniu zmiennej długości pola rekordu przechowującego *Imie* i *Nazwisko*. Struktura przyjętego rozwiązania przedstawia się następująco:

- pierwszy bajt danych służy do określenia długości tekstu,
- kolejne bajty danych zawierają odpowiednio poszczególne znaki tekstu.

Na rysunku 2 przedstawiono graficzną reprezentację zapisu danych tekstowych.



Rys. 2. Reprezentacja zapisu danych tekstowych na dysku
 Fig. 2. Representation of a text data on a hard disk

4 Podsumowanie i wnioski

Końcowy wynik optymalizacji jaki uzyskano, będący złożeniem opisywanych jednostkowych rozwiązań, pozwolił na znaczne oszczędności pamięci dyskowej dla przechowania danych aplikacji. Część stałą danych, na którą składają się pola: *Dzien, Miesiac, Rok, Zonaty, Plec, NIP, Pesel* oraz *Kod_pocztowy* udało się zredukować z 41 do 14 bajtów. Jest to blisko trzykrotne zmniejszenie zapotrzebowania na pamięć dla przechowania dokładnie tych samych informacji. Część zmienna danych, czyli pola *Imie* i *Nazwisko* również poddały się optymalizacji, przy średnim wykorzystaniu 47% oryginalnego miejsca przeznaczonego na dane. Biorąc pod uwagę rekord jako całość, średnia oszczędność pamięci wyniosła 238%.

Należy zwrócić także uwagę na to, że zastosowanie opracowanych algorytmów spowalnia nieznacznie czas dokonywanych transmisji danych na linii dysk - pamięć oraz pamięć - dysk. Jest to niewątpliwie wada przyjętego rozwiązania. Spowodowane jest to koniecznością konwersji danych. Z drugiej jednak strony, ponad dwukrotnie mniejszy plik z danymi przyspiesza sam proces zapisu i odczytu, co po części rekompensuje omawianą niedogodność. Minusem zastosowanych algorytmów jest również podwyższona złożoność logiki aplikacji, co zwiększa prawdopodobieństwo występowania błędów w działaniu programu.

Literatura

1. Biernat J.: *Arytmetyka komputerów*. PWN, Warszawa 1996.
2. Biernat J.: *Architektura komputerów*. OWOP, Wrocław 1999.
3. Marciniak A.: *Borland Pascal 7.0*. Nakom, Poznań 1995.
4. Chalk B.S.: *Organizacja i architektura komputerów*. WNT, Warszawa 1998.

Streszczenie

W artykule omówiono metody minimalizacji wielkości danych programu przechowywanych w pamięci dyskowej. Zaprezentowano kilka algorytmów mających na celu zmniejszenie zapotrzebowania aplikacji na przestrzeń dyskową. Przedstawiono osiągnięte rezultaty oraz zapoznano czytelnika z praktyczną realizacją omówionych metod optymalizacji. W szczególności zaprezentowano mechanizmy bardziej optymalnego formatu przechowywania daty, danych numerycznych oraz dowolnego ciągu znaków. Przedstawiono potencjalne wady, jakie niesie za sobą praktyczna realizacja opisanych zagadnień.

A METHOD OF OPTIMISATION OF DATA STRUCTURES STORED IN COMPUTER MEMORY

Summary

In the paper methods of data size minimization are presented. A set of algorithms to decrease the amount of memory is described and achieved results are presented and illustrated by a practical application. Mechanisms of optimisation of date, numerical data and text storage are discussed. The potential disadvantages of the proposed solution are presented.

Referat recenzowany

ZASTOSOWANIE ALGORYTMU GENETYCZNEGO W ZAGADNIENIU SELEKCJI STRATEGII POSTĘPOWANIA

1. Wprowadzenie

W systemach sztucznej inteligencji jednym z głównych zagadnień jest implementacja wyboru właściwej strategii postępowania prowadzącej do osiągnięcia założonego celu. Przykładem takiego zagadnienia jest zarówno pokonywanie przeszkód stojących na drodze poruszającego się robota jak i strategia gry w pokera. W przypadku problemów dla których nie jesteśmy w stanie określić rozwiązania analitycznego, wybór właściwej strategii postępowania można osiągnąć poprzez analizę danych i stawianie właściwych hipotez dotyczących zależności pomiędzy nimi. W przypadku dużej ilości danych podstawowym problemem jest zarówno wybór właściwych danych do analizy jak i sposób stawiania właściwych hipotez. Realizacja programowa postępowania, w którym następowałyby analiza samodzielnie wybranych danych oraz wybór skutecznej metody postępowania wobec nowych warunków i zadań jest niezmiernie trudna i praktycznie nie realizowalna poprzez standardowe, sekwencyjne przetwarzanie danych. W sytuacjach tych w większości przypadków stosuje się inne metody algorytmiczne takie jak sieci neuronowe, zbiory rozmyte czy algorytmy genetyczne.

W pracy przedstawiono badania nad wyborem strategii postępowania w grze hazardowej o nazwie : „Kamień – nożyczki – papier”. Do wyboru strategii gry wykorzystano odpowiednio zaprojektowany algorytm genetyczny wzbogacony wiedzą. Dla potrzeb badań zrealizowano programową symulację środowiska gry, w którym przeprowadzono symulację współzawodnictwa pomiędzy poszczególnymi strategiami gry.

2. Gra „Kamień – nożyczki – papier”

Problem wyboru strategii gry jest jednym z głównych zagadnień jakimi zajmuje się teoria gier. W większości gier prawie każda forma działalności jednostki wywołuje skutki globalne dla danej rozgrywki i wpływa na postępowanie innych uczestników gry, w większości przypadków modyfikując ich wcześniejszą decyzję o sposobie osiągnięcia założonego celu. Z kolei reakcja innych graczy powoduje generowanie nowych faktów, co z kolei wpływa na naszą decyzję o sposobie dalszego postępowania. Mamy tu więc do czynienia z działaniem w interaktywnym środowisku, w którym zachodzą zmiany w większości przypadków trudne do przewidzenia, ze względu na dynamiczny charakter zmiennych opisujących środowisko oraz brak pełnych informacji o jego charakterystyce [1]. System, który byłoby zdolny do funkcjonowania w takich warunkach musi opierać się na analizie napływających faktów i trafnym stawianiu hipotez dotyczących stanu środowiska i optymalnej strategii postępowania [6]. W przypadku większości gier dodatkową trudnością jest fakt, iż poszczególni uczestnicy gry mają do osiągnięcia wspólny cel, co owocuje współzawodnictwem i zmusza gracza do aktywnego udziału w rozgrywce powodując często konflikt interesów [1].

Przykładem takiej gry jest gra o nazwie : „Kamień – nożyczki – papier”. Gra ta ma dość długą historię i mimo swojej prostoty jest interesującym przedmiotem badań, zwłaszcza

* Politechnika Zielonogórska, Instytut Informatyki i Elektroniki, E-Mail : r.puscian@iie.pz.zgora.pl

jeśli chodzi o zagadnienie strategii postępowania. Grać w nią może dowolna liczba osób, a jej cechą charakterystyczną jest brak możliwości kooperacji (po za grą na remis, która nie wyłącza zwycięzcy). W grze tej każdy uczestnik gra przeciwko wszystkim pozostałym, wybierając wg własnego uznania jeden z trzech symboli, który reprezentuje gracza w danej rozgrywce. Symbolami tymi są: kamień, nożyczki i papier, zaś o przydzieleniu punktów graczowi w danej rozgrywce decyduje zależność pomiędzy symbolami gracza i pozostałych uczestników gry. Liczba rozgrywek jest dowolna, a pomiędzy kolejnymi rozgrywkami gracze mogą zmieniać swoje symbole. Zależności pomiędzy symbolami są następujące :

- papier owija kamień – starcie wygrywa papier,
- kamień tępi nożyczki – starcie wygrywa kamień,
- nożyczki tną papier – starcie wygrywają nożyczki.

Z zależności tych wynika, że każdy symbol ma szansę na wygraną, przegraną lub remis, jednocześnie nie ma takiego symbolu, który zawsze wygrywa, zawsze przegrywa lub zawsze remisuje. Spośród wszystkich graczy wygrywa ten, który w zaplanowanej liczbie rund uzyska największą liczbę punktów.

Tab. 1. Tabela punktacji gry

Tab. 1. Awards table game

	Kamień	Papier	Nożyczki
Kamień	Remis – po 1 pkt.		
Papier	Papier : 2 pkt. Kamień : 0 pkt.	Remis – po 1 pkt.	
Nożyczki	Nożyczki : 0 pkt. Kamień : 2 pkt.	Nożyczki : 2 pkt. Papier : 0 pkt.	Remis – po 1 pkt.

Zmieniając zależności pomiędzy ilością punktów za wygraną, remis i przegraną można stworzyć kilka modeli gry, z których najbardziej interesujące wydają się trzy :

1. Pełna rywalizacja – wygrana : 5 punktów, remis : 2 punkty, przegrana : 0 punktów. Bonus za zwycięstwo : 2 punkty * liczba graczy.
2. Duży zysk, małe ryzyko, – wygrana : 1 punkt, remis : 1 punkt, przegrana : 0 punktów. Bonus za zwycięstwo : 2 punkty * liczba graczy.
3. Duży zysk, duże ryzyko, – wygrana : 5 punktów, remis : 5 punktów, przegrana : -1 punkt. Bonus za zwycięstwo: 2 punkty * liczba graczy.

W grze toczonej wg schematu pierwszego mamy do czynienia z pełną rywalizacją pomiędzy poszczególnymi zawodnikami. Gra na remis jest tutaj nieopłacalna, zaś zwycięzca jest dodatkowo nagradzany, powiększając swoją przewagę nad konkurentami. W drugim wariantcie gry zarówno remis jak i zwycięstwo są jednakowo nagradzane. W grze tej najbardziej optymalny jest najmniej ryzykowny sposób zdobywania punktów, a mianowicie gra na remis. Jednak gracz który się wyłamie z tej „umowy” ma szansę na dużą nagrodę w przypadku zwycięskiego ruchu, niewiele przy tym ryzykując bowiem gdy przegra ze zwycięską koalicją remisową, każdemu z nich doda tylko jeden punkt, natomiast nagroda główna przypadnie z braku pojedynczego zwycięzcy. W grze toczonej według trzeciego schematu również należy spodziewać się przewagi gry na remis z tą jednak różnicą, że gracz, który się wyłamie z „koalicji

remisowej” i przegra, straci wiele punktów bowiem każda przegrana po za dodaniem punktów przeciwnikowi skutkuje także utratą jednego punktu własnego co przy dużej liczbie przeciwników (i małej liczbie rund) można ostatecznie przekreślić szansę tego gracza na wygraną całej gry.

Nie zależnie od modelu punktacji wybór zwycięskiej strategii postępowania w tej grze jest zagadnieniem trudnym do rozwiązania ze względu na brak możliwości analitycznego przewidywania ruchów przeciwnika. Mamy tu więc do czynienia z zadaniem które jest nierozwiązywalne analitycznie ze względu na brak danych. Jedną z metod rozwiązania tego problemu jest zastosowanie algorytmów, które na podstawie danych „z przeszłości” będą stawiać właściwe hipotezy co do wartości tych danych w „przyszłości” i na ich podstawie podejmą właściwą decyzję co strategii postępowania. Innym podejściem może być podejmowanie decyzję na podstawie analizy skutków swojego poprzedniego postępowania. Tę ostatnią metodę można zrealizować za pomocą algorytmów genetycznych.

3. Algorytm genetyczny

Klasyfikując różne zagadnienia na podstawie teorii matematycznych możemy z analitycznego punktu widzenia podzielić je na rozwiązywalne i nierozwiązywalne. Problemy których nie jesteśmy w stanie rozwiązać metodami analitycznymi ze względu na zbyt dużą ilość niewiadomych lub ze względu na nieznaną model matematyczny rozwiązuje się metodą przeszukiwania przestrzeni rozwiązań. Jedną z najpopularniejszych takich metod jest metoda generowania i sprawdzania (*generate-and-test*). Polega ona na generowaniu wartości parametrów wejściowych i sprawdzeniu czy osiągnięte przez te wartości rozwiązanie odpowiada naszym wymaganiom. W metodach poszukiwania rozwiązania metodą generowania i sprawdzania wyróżnia się dwa odmienne podejścia. Pierwsze podejście określa się strategią poszukiwania w głąb i polega na sprawdzeniu wszystkich kombinacji parametrów wejściowych. Stosuje się je do zadań dla których jesteśmy w stanie przeprowadzić taką procedurę w rozsądnym (akceptowalnym) przedziale czasu. W przypadku zadań dla których tego typu strategia jest zbyt długotrwała (nie akceptowalna ze względu na koszt numeryczny) stosuje się strategię zachłanną. Polega ona na generowaniu grupy rozwiązań (ciągu parametrów wejściowych), z której po oceniu zostaną wybrani najlepsi reprezentanci celem dalszej eksploracji przestrzeni rozwiązań. Jedną z odmian tej strategii jest strategia poszukiwania rozwiązania oparta o algorytm genetyczny.

Algorytmy genetyczne są oparte o iterowany, probabilistyczny model przeszukiwania przestrzeni rozwiązań, jednak tym co je odróżnia od przeszukiwania czysto losowego jest uwzględnianie w bieżącym cyklu przeszukiwania, informacji o wynikach przeszukiwania z cyklu poprzedniego [2], [4]. Algorytmy genetyczne swą nazwę zawdzięczają faktowi wzorowania się na mechanizmach ewolucji organizmów żywych i nazewnictwo w nich obowiązujące zaczerpnięte jest z genetyki [2], [3].

Kodowanie chromosomu

Podstawowym elementem algorytmu genetycznego jest chromosom, który koduje wartości parametrów wejściowych dla danego problemu (zadania). Jeden parametr wejściowy reprezentowany jest przez część chromosomu zwaną genem. Sposób kodowania parametrów wejściowych w postaci zbioru genów ma istotny wpływ na efektywność algorytmu genetycznego i jest kluczowym zagadnieniem w procesie jego implementacji [2], [4].

Do wyboru strategii postępowania w rozważanej grze został wykorzystany algorytm genetyczny wzbogacony wiedzą, oparty o 8 bitowy chromosom binarny [2], który koduje strategię gry. Baza wiedzy jest unikalna dla każdego osobnika (reprezentanta

chromosomu) i składa się z 4 informacji : ostatni własny symbol wygrywający (kamień–nożyczki–papier), ostatni własny symbol przegrywający, wygrywający symbol ostatniego przeciwnika, przegrywający symbol ostatniego przeciwnika. Informacje te są wykorzystywane przez niektóre strategie gry.

Strategie gry z logicznego punktu widzenia są kodowane dwustopniowo. Pierwszy stopień określa rodzaj zagrywki :

1. zawsze symbol : „kamień”,
2. zawsze symbol : „nożyczki”,
3. zawsze symbol : „papier”,
4. ostatni własny symbol wygrywający,
5. ostatni własny symbol przegrywający,
6. wygrywający symbol ostatniego przeciwnika,
7. przegrywający symbol ostatniego przeciwnika,
8. losowy wybór symbolu.

Strategie te (po jednej) są zapamiętane w dwóch rozłącznych miejscach w chromosomie, tworząc w ten sposób parę zagran „A” i „B”. O wyborze zagrania ‘A’ lub ‘B’ decyduje drugi stopień kodowania, który określa cztery następujące możliwości :

1. używaj zawsze zagrania ‘A’,
2. używaj na przemian zagrania ‘A’ i ‘B’,
3. dokonaj wyboru losowego pomiędzy zagranieniem ‘A’ i ‘B’,
4. jeżeli ostatnie starcie wygrałeś lub zremisowałeś to następnym razem użyj zagrania ‘A’, w przeciwnym wypadku użyj zagrania ‘B’.

Wybór strategii gry odbywa się poprzez wybranie na podstawie wartości chromosomu zagrania ‘A’ lub ‘B’, a następnie wyborze odpowiedniego symbolu (również na podstawie wartości chromosomu).

W konstrukcji chromosomu możemy więc wyróżnić z trzy logiczne części. Część kodującą zagranie ‘A’, część kodującą zagranie ‘B’ oraz część kodującą wybór zagrania ‘A’ lub ‘B’. Części ‘A’ i ‘B’ są kodowane przez 6 bitów (3 dla ‘A’ i 3 dla ‘B’), zaś część kodującą wybór zagrania ma rozmiar 2 bitów. Struktura przedstawiająca chromosom wygląda następująco :

```
struct geny_s
{
    char    symbol_A    : 3;
    char    symbol_B    : 3;
    char    strategia   : 2;
};
/*****/
union chromosom_s
{
    char    chromosom;
    geny_s gen;
};
```

Krzyżowanie

Poszczególne osobniki przedstawiające wybrane strategie gry oceniane są na podstawie liczby zdobytych punktów w rozgrywce typu „każdy z każdym”. Po każdej takiej rozgrywce i ocenieniu osobników na podstawie uzyskanych rezultatów następuje generowanie nowej populacji osobników (reprezentantów strategii gry). Nowa populacja generowana jest w wyniku operacji krzyżowania, do której osobniki są wybierane wg reguły ruletki. Krzyżowanie jest jednopunktowe z losowym punktem krzyżowania i odbywa się z zadaniem na starcie symulacji prawdopodobieństwem [2], [4]. Uzupełnieniem operacji reprodukcji jest operacja mutacji polegająca na losowej zmianie losowo wybranych bitów chromosomu [2], [4]. Również ta operacja odbywa się z prawdopodobieństwem zadaniem na starcie symulacji.

4. Wyniki symulacji

Symulacje, których wyniki zostały przytoczone w niniejszej pracy zostały przeprowadzone przy prawdopodobieństwie krzyżowania równym 90% i prawdopodobieństwie mutacji równym 0.5%. Populacja składała się z 20 osobników. Pomiedzy poszczególnymi rozgrywkami następowało tworzenie nowej populacji poprzez operację krzyżowania. Rozegrane zostały trzy turnieje składające się z 10 gier po 20 rozgrywek każda. Rozpoczęcie nowej gry rozpoczynało się od losowego stworzenia nowej populacji. Poszczególne turnieje różniły się między sobą strukturą punktacji. Za zwycięską strategię została uznana ta, która w całym turnieju zdobyła najwięcej punktów.

Pierwsze symulacje zostały przeprowadzone dla modelu gry określonego wcześniej jako „pełna rywalizacja”. Struktura punktacji była zgodna z przedstawioną wcześniej i wyglądała następująco: wygrana = 5 punktów, remis = 2 punkty, przegrana = 0 punktów, dodatkowa nagroda za zwycięstwo w rozgrywce: po 2 punkty od każdego gracza. W turnieju tym zdecydowanie wygrała strategia wyboru ruchu na podstawie rezultatu ostatniej rozgrywki. Zyskała ona aż prawie dwa razy więcej punktów niż następna w kolejności. Jej algorytm był następujący: w przypadku gdy ostatnia rozgrywka była wygrana wybierany był zestaw ‘A’, który nakazywał powtórzyć ostatni wygrany symbol. Natomiast w przypadku gdy ostatni pojedynek był przegrany symbol zagrywki wybierany był losowo.

Druga symulacja została przeprowadzona dla modelu gry „duży zysk, małe ryzyko” i odbyła się dla następującej punktacji: wygrana = 1 punkt, remis = 1 punkt, przegrana = 0 punktów, dodatkowa nagroda za zwycięstwo jak w poprzednim turnieju. Ten turniej z niewielką przewagą (~110% rezultatu następnego w kolejności) wygrała strategia „naprzemienna” – raz wybierany był symbol „nożyczki”, raz symbol „papier”.

Turniej trzeci przeprowadzony został w oparciu o model gry „duży zysk, duże ryzyko”. W tym modelu struktura punktacji rozkładała się następująco : wygrana = 5 punktów, remis = 5 punktów, przegrana = -1 punkt, dodatkowa nagroda za zwycięstwo w rozgrywce: po 2 punkty od każdego gracza. Turniej wygrała strategia wyboru ruchu na podstawie rezultatu ostatniej rozgrywki. W przypadku gdy ostatnia rozgrywka była wygrana wybierany był zestaw ‘A’, który reprezentował zwycięski symbol z ostatniego pojedynku. Natomiast w przypadku gdy ostatni pojedynek był przegrany symbol zagrywki wybierany był losowo. Również w tym przypadku przewaga zwycięskiej strategii nad następną w kolejności była nie duża i wynosiła ~ 15%.

5. Wnioski

Analizując wyniki symulacji przytoczone w niniejszej pracy należy zwrócić uwagę na fakt, iż aż w dwóch turniejach na trzy zwycięska strategia „posiłkowała się” wyborem losowym. Może to świadczyć o trudności w stawianiu hipotez co przyszłego stanu środowiska rozgrywki, a tym samym wyboru strategii gry.

Podsumowując wyniki badań należy stwierdzić, iż zagadnienie wyboru strategii gry, przedstawione na przykładzie gry „kamień–nożyczki–papier” jest zagadnieniem trudnym i wymagającym zastosowania algorytmów przetwarzających wiedzę (w tym przypadku w postaci wyników poprzednich rozgrywek). Przedstawiona w pracy implementacja algorytmu genetycznego wykazała swoją przydatność w wyborze strategii gry tylko w niektórych modelach rozgrywek. Należy się spodziewać podniesienia efektywności algorytmu poprzez połączenie go z algorytmami przetwarzającymi wiedzę, tworząc w ten sposób system o konstrukcji hybrydowej [6].

Literatura

1. Tyszka T.: *Konflikty i strategie. Niektóre zastosowania teorii gier*, WNT Warszawa 1978.
2. Goldberg D.E.: *Algorytmy genetyczne i ich zastosowania*. WNT Warszawa 1995.
3. Gwiazda T.D., : *Algorytmy genetyczne – wstęp do teorii*. T.D.G Warszawa 1995.
4. Michalewicz Z.: *Algorytmy genetyczne+struktury danych=programy ewolucyjne*. WNT Warszawa 1995.
5. Rudkowska D., Piliński M., Rutkowski L. : *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. PWN Warszawa 1997.
6. Muławka J.J.: *Systemy ekspertowe*. WNT Warszawa 1996.

Streszczenie

W artykule opisane zostały badania nad zastosowaniem algorytmów genetycznych do wyboru strategii gry typu „kamień–nożyczki–papier”. Dla potrzeb badań zostało zrealizowane programowo środowisko testowe, w którym został przetestowany genetyczny algorytm wyboru strategii postępowania.

APPLICATION OF GENETICAL ALGORITHM TO SOLVE THE STRATEGY SELECTION PROBLEM

Summary

In the paper there are presented results of research on application the genetical algorithm to select the strategy of the game "stone–scissors–paper". The results of the research were obtained from a computer programme written especially for this purpose.

Referat recenzowany

ANALIZA WPŁYWU AERODYNAMIKI/HYDRAULIKI NA EFEKTYWNOŚĆ PRACY URZĄDZEŃ PRZEMYSŁOWYCH

1 Uwagi wstępne

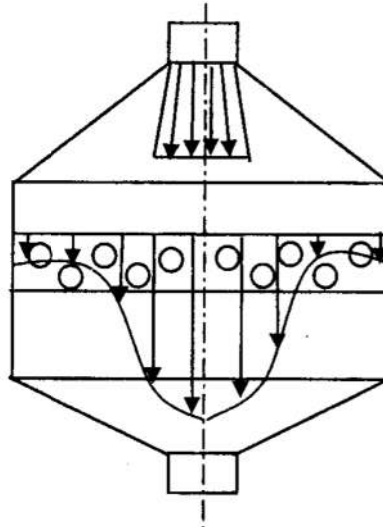
Nowoczesne aparaty technologiczne pracujące w instalacjach wielkotonazowych, niezależnie od tego czy są reaktorami, wymiennikami ciepła, czy specjalizowanymi kolumnami, bardzo często posiadają wymiary poprzeczne porównywalne z ich długością. Często zdarza się również, że długość warstwy roboczej (wysokość złoża katalizatora, rur wymiennika, warstwy cieczy) bywa niekiedy znacznie mniejsza niż jego szerokość. Wpływ nierównomiernego obciążenia powierzchni roboczej aparatu na sprawność lub wydajność pracy tych urządzeń jest powszechnie znany. Pomimo tej wiedzy w praktyce ciągle jeszcze wiele aparatów przemysłowych pracuje przy nierównomiernym obciążeniu. Jednym z ważnych powodów tego stanu rzeczy jest fakt, że wraz ze wzrostem wydajności aparatów rosną wymiary poprzeczne tych urządzeń. Osiągają one w przekroju roboczym średnice rzędu kilku lub nawet kilkunastu metrów. Przy tak znacznych wymiarach trudniejsze jest zapewnienie równomiernego przepływu przez całą powierzchnię roboczą. Efekty uzyskane na wyjściu aparatu przy nierównomiernym obciążeniu wypełnienia są inne niż przy obciążeniu równomiernym i są na ogół obniżone. Często w oparciu o uzyskiwane wyjście prowadzi się proces i uzyskuje trudne do wytłumaczenia efekty polegające na innym od oczekiwanych poziomach sygnałów wyjściowych. Rozwiązania przemysłowe są zabezpieczane przed wpływem nierównomierności w ogólności poprzez zwiększanie objętości wsadu. Oczywiście rozwiązania takie są droższe i mniej konkurencyjne. Celem niniejszej pracy jest przedstawienie wpływu nierównomiernego obciążenia wypełnienia w aparatach zbiornikowych na efektywność ich pracy. Z tym problemem wiąże się również sprawa pomiarów sygnałów wyjściowych z tych urządzeń lub raczej poziomu tych sygnałów. Bardzo często poszukuje się błędów w pomiarach, podczas gdy problem leży niekoniecznie w jakości tych pomiarów. Obniżenie sprawności pracy reaktora, wymiennika ciepła lub innego urządzenia może mieć wpływ na przykład na obniżenie stężenia produktu, niekorzystną zmianą temperatury czynnika procesowego, wzrostem zanieczyszczeń lub niepożądanych komponentów w strumieniu wylotowym. Efekty nierównomiernego obciążenia wypełnień są znane zwłaszcza technologom i operatorom procesów pracującym na instalacjach o wysokiej wydajności i dużych wymiarach geometrycznych.

2 Ocena nierównomierności przepływu płynu przez warstwę wypełnienia

Nierównomierny przepływ płynu (gazu lub cieczy) przez warstwę wypełnienia generowany jest przez nierównomierne naprowadzenie płynu na powierzchnię roboczą, przez geometrię wypełnienia oraz przez elementy umieszczone pod warstwą [1], [2]. Charakter geometrii napływu można kształtować przy pomocy dodatkowych urządzeń rozdzielających lub naprowadzających płyn do przestrzeni roboczej. Poprawne zaprojektowanie urządzeń dystrybucyjnych jest trudne lub niemożliwe przy pomocy tylko metod obliczeniowych, ale możliwe przy łącznym wykorzystaniu metod obliczeniowych i eksperymentalnych.

Na geometrię wypełnienia wpływ ma jego producent (rury wymiennika, tkana siatka) lub w przypadku złożów ziarnistych charakter ten jest przypadkowy zależny od sposobu zasypania lub ułożenia wypełnienia. Na elementy rusztu pod wypełnieniem lub wylotem wpływ ma projektant na etapie projektowania urządzenia.

* Instytut Nawozów Sztucznych, 24-110 Puławy, ul. Al. 1000-lecia PP13A, e-mail: eroj@atena.ins.pulawy.pl



Rys. 1. Schemat strumienia płynu w aparacie zbiornikowym z wypełnieniem
 Fig. 1. Diagram of fluid flow stream in a vessel with packing

Ocenę stopnia nierównomierności przepływu płynu przez powierzchnię roboczą można przedstawić na kilka sposobów. Spotykane w literaturze sposoby obejmują wykorzystanie metod analizy statystycznej [3] lub wykorzystanie pojęcia pędu lub energii kinetycznej stosowanego w mechanice płynów [1], [2]. Dla celów niniejszej pracy wykorzystano arbitralnie podejście stosowane w mechanice płynów. Przedmiotem dalszego zainteresowania będzie więc pole prędkości w przestrzeni roboczej opisane przy pomocy pędu strumienia płynu. W celu obliczenia rzeczywistego pędu strumienia płynu w całym przekroju roboczym aparatu sumuje się pęd elementarnych strumieni płynu, co można zapisać przy pomocy zależności:

$$K_r = \int_F w dm \quad (1)$$

Uwzględniając gęstość płynu ρ oraz zakładając jednorodność i nieściśliwość płynu element masy płynu przepływający przez elementarną powierzchnię można wyrazić zależnością:

$$dm = \rho w dF \quad (2)$$

Podstawiając otrzymaną zależność do wyrażenia na rzeczywisty pęd strumienia płynu otrzymuje się zależność:

$$K_r = \int_F \rho w^2 dF \quad (3)$$

Obliczając średnią wartość pędu strumienia płynu dla przekroju w przestrzeni roboczej można otrzymać zależność:

$$K_{sr} = \rho w_{sr}^2 F \quad (4)$$

Obliczając stosunek pędu rzeczywistego do pędu średniego otrzymuje się wielkość znaną w literaturze pod nazwą współczynnika pola prędkości k_v . Współczynnik pola prędkości można zapisać przy pomocy zależności:

$$k_v = \frac{1}{F} \int_F \left(\frac{w}{w_{sr}} \right)^2 dF \quad (5)$$

Tak zdefiniowany współczynnik osiąga wartość równą jedności przy przepływie równomiernym (łukowym) oraz przyjmuje wartości większe od jedności przy przepływie różnym od przepływu łukowego. W spotykanych warunkach przemysłowych, gdy na etapie projektowania aparatu przepływowego uwzględniano istnienie nierównomierności i przeciwdziałano jej występowaniu, wartość współczynnika pola prędkości zawiera się w przedziale 1.2 – 1.5. W innych przypadkach wartość współczynnika pola prędkości może przyjmować nawet wartości większe od 2, co oznacza istnienie obszarów o charakterze zastoin.

3 Reaktor z reakcją chemiczną

Rozpatrzmy reaktor z reakcją chemiczną rzędu I zachodzącą w obszarze dyfuzyjnym w złożu katalizatora o wysokości H . Zależność stężenia produktu na wylocie z reaktora można przedstawić przy pomocy zależności:

$$C = C_0(1 - e^{-kt}) \quad (6)$$

Zakładając w uproszczeniu, że rozkład prędkości w złożu jest nierównomierny i stały dla całej objętości złoża, można przedstawić czas kontaktu w złożu jako funkcję prędkości zależną od położenia, czyli $t(F) = H/w(F)$. Podstawiając $t(F)$ do zależności (6), otrzymuje się zależność:

$$C = C_0(1 - e^{-k \frac{H}{w(F)}}) \quad (7)$$

Całkowity efekt na wyjściu z reaktora można obliczyć sumując efekty od poszczególnych udziałów elementarnych reaktorów o przekroju dF , czyli:

$$C_{wy-n} = \frac{1}{F} \int_F C(F) dF = \frac{1}{F} \int_F C_0(1 - e^{-k \frac{H}{w(F)}}) dF \quad (8)$$

W miejsce funkcji rozkładu prędkości $w(F)$ w równaniu (9) można podstawić wyrażenie otrzymane poprzez wprowadzenie średniej według wydatku (nie według powierzchni) zgodnie z zależnością:

$$w_k = \frac{\int_F w^2 dF}{\int_F w dF} = \frac{\int_F w^2 dF}{w_{sr} F} = w_{sr} k_v \quad (9)$$

Uwzględniając zależność (9) i zależności (8), otrzymuje się zależność na efekt reakcji w obecności nierównomierności:

$$C_{wy-n} = \frac{1}{F} \int_F C_0(1 - e^{-k \frac{H}{w_{sr} k_v}}) dF = C_0(1 - e^{-k \frac{H}{w_{sr} k_v}}) \quad (10)$$

Gdyby przepływ w złożu był równomierny, wówczas efekt na wyjściu z reaktora byłby zgodny z zależnością:

$$C_{wy-r} = C_0(1 - e^{-k \frac{H}{w_{sr}}}) \quad (11)$$

Stosunek stężeń obliczonych według zależności (10) i (11) jest nie większy od jedności, czyli $C_{wy-n}/C_{wy-r} \leq 1$. Oznacza to, że przy nierównomiernym obciążeniu wypełnienia obserwuje się obniżenie stężenia wyjściowego w stosunku do pracy przy obciążeniu równomiernym.

Reaktor utleniania amoniaku

W procesie wytwarzania kwasu azotowego występuje węzeł utleniania amoniaku. W reaktorze należącym do tego węzła zachodzi proces utleniania amoniaku do tlenku azotu według reakcji:



Współczesne reaktory do utleniania amoniaku są aparatami o dużej średnicy rzędu kilku metrów. Katalizator do utleniania amoniaku stanowi pakiet siatek tkanych lub dzianych wykonanych z drutu o średnicy 60 μm lub 70 μm (rzadziej stosuje się siatki o większych średnicach drutu). Pakiet siatek umieszczony jest w przestrzeni roboczej reaktora na powierzchni usypanej z pierścieni ceramicznych. Przepływająca przez pakiet siatek mieszanina gazowa składająca się z amoniaku i powietrza reaguje na powierzchni katalizatora tworząc tlenki azotu.

Duża średnica reaktora stwarza problemy aerodynamiczne związane z nierównomiernym rozproszaniem płynu na powierzchni katalizatora. Nierównomierność napływu generowane są przez elementy doprowadzające gaz (nagła zmiana średnic rur doprowadzających gaz w stosunku do przestrzeni roboczej, kolana), a także różne dystrybutory umieszczone nad powierzchnią katalizatora. Elementem generującym nierównomierny napływ może być również niejednorodna struktura złoża umieszczona pod pakietem siatek katalitycznych. Nierównomierny przepływ przez warstwę katalizatora powoduje jego nierównomierne obciążenie. Ma to wpływ na sprawność pracy węzła utleniania a także na straty katalizatora. Ze względu na duże koszty amoniaku celowym jest uzyskiwanie możliwie najwyższych sprawności, a ze względu na koszty katalizatora – zapewnienie możliwie najniższych jego strat. Zapewnienie równomiernego obciążenia siatek katalizatora jest ważnym zadaniem w ramach działań zmierzających do intensyfikacji i obniżenia kosztów produkcji kwasu azotowego.

Wpływ nierównomiernego przepływu na sprawność utleniania amoniaku

Do rozważań wykorzystano model reaktora obejmujący między innymi opis sprawności utleniania amoniaku z uwzględnieniem istotnych parametrów wpływających na jej wartość: stężenia amoniaku w mieszaninie NH_3 –powietrze, obciążenia siatek katalizatora, temperatury drutów siatek, ilości siatek w pakiecie, a także straty metali szlachetnych związane z procesem utleniania amoniaku przy wysokiej temperaturze [7]. Do celów badania wpływu nierównomiernego obciążenia katalizatora na sprawność utleniania amoniaku wykorzystano korelację zaproponowaną w pracy [4] o postaci:

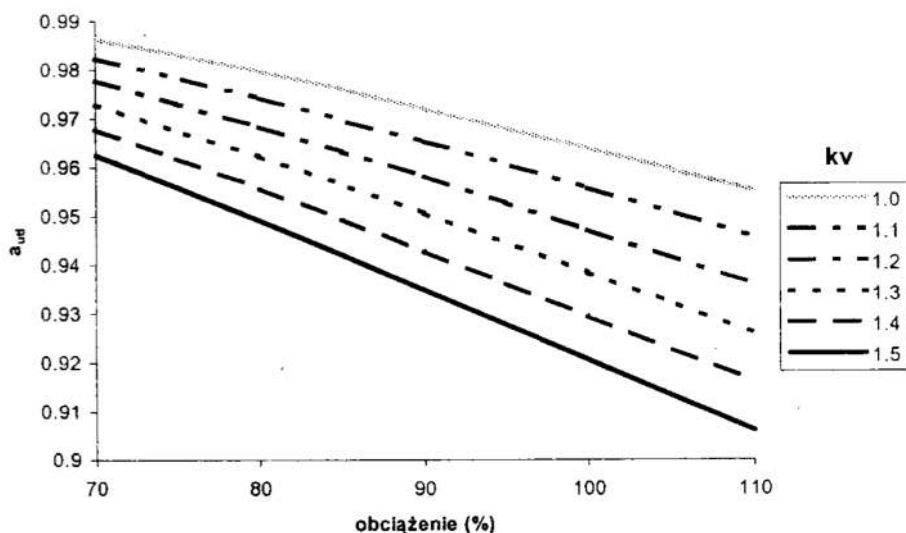
$$\alpha_{\text{utl}} = 1 - \exp\left(-83,149 \cdot k \cdot 10^{-6} \sqrt[3]{\frac{n_s^2 \cdot T_g^{1.56} \cdot D^4 \cdot C_{\text{NH}_3}^2}{m_{\text{NH}_3}^2}}\right) \quad (13)$$

Wykorzystując zależność (13) opisującą sprawność utleniania amoniaku, zależność opisującą współczynnik pola prędkości (5) oraz zależność (9), można po prostych przekształceniach otrzymać zależność:

$$\alpha_{\text{utl}} = 1 - \exp\left(-\frac{A}{\sqrt[3]{(k_v m_{\text{NH}_3})^2}}\right) \quad (14)$$

Otrzymana zależność pozwala obliczyć obniżenie sprawności utleniania amoniaku w warunkach istnienia nierównomierności. Pozwala również badać kompensujący wpływ pozostałych zmiennych na utrzymanie sprawności utleniania na stałym poziomie. Na podstawie uzyskanej zależności można stwierdzić, że nierównomierne obciążenie wpływa ujemnie na sprawność utleniania amoniaku, a tym samym na wzrost kosztów wytwarzania kwasu azotowego. Oczywiście można wpływać na zachowanie sprawności

utleniania amoniaku na zadanym poziomie. Obliczenia wykazują, że na istniejącym reaktorze można zachować stałą sprawność poprzez zwiększenie ilości siatek lub poprzez zwiększenie stężenia amoniaku w mieszaninie amoniakalno-powietrznej. Oba sposoby wpływają na wzrost kosztów produkcji. Pierwszy wpływa na wzrost zamrożonego kapitału, a drugi poprzez wzrost temperatury, wpływa na wzrost strat katalizatora. Na załączonym rysunku pokazano wpływ stopnia nierównomierności przepływu gazu przez pakiet katalityczny na spadek sprawności utleniania amoniaku.



Rys. 2. Wykres sprawności utleniania amoniaku w zależności od obciążenia reaktora i współczynnika pola prędkości

Fig. 2. Ammonia oxidation efficiency vs reactor load and coefficient of flow velocity

Wpływ nierównomiernego przepływu gazu na straty katalizatora

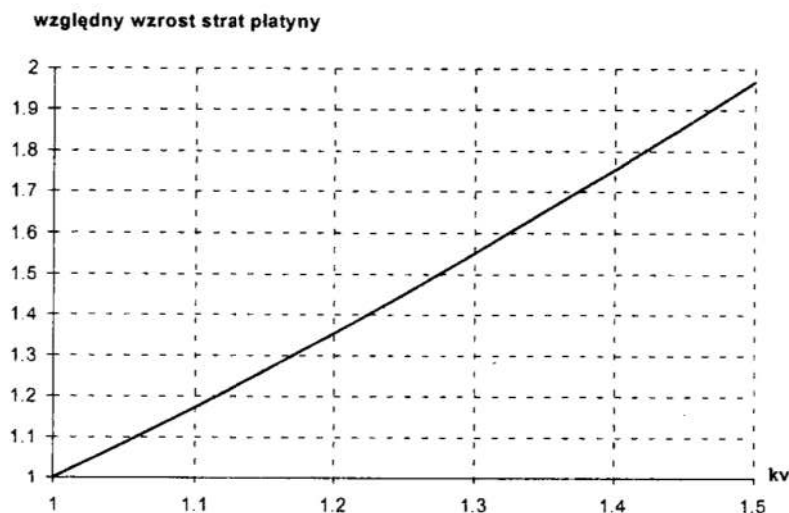
Do obliczenia strat platyny wykorzystano równanie [6] o postaci:

$$\frac{dG_{Pt}}{d\tau} = 19,0476 \frac{T \cdot M \cdot G_{NH_3} \cdot K \cdot y_{O_2}}{P \cdot y_{NH_3}} Re^{0,67} \quad (15)$$

W celu określenia wpływu nierównomiernego przepływu gazu przez pakiet katalityczny na straty metali szlachetnych wykorzystano zależność podaną wyżej. Wykorzystując podane wyżej równania (15) oraz równania (5) i (9), otrzymuje się zależność:

$$\frac{dG_{Pt}}{d\tau} = B \cdot (k_v \cdot w_{sr})^{1,67} \quad (16)$$

Przedstawiona zależność informuje, że straty katalizatora zależą dość silnie od współczynnika pola prędkości. Istnieje jednakże możliwość przeciwdziałania tym wpływom poprzez zmniejszenie temperatury drutów siatek czyli poprzez obniżenie stężenia amoniaku w mieszaninie amoniakalno-powietrznej oraz poprzez wzrost ciśnienia w instalacji lub także poprzez możliwość zmiany zawartości tlenu i amoniaku w mieszaninie reakcyjnej.



Rys. 3. Wykres względnego wzrostu strat katalizatora platynowego w zależności od współczynnika pola prędkości

Fig. 3. Diagram of relative platinum catalyst loses vs flow velocity coefficient

4 Wymiana ciepła

W procesach wymiany ciepła współczynniki wnikania ciepła zależą między innymi od prędkości strumienia oraz własności badanego procesu [5]. Ogólnie można zapisać, że współczynnik wnikania ciepła jest proporcjonalny do prędkości liniowej w potęgze mniejszej od jedności, czyli $\alpha = kw^n$. Zwykle wartość wykładnika potęgi zmienia się w zakresie 0,5–0,9 i zależy od rodzaju wypełnienia. W przypadku nierównomiernego rozkładu prędkości średnia wartość współczynnika wnikania ciepła wynosi zgodnie z zależnością:

$$\alpha_n = \frac{k}{F} \int_F w^n dF = \frac{k w_{sr}^n}{F} \int_F \left(\frac{w}{w_{sr}} \right)^n dF \quad (17)$$

Współczynnik obniżenia efektywności wymiany ciepła można wyrazić poprzez stosunek współczynnika wnikania obliczonego dla przepływu równomiernego przy prędkości równej prędkości średniej dla całego przekroju roboczego oraz współczynnika wnikania obliczonego przy przepływie nierównomiernym α_n , czyli $\eta = \alpha_n / \alpha$. Podstawiając powyższe wyrażenia do zależności (17), otrzymuje się zależność:

$$\eta = \frac{1}{F} \int_F \left(\frac{w}{w_{sr}} \right)^n dF \quad (18)$$

Oczywiście proces wymiany ciepła pogorszy się w zależności od stopnia nierównomierności pola prędkości oraz od rodzaju wypełnienia, rodzaju przepływu, które określa wartość wykładnika potęgi n . W przypadku wymiennika ciepła może to wpłynąć zarówno na rozkład temperatur w wymienniku jak również na poziomy temperatur czynnika chłodzącego i grzejącego. W przypadku złożeń ziarnistych wpływ ten uwidoczni się lub może uwidocznic na przykład poprzez zmianę efektywności wymiany ciepła w procesie nagrzewania złoża w procesie rozruchu reaktora złożowego.

Z przypadkiem takim można się spotkać dla przykładu podczas pierwszego rozruchu reaktora do syntezy amoniaku po wymianie katalizatora, podczas którego zachodzi proces redukcji złoża. Złe przeprowadzenie redukcji ma później istotny wpływ na cały okres pracy katalizatora.

5 Podsumowanie

Przeprowadzone rozważania pokazują na możliwość wystąpienia niekorzystnych zjawisk w procesach przemysłowych spowodowanych poprzez nierównomierne obciążenia wsadów aparatów przemysłowych o dużych wymiarach geometrycznych zwłaszcza poprzecznych. W zależności od rodzaju rozważanego procesu oraz rodzaju wypełnienia wpływ ten może być silniejszy lub słabszy. Przeprowadzone obliczenia symulacyjne wykazały, że dla spotykanych w dużych reaktorach utleniania amoniaku nierównomierności, możliwe jest obniżenie sprawności utleniania nawet o kilka procent. Jednocześnie straty katalizatora mogą osiągnąć nawet kilkudziesięciu procentowy wzrost w stosunku do strat przy obciążeniu równomiernym. Można przeciwdziałać skutkom nierównomiernego obciążenia wypełnienia aparatu poprzez zwiększenie objętości wsadu, co jest powszechnie praktykowane w postaci różnych współczynników bezpieczeństwa. W przypadku katalizatora do utleniania amoniaku oznacza to zwiększenie ilości siatek. Jednakże zwiększenie ilości siatek wiąże się ze wzrostem zamrożonego kapitału z jednej strony oraz należy się liczyć z wystąpieniem niekorzystnych efektów technologicznych z drugiej strony (reakcja utleniania amoniaku może pójść w niepożądanym kierunku). Korzystne jest zatem zapewnienie możliwie równomiernego obciążenia wypełnienia aparatów przemysłowych.

Wykaz symboli

- A, B – stałe,
C, C₀ – stężenie na wylocie i stężenie równowagowe, %,
C_{NH₃} – stężenie mieszanki amoniakalno-powietrznej, % obj.,
dF – elementarna powierzchnia, m².
dm – masa płynu przepływająca w jednostce czasu przez elementarną powierzchnię, kg/s.
D – średnica reaktora, m,
G_{NH₃} – natężenie przepływu amoniaku przez reaktor, mol/s,
G_{Pt} – straty platyny, g/t_{HNO₃},
H – wysokość złoża katalizatora, m
k – współczynnik charakteryzujący geometrię siatki,
k_v – współczynnik pola prędkości,
K – współczynnik określający parametry konstrukcyjne reaktora,
K_r – rzeczywisty pęd strumienia płynu, kg m/s²,
K_{śr} – średnia wartość pędu strumienia płynu, kg m/s²,
m_{NH₃} – obciążenie reaktora w przeliczeniu na amoniak, t_{NH₃}/d,
M – średnia masa molowa materiału siatki, kg/kmol,
n_s – ilość siatek,
P – ciśnienie, ata,
t – czas, s,
T – temperatura drutów siatki, K,
T_g – temperatura gazów poreakcyjnych, K,
Re = $\frac{68.12208G_{NH_3}}{\pi D \eta_G y_{NH_3}}$, liczba Reynoldsa,
w, w_{śr} – prędkość strumienia płynu przepływającego przez elementarną powierzchnię dF, m/s, oraz średnia prędkość przepływu płynu przepływającego przez powierzchnię roboczą, m/s,
Y_{O₂}, Y_{NH₃} – stężenie tlenu i amoniaku w mieszaninie NH₃-powietrze, ułamek molowy,

- α, α_n – współczynniki wnikania ciepła, W/m²K,
 α_{utl} – sprawność utleniania amoniaku, %,
 η – współczynnik efektywności wymiany ciepła,
 η_G – lepkość dynamiczna mieszaniny gazowej w temperaturze T, kg/ms,
 ρ – gęstość płynu, kg/m³.

Literatura

1. Idelcik I.E.: *Aerodinamika promyshlennyh apparatov*. Energia, Moskva, 1964.
2. Idelcik I.E.: *Aerogirodinamika tehnologicheskikh apparatov*. Mashinostrojenie, Moskva, 1983.
3. Ajzenbud M.B., Krasnushkina N. V.: *Teor. osnovy khim. tekhnol.* No 4, 616, 1977.
4. Borisova R.G., Hamaha L.N.: *Khim. Prom.* 11, s. 44–46, 1976.
5. Hobler T.: *Ruch ciepła i wymienniki*. WNT, 1979.
6. Vjatkin Ju. A., Savenkov A.S. i inni: *Khim. Prom.* 10, s. 42–44, 1979.
7. Rój E.: *Modelowanie wpływu nierównomiernego przepływu gazu na efektywność pracy reaktora utleniania amoniaku w instalacji kwasu azotowego*. VII Sympozjum Naukowo–Techniczne na temat „Inżynieria reaktorów chemicznych”, Ustroń, 1996.

Streszczenie

W pracy omówiono wpływ nierównomiernego obciążenia aparatów przemysłu chemicznego o dużych wymiarach poprzecznych w stosunku do wymiarów poosiowych na efektywność ich pracy. Problem zilustrowano na prostym przykładzie reaktora z reakcją I rzędu, reaktora do utleniania amoniaku, prostego procesu wymiany ciepła. Uzyskane wyniki pokazują, że istnieje wpływ nierównomiernego obciążenia wsadu aparatu na efektywność jego pracy. Na załączonych rysunkach pokazano wpływ nierównomiernego pola prędkości na sprawność utleniania amoniaku oraz na bezpowrotne straty katalizatora.

ANALYSIS OF THE AERODYNAMIC/HYDRAULIC INFLUENCE ON THE EFFICIENCY OF INDUSTRIAL EQUIPMENT PERFORMANCE

Summary

Nonuniform flow of fluid through industrial vessels of high diameters comparing their longitudinal dimensions and its influence on performance of the vessels has been discussed. The problem has been illustrated by means of simple examples: reactor with first order reaction, ammonia oxidation reactor and heat transfer model. Obtained results shows that nonuniform flow of fluid through packing strongly influence its efficiency and depends on type of the packing. Some simulation results demonstrating an influence of nonuniform gas flow on ammonia efficiency oxidation and irrecoverable catalyst losses have been presented in the attached figures.

Referat recenzowany

KOMPUTEROWE WSPOMAGANIE DOBORU PARAMETRÓW ŻEGLUGI STATKÓW ŻAGLOWO-MOTOROWYCH

1 Wprowadzenie

Wiatr od wieków był stosowany jako siła napędowa w statkach morskich. W ostatnich latach wraca się do napędu żaglowego podejmując próby budowy i użytkowania statków o napędzie żaglowym. Jak na razie tylko na niewielką skalę, budując statki szkolne lub badawcze np. statek badawczy Polskiej Akademii Nauk „OCEANIA”, czy statek szkolny Wyższej Szkoły Morskiej w Gdyni „Dar Młodzieży”.

Ze względów bezpieczeństwa ludzi, statku i ładunku żagle nie mogą być jedynym źródłem napędu statku. Przy niekorzystnych warunkach pogodowych statki żaglowe nie mogą czekać na zmianę kierunku czy siły wiatru. Muszą więc wspomagać się silnikami spalinowymi. Ponadto, w sprzyjających warunkach, podczas żeglugi na silniku, stosuje się jednocześnie wspomaganie żaglami. W tych różnych sytuacjach żeglugowych podjęcie decyzji, jaki przyjąć kurs, jakie żagle postawić oraz jak dobrać optymalne nastawy parametry pracy silnika jest zadaniem wyjątkowo trudnym. Do tej pory nie zajmowano się optymalizacją ruchu tego typu statków, ze względu na stosunkowo niewielką ich liczbę. Natomiast optymalizacją doboru kursu i odpowiednich żagli dla danych warunków pogodowych zajmowano się tylko w stosunku do jachtów regatowych, biorących udział w prestiżowych regatach, na przykład w Regatach o Puchar Ameryki. Istniejące komercyjne programy komputerowe służące do doboru odpowiedniego kursu względem wiatru i odpowiedniego zestawu żagli na jachtach regatowych uwzględniają tylko parametry konstrukcyjne jachtu. Ich istota zasadza się na tabelach zawierających zestawienie teoretycznej maksymalnej prędkości jachtu, wyliczonej dla danych warunków pogodowych przez konstruktora jednostki.

Metody stosowane do optymalizacji kosztów ruchu statków z klasycznym napędem, jak również metody i narzędzia stosowane w żeglarskim regatowym nie nadają się do zastosowania dla statków żaglowych. Optymalizacja ich ruchu wymaga opracowania nowych specjalnych metod.

W artykule przedstawiono wyniki badań przeprowadzonych w czasie realizacji grantu uczelnianego na statku żaglowym sts „POGORIA”. Statek ten ma wyporności 342 tony, długość 47 metrów i szerokość 8 metrów, posiada ożaglowanie typu barkentyna (3 maszty w tym 1 z żaglami rejowymi) o powierzchni 1000 m² (razem 15 żagli). Pomocniczy silnik napędowy ma moc 228 kW. Tego rodzaju statek jest jednostką specyficzną. Charakteryzuje się on następującymi cechami:

- krótkim czasem żeglugi na silniku (do 48 godzin ciągłego pływania),
- częstym wspomaganiami pracy silnika żaglami,
- dużym wpływem wysokości i kierunku fali na prędkość statku,
- dużym wpływem siły i kierunku wiatru na prędkość statku,
- specyficznym rodzajem żeglugi (szkolenie młodzieży).

* Wyższa Szkoła Morska w Gdyni, Katedra Podstaw Techniki, 81-225 Gdynia, ul. Morska 83, tel. (58) 62 89 331, e-mail: rudz@wsm.gdynia.pl.

Na podstawie wyników pomiarów podjęto próbę stworzenia modelu matematycznego układu napędowego statku żaglowego z pomocniczym napędem silnikiem spalinowym ze śrubą nastawną. Model ten powinien pozwolić na optymalizację wielokryterialną współpracy wymienionego układu, w tym optymalizację zużycia paliwa a także drogi lub czasu żeglugi w określonych warunkach pływania.

Model ten będzie wykorzystany w optymalizatorze będącym częścią komputerowego systemu do wspomagania decyzji oficera wachtowego podczas normalnej eksploatacji statku.

2 Model układu napędowego ze śrubą nastawną

Jednym z zadań prowadzonych badań jest stworzenie prototypu komputerowego systemu wspomagającego oficera wachtowego w podejmowaniu decyzji m.in. w doborze najkorzystniejszych nastaw listwy paliwowej silnika i odpowiedniego kąta nachyleniach płatów śruby nastawnej w sytuacji, kiedy zastosowanie napędu spalinowego będzie zalecane dla panujących warunków pogodowych.

W badaniach założono, że model matematyczny napędu silnikiem i napędu silnikiem z wspomaganiami żaglami, zostanie zbudowany w oparciu o sieć neuronową. Aby tego dokonać należało pozyskać i przygotować odpowiednie zbiory uczące. W tym celu wybrano szereg parametrów niezbędnych do budowy tej sieci.

W optymalizacji zużycia paliwa a także drogi lub czasu żeglugi w określonych warunkach pływania przyjęto dwa kryteria: prędkość jednostki v_s i zużycie paliwa g_p . Prędkość statku i zużycie paliwa są funkcjami wielu zmiennych, takich np. jak wysokość fali, prędkość i kierunek wiatru, postawione żagle itp. W sumie przyjęto 26 wielkości, od których zależą kryteria i które są możliwe do zaobserwowania.

$$v_s = f(x_1, x_2, \dots, x_i); \quad i = 26 \quad (1)$$

$$g_p = f(x_1, x_2, \dots, x_i); \quad i = 26 \quad (2)$$

Zbiory uczące pozyskano podczas rejsu badawczego gdzie mierzono i obserwowano następujące parametry:

- moment skręcający na wale napędowym [działki] (później przeliczano na Nm),
- prędkość obrotową silnika [obr/min],
- nastawy listwy paliwowej [działki],
- nastawy skoku śruby [działki],
- prędkość statku [węzły],
- prędkość i kierunek wiatru [m/s] i [°],
- ocena stanu morza (wg. skali Beauforta) [°B],
- wyliczona prędkość i kierunek prądu [m/s] i [°],
- kurs kompasowy i kurs względem dna [°],
- chwilowe zużycie paliwa [dm³/min],
- postawione żagle.

W sumie rejestrowano 28 wielkości (w tabeli nr 1 pokazano przykładowe wartości rejestrowanych wielkości).

Tabela 1. Przykładowe dane eksperymentalne
Table 1. Example of experimental data

Moment [dziaki]	Prędkość obr. [obr./min]	Listwa paliwowa [dziaki]	Skok śruby [dziaki]	v [w]	v wiatru [w]	Kierunek wiatru [°]	Stan morza	v prądu [w]	Kierunek prądu [°]	KDD [°]	KK [°]	Zużycie paliwa [l/h]	Postawione żagle (w kolejności od dziobu)																		
45	1200	10,0	16	6,9	18,4	-36	1	0	0	285	279	25,8	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	
45	1200	10,0	16	7,1	18,9	-36	1	0	0	285	281	25,8	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	
44	1200	10,0	16	6,7	18,7	-37	1	0	0	285	280	25,5	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	
30	1200	10,0	13	8,0	15,3	-63	1	0	0	245	241	24,5	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	
43	1200	10,0	17	9,0	14,0	130	2	0	0	180	171	32,0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
23	1200	8,0	10	4,9	3,0	-30	1	0	0	90	84	20,0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	1100	8,0	12	5,0	4,0	-30	1	0	0	70	78	21,2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	1350	12,0	14	6,5	6,5	-20	2	0	0	65	70	26,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	1350	12,0	14	5,5	10,2	-20	3	0	0	65	70	28,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...																			

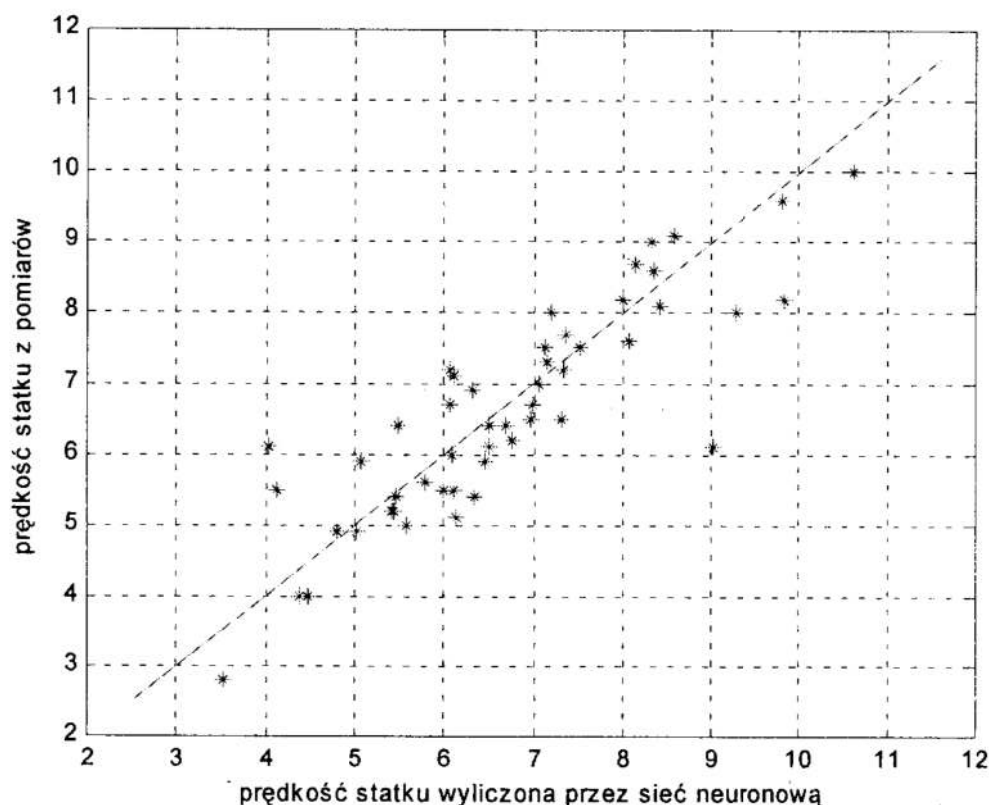
W czasie rejsu na sts „POGORIA” dokonano ponad 800 obserwacji pomiarowych w różnych warunkach pogodowych, na różnych akwenach i przy różnych obciążeniach silnika głównego.

Informacje uzyskane w trakcie badań zastosowano w procesie nauczania sieci neuronowej. Dane poddano liniowej normalizacji w przedziale 0.1 do 0.9. Postawione żagle opisywano w sposób dyskretny, osobno dla każdego z 15 żagli (1 – dany żagiel postawiony, 0 – bez tego żagla), co miało istotny wpływ na wyniki uzyskiwane podczas uczenia sieci. Analiza uzyskanych wyników pokazała, że lepsze rezultaty w nauczaniu sieci uzyskano po wzbogaceniu danych o powierzchnię postawionych żagli. Być może wskazane będzie również uwzględnienie efektywności poszczególnych żagli lub wyznaczenie siły uciągu dla poszczególnych żagli. Eksperymentowano z różnymi konfiguracjami sieci. Ostatecznie zamodelowano dwie sieci neuronowe o różnej strukturze, jedną do wyznaczania zużycia paliwa a drugą do wyznaczania prędkości statku. W warstwie wejściowej było po 26 neuronów a w warstwie wyjściowej po 1 neuronie.

Podjęto również próbę stworzenia modelu matematycznego metodami statystycznymi (analiza regresji) jednak metody te nie pozwoliły uzyskać wystarczająco dokładnego modelu. Natomiast sieć neuronowa dała zadowalającą dokładność obliczania prędkości i zużycia paliwa na badanej jednostce.

Zestawienie wyników uzyskanych z pomiarów i wyników uzyskiwanych przez sieć przy obliczaniu prędkości statku przedstawiono na rysunku 1.

Oszacowana dokładność wyników uzyskanych przez sieć neuronową przy wyznaczaniu prędkości statku wyniosła średnio ok. 17%, a dla zużycia paliwa ok. 14%.



Rys. 1. Wyniki uzyskane podczas uczenia sieci neuronowej (dla 50 punktów pomiarowych)
 Fig. 1. Outputs resulting from teaching of neural nets (for 50 test points)

3 Optymalizacja kursu

Informacje uzyskane w trakcie badań zastosowano również do budowy modelu optymalizacyjnego doboru kursu i parametrów układów napędowych statku żaglowego. Wykorzystano do tego celu metodę programowania dynamicznego i metodykę sieci neuronowych.

Dla opracowania algorytmu optymalizacji kursu i nastaw układów napędowych, przy tworzeniu systemu doradczego, przyjęto następujące założenia:

1. optymalizowana jest tylko żegluga na wiatr (od bajdewindu do półwiatru),
2. kąty martwe (kąty pomiędzy kursem statku a kierunkiem wiatru, w których żagle nie dają ciągu) dla lewego i prawego halsu mogą się różnić,
3. przy żegludze wspomaganej silnikiem kąt martwy zmienia się w zależności od siły wiatru i nastaw silnika i śruby,
4. uwzględniane są prognozy pogody,
5. mogą być narzucane niektóre punkty, przez które mamy płynąć,
6. wykluczane są pewne rejony żeglugi (zabronione dla żeglugi lub omijane z jakiś innych względów),
7. uwzględnia się możliwości fizyczne załogi (czasami jako załoga szkolna pływa młodzież niepełnosprawna).

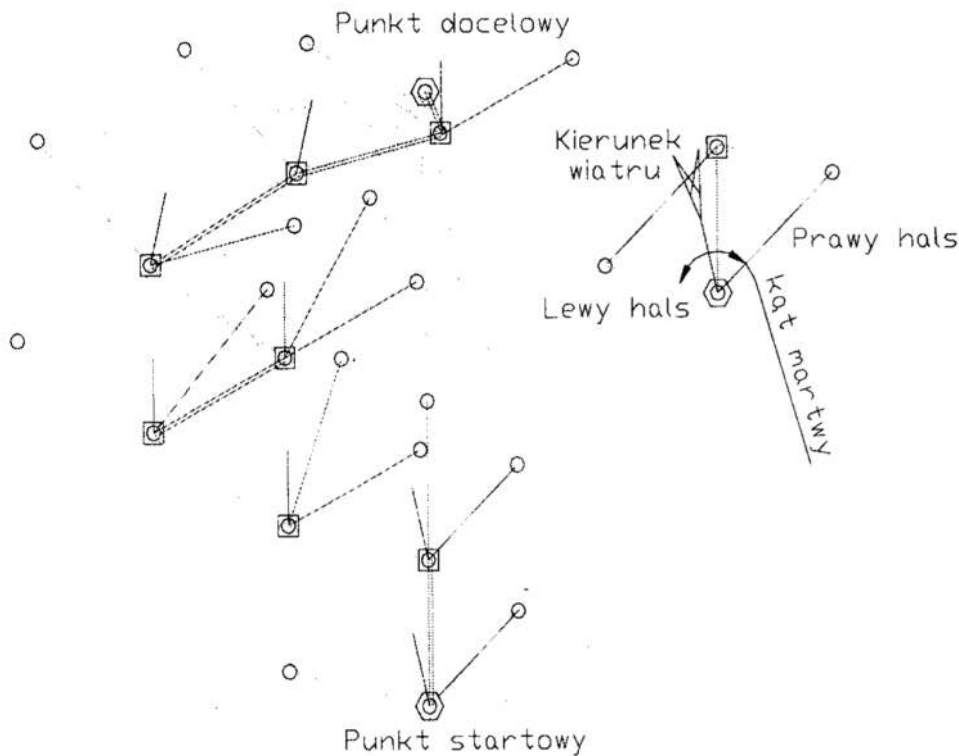
8. zmiany kursu na bezpiecznym akwenuje wykonuje się nie częściej, niż co T minut (przy żegludze na żaglach minimalny czas zwrotu nie mniejszy niż $T=20$ minut).
9. uwzględnia się trzy warianty napędu:
 - tylko żagle,
 - żagle i silnik,
 - tylko silnik.

Do budowy prototypu systemu doradczego zastosowano dwa języki programowania. Sieć neuronowa została zamodelowana w języku MATLAB, natomiast moduł dialogowy i system ekspertowy będą oprogramowane w języku PROLOG LPA.

Przy rozpatrywaniu żeglugi pod wiatr uwzględniano w każdym punkcie decyzyjnym 5 możliwych wariantów żeglugi:

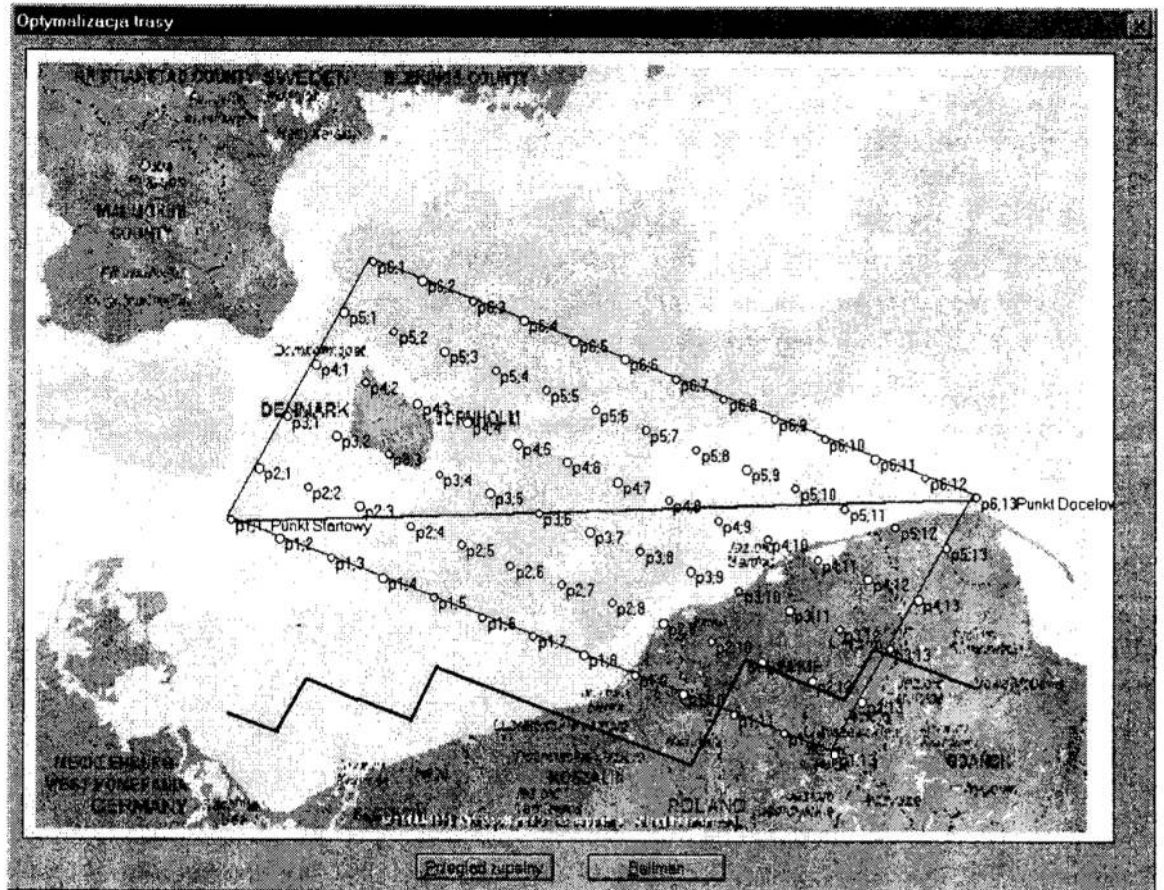
1. żegluga na silniku bez żagli najkrótszą drogą do celu,
2. żegluga lewym halssem wspomagana silnikiem,
3. żegluga prawym halssem wspomagana silnikiem,
4. żegluga tylko na żaglach lewym halssem,
5. żegluga tylko na żaglach prawym halssem.

Rysunek nr 2 pokazuje punkty decyzyjne z możliwymi do wybrania kursami przy ograniczeniu możliwości do 3 wariantów (żegluga na silniku najkrótszą drogą do celu, żegluga na żaglach lewym halssem lub prawym halssem).



Rys. 2. Punkty decyzyjne z siatką przykładowej drogi od startu do celu
 Fig. 2. Decision-making points together with nets of a hypothetical way from a start to a finish

Do tej pory został zbudowany wstępny moduł znajdujący jedynie drogę metodą przeglądu zupełnego przy niezbyt dużej ilości punktów decyzyjnych. Rysunek 3 pokazuje wygląd ekranu programu komputerowego wykonanego w języku prolog, z naniesioną na tle mapy siatką punktów decyzyjnych i drogą optymalną wyznaczoną metodą przeglądu zupełnego.



Rys. 3. Ekran systemu doradczego z mapą rejonu pływania (wersja wstępna).
 Fig. 3. A screen of the expert system including a chart of sailing region (preliminary version)

4 Wnioski

1. W początkowej fazie prac nad optymalizacją żeglugi statków żaglowych nie zakładano konieczności stosowania systemu doradczego, ale prace dotychczas wykonane potwierdziły konieczność zastosowania takiego systemu. System ekspertowy musi być wyposażony w dość rozbudowaną bazę wiedzy, konieczne jest uwzględnianie bardzo wielu czynników np. dobór odpowiedniego zestawu żagli nie może być warunkowany tylko względami ekonomicznymi tj. czasem żeglugi do celu i zużyciem paliwa. Trzeba uwzględniać też kwalifikacje załogi i jej formę fizyczną (wpływ choroby morskiej, zmęczenie itp.) np. postawienie grot żagla wymaga od załogi dużego wysiłku fizycznego, a zrzucenie tego żagla, przy wzroście siły wiatru powyżej 6°B może być bardzo trudne, a nawet niebezpieczne dla załogi lub może zakończyć się zniszczeniem żagla. Wiele decyzji na żaglowcach jest podejmowanych

na podstawie doświadczenia kapitana lub oficerów wachtowych zgodnie z tzw. „dobrą praktyką morską”. Konieczne jest więc zgromadzenie w komputerze wiedzy wielu doświadczonych żeglarzy i stworzenie modułu pozwalającego na efektywne wykorzystanie tej wiedzy przez użytkownika.

2. Osobny problem stanowi wiarygodność prognoz pogody, które trzeba brać pod uwagę przy obliczaniu kosztów żeglugi. Wiarygodność komunikatów meteorologicznych zależy w dużym stopniu od stacji, z których się je odbiera, od długości okresu, jaki obejmują, od rejonu pływania i od pory roku. Na niektórych akwenach zamiast prognozy pogody odbierane są mapy synoptyczne i oficer lub kapitan musi samodzielnie przeanalizować mapę i przewidzieć pogodę. Interpretacje mogą być różnie w zależności od wiedzy i doświadczenia dowodzącego, więc konieczne jest uwzględnienie jego doświadczenia (np. korzystanie przez program komputerowy ze stworzonej przez armatora bazy informacji o kapitanach i oficerach, o ich kompetencjach i doświadczeniu).
3. Dużym problemem jest zastosowanie metod optymalizacji dynamicznej (np. metody Bellmana) ze względu na zmienność w czasie kosztów przejścia pomiędzy punktami decyzyjnymi np. prędkość i kierunki prądów pływowych zależą od czasu i miejsca, w którym się znajdziemy a to jest zależne od wcześniej podejmowanych decyzji.
4. Optymalizację przeprowadza się ze względu na dwa kryteria: koszt żeglugi i czas jej trwania i użytkownik systemu musi mieć możliwość decydowania, w trakcie optymalizacji, o współczynniku wagi poszczególnych kryteriów. Ze względu na rodzaj uprawianej żeglugi przez badany statek przydatna byłby możliwość automatycznego dobierania tych współczynników, po podaniu przez użytkownika przedziału czasu, jakim dysponuje na przebycie danego odcinka drogi.

Praca nad optymalizacją żeglugi statków z napędem żaglowym jest kontynuowana.

Literatura

1. Bellman R.: *Adaptacyjne procesy sterowania*. Państwowe Wydawnictwo Naukowe, Warszawa 1965.
2. Brzózka J., Dorobczyński L.: *Programowanie w Matlab*. Zakład Nauczania Informatyki "MIKOM", Warszawa 1998.
3. Jurdziński M.: *Nawigacyjne planowanie podróży*. Wydawnictwo Morskie, Gdańsk 1989.
4. Zalewski A., Cegiela R.: *Matlab - obliczenia numeryczne i ich zastosowania*. Wydawnictwo Nakom, Poznań 1996.

Streszczenie

W artykule przedstawiono badania dotyczące parametrów żeglugowych statków żaglowych wspomaganych silnikiem spalinowym. Podjęto próbę budowy modelu matematycznego układu napędowego takiego statku. Model ten będzie zastosowany do optymalizacji procesu podejmowania decyzji oficera wachtowego podczas eksploatacji tego typu statków.

COMPUTER-AIDED SELECTION OF SAILING PARAMETERS FOR SAIL-MOTOR VESSELS

Summary

Sailing parameters of sail-motor vessels has been studied. An attempt of building the mathematical model for their propulsion system has been carried out. The model will be used to optimize decision-making process of the officer-in-watch during operational use of these kind of vessels.

Referat recenzowany

EWOLUCYJNA METODA ZNAJDOWANIA ROZWIĄZAŃ DOPUSZCZALNYCH

1 Sformułowanie problemu

Dla szerokiej klasy projektowanych obiektów zbiór dopuszczalnych rozwiązań jest opisany modelem matematycznym, na który składają się: podzbiór zmiennych decyzyjnych, podzbiór parametrów stałych, podzbiór ograniczeń nierównościowych zwany *układem warunków sprawności* [1, 2]. W tym ujęciu projektowanie obiektu polega na wyznaczeniu wszystkich kombinacji wartości zmiennych decyzyjnych, dla których, przy założonych wartościach stałych, spełnione są warunki sprawności. Ten zbiór kombinacji zmiennych decyzyjnych spełniających przyjęte ograniczenia jest nazywany *zbiorem wariantów poprawnych* [2] lub *obszarem sprawności* [1]. Po jego wyznaczeniu można dokonać optymalizacji układu, wybierając taki element obszaru sprawności, który maksymalizuje określoną funkcję kryterialną.

W przypadku opisywanej tu metody danymi zadania są:

- zmienne decyzyjne: x_1, x_2, \dots, x_n
- stałe rozwiązania: c_1, c_2, \dots, c_m
- zbiór ograniczeń nierównościowych, o ogólnej postaci:

$$F_1(x_1, x_2, \dots, x_n, c_1, c_2, \dots, c_m) \geq 0$$

$$F_2(x_1, x_2, \dots, x_n, c_1, c_2, \dots, c_m) \geq 0$$

- - - - -

$$F_g(x_1, x_2, \dots, x_n, c_1, c_2, \dots, c_m) \geq 0$$

Oznaczmy rozwiązanie powyższego układu relacji (które jest zbiorem dopuszczalnych rozwiązań) jako D_R . Wielkościami *szukanymi*, wyznaczanymi przez opisaną metodę ewolucyjną, będą takie wektory zmiennych decyzyjnych R_i , które spełniają inkluzję:

$$R_i(x_1, x_2, \dots, x_n) \in D_R$$

2 Ogólny opis metody

W opisywanej metodzie wartości zmiennych decyzyjnych są kodowane binarnie w postaci liczb dwójkowych zwanych *chromosomami*. Założona długość chromosomu zależy od zakresu wartości zmiennej decyzyjnej i wymaganej dokładności obliczeń. Jednak w każdym przypadku taki sposób reprezentacji zmiennych prowadzi do dyskretyzacji odpowiedniej przestrzeni. Oznacza to, że liczba możliwych rozwiązań jest skończona. Do generowania kolejnych populacji rozwiązań zastosowano klasyczny algorytm genetyczny o postaci przedstawionej poniżej. W strukturze tego algorytmu można wyróżnić *fazę inicjacji* oraz *fazę ewolucji*.

* Politechnika Zielonogórska, Instytut Informatyki i Elektroniki, ul. Podgórna 50, 65-246 Zielona Góra,
e-mail: a.sielicki@pz.zgora.pl

Faza inicjacji polega na losowej generacji a następnie ocenie początkowego pokolenia rozwiązań. Funkcja `initpop()` wypełnia losowo tablicę `oldpop` danymi osobników

```
//ogólna postać użytego algorytmu genetycznego
popnumber=0;
initpop(oldpop);
assesspop(oldpop, sumfitness, inratio);
for (popnumber=1; popnumber<=N; popnumber++)
{
    makepop(oldpop, newpop);
    assesspop(newpop, sumfitness, inratio);
    copypop(oldpop, newpop);
}
```

początkowej populacji rozwiązań. Populacja ta jest oceniana – funkcja `assesspop()` oblicza wartości dwóch jej charakterystyk: *sumy miar przystosowania* wszystkich osobników populacji (`sumfitness`) i *współczynnika inkluzji* (`inratio`). Pierwsza z nich jest potrzebna w procesie wyboru rodziców do reprodukcji w celu utworzenia kolejnego pokolenia rozwiązań. Współczynnik `inratio` służy natomiast do badania zbieżności algorytmu i jest definiowany jako $L_R / \text{POPSIZE}$, gdzie L_R jest liczbą osobników reprezentujących rozwiązanie dopuszczalne a `POPSIZE` oznacza rozmiar populacji (liczbę jej osobników). Jeżeli algorytm jest zbieżny, to wartość `inratio` kolejnych pokoleń wzrasta z pokolenia na pokolenie, by ustalić się po pewnym czasie na poziomie bliskim jedności.

Faza ewolucji to proces sekwencyjnego generowania i oceny kolejnych pokoleń. W każdym cyklu wywołuje się kolejno trzy funkcje: Pierwsza z nich, `makepop()` przeprowadza reprodukcję, dokonując operacji genetycznych na wybranych parach chromosomów. W wyniku tej reprodukcji, z poprzedniego pokolenia `oldpop` jest tworzone nowe pokolenie `newpop` o tej samej liczbie osobników, lecz lepszych charakterystykach. Jest ono oceniane przez funkcję `assesspop()`. W każdym cyklu pamiętane są jedynie dwie populacje; na końcu cyklu funkcja `copypop()` zapamiętuje populację bieżącą jako populację poprzednią. Warunkiem zakończenia procesu jest uzyskanie założonej liczby pokoleń `N`.

3 Przykład implementacji metody

Stałe i struktury danych przyjęte w implementacji

Poniżej podano definicje stałych użytych w eksperymentalnej implementacji metody. Przyjęta długość chromosomu `STRLEN` równa 7 określa bezpośrednio zakres możliwych wartości każdej z dwóch zmiennych decyzyjnych `x1, x2`. Najmniejsza z tych wartości to w zapisie binarnym `0000000`; największą z nich jest `1111111`. Przy tego typu odwzorowaniu zmienne decyzyjne mogą przyjmować dyskretne wartości dziesiętne z zakresu od 0 do 127. W dyskretnej przestrzeni `x1, x2` mamy zatem 16384 możliwe położenia wektora zmiennych decyzyjnych. Znaczenie stałych `POPSIZE`, `N`, `CPROB`, `MPROB` nie wymaga komentarzy. Trzy ostatnie stałe `M`, `K`, `R` mają związek

z wyznaczaniem wartości miary jakości (przystosowania) danego osobnika populacji. Ich znaczenie zostanie wyjaśnione poniżej.

```
//stałe przyjęte w implementacji algorytmu
#define STRLEN 7 //długość chromosomu
#define POPSIZE 500 //liczebność populacji
#define N 20 //liczba pokoleń
#define CPROB 0.7F //prawdopodobieństwo krzyżowania
#define MPROB 0.001F //prawdopodobieństwo mutacji
#define M 10.0F //kwant miary przystosowania
#define K 4.0F //liczba ograniczeń
#define W 2.0F //współczynnik preferencji
```

Jak już wspomniano, algorytm wymaga, by w każdym cyklu były pamiętane dwie populacje. Do ich przechowania służą dwie zdefiniowane poniżej tablice struktur o nazwach `oldpop` oraz `newpop`.

```
struct member
{
    int ch1[STRLEN]; //chromosom X
    int ch2[STRLEN]; //chromosom Y
    float x1; //wartość dziesiętna zmiennej x1
    float x2; //wartość dziesiętna zmiennej x2
    float fitness; //miara przystosowania
};
```

```
member oldpop[POPSIZE], newpop[POPSIZE];
```

Każda ze struktur składowych obu tablic przechowuje dane odpowiedniego rozwiązania. Struktury zawierają pięć pól. Dwa z nich to tablice binarne: `ch1` dla chromosomu `x1` oraz `ch2` dla chromosomu `x2`. Następne dwa pola, `x1` oraz `x2`, przechowują dziesiętne wartości zmiennych rozwiązania, odpowiadające binarnym wartościom chromosomów. Ostatnie pole `fitness` służy do przechowania miary przystosowania osobnika reprezentowanego przez daną strukturę. (Wartość tej miary określają funkcje `model()` oraz `fun()`, opisane poniżej.)

Model obszaru rozwiązań i definicja miary przystosowania

W eksperymentalnej implementacji szukamy takich wektorów zmiennych `x1, x2`, które spełniają cztery założone ograniczenia – jedno nieliniowe i trzy liniowe. Ograniczenia te zapisano w pokazanej poniżej funkcji `model(x1, x2)`, w postaci warunków wykonania trzech kolejnych instrukcji `if`. Jak widać, funkcja `model(x1, x2)` zwraca jako wynik liczbę ograniczeń, spełnianych przez dany wariant rozwiązania.

Niech `fitness` oznacza wartość miary przystosowania, `n` – liczbę spełnionych przez danego osobnika ograniczeń, zaś `K` – całkowitą liczbę ograniczeń zadania. Wtedy:

$$\begin{aligned} \text{fitness} &= M \cdot n && \text{dla } n < K \\ \text{fitness} &= W \cdot M \cdot n && \text{dla } n = K \end{aligned}$$

gdzie M jest *kwantem miary przystosowania*, przydzielanym za spełnienie jednego ograniczenia, a W – *współczynnikiem preferencji*, tj, wartością, przez którą mnoży się fitness przy spełnieniu przez dane rozwiązanie *wszystkich* warunków zadania. Według tej zasady działa funkcja `fun()` przedstawiona poniżej. Funkcja ta zwraca po `return` wartość miary przystosowania dla danej pary chromosomów `ch1`, `ch2`. Oblicza ponadto dziesiętne wartości zmiennych decyzyjnych, przekazując je przez argumenty wyjściowe `x1`, `x2`.

```
// funkcja zwracająca liczbę spełnionych ograniczeń
float model(float x1, float x2)
{
    float m=0.0;
    if (x2>=10.0-x1/6.0) m++;
    if (x2<18.0*sin((x1-10.0)/11.0)+6.0*sin((x1-10.0)/2.5))
        m++;

    if (x1>=10) m++;
    if (x1<=40) m++;
    return m;
}

//funkcja licząca miarę przystosowania i dziesiętne
//wartości zmiennych decyzyjnych
float fun(int ch1[],int ch2[],float &x1,float &x2)
{
    float w=1.0,m;
    x1=0.0; x2=0.0;
    for (int i=0;i<STRLEN;i++)
    {
        x1=x+w*ch1[i];
        x2=y+w*ch2[i];
        w=w*2.0;
    }
    m=model(x1,x2);
    if (m<K) return (m*M);
    else return (m*M*W);
}
```

Tworzenie kolejnej populacji rozwiązań

Do wykonania tego zadania służy funkcja `makepop()`, pokazana poniżej. W każdym cyklu pierwszej pętli `for` funkcja `select()` wybiera do reprodukcji parę rodziców: `oldpop[p1]`, `oldpop[p2]`. Następnie, w wyniku procesów krzyżowania i mutacji, realizowanych przez funkcję `cross_mut()`, otrzymuje się chromosomy pary potomków, a więc `newpop[i]`, `newpop[i+1]`. Druga pętla `for` uzupełnia pozostałe pola struktur tych potomków.

Funkcja `select()` nie jest tutaj opisana. Działa ona zgodnie z regułą ruletki [4], według której prawdopodobieństwo wyboru elementu jako rodzica jest proporcjonalne do miary przystosowania tego elementu. Z braku miejsca nie podajemy również kodu

funkcji `cross_mut()`, która z założonym prawdopodobieństwem realizuje na wybranej parze chromosomów dwie operacje genetyczne: jednopunktowego krzyżowania i jednobitowej mutacji (rys.1).

```
void makepop(float sumfitness,
            member oldpop[],member newpop[])
{
    int i,p1,p2;
    float v1,v2;
    for (i=0;i<POPSIZE-1;i+=2)
    {
        p1=select(sumfitness,oldpop);
        p2=select(sumfitness,oldpop);
        cross_mut(oldpop[p1].ch1,oldpop[p2].ch1,
                 newpop[i].ch1,newpop[i+1].ch1);
        cross_mut(oldpop[p1].ch2,oldpop[p2].ch2,
                 newpop[i].ch2,newpop[i+1].ch2);
    }
    for(i=0;i<POPSIZE;i++)
    {
        newpop[i].fitness=fun(newpop[i].ch1,
                             newpop[i].ch2,v1,v2);

        newpop[i].x=v1;
        newpop[i].y=v2;
    }
}
```

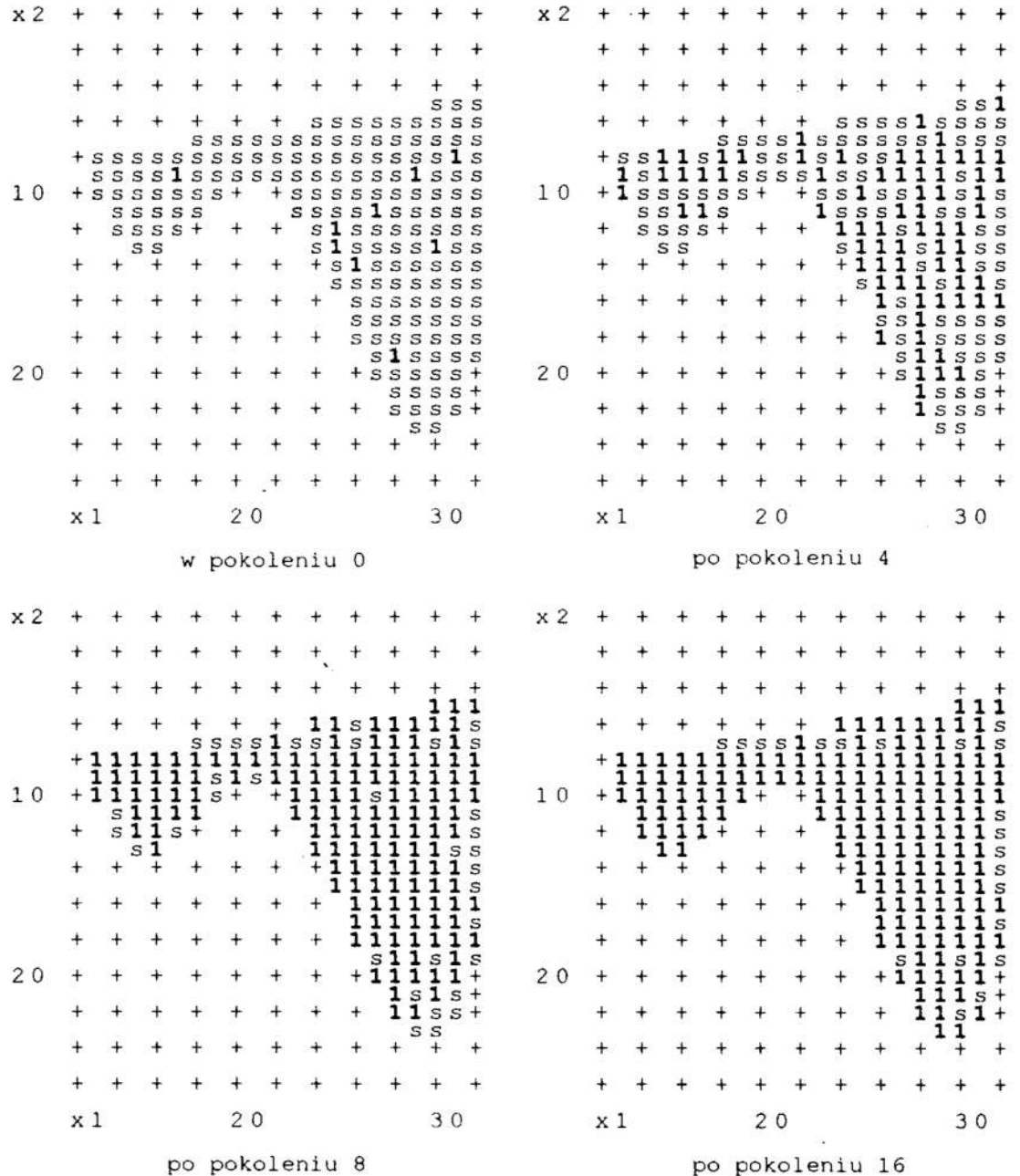


Rys.1. Przykład działania operatorów krzyżowania i mutacji
Fig.1. Example of crossover and mutation

4 Niektóre wyniki eksperymentu komputerowego

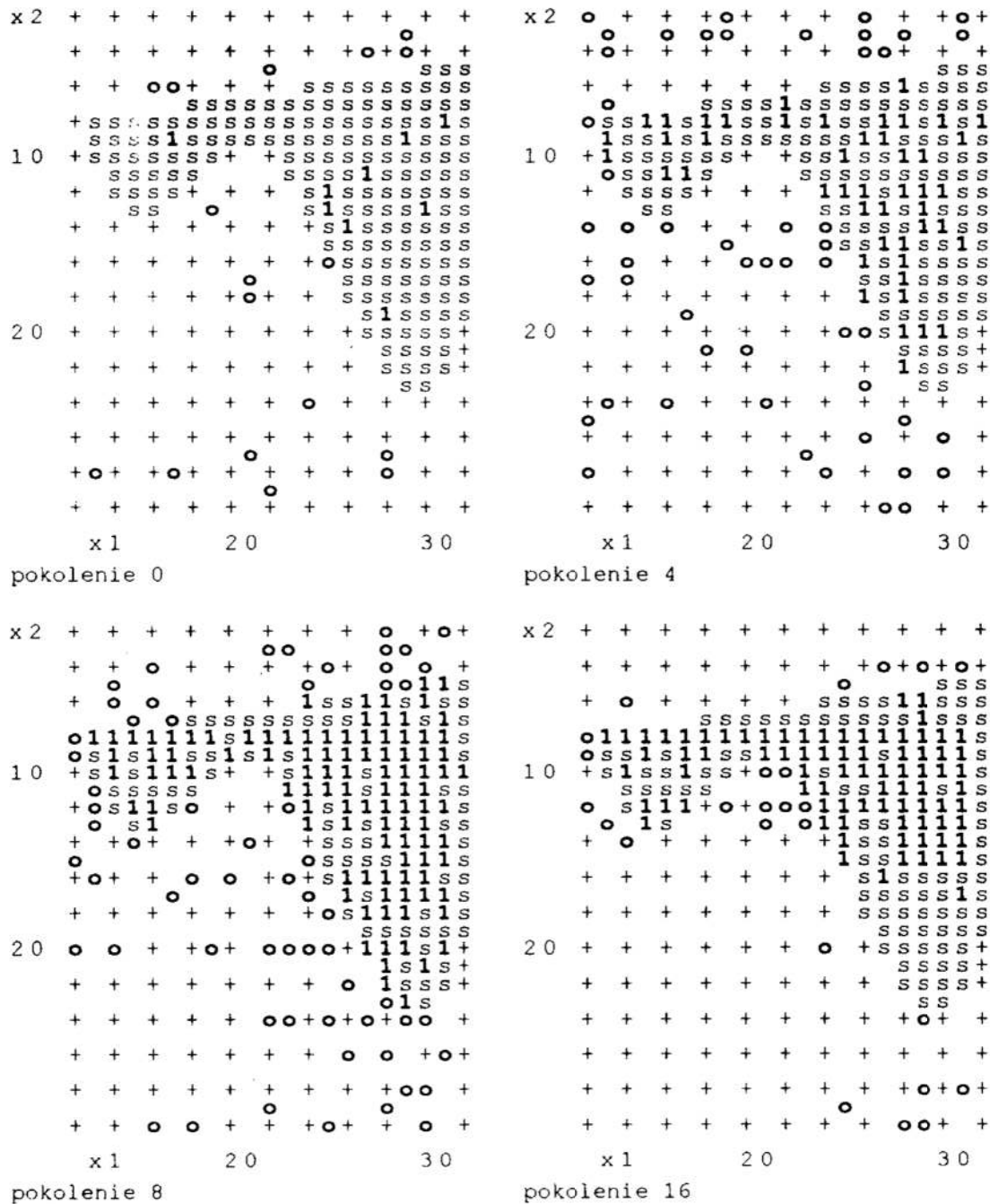
Rysunek 2 pokazuje, dla czterech różnych pokoleń, wyprowadzone na ekran obrazy fragmentu przestrzeni zmiennych decyzyjnych x_1 , x_2 , w której leży obszar dopuszczalnych rozwiązań, wyznaczony przez cztery założone w zadaniu ograniczenia.

Punkty spełniające warunki zadania oznaczono symbolem 's'. Te spośród nich, które algorytm znalazł po pewnej liczbie kolejnych pokoleń, są wyróżnione na rysunku symbolem '1'. Pokazano rozwiązania znalezione po czterech wybranych pokoleniach: początkowym, czwartym, ósmym i szesnastym. Zaobserwowano szybki przyrost liczby znajdujących rozwiązań w początkowych pokoleniach. W dalszych pokoleniach, po odnalezieniu przeszło 90% możliwych rozwiązań, działanie algorytmu staje się mniej efektywne.



Rys. 2. Rozwiązania odszukane po niektórych pokoleniach
 Fig. 2. Solutions found after some generations

Rysunek 3 przedstawia zmiany kolejnych pokoleń rozwiązań. Pokazano cztery pokolenia o numerach 0, 4, 8, 16. Rozwiązania dopuszczalne oznaczono jako '1', a pozostałe jako '0'. Widać, że w miarę ewolucji elementy populacji skupiają się coraz bardziej w obszarze rozwiązań dopuszczalnych i jego otoczeniu. Rysunek nie pokazuje wszystkich elementów, ponieważ obejmuje niewielką część przestrzeni x_1, x_2 . Rozwiązania o tych samych współrzędnych są reprezentowane przez pojedynczy symbol.



Rys.3. Niektóre pokolenia rozwiązań w procesie ewolucji
 Fig.3. Some generations of solutions during the evolution

Literatura

1. Sielicki A.: *Metodologia projektowania*. Wyd. Politechniki Wrocławskiej, 1985
2. Tarnowski W.: *Podstawy projektowania technicznego*. Wydawnictwa Naukowo-Techniczne, Warszawa 1997
3. Michalewicz Z.: *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin-Heidelberg, 1996
4. Goldberg D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Inc., Reading, Massachusetts, USA, 1989
5. Fonseca C.M., Fleming P.J.: *An Overview of Evolutionary Algorithms in Multiobjective Optimization*. *Evolutionary Computation*, Vol. 3, No 1, 1995, p. 165-180
6. Srinivas N., Patnaik L.M.: *Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms*. *Evolutionary Computation*, Vol. 2, No. 3, 1994, p. 221-248

Streszczenie

W pracy przedstawiono propozycję ewolucyjnej metody, przeznaczonej do wyznaczania zbioru dopuszczalnych rozwiązań wyznaczonego przez układ ograniczeń nierównościowych. Metoda stosuje algorytm genetyczny, w którym wykorzystano operatory krzyżowania i mutacji, a miara przystosowania osobnika populacji jest proporcjonalna do liczby spełnionych ograniczeń, wzrastając skokowo przy spełnieniu wszystkich ograniczeń. Pokazano przykładową implementację metody w języku C++ i wyniki eksperymentu, wskazujące na skuteczność zastosowanego algorytmu.

EVOLUTIONARY METHOD OF SEARCHING FOR ACCEPTABLE SOLUTIONS

Summary

In the paper an evolutionary method of computer-aided searching for acceptable solutions in the case of constrained design problems is proposed. In the method, a genetic algorithm implemented in C++ has been used, applying the crossover and mutation operators. Some of the algorithm subroutines are discussed. An example of experimental results is shown to demonstrate effectiveness of the proposed method.

Referat recenzowany

DYNAMIC RESPONSE OF AN ELASTIC HALF-SPACE SOIL MEDIUM UNDER IMPULSIVE LOADINGS

1 Introduction

An elastic half-space bounded by one plane surface is the simplest and useful model for observing elastic waves in solids being applicable in seismic prospecting [1], testing of underground explosions [2], [3], design of blast-resistant structures [4], dynamics of structural foundations [5] as well as basic studies on dynamics of material defects [6]. In the field of geophysics it is desirable to calculate the full response of the free surface of the half-space to a short pulse of an arbitrary shape [7]. The time history of the free surface displacement at a given epicentral distance from the source is called a theoretical seismogram.

In this paper we shall deal with problem of calculating a theoretical seismogram for a given source by numerical modelling of the wave propagation problem. In the last years, significant advances have been made in the development and application of numerical methods to the solution of dynamic transient problems with the parallel development of hardware. This has resulted in the appearing of many commercially available and special purpose codes. For the modelling and solution of the present problem, the ABAQUS/Explicit finite element code is used [8]. However, the finite element method can only deal with finite domains possessing well-defined boundaries. If such boundaries do not exist naturally, waves will be reflected back into the model. Then so-called 'transmitting' boundaries must be used to simulate the effect of the truncated exterior infinite domain. Applying the finite element spatial discretizations with 'silent' boundaries to preserve the real physical behaviour of semi-infinite medium leads to the governing ordinary second order differential equations in time, which in turn must be integrated by a numerical step-by-step procedure [9]. The wave propagation problems are best solved by so-called explicit time-integration techniques. In the presented approach the central difference integration rule is used being particularly well suited for short duration dynamical problems.

2 Statement of the problem

Consider a linearly elastic soil medium occupying in its natural state a semi-infinite region Ω of a three-dimensional space \mathcal{R}^3 , bounded by a free surface $\partial\Omega$ ($z \geq 0$), Figure 1.

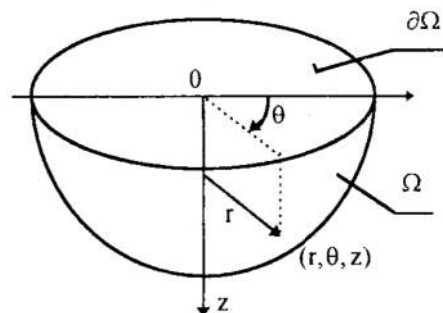


Fig. 1. An elastic half-space model of soil medium and co-ordinate system

Rys. 1. Ośrodek gruntowy w postaci półprzestrzeni sprężystej i układ współrzędnych

* Politechnika Koszalińska, Katedra Mech. Budowli, sien@lew.tu.koszalin.pl, domski@lew.tu.koszalin.pl

The motion of the medium is referred to an inertial frame of reference and assuming infinitesimal deformations, the Cauchy equation of motion is given by

$$\nabla \bullet \sigma(\mathbf{x}, t) + \rho \mathbf{b}(\mathbf{x}, t) = \rho \ddot{\mathbf{u}}(\mathbf{x}, t) \quad , \quad (\mathbf{x}, t) \in \mathcal{R}^3 \times T \quad , \quad T = (-\infty, \infty) \quad (1)$$

where σ , \mathbf{b} and \mathbf{u} stand for the stress tensor, body force vector and displacement vector, respectively, ∇ denotes the del operator, $\nabla \bullet ()$ implies the divergence of (), and the dot denotes differentiation with respect to time, [6]. In addition the balance of angular momentum provides the symmetry of the tensor σ .

The kinematical relation for the infinitesimal deformations is expressed as

$$\varepsilon(\mathbf{x}, t) = \frac{1}{2} \{ \nabla \mathbf{u}(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t) \nabla \} \quad (2)$$

and the constitutive equation for the linear elastic material is

$$\sigma(\mathbf{x}, t) = \mathbf{C} : \varepsilon(\mathbf{x}, t) \quad (3)$$

where ε , \mathbf{C} stand for the strain tensor and linear elasticity tensor, respectively, and the double dot implies the double inner product. Substitution of eq. (2) into (3) and further into (1) leads to the Navier–Cauchy equation of motion

$$\nabla \bullet (\mathbf{C} : \nabla \mathbf{u}) + \rho \mathbf{b} = \rho \ddot{\mathbf{u}} \quad (4)$$

that for an isotropic and homogeneous medium reduces to

$$\mu \nabla^2 \mathbf{u} + (\lambda + \mu) \nabla \nabla \bullet \mathbf{u} + \rho \mathbf{b} = \rho \ddot{\mathbf{u}} \quad (4a)$$

where λ and μ are the Lamé constants, [6].

The problem is defined such as to find a solution of the equation of motion (4) for a semi-infinite region Ω subjected to a concentrated time-dependent body force acting at a point ξ , in a fixed direction \mathbf{e}

$$\rho \mathbf{b} = \mathbf{P}(t) \delta(\mathbf{x} - \xi) \mathbf{e} \quad (5a)$$

where $\delta(\mathbf{x})$ is the three-dimensional delta distribution.

The solution must satisfy the initial conditions

$$\mathbf{u}(\mathbf{x}, 0) = \dot{\mathbf{u}}(\mathbf{x}, 0) = \mathbf{0} \quad , \quad \mathbf{x} \in \Omega \quad , \quad (5b)$$

the condition of vanishing traction on the free surface of the half-space

$$\mathbf{n} \bullet \sigma = \mathbf{n} \bullet \mathbf{C} : \nabla \mathbf{u} = \mathbf{0} \quad , \quad (\mathbf{x}, t) \in \partial\Omega \times T \quad , \quad (5c)$$

where \mathbf{n} denotes the unit outward normal vector to $\partial\Omega$, and a suitable radiation conditions at infinity, [10].

3 Discrete formulation

To solve the transient continuum mechanics problem numerically, the governing hyperbolic partial differential equations are first discretized in space by a finite element method [9]. The semi-infinite medium is divided into a finite interior region and the remaining exterior region which includes the boundary extended to infinity. First-order uniform strain finite elements are used to model the interior region and special infinite elements are selected to represent the exterior unbounded medium [9], [11], [12], [13]. This procedure called a semi-discretization leads to a system of ordinary differential equation in time. For a linear finite element model the equation of motion is

$$\mathbf{M} \ddot{\mathbf{d}}(t) + \mathbf{K} \mathbf{d}(t) = \mathbf{F}(t) \quad (6)$$

where \mathbf{M} is the mass matrix, \mathbf{K} is the stiffness matrix and $\mathbf{F}(t)$, $\mathbf{d}(t)$, $\ddot{\mathbf{d}}(t)$ are time-dependent vector of nodal forces, displacements and acceleration, respectively. The equation (6) is solved using the explicit central difference integration rule the effectiveness of which is strengthened by the use of a diagonal mass matrix.

The explicit time integration scheme is only conditionally stable and generally small time steps are necessary to insure numerical stability, [14]. However, the explicit time-integration technique is commonly used in wave propagation studies, [15].

4 Numerical experiments

The dynamic characteristics of an elastic semi-infinite soil medium have been studied on axially symmetric boundary-value problems. Consider an isotropic homogeneous half-space with the elastic constants $\lambda=\mu=40$ MPa, and mass density $\rho=1850$ kg/m³; it gives S-wave (secondary wave, distortional wave, shear wave) propagation velocity $c_2=(\mu/\rho)^{1/2}=147$ m/s, and P-wave (primary wave, dilatational wave, compression wave) propagation velocity $c_1=c_2 \cdot 3^{1/2}$. The axisymmetric interior region under consideration is the domain $\Omega \ni \mathbf{x} = (r, \theta, z)$ where $0 \leq r \leq 0.2$ [m], $0 \leq \theta \leq 2\pi$, $0 \leq z \leq 0.2$ [m].

Three meshes of 4-node bilinear axisymmetric ring elements have been used to model the region of interest: 90×90 (Figure 2), 180×180 and 360×360 elements, respectively.

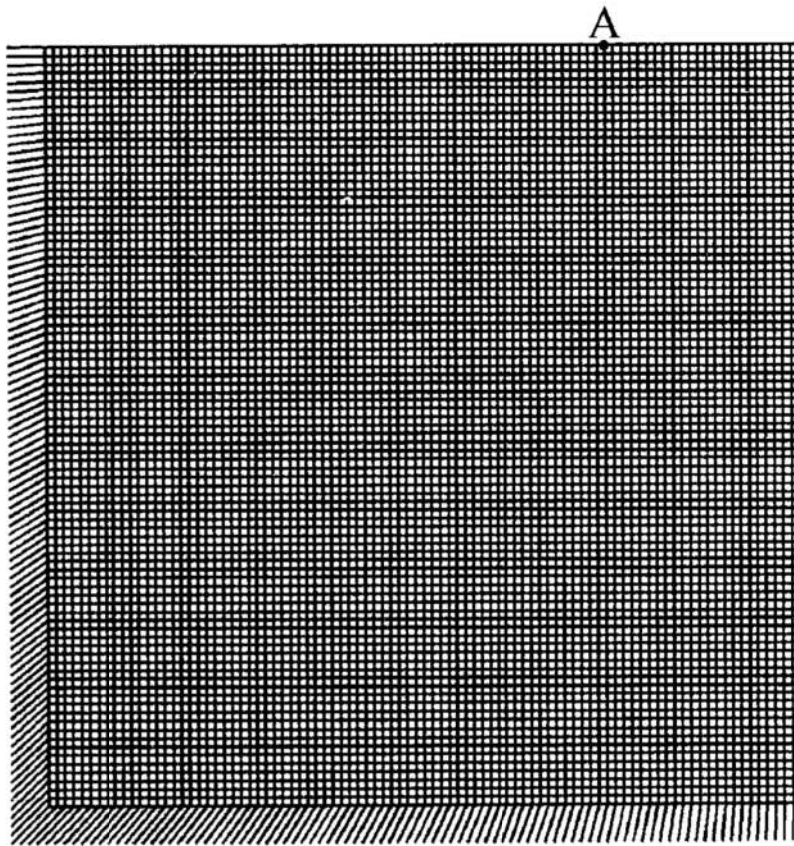


Fig. 2. Finite element mesh for an axisymmetric problem; interior region: 90×90 4-node bilinear ring elements; exterior region: 180 4-node linear infinite elements; A—receiver point at a radial distance $r_A=0.051$ m

Rys. 2. Siatka elementów skończonych w zagadnieniu osiowosymetrycznym; obszar wewnętrzny: 90×90 czterowzłowych dwuliniowych elementów pierścieniowych; obszar zewnętrzny: 180 czterowzłowych liniowych elementów nieskończonych; A—punkt rejestracji drgań w odległości $r_A=0.051$ m

The number of 4-node linear axisymmetric infinite elements modelling the far-field region was 180 (Figure 2), 360 and 720, respectively.

First, the three-dimensional Lamb's problem is considered. The uniform half-space is subjected to a concentrated load $P(t)=P \times H(t)$ normal to the surface $z=0$ at the origin of

the cylindrical coordinates (r, θ, z) at $t=0$, where $H(t)$ is the Heaviside unit step function of time t . The closed form solution of the Lamb's problem for a Poissonian material (Lame' constants $\lambda=\mu$) was obtained by Pekeris [6]. Figure 3 shows the time history of the vertical displacement response at the surface point A obtained from the first finite element model. The results from the second model and from the third one are presented in Fig. 4 and 5, respectively. In the Figures 3, 4 and 5 P, S, and R denote times of arrival primary, secondary and Rayleigh waves respectively. It is seen that first a P-wave arrives, then S-wave marked by a change in the slope of the curve, followed by the R-wave. Near the time of arrival of the R-wave, vertical u_z as well as horizontal u_r displacements become infinite according to $1/(r-c_2t)^{1/2}$. Immediately, thereafter, the steady-state condition is reached (Boussinesq's problem). It is seen from the Figures 3, 4 and 5 that finite elements act like 'low-pass' filters with a certain cut-off frequency that depends on the size of the element—the larger the element, the lower its cut-off frequency. The accuracy of the dynamic finite element solution from the third model is satisfactory, at least from an engineering point of view.

Consider next the case of a concentrated pulse $P(t)=P\delta(t)$ acting at a point $(r, \theta, z)=(0, 0, h)$, where $\delta(t)$ is the Dirac delta distribution of time and h is the depth at which the pulse is buried. This substitution for $P=1$ leads to Green's function in a semi-infinite medium, [10]. Time histories of the vertical and horizontal displacements at surface point A are presented in the Figures 6 and 7 respectively, for different values of the embedment of the source point. The results are obtained from the most refined finite element model.

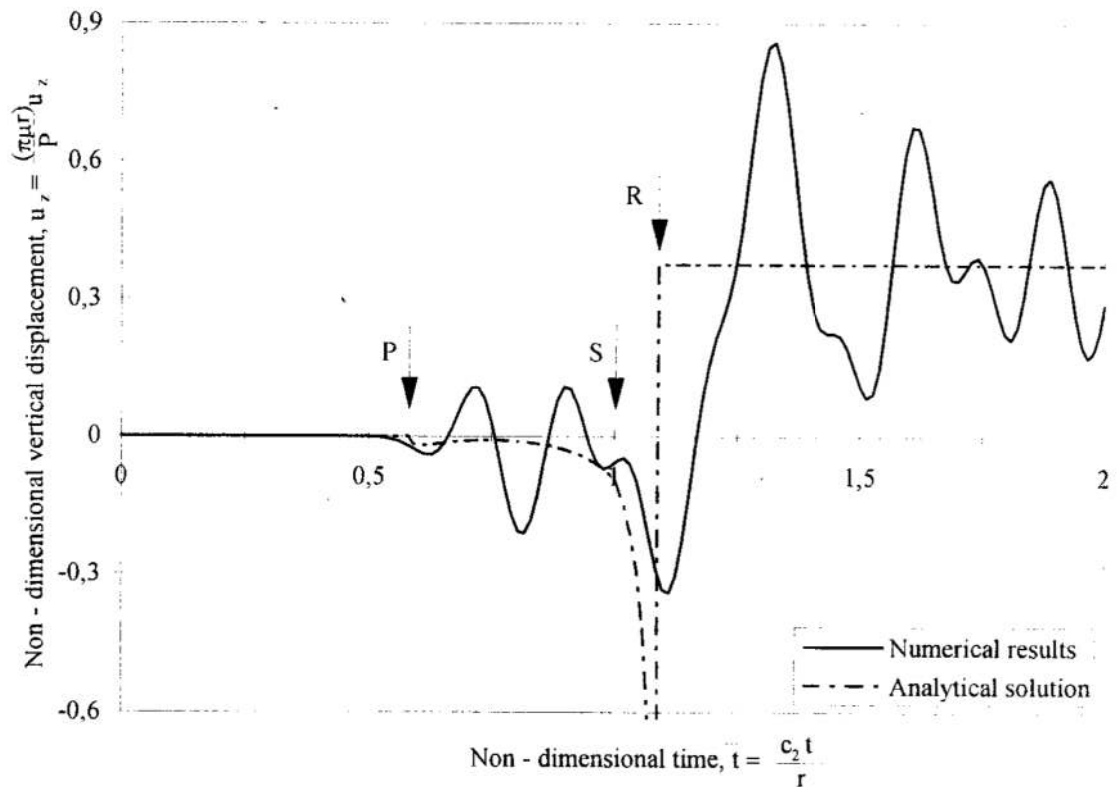


Fig. 3. Step-load on an elastic uniform half-space: vertical displacement history at the point A; the first finite element model; time step $\Delta t=4.659 \times 10^{-6}$ sec.

Rys. 3. Nagłe obciążenie jednorodnej półprzestrzeni sprężystej: historia przemieszczenia pionowego w punkcie A; pierwszy model dyskretny; krok całkowania $\Delta t=4.659 \times 10^{-6}$ s

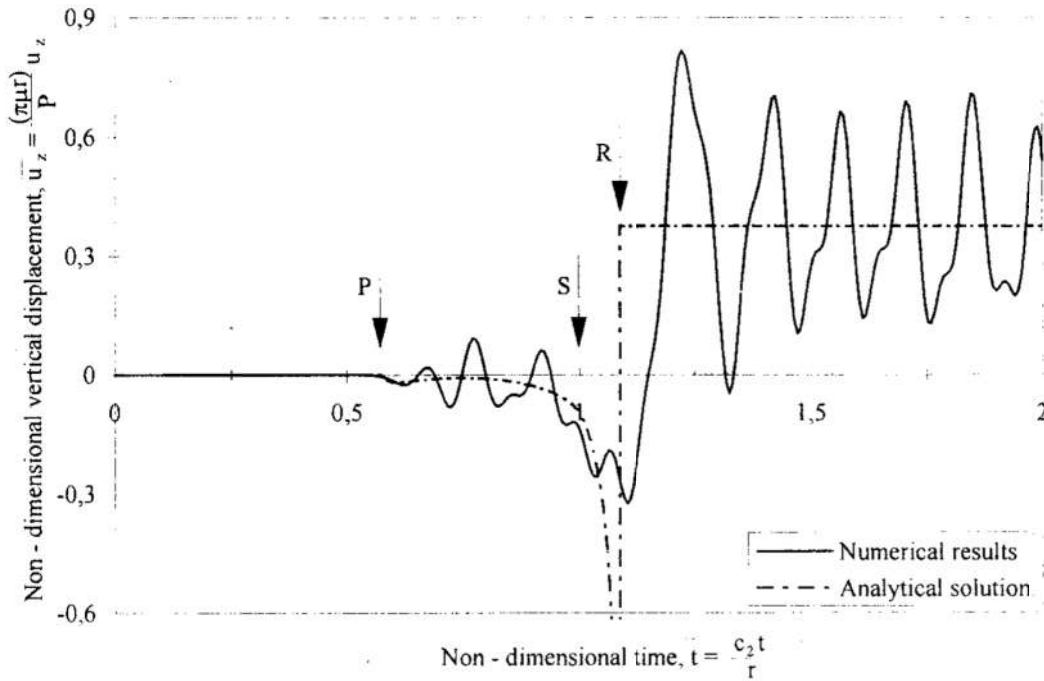


Fig. 4. Step-load on an elastic uniform half-space: vertical displacement history at the point A; the second finite element model; time step $\Delta t = 2.329 \times 10^{-6}$ sec.
 Rys. 4. Nagłe obciążenie jednorodnej półprzestrzeni sprężystej: historia przemieszczenia pionowego w punkcie A; drugi model dyskretny; krok całkowania $\Delta t = 2.329 \times 10^{-6}$ s

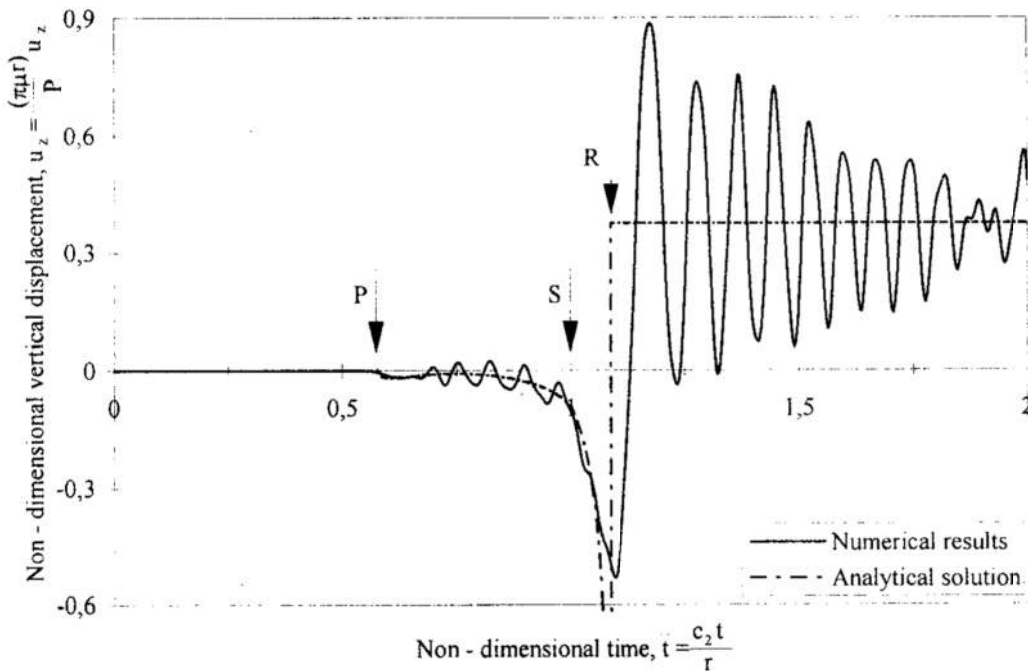


Fig. 5. Step-load on an elastic uniform half-space: vertical displacement history at the point A; the third finite element model; time step $\Delta t = 1.165 \times 10^{-6}$ sec.
 Rys. 5. Nagłe obciążenie jednorodnej półprzestrzeni sprężystej: historia przemieszczenia pionowego w punkcie A; trzeci model dyskretny; krok całkowania $\Delta t = 1.165 \times 10^{-6}$ s

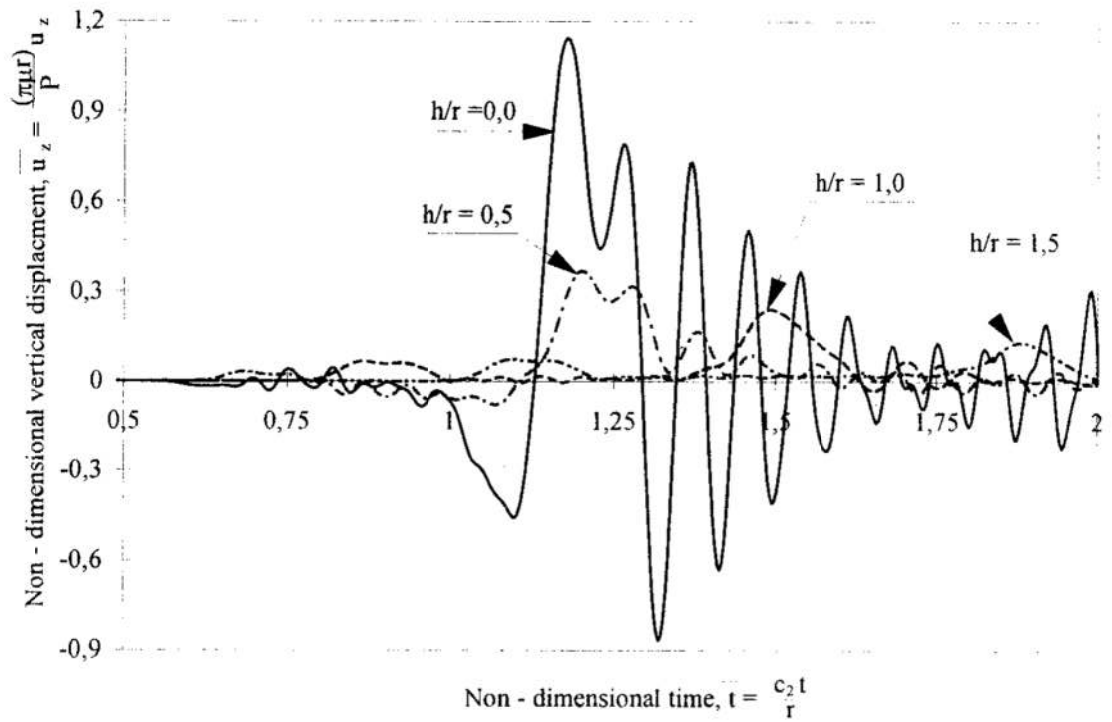


Fig. 6. Buried concentrated pulse: vertical displacement history at surface point A
 Rys. 6. Impuls podziemny: historia przemieszczenia pionowego w punkcie A

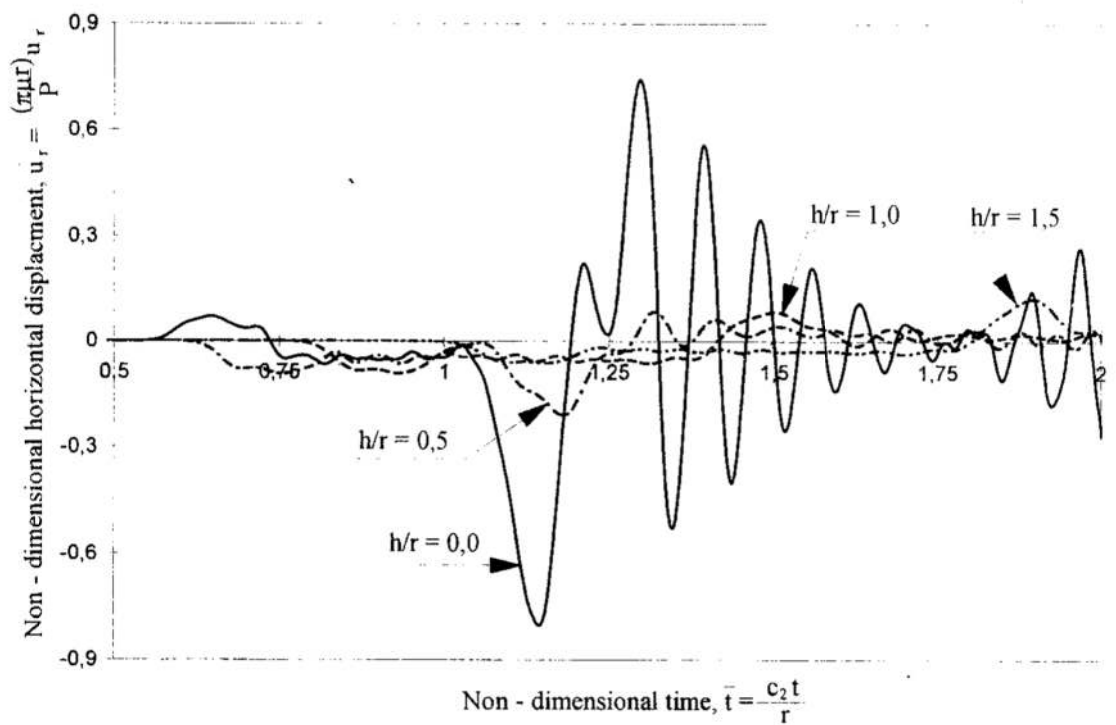


Fig. 7. Buried concentrated pulse: horizontal displacement history at surface point A
 Rys. 7. Impuls podziemny: historia przemieszczenia poziomego w punkcie A

5. Concluding remarks

Most of the information about the wavefields in soil medium is provided by experimental seismograms generated by earthquakes and explosions. The wavefields are composed of various types of reflected, refracted and diffracted waves originating at velocity and density discontinuities in the medium. For interpreting the experimental seismograms it is necessary to know the correlations between characteristics of wave motion and various physical factors associated with the wave source and propagation path. It can be obtained from numerical seismograms provided by the transient dynamic finite element model that allow the process of wave motion to be displayed through computer simulation. Therefore, the finite element technique is an important tool for studying the wave propagation problems.

In the case of three-dimensional Lamb's problem it is possible to compare the results predicted by the finite element models with the explicit solution. The low pass filtering action of even the finest finite element mesh should be taken into account when the quality of numerical results is assessed.

To overcome the problem of waves reflected at the model's boundaries the time period over which the analysis is performed should be limited to the range within the waves do not arrive at the point of interest. The approach leads to use of a large model and may not be economical because of its storage and computational requirements to preserve a high degree of accuracy of numerical results in the time domain. Generally, the radiation energy through outer boundaries of the finite element model must be properly modelled. The absorbing boundaries can be classified as local or non-local, depending on the degree of coupling introduced by the corresponding generalised radiation condition between the individual boundary points. On a local boundary either weak or no coupling exists between points whereas the radiation condition at any point on a non-local boundary involves every other point on that boundary. Although the non-local boundaries are more efficient in minimising artificial wave reflections than the local ones, the local approach used in this paper is easier to construct for transient problems, involves less computational effort and provides acceptable modelling for most practical cases.

References

1. Ben-Menahem A., Singh S.J.: *Seismic waves and sources*. Springer-Verlag, New York 1981
2. Henrych J.: *The dynamics of explosions and its use*. Elsevier, Amsterdam 1979
3. Awojobi A.O., Sobayo O.A.: *Ground vibrations due to seismic detonation of buried source*. *Earthquake Eng. Struct. Dyn.*, 2, 131-143, 1977
4. Bangash M.Y.H.: *Impact and explosion-analysis and design*. Blackwell Scientific Publications, Cambridge, 1993
5. Sienkiewicz Z.: *Vibrations of a rigid body on a linear deformable subsoil*. (In Polish), Technical University of Koszalin, No. 55: pp. 221, Koszalin 1995
6. Achenbach J.D.: *Wave propagation in elastic solids*. North-Holland, Amsterdam 1973
7. Hanyga A.(ed.): *Seismic wave propagation in the earth*. PWN-Warszawa, Elsevier-Amsterdam 1984
8. *Abaqus/explicit version 5.8-user's manual*, HKS, Inc., Pawtucket 1998

9. Zienkiewicz O.C., Taylor R.L.: *The finite element method. Basic formulation and linear problems*. Vol. 1, 4th edition, McGraw-Hill, London 1989
10. Eringen A.C., Suhubi E.S.: *Elastodynamics. Linear theory*. Vol. 2. Academic Press, New York, 1981
11. Lysmer J., Kuhlemeyer R.L.: *Finite dynamic model for infinite media*. J. Engng Mech. Div., ASCE 95, 859-877, 1969
12. White W., Valliappan S., Lee I.K.: *A unified boundary for finite dynamic models*. J. Engng Mech. Div., Proc. Am. Soc. Civ. Eng., 103, 949-964, 1977
13. Kausel E.: *Local transmitting boundaries*. J. Engng Mech. 114, 1011-1027, 1988
14. Dokainish M.A., Subbaraj K.: *A survey of direct time-integration methods in computational structural dynamics—I. Explicit methods*. Computers & Structures, Vol. 32, No.6, 1371-1386, 1989
15. Osama Al-Hunaidi M.: *Analysis of wave propagation in unbounded media*. Computers & Structures, Vol. 33, No. 4, 1037-1045, 1989

Summary

The finite element technique in the time domain is applied to study the wave motion in an elastic soil medium. The transient wavefield is generated by an impulsive vertical load, expressed in time by the Heaviside unit step function and Dirac delta distribution. The step load is suddenly applied and maintained on the surface of an elastic uniform half-space while the concentrated pulse is buried to a some depth. To model the radiation condition in a semi-infinite domain, a local transmitting boundary has been used. The time histories of displacements at the selected surface point are obtained using different finite element meshes. The validity and accuracy of the numerical results are assessed by comparison with an explicit solution of a three-dimensional Lamb's problem.

ODPOWIEDŹ DYNAMICZNA OŚRODKA GRUNTOWEGO W POSTACI PÓŁPRZESTRZENI SPRĘŻYSTEJ OBCIĄŻONEJ IMPULSOWO

Streszczenie

Zastosowano metodę elementów skończonych w dziedzinie czasu do analizy ruchu falowego w sprężystym ośrodku gruntowym. Pole falowe generowane jest przez pionowe obciążenie impulsowe, opisane w czasie funkcją Heaviside'a lub deltą Diraca. Siła skupiona przykładana jest nagle i utrzymywana na powierzchni jednorodnej półprzestrzeni sprężystej (przestrzenne zagadnienie Lamba) podczas gdy oddziaływanie o charakterze impulsu w postaci delty Diraca zadawane jest w ustalonych punktach wewnątrz półprzestrzeni. Warunki wypromieniowania w obszarze półnieskończonym opisano za pomocą lokalnych warunków brzegowych, które zapewniają transmisję fal sprężystych przez brzeg obszaru skończonego. Wyznaczono przemieszczeniowe historie odpowiedzi półprzestrzeni sprężystej w ustalonym punkcie na jej powierzchni, stosując różne siatki elementów skończonych. W przypadku przestrzennego zagadnienia Lamba, wyniki numeryczne porównano z rozwiązaniem analitycznym.

Referat recenzowany

MODEL SYMULACYJNY TŁOCZNI LEKKIEJ W FABRYCE SAMOCHODÓW OSOBOWYCH

1 Wprowadzenie

Dla celów edukacyjnych i badawczych opracowywany jest informatyczny system symulacji sterowania produkcją przykładowej fabryki samochodów osobowych [1]. Przy projektowaniu struktury informacyjnej i funkcjonalnej wzorowano się w znacznym stopniu na standardzie MRP II [5]. Zaletą omawianego systemu symulacji sterowania produkcją jest możliwość prostej integracji w jego strukturach funkcji standardu MRP II oraz autorskich modeli i algorytmów sterowania i planowania produkcji [2], [3], [4], jak również współpracy systemu z komercyjnymi narzędziami do optymalizacji i symulacji, jak np. systemem optymalizacji dyskretnej „Lingo”, czy systemem symulacji Taylor II.

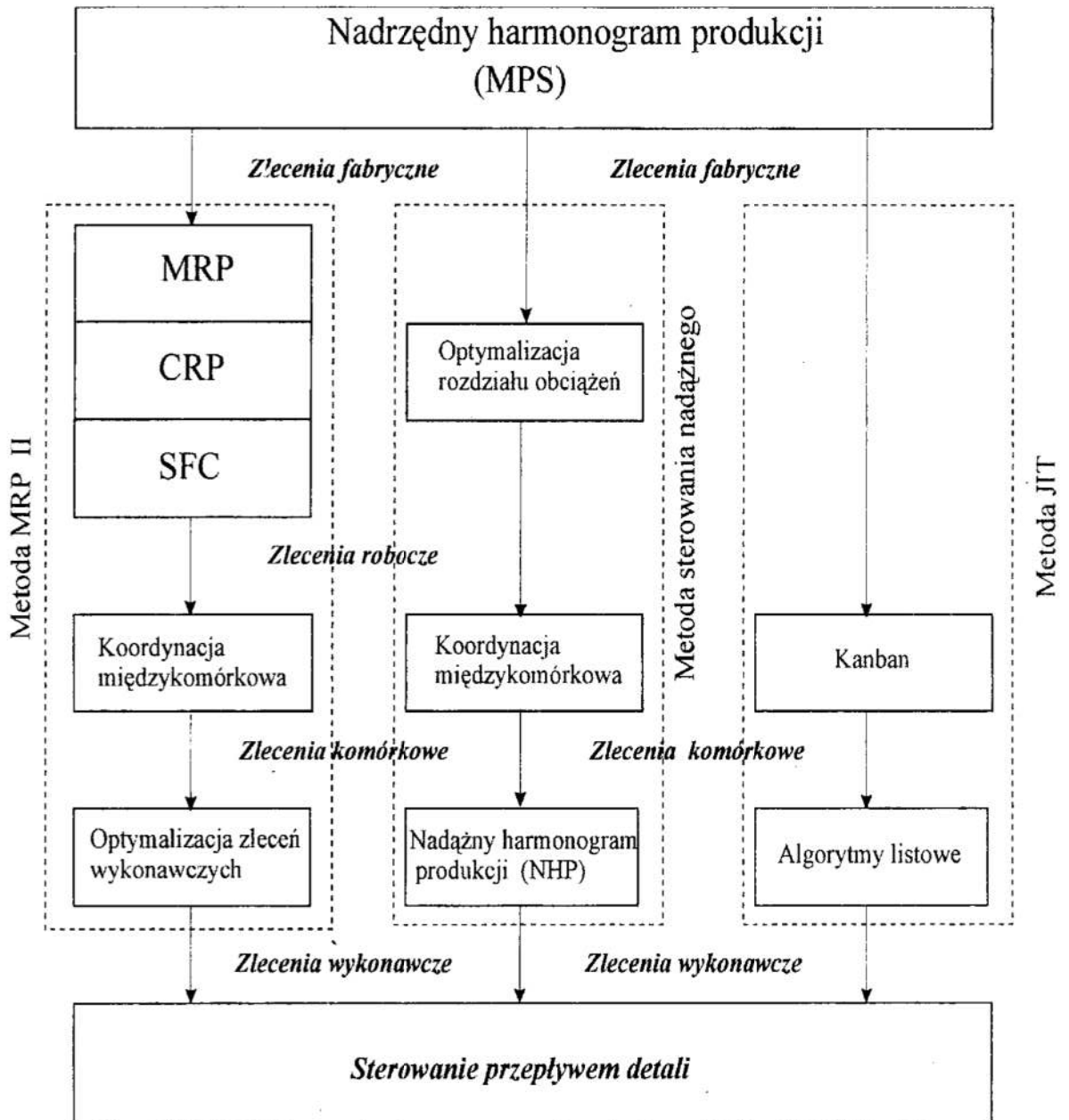
2 Sterowanie produkcją w tłoczni lekkiej

Do wydziałów przykładowej fabryki samochodów należą między innymi dwie tłocznie blach karoseryjnych – ciężka i lekka. Proces tłoczenia blach karoseryjnych przebiega w tłoczni w liniach pras (rys.2). Jego produktami są wytłoczki o rozmaitych kształtach, z których można zmontować karoserię samochodu. Każda wytłoczka powstaje z odpowiednio przyciętego kawałka blachy, tzw. przygotówki, w wyniku jednej lub kilku operacji tłoczenia wykonywanych na szeregowo ustawionych prasach linii pras. Typ wytłoczki zależy od operacji wykonywanych na kolejnych prasach, a typ operacji jest określony jednoznacznie przez typ tłoczniaka zamontowanego w prasie. Każdy tłoczniak składa się z dwu części – stempla zamocowanego do suwaka prasy i matrycy przytwierdzonej do stołu prasy. Wymiana tłoczniaków w linii pras, czyli tzw. przebrojenie linii, trwa wielokrotnie dłużej niż czas jednej operacji tłoczenia. Ponadto koszty przebrojenia są znaczne, ponieważ konieczne jest zaangażowanie pracowników o wysokich kwalifikacjach. Z tego powodu dąży się do wydłużania okresów czasu między przebrojeniami, czyli do zwiększania wielkości partii produkowanych wytłoczek. Z drugiej strony, wzrost wielkości partii skutkuje wzrostem poziomu zapasów, co prowadzi do zwiększenia kosztów magazynowania. Znane są wzory na tzw. optymalną wielkość partii, która minimalizuje sumę kosztów przebrojeń i magazynowania, lecz w praktyce bieżące sterowanie przebrojeniami daje wielkości partii znacznie różniące się od optymalnych. Celem tego sterowania jest bowiem przede wszystkim dostosowanie obciążenia każdej z linii produkcyjnych systemu do zmieniającego się zapotrzebowania na jej produkty. Produkcja wytłoczek w liniach pras tłoczni blach karoseryjnych należy do klasy produkcji powtarzalnej, ponieważ każda linia charakteryzuje się skończoną liczbą wariantów uzbrojenia pras i odpowiadającego mu asortymentu produktów, czyli skończoną liczbą tzw. wariantów produkcyjnych, powtarzających się w na ogół nieregularnych odstępach czasu.

Do sterowania przepływem produkcji między liniami produkcyjnymi coraz częściej stosuje się tzw. metodę Just-In-Time, a ściślej algorytm Kanban [8]. Podobne funkcje może spełniać algorytm optymalizacji rozdziału obciążeń między linie, jak również odpowiednio uzupełnione moduły MRP, CRP, SFC systemu MRP II [5]. Niezależnie od metody generacji zleceń komórkowych (rys.1) dla poszczególnych linii są potrzebne algorytmy szeregowania zleceń wykonawczych. Może to być algorytm harmonogramowania nadążnego [6], algorytm optymalizacji dynamicznej [10] lub proste algorytmy listowe. Nie wchodząc w szczegóły realizacyjne można stwierdzić, że

* Politechnika Świętokrzyska, Samodzielny Zakład Systemów Sterowania i Zarządzania, Al.1000-PP 7, 25-314 Kielce, E-Mail: sitek@sabat.tu.kielce.pl, wikarek@sabat.tu.kielce.pl

** Politechnika Śląska, Instytut Automatyki, ul. Akademicka 16, 44-100 Gliwice, E-Mail: mzaborowski@zeus.polsl.gliwice.pl



Rys. 1. Alternatywne struktury funkcjonalne systemu sterowania produkcją dla wytwarzania synchronicznego

Fig. 1. Alternative functional structures of production control for synchronized manufacturing

w każdym przypadku zlecenia wykonawcze dla sekcji linii potokowych nadążają za wielkościami wiodącymi całego systemu produkcyjnego, którymi są zlecenia fabryczne reprezentujące bieżące zapotrzebowanie na wyroby danego zakładu (rys. 1).

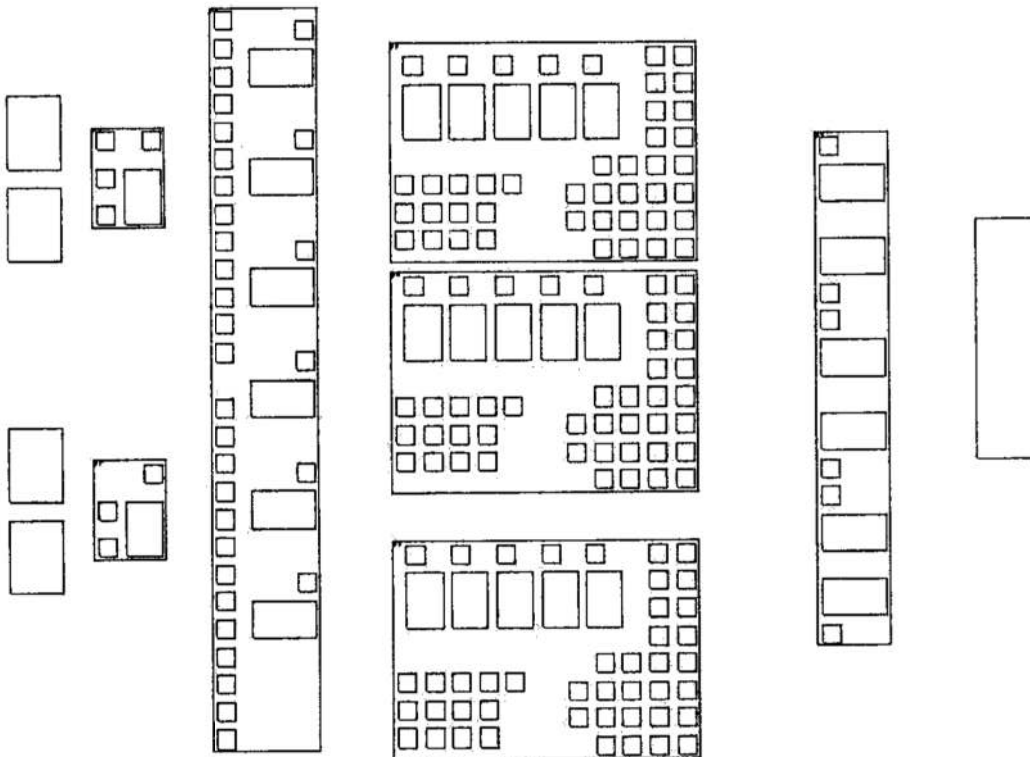
Tłoczenie blach karoseryjnych na kolejnych prasach przydzielonych do danej wyłóczki jest zsynchronizowane. Liczba możliwych wariantów uzbrojenia pras, determinujących przydział produktów do ciągów pras, jest skończona. W praktyce przełączeń dokonuje się w niezbyt licznych zbiorze wariantów produkcyjnych, wystarczającym jednak do pokrycia pełnego asortymentu wyłóczek. Dlatego konfiguracja wariantów produkcyjnych musi powtarzać się w czasie, choć niekoniecznie cyklicznie. Niezależnie od metody bieżącego

sterowania produkcją powtarzalną duży wpływ na jakość sterowania wywierają podjęte wstępnie decyzje o konfiguracji wariantów produkcyjnych poszczególnych linii [3]. Z problemem tym autorzy zetknęli się m.in. w trakcie prób wdrożenia systemu informatycznego wspomagającego sterowanie produkcją w jednej z krajowych fabryk samochodów [7]. Przed zastosowaniem algorytmu harmonogramowania nadeżnego należało bowiem przyporządkować produkowane wytłoczki do jednej lub kilku spośród kilkunastu linii pras i do jednego lub kilku wariantów produkcyjnych wybranych linii. W przypadku harmonogramowania nadeżnego (NHP) zlecenia komórkowe danej linii są przetwarzane na decyzje o:

- chwilach przebrojeń linii produkcyjnych,
- długościach okresów pomiędzy przebrojeniami,
- wytwarzanych produktach (wybór wariantu produkcyjnego),
- wielkościach partii produkcyjnych.

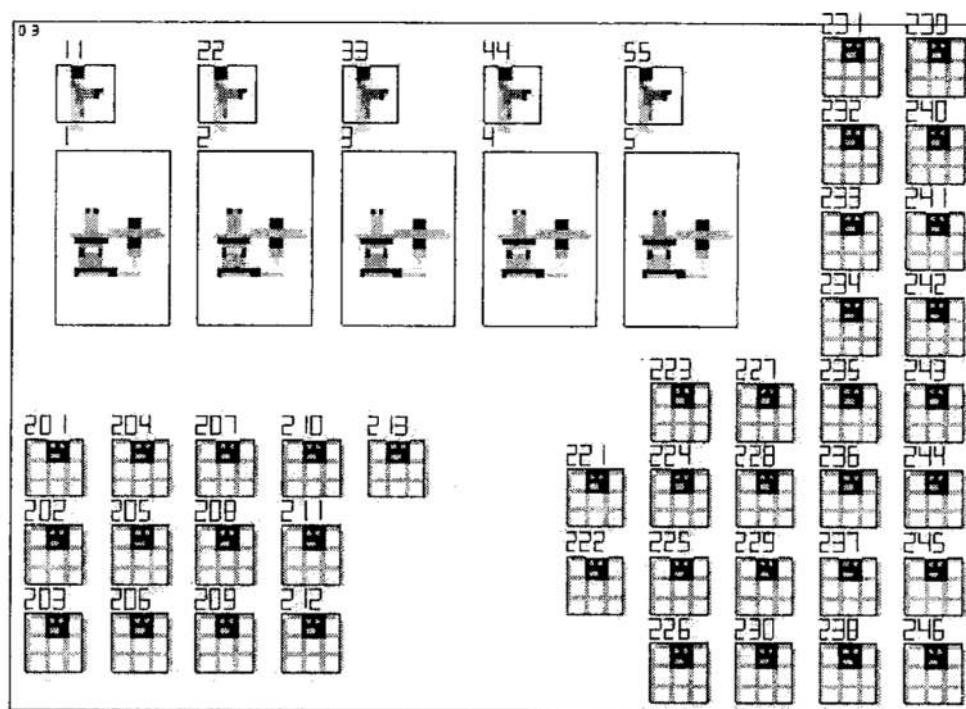
3 Model symulacyjny tłoczni lekkiej w systemie Taylor II

W rzeczywistym systemie produkcyjnym decyzje generowane przez algorytm NHP mogą zostać zakłócone przez występujące awarie maszyn, opóźnienia w transporcie czy wybór innego wariantu produkcyjnego na skutek zmian decyzji wykonawczych itd. W celu uwzględnienia tego typu zakłóceń dokonano symulacji przepływu detali przez system produkcyjny, przy wykorzystaniu systemu symulacji Taylor II. Tłocznia lekka przykładowej fabryki samochodów osobowych posiada trzy ciągi pras, dwie krajalnice oraz dwa oddziały wózków do transportu międzyoperacyjnego. Każdy ciąg składa się z pięciu identycznych pras. Tłocznia lekka została zamodelowana w systemie symulacji Taylor II (rys.2 i rys.3). Na rysunku 3 przedstawiono w powiększeniu jeden ciąg pras wraz z magazynami wytłoczek, które mogą być w nim wytwarzane.



Rys. 2. Uproszczony schemat rozmieszczenia elementów tłoczni lekkiej w systemie Taylor II

Fig. 2. Simplified layout of light press department in Taylor II system



Rys. 3. Schemat rozmieszczenia elementów pierwszego ciągu pras w systemie Taylor II
 Fig. 3. Layout of the first line of presses in Taylor II system

4 Przykład symulacji procesu produkcyjnego

Dla zamodelowanej tłoczni lekkiej (rys.2) wygenerowane zostały harmonogramy przy wykorzystaniu algorytmu NHP zaimplementowanego w informatycznym systemie symulacji sterowania produkcją przykładowej fabryki samochodów osobowych [1]. Przy generacji harmonogramów wykorzystano wyznaczony wcześniej przydział produktów do wariantów produkcyjnych linii. Został on przedstawiony w tabeli I, gdzie produkty (wyłoczki) są reprezentowane swoimi numerami z Indeksu Materiałowego. W tabeli II przedstawiono plany produkcji stanowiące podstawę wymuszeń przepływów w tłoczni lekkiej, które są również danymi wejściowymi dla algorytmu NHP. Harmonogramy powtarzalne dla poszczególnych ciągów pras zawierają decyzje m.in. o chwilach początkowych i końcowych okresów harmonogramowania, wybranym wariantcie pracy czy wielkości partii produkcyjnej. W eksperymencie symulacyjnym wykorzystano numery wariantów wynikające z harmonogramu nadążnego jako formę podpowiedzi dla operatora (rys. 4). Dane o planach widoczne w tabeli II umieszczono w tablicy *Schedule* systemu Taylor II, stanowiącej zbiór wymuszeń, przy czym w poszczególne komórki tej tablicy wpisano zapotrzebowania na konkretne wyłoczki podając indeks wyłoczki, okres planistyczny przeliczony na takty systemu Taylor (tab.III), w którym występuje zapotrzebowanie oraz wielkość zapotrzebowania. W tablicy *Schedule* można umieścić również harmonogramy nadążne. W takim przypadku symulowana będzie metoda NHP. Wykorzystanie harmonogramów nadążnych jako zbioru wymuszeń będzie przedmiotem dalszych eksperymentów symulacyjnych. System Taylor korzystając z danych stałych opisujących strukturę wyrobu (fragment pliku *config.tli*) w okresie planistycznym, w którym występuje zapotrzebowanie na wybraną wyłoczkę generuje zapotrzebowanie na odpowiednie przygotówki i symuluje ich wykonanie na nożycach stosując przy tym porządek leksykograficzny przy wyborze kolejności cięcia przygotówek. Przyjęto, że docięcie przygotówki zajmuje pięć taktów, a czasy przebrożeń nożyc podano w tabeli przez_k.

Po wykonaniu partii przygotówek (wielkość partii określono w tabeli partii p), są one przewożone wózkami, zasymulowanymi jako maszyny w systemie Taylor, do magazynów przed liniami pras. Przewóz składa się z trzech czynności: załadunku, transportu i wyładunku. Kiedy przygotówki znajdują się przed linią pras, system symulacji stwierdza, że można rozpocząć pracę na danej linii. Na początku kolejnego okresu harmonogramowania system zadaje pytanie operatorowi, do jakiego wariantu pracy linia ma być uzbrojona (podpowiada wariant wynikający z harmonogramu nadążnego). Po akceptacji wariantu linia jest uzbrajana do produkcji wytłoczek przydzielonych do tego wariantu (zbiór wariantów, plik var_xy.tli). Czas przebrojenia zapisany jest w tabeli przez_p. Następnie uruchamiana jest produkcja wytłoczek. Po wykonaniu całej partii produkcyjnej wytłoczki transportowane są do magazynu. Schemat pracy systemu symulacji oraz wymuszenia pokazano na rysunku 4.

Tab. I. Optymalna konfiguracja

Tab. I. Optimal configuration

G	H	J	J	J
3	0	-----	-----	-----
3	1	225	228	229
3	2	231	233	-----
3	3	245	-----	-----
3	4	221	237	244
3	5	227	230	240
4	0	-----	-----	-----
4	1	238	240	246
4	2	222	224	236
4	3	221	224	232
4	4	223	227	235
4	5	221	234	243
4	6	223	234	237
5	0	-----	-----	-----
5	1	222	226	242
5	2	230	242	243
5	3	230	234	246
5	4	221	236	241
5	5	234	236	239

Tab. II. Plany produkcyjne

Tab. II. Production plans

J	K	L	K	L
222	3	500	8	500
224	4	500	7	500
225	3	800	8	200
226	5	600	10	400
228	4	400	9	400
230	2	500	6	500
232	2	500	6	500
233	1	600	6	400
234	2	800	7	800
236	3	800	8	600
238	4	600	9	600
241	5	700	10	300
242	5	700	10	300
245	6	600	8	400

Tab. III Czasy symulacji

Tab. III Times of simulation

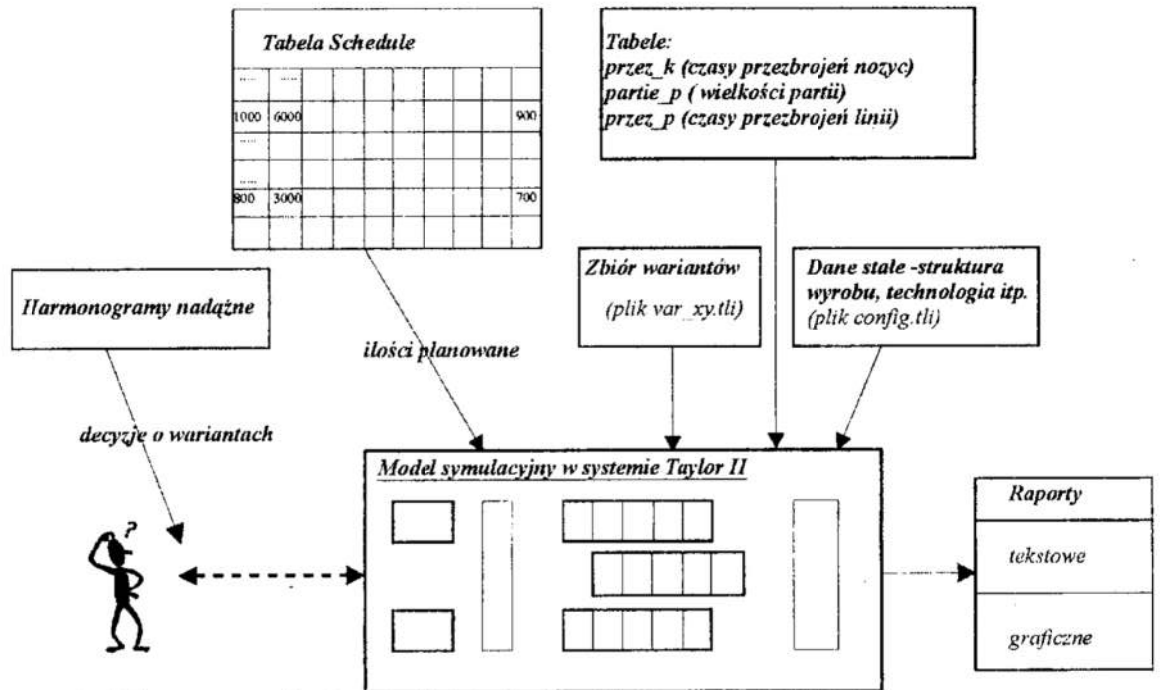
Gdzie:

K – numer okresu planistycznego
 L – planowana wielkość produkcji
 H – numer wariantu produkcyjnego
 G – indeks ciągu pras
 J – numer produktu

	Okres Planowania	Okres Harmonogram.
Minuty	480	15
Takty	2880	90

Aby dokonać symulacji należało przyjąć odpowiednie skale czasu, tzn. okresy planowania, harmonogramowania i odnieść je do sposobu reprezentacji czasu w systemie Taylor II, gdzie czas przedstawiany jest w postaci taktów symulacji. W przypadku symulacji przepływu detali w tłoczni lekkiej przyjęto, że każda minuta zawierać będzie 6 taktów, okres harmonogramowania 15 minut natomiast okres planowania 8 godzin (jedna zmiana). Zależności pomiędzy okresami wyrażonymi

w minutach i taktach pokazano w tabeli III. Przyjęto okres symulacji obejmujący 20 dni roboczych. Każdy dzień roboczy składa się z dwóch zmian roboczych co spowodowało, że okres symulacji obejmuje 1920 okresów harmonogramowania. Dokonano symulacji dla danych z tabeli I, II, III. Można było zaobserwować w postaci animacji detale przepływające w poszczególnych liniach i na poszczególnych maszynach. System Taylor II umożliwia raportowanie uzyskanych wyników przy wcześniejszym zadeklarowaniu odpowiednich parametrów w procedurach języka TLI. W tabeli IV przedstawiono wyniki symulacji.



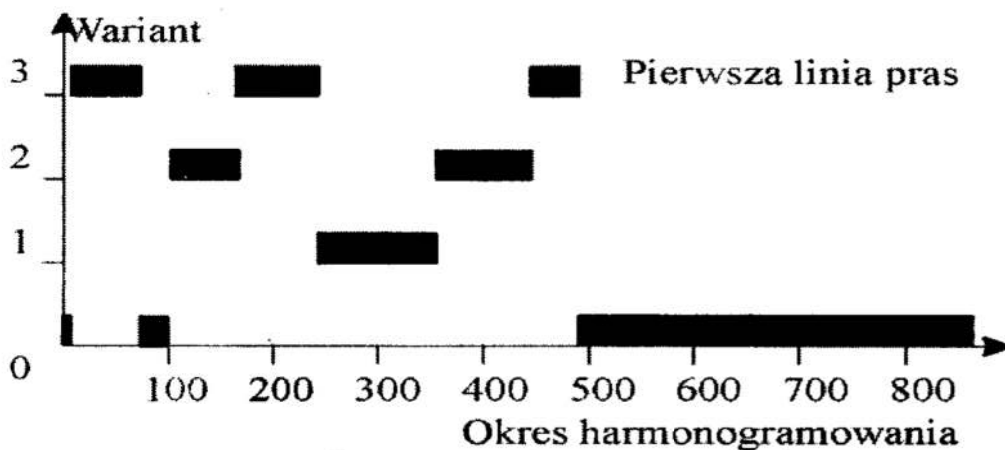
Rys. 4. Schemat symulacji w systemie Taylor II

Fig. 4. Schema of simulation in system Taylor II

Tab.IV Raport symulacji

Tab.IV Report of simulation

Produkt	Wielkość	Czas	Produkt	Wielkość	Czas
221			234	1600	32910
222	1000	59010	235		38975
223			236	1400	
224	1000	47485	237		44015
225	1000	44495	238	1200	
226	1000	48485	239		42260
227			240		
228	800	43615	241	1000	
229			242	1000	29645
230	1000	38015	243		35925
231			244		
232	1000	36025	245	1000	
233	1000	32910	246		25845



Rys. 5. Warianty pracy linii pierwszej (raport systemu Taylor II)

Fig. 5. Production variants of the first line (report from Taylor II)

Na rysunku 5 pokazano wykres z symulacji linii pierwszej pokazujący pracę w poszczególnych wariantach produkcyjnych w okresie symulacji (1920 okresów harmonogramowania).

5 Podsumowanie

System Taylor II okazał się przydatnym narzędziem do prowadzenia badań symulacyjnych wybranych systemów sterowania produkcją. Dużą zaletą jest przyjazne przedstawianie przy pomocy edytora graficznego poszczególnych elementów symulowanego systemu. Sam proces symulacji w postaci animacji jest bardzo wygodny szczególnie dla celów edukacyjnych. Okazało się jednak, że współpraca tego systemu z systemami zarządzania i sterowania produkcją naraża pewne problemy. Dane generowane przez te systemy zapisywane są bowiem w określonym formacie w bazie danych. System Taylor II wymaga zapisania tych danych we własnych predefiniowanych tabelach jak np. *schedule* lub w tabelach definiowanych przez użytkownika np. *przez_k*, *partie_p* itd. Możliwe jest również zasilanie danymi systemu Taylor II poprzez pliki z instrukcjami języka TLI np. *config.tli*, *var_xy.tli*. Zwiększa to elastyczność systemu, nie mniej jednak wymaga dużych nakładów programistycznych.

Literatura

1. Sitek P., Wikarek J., Zaborowski M. *Edukacyjny system symulacji sterowania produkcją zgodny ze standardem MRP II Materiały III K.K. Komputerowo Zintegrowane Zarządzanie, WNT 2000, strona. 160 – 170*
2. Zaborowski M.: *Zarys struktury uniwersalnego systemu zarządzania i sterowania produkcją. Materiały II K.K. „Komputerowo Zintegrowane Zarządzanie”, WNT 1999, str. 491-498.*
3. Zaborowski M., Sitek P. *Optymalizacja konfiguracji wariantów produkcyjnych w liniach pras tłoczni blach karoseryjnych Materiały XV Ogólnopolskiej Konferencji Polioptymalizacji Komputerowe Wspomaganie Projektowania, Mielno 1997, strona 317 – 322*
4. Zaborowski M., Wikarek J., *Model planowania potrzeb materiałowych z optymalizacją rozdziału obciążeń, Materiały XV Ogólnopolskiej Konferencji Polioptymalizacji Komputerowe Wspomaganie Projektowania Mielno'97, strona 323 – 330*

5. Landvater D.V., Gray C.D.: *MRP II Standard System*. Oliver Wight Publications, 1989.
6. Zaborowski M.: *Harmonogramowanie nadążne jako metoda dekompozycji złożonych zadań harmonogramowania produkcji*. Zeszyty Naukowe Politechniki Śląskiej, ser. Automatyka, z.109, 1992,s.345-352
7. Zaborowski M.(red.): *Modyfikacja i rozwój systemu sterowania produkcją ISTEP dla potrzeb FSS „POLMO-SHL” w Kielcach*. Raport z etapu I projektu celowego KBN Nr 1066/CSS-8/94. Kielce 1994 (niepublikowane).
8. Scheer A.-W.: *CIM (Computer Integrated Manufacturing) – Towards the Factory of the Future*. Springer-Verlag, 1994.
9. Sitek P. *Algorytmy obliczeniowe optymalizacji wariantów uzbrojenia linii produkcyjnych* Zeszyty Naukowe Politechniki Śląskiej, seria Automatyka, zeszyt 124, Gliwice 1998, strona 143 – 152
10. Zaborowski M., Jagodziński M.: *An execution planning model for complex manufacturing systems composed of production lines*. Dynamic Control and Management Systems in Manufacturing Processes, DYCOMANS NETWORK, WORKSHOP IV. Zakopane 1997, pp. 71 – 76.

Streszczenie

W artykule przedstawiono model symulacyjny tłoczni lekkiej w przykładowej fabryce samochodów osobowych. Dane do symulacji zaczerpnięto z systemu informatycznego sterowania tą fabryką. Dokonano przykładowych symulacji, których wyniki zaprezentowano w postaci raportów.

SIMULATION MODEL OF LIGHT PRESS IN A CAR FACTORY

Summary

The simulation model of light press department of a car factory has been presented in the paper. Data to simulation are taken from informatic system of this factory production control. The results of sample simulation experiment have been presented as a report from Taylor II.

Referat recenzowany

KOMPUTEROWE WSPOMAGANIE PODEJMOWANIA DECYZJI O PRZYJMOWANIU ZLECEŃ PRODUKCYJNYCH

1 Wprowadzenie

W zakresie planowania produkcji, spośród rozwijających się technik programowania matematycznego, symulacji komputerowej i metod sztucznej inteligencji największą popularność zdobyły metody symulacji komputerowej [1], [2]. Niedostatkami badań symulacyjnych jest długi czas trwania tego typu badań oraz ich koszt. Wiarygodność decyzji bazujących na badaniach symulacyjnych w praktyce ogranicza się do uprzednio sprawdzonych wariantów.

W warunkach ostrej konkurencji liczy się zdolność do podejmowania szybkich i trafnych decyzji, gdyż błędna ocena może prowadzić nie tylko do wymiernych strat finansowych, ale również do utraty klientów, a przez to potencjalnych korzyści. Dlatego też producenta nurtuje pytanie „*Czy to, co mam na myśli, jest akceptowalne?*”.

Wyznaczenie parametrów pracy systemu na poziomie planowania jest możliwe jedynie w przypadku produkcji rytmicznej. Jednakże ze względu na wymagania rynku co do zróżnicowanego asortymentu rozpatrywane będą systemy produkcji wieloasortymentowej.

W artykule przedstawiono metodę rozwiązywania zadań z zakresu planowania, a w szczególności decyzji o przyjmowaniu zleceń do realizacji polegającą na syntezie systemu poprzez sprawdzanie „łańcucha” warunków wystarczających, których spełnienie zapewnia osiągnięcie rozwiązania dopuszczalnego ze względu na ograniczenia, które musi spełnić. Przedstawiono program komputerowy, wspomagający planistę przy podejmowaniu decyzji oraz przedstawiono przykład jego zastosowania.

2 Metodyka weryfikacji zleceń

Zaproponowane podejście sprowadza się do etapowego podejmowania decyzji kolejno ograniczających zbiór rozwiązań dopuszczalnych. Umożliwia analizę zleceń bez potrzeby symulacji. Polega ono na tym, że w miejsce problemów optymalnego planowania i harmonogramowania produkcji, optymalnego rozmieszczenia komórek produkcyjnych, to znaczy problemów optymalizacji kombinatorycznej, które są problemami NP-trudnymi, rozważa się problemy syntezy (o wielomianowej złożoności obliczeniowej) rozwiązań dopuszczalnych. Podejście to, warunkujące zastosowania inżynierskie, okupione jest „niebezpieczeństwem suboptymalności” rozwiązań, gdyż rozwiązanie optymalne może być poza wyznaczonym podzbiorem rozwiązań dopuszczalnych.

Uwzględniając kryteria o charakterze ilościowym, przechodzi się do problemu, gdzie na danym (niepustym) zbiorze rozwiązań osiągalnych (być może niezawierającym już rozwiązania optymalnego) poszukiwane jest rozwiązanie najlepsze w określonym, ilościowym sensie. Jest to osiągnięte poprzez dobór wielkości partii produkcyjnych

* Katedra Technologii Maszyn i Zintegrowanych Systemów Wytwarzania, Politechnika Śląska,
ul. Konarskiego 18a, 44-100 Gliwice, e-mail: skolud@zeus.polsl.gliwice.pl

poszczególnych zleceń, przydział reguł rozstrzygania konfliktów zasobowych oraz przydział pojemności magazynów.

Przedstawiono dwa podejścia do problemu szybkiego wariantowania zleceń. Pierwsze z nich rozwiązuje problem przyjmowania pakietu zleceń do współbieżnej realizacji w systemie. Podejście to znajduje zastosowanie, w szczególności, na etapie planowania produkcji w nowym przedsiębiorstwie. Może być zastosowane również przy współbieżnej produkcji elementów, takich że wymuszona jest realizacja asortymentów w określonych proporcjach ilościowych i nie ma możliwości rezygnacji z któregoś z nich. W drugim podejściu zakłada się, że w systemie już są realizowane zadania w pewnym określonym rytmie, który nie może ulec zmianie [3], [4].

Zaprezentowana metodyka sterowania przepływem produkcji jest podstawą utworzenia systemu wspomagającego planistę w podejmowaniu decyzji o przyjmowaniu zleceń do systemu produkcji rytmicznej.

3 Aplikacja i przykład jej zastosowania

Różnorodność użytkowników (planiści, obsługa biur kontaktu z klientem, projektanci systemu) powoduje konieczność przygotowania aplikacji w powszechnie dostępnym środowisku, charakteryzującym się prostą budową i łatwą obsługą. Spełnienie tych wymogów gwarantuje implementacja w języku programowania Visual Basic w środowisku MS Windows. Programowanie obiektowe umożliwi rozbudowę pakietu bez potrzeby modyfikacji wcześniej utworzonych modułów.

Rozpatrywany pakiet pod nazwą System Weryfikacji Zleceń (SWZ) jest dostępny przez internet na stronie <http://zeus.polsl.gliwice.pl/~skolud>. Może on znaleźć zastosowanie w komórkach zajmujących się bezpośrednio planowaniem, w dziale sprzedaży do szybkiego szacowania możliwej terminowej realizacji zlecenia oraz na etapie projektowania nowych komórek produkcyjnych (wyznaczanie pojemności magazynów).

Obsługę SWZ ułatwiają okna dialogowe do wprowadzania danych wejściowych.

Na strukturę SWZ składają się dwie grupy okien dialogowych:

- okna dialogowe wprowadzania danych – *dane wejściowe*,
- okna, zawierające wyniki przeprowadzonych eksperymentów – *wyniki*.

Specyfikacja systemu wytwórczego obejmuje informacje o zasobach systemu, wprowadzonych jako niezależne pliki zewnętrzne do SWZ oraz dane o procesach już w nim realizowanych. Dane specyfikujące system to: liczba zasobów w systemie wytwórczym, liczba zleceń planowanych do realizacji, rodzaj magazynu oraz jego dysponowana pojemność. Ponadto użytkownik może wprowadzić dodatkowe informacje o stanie każdego z zasobów w każdej jednostce czasu trwania cyklu.

Specyfikacja zleceń zgłaszających się do realizacji w systemie produkcyjnym obejmuje następujące dane:

- wielkość zlecenia,
- wymagany (dyrektywny lub planowany) termin realizacji zlecenia,
- liczba operacji dla każdego procesu,
- marszruty alternatywne,
- czasy jednostkowe operacji oraz czasy przygotowawczo-zakończeniowe.

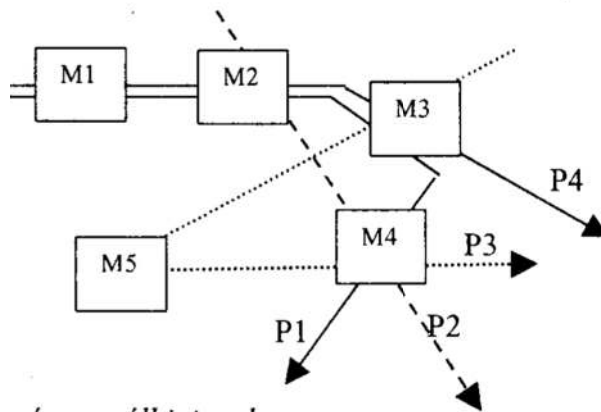
Struktura i działanie systemu

Podstawowym celem zastosowania SWZ jest udzielenie odpowiedzi na pytanie: czy dane zlecenie (pakiet zleceń) może być zrealizowane w danym systemie wytwórczym, nie zakłócając przepływu produkcji aktualnie w nim realizowanej. Pozytywnej odpowiedzi systemu towarzyszy wykaz alokowanych reguł wyboru priorytetu, stanowiący zapis zmodyfikowanej procedury sterowania rozproszonego. Ponadto SWZ umożliwia wyznaczenie: terminu realizacji zlecenia, ostatecznego terminu uruchamiania zlecenia gwarantującego dotrzymanie planowanego terminu realizacji, wskaźnika efektywności wykorzystania zasobów i wymaganych pojemności magazynów, jakie gwarantują realizację zleceń w danym systemie.

Interfejs użytkownika przedstawiony zostanie w trakcie prezentacji poniższego przykładu.

Przykład ilustrujący

Dany jest system 5 zasobowy, do systemu zgłosiły się 4 zlecenia produkcyjne. Przebieg procesów ilustruje rysunek 1, a formularz wprowadzania danych o zasobach systemu rysunek 2.



Rys. 1. System procesów współbieżnych
Fig. 1. The system of concurrent processes

Nr zasobu	Nazwa zasobu
1	Maszyna 1
2	Maszyna 2
3	Maszyna 3
4	Maszyna 4
5	Maszyna 5

Rys. 2. Formularz wprowadzania danych o zasobach
Fig. 2. Specification of resources

Następnie wprowadzana jest informacja o stanie systemu w momencie napływu zleceń. W komórce tej zaproponowana jest domyślna opcja stan **PUSTY** oznaczająca, że do

systemu do tej pory nie przyjęto żadnego zlecenia, możliwe jest też wybranie opcji **ZAJĘTY**, która jest wybierana w przypadku, gdy w systemie albo operują wcześniej przyjęte zlecenia, a zadanie sprowadza się do podjęcia decyzji o akceptacji kolejnego, albo zlecenia przyjmowane są jedno po drugim (dobierane). Do tego samego formularza zostaje wpisana liczba zleceń oczekujących na przyjęcie do realizacji w rozpatrywanym systemie wytwórczym oraz rodzaj i pojemność magazynu centralnego lub magazynów międzyoperacyjnych (rysunek 3).

Dane wejściowe o systemie

Stan systemu
 PUSTY
 ZAJĘTY

Liczba zleceń i zasobów
 Liczba zleceń w kolejce: 4
 Dane o zasobach: zasG-poliapt.

Magazyn
 CENTRALNY
 MIĘDZYOPERACYJNY

Pojemności magazynów międzyoperacyjnych

	M1	M2	M3	M4	M5
M1		4	0	0	0
M2	0		4	5	0
M3	0	0		2	0
M4	0	0	0		5
M5	0	0	0	0	

Maszyna 2

Rys. 3. Formularz elektroniczny wprowadzania danych o systemie

Fig. 3. Specification of the system

Specyfikacja zlecenia

Informacje o zleceniach oczekujących na realizację w systemie są wprowadzane przez operatora do tabeli „dane o zleceniach” (rys.4). Liczba operacji determinuje liczbę zasobów, jaką należy uwzględnić przy realizacji każdego ze zleceń.

Dane o zleceniu

Zlecenia	Wielkość zlecenia	Termin planowany	l. oper. = l. zas.
1	20	150	4
2	10	100	2
3	20	100	3
4	30	120	2

Rys. 4. Formularz elektroniczny wprowadzania danych o zleceniu

Fig. 4. Production order specification

Ponadto wypełniana jest **macierz marszrut alternatywnych** M_{Aj} która ma wymiary $m \times k$, gdzie m jest liczbą zasobów w systemie, a k jest liczbą operacji w j -tym procesie. Elementami macierzy są czasy realizacji poszczególnych operacji na zasobach.

System zgodnie z procedurą wybiera marszrutę i do pierwszego wiersza wprowadza numery zasobów, a do drugiego wiersza – czasy jednostkowe. Do trzeciego wiersza operator wprowadza czasy przygotowawczo-zakończeniowe.

The screenshot shows a graphical user interface for defining matrices. It consists of several overlapping windows:

- Macierz marszrut alternatywnych (MA₁):** A window with a grid for defining alternative routes. The grid has columns labeled 1, 2, 3 and rows labeled 1, 2.
- Macierz procesu (MP₁):** A window with a grid for defining process matrices. The grid has columns labeled 1, 2, 3, 4 and rows labeled t_{ij} and tpz_{ij}.
- Macierz procesu (MP₂):** A window with a grid for defining process matrices. The grid has columns labeled 2, 4 and rows labeled t_{ij} and tpz_{ij}.
- Macierz procesu (MP₃):** A window with a grid for defining process matrices. The grid has columns labeled 3, 5, 4 and rows labeled t_{ij} and tpz_{ij}.
- Macierz procesu (MP₄):** A window with a grid for defining process matrices. The grid has columns labeled 1, 2, 3 and rows labeled t_{ij} and tpz_{ij}.

At the bottom of the interface, there are three buttons: "Odczyt", "Zapis", and "Zatwierdź".

Rys. 5. Formularze elektroniczne: macierz marszrut alternatywnych, macierze procesu
 Fig.5. Specifications of matrix of the alternative routes and matrixes of the process

Dla rozpatrywanego przykładu – macierze procesów przyjmują postać:

$$MP_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 2 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad MP_2 = \begin{bmatrix} 2 & 4 \\ 2 & 2 \\ 1 & 1 \end{bmatrix} \quad MP_3 = \begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad MP_4 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Macierz struktury systemu

Kolejne wiersze tej macierzy odpowiadają zasobom kolumny - procesom realizowanym w systemie. Elementy macierzy odpowiadają liczbie wystąpień danego procesu w regule alokowanej na danym zasobie. Odzwierciedlają tym samym powtarzalność danego procesu na zasobie w trakcie jednokrotnej realizacji reguły. Wypełniane są tylko te pozycje w macierzy, które odpowiadają wybranym marszrutom. System proponuje wartości jeden. Oznacza to, że każdy z procesów przechodzi w trakcie realizacji reguły tylko jeden raz przez każdy z zasobów swojej marszruty. Operator może zmienić proponowane przez system wartości.

Macierz struktury systemu				
	1	2	3	4
1	1			1
2	1	1		1
3	1		1	1
4	1	1	1	
5			1	

Rys. 6. Formularz elektroniczny wprowadzania powtarzalności procesów w regułach alokowanych na zasobach systemu

Fig. 6. Electronic specification of the input data regarding repetitiveness of processes caused by dispatching rules allocated

Potwierdzenie poprawności wprowadzonych danych jest równoznaczne z uruchomieniem realizacji programu. W przypadku wystąpienia stanów, które wymagają decyzji operatora, SWZ generuje komunikat i oczekuje na jego decyzję.

Zakończenie działania SWZ poprzedzone jest wygenerowaniem raportu końcowego, który składa się z dwóch części:

- specyfikacja systemu i zlecenia podane przez operatora,
- informacje wygenerowane przez system

Wnioski

Dla analizowanego przykładu wygenerowany został następujący raport oraz wykres Gantta (rys. 7) przedstawiający cykliczną pracę systemu.

S W R --- Raport

Liczba zasobów w systemie : 5

Liczba zleceń w kolejce : 4

Dane o zleceniach :

Zlecenie 1: Wielkość zlecenia: 20. Termin planowany: 150. L. operacji: 4

Zlecenie 2: Wielkość zlecenia: 10. Termin planowany: 100. L. operacji: 2

Zlecenie 3: Wielkość zlecenia: 20. Termin planowany: 100. L. operacji: 3

Zlecenie 4: Wielkość zlecenia: 30. Termin planowany: 120. L. operacji: 3

Reguły priorytetu:

R1 (1, 4) R2 (1, 2, 4) R3 (1, 3, 4) R4 (1, 2, 3) R5 (3)

Powtarzalności reguł:

Na zasobie 1 - 1 razy. Na zasobie 2 - 1 razy. Na zasobie 3 - 1 razy.

Na zasobie 4 - 1 razy. Na zasobie 5 - 1 razy.

Dane o magazynie :

Wymagana pojemność magazynu

między zasobami 1 i 2 wynosi 1. Rzeczywista pojemność wynosi 4

między zasobami 2 i 3 wynosi 1. Rzeczywista pojemność wynosi 4

między zasobami 2 i 4 wynosi 1. Rzeczywista pojemność wynosi 5

między zasobami 3 i 4 wynosi 1. Rzeczywista pojemność wynosi 2

między zasobami 5 i 4 wynosi 1. Rzeczywista pojemność wynosi 5

między zasobami 3 i 5 wynosi 1. Rzeczywista pojemność wynosi 0

Czasy realizacji reguł alokowanych na zasobach :

Zasób 1 = 5

Zasób 2 = 10

Zasób 3 = 7

Zasób 4 = 6

Zasób 5 = 3

Okres pracy systemu wynosi 10

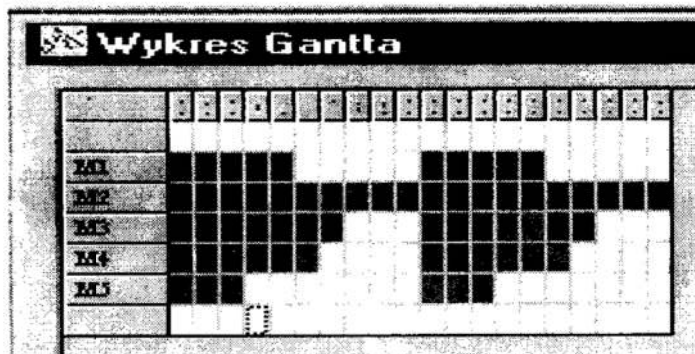
Terminowości realizacji zleceń :

Zlecenie 1 = 200 Zlecenie 2 = 100

Zlecenie 3 = 200 Zlecenie 4 = 300

Współczynnik wykorzystania zasobów wynosi 0,52

KONIEC raportu



Rys. 7. Wykres Gantta

Fig. 7. The Gantt's chart

4 Podsumowanie i kierunki dalszych badań

W pracy przedstawiono system SWZ wspomagający podejmowanie decyzji o przyjmowaniu zleceń do systemu wytwórczego. SWZ znajduje zastosowanie na etapie przyjmowania zleceń do systemu oraz przy projektowaniu produkcji w szczególności zaś wieloasortymentowej produkcji rytmicznej realizowanej w długich seriach, co jest charakterystyczne dla przemysłu samochodowego, przedsiębiorstw produkujących sprzęt AGD i wielu innych. Zastosowane podejście, polegające na propagacji ograniczeń umożliwia natychmiastową odpowiedź o możliwości realizacji zlecenia (zleceń) w danym systemie wytwórczym.

Badania wykonano w ramach projektu badawczego nr 7 T07D 025 15 pt: „Metoda harmonogramowania zleceń w dedykowanych systemach wytwarzania o charakterze cyklicznym” finansowanego przez Komitet Badań Naukowych.

Literatura

1. Chang F.R.: *A Study of Factors Affecting Due-Date Predictability in a Simulated Dynamic Job Shop*. Journ. of Manufacturing Sys. vol.13, no.6, 1994, s.393–400.
2. Gregor M., Haluskova M., Hromada J., Kosturiak J., Matuszek J.: *Simulation Manufacturing Systems*. Wydawnictwo Politechniki Łódzkiej, 1998.
3. Kłós S., Skołod B., Gattner D.: *Terminowość realizacji zleceń w systemie współbieżnych procesów produkcyjnych* Materiały II krajowej konferencji „Komputerowo zintegrowane zarządzanie”, Zakopane 12-14.01.1998, s. 161–168.
4. Skołod B.: *Production Order Control: Constraints Approach*. Proceedings of the International Conference on Computer Integrated Manufacturing, vol. II, WNT, Warszawa 1999, s. 219–226.

Streszczenie

W artykule przedstawiono podejście do rozwiązywania zadań z zakresu planowania, a w szczególności podejmowania decyzji o przyjmowaniu zleceń do realizacji, polegające na sprawdzaniu „łańcucha” warunków wystarczających, których spełnienie zapewnia osiągnięcie rozwiązania dopuszczalnego ze względu na ograniczenia. Metodologia ta jest podstawą utworzenia programu komputerowego „System Weryfikacji Zleceń”, wspomagającego planistę przy podejmowaniu decyzji. Przedstawiono interfejs użytkownika oraz przykład jego zastosowania.

COMPUTER AIDED DECISION MAKING ABOUT PRODUCTION ORDER VALIDATION

Summary

The approach to the validation of production orders is presented in the paper. The approach consists in checking of the sufficient conditions. Fulfilling of the conditions assure obtaining of the permissible solution. The methodology is the base of the computer aided tool (pol.: System Weryfikacji Zleceń”), assistance the planist in decision making. The user’s interface is presented and the illustrative example is presented, as well.

Referat recenzowany

DYCHOTOMICZNA KLASYFIKACJA KREDYTOBIORCÓW PRZY UŻYCIU SZTUCZNYCH SIECI NEURONOWYCH

Wykorzystanie sztucznych sieci neuronowych do oceny wiarygodności kredytowej zajmuje w ostatnim okresie czołową pozycję w literaturze przedmiotu. W Stanach Zjednoczonych, Niemczech i Wielkiej Brytanii tego rodzaju systemy znajdują się już w użyciu¹. Bogata literatura na ten temat stanowi o aktualności problemu, należy tutaj wymienić przede wszystkim prace: H. Rehkugler, A. Schmidt-von Rhein [4], R. F. Walker, E. W. Haasdijk, M. C. Gerrets [9], M. Schumann, S. A. Shumsky i A. V. Yaravoy [6] oraz D. Witkowska [10].

1 Struktura zbioru danych

W badaniach wykorzystano informacje dotyczące 110 przedsiębiorstw ubiegających się w latach 1994-1998 o przyznanie kredytu w jednym z banków w regionie łódzkim. Na ich podstawie obliczono podstawowe wskaźniki finansowe, który stanowiły pierwotną bazę danych w eksperymentach dotyczących dychotomicznej klasyfikacji firm za pomocą liniowej i logistycznej funkcji dyskryminacji. Każde przedsiębiorstwo zostało opisane przez wskaźniki finansowe wyznaczone na podstawie sprawozdań finansowych dla roku poprzedzającego składanie wniosku kredytowego ($t-1$) oraz dla okresu bieżącego t (tzn. od początku roku do momentu ubiegania się o kredyt) oraz przez dwie cechy jakościowe. Te ostatnie tj. rodzaj branży, decyzję odnośnie do przyznania (lub nie) kredytu. Zatem podstawowy zbiór danych wykorzystywanych w eksperymentach zawierał 50 zmiennych dla każdego ze 110 przedsiębiorstw, które zostały zdefiniowane następująco:

- x_1 i x_{25} - wskaźnik rentowności aktywów (ROA) [w %] odp. w okresie $t-1$ i t ;
- x_2 i x_{26} - wskaźnik rentowności kapitału własnego (ROE) [w %] odp. w okresie $t-1$ i t ;
- x_3 i x_{27} - wskaźnik rentowności sprzedaży (ROS) [w %] odp. w okresie $t-1$ i t ;
- x_4 i x_{28} - wskaźnik rentowności brutto [w %] odp. w okresie $t-1$ i t ;
- x_5 i x_{29} - wskaźnik rentowności netto [w %] odp. w okresie $t-1$ i t ;
- x_6 i x_{30} - wskaźnik płynności bieżącej odp. w okresie $t-1$ i t ;
- x_7 i x_{31} - wskaźnik krótkoterminowej płynności finansowej (szybkiej) odp. w okresie $t-1$ i t ;
- x_8 i x_{32} - wskaźnik długoterminowej płynności finansowej odp. w okresie $t-1$ i t ;
- x_9 i x_{33} - wskaźnik rotacji należności w dniach odp. w okresie $t-1$ i t ;
- x_{10} i x_{34} - wskaźnik rotacji zapasów w dniach odp. w okresie $t-1$ i t ;
- x_{11} i x_{35} - wskaźnik produktywności aktywów odp. w okresie $t-1$ i t ;
- x_{12} i x_{36} - wskaźnik poziomu kosztów odp. w okresie $t-1$ i t ;
- x_{13} i x_{37} - okres płacenia zobowiązań w dniach odp. w okresie $t-1$ i t ;
- x_{14} i x_{38} - wskaźnik rotacji majątku trwałego odp. w okresie $t-1$ i t ;
- x_{15} i x_{39} - wskaźnik rotacji majątku obrotowego odp. w okresie $t-1$ i t ;

* Zakład Metod Ilościowych w Zarządzaniu, Instytutu Zarządzania Politechniki Łódzkiej,
istan@ck-sg.p.lodz.pl.

¹ Porównaj T. Gwiazda [2], s. 131-143.

$x16$ i $x40$ - wskaźnik ryzyka aktywów odp. w okresie $t-1$ i t ;
 $x17$ i $x41$ - wskaźnik ogólnego zadłużenia odp. w okresie $t-1$ i t ;
 $x18$ i $x42$ - wskaźnik pokrycia majątku trwałego kapitałem stałym odp. w okresie $t-1$ i t ;
 $x19$ i $x43$ - wskaźnik długu (dźwignia finansowa) odp. w okresie $t-1$ i t ;
 $x20$ i $x44$ - wskaźnik zadłużenia kapitału własnego odp. w okresie $t-1$ i t ;
 $x21$ i $x45$ - wskaźnik pokrycia obsługi długu odp. w okresie $t-1$ i t ;
 $x22$ i $x46$ - wskaźnik pokrycia odsetek odp. w okresie $t-1$ i t ;
 $x23$ i $x47$ - wskaźnik zadłużenia środków trwałych odp. w okresie $t-1$ i t ;
 $x24$ i $x48$ - stopa zadłużenia odp. w okresie $t-1$ i t ;
 $x49$ - rodzaj branży;
 $x50$ - decyzja kredytowa (*wiarygodny* lub *niewiarygodny* klient).

Oprócz prezentowanego zbioru zmiennych eksperymenty numeryczne przeprowadzono wykorzystując w tym celu przyrosty wskaźników finansowych postaci $\Delta x_t = x_t - x_{t-1}$ wyrażających różnicę między wartościami wskaźników w okresach t i $t-1$. W ten sposób powstała nowa baza danych, które uzupełniono zmiennymi $x49$ i $x50$ zawierającą 26 zmiennych.

Dokonując klasyfikacji ze wzorcem istnieje konieczność podziału bazy danych na zbiór uczący (treningowy) i testowy. Strukturę obu zbiorów podano w tabeli 1.

Tabela 1. Struktura zbioru danych

Table 1. Structure of data set

Próba treningowa (ucząca)		Próba testowa	
Niewiarygodnych	Wiarygodnych	Wiarygodnych	Niewiarygodnych
45	45	10	10

Źródło: Opracowanie własne.

2 Metody doboru zmiennych diagnostycznych

Dobór cech diagnostycznych należy do zadań szczególnie ważnych, jako że w znacznym stopniu zależą od niego ostateczne wyniki badania. Zestaw cech diagnostycznych powinien być tak sporządzony, by w sposób możliwie pełny charakteryzował najważniejsze aspekty badanego zjawiska. Wybór cech odbywa się w drodze przetwarzania i analizy informacji statystycznych za pomocą odpowiednich procedur formalnych. Podstawą do wyboru cech diagnostycznych jest tzw. wstępna lista cech zaproponowana przez badacza na podstawie ogólnej znajomości zjawiska².

W wielu badaniach ekonomicznych istnieje potrzeba redukcji liczby zmiennych opisujących badany wycinek rzeczywistości. Potrzeba ta może wynikać z faktu posiadania mało licznej próby i jednakowo dużej liczby szacowanych parametrów lub występowania zmiennych powielających tę samą informację. Przeprowadzana redukcja musi odpowiadać pewnym wymaganiom, aby uzyskany opis nie fałszował rzeczywistości.

² Porównaj B. Podolec, K. Zajac [3], s. 20.

Tabela 2. Specyfikacja zmiennych diagnostycznych dla danych w formie wskaźników finansowych

Table 2. Diagnostic variables specification for data in the form of finance index

Specyfikacja zmiennych		Zestaw zmiennych diagnostycznych
Metoda Nowaka	wszystkie	x25, x40, x42
	podzielne	x25, x40
	wszystkie + rodzaj branży	x25, x40, x42, x49
	podzielne + rodzaj branży	x25, x40, x49
Metoda Hellwiga	wszystkie	x2, x3, x4, x6, x7, x8, x14, x18, x19, x22, x24, x25, x27, x30, x33, x38, x35, x47
	wszystkie podzielne	x2, x3, x4, x6, x19, x22, x24, x25, x27, x30, x33
	centralne	x3, x4, x7, x8, x18, x19, x24, x25, x27, x30, x33, x35, x47
	centralne podzielne	x3, x4, x19, x24, x25, x27, x30, x33
	izolowane	x2, x14, x22, x6, x38
	izolowane podzielne	x2, x22, x6
Algorytm genetyczny		x1, x3, x6, x7, x9, x10, x14, x18, x25, x28, x29, x30, x32, x42, x49

Źródło: Opracowanie własne.

Podstawowe metody doboru zmiennych diagnostycznych to:

- metoda analizy macierzy współczynników korelacji (tzw. Nowaka i Hellwiga);
- metoda algorytmu genetycznego.

Analizując tablicę 2 można zauważyć, że niemal wszystkie zestawy zmiennych zawierają wskaźnik rentowności aktywów (x25), który nie występuje jedynie w zestawie izolowanych i izolowanych podzielnych zmiennych wyznaczonych metodą Hellwiga.

W przypadku zmiennych zdefiniowanych w postaci przyrostów wskaźników finansowych we wszystkich zestawach zmiennych (wykluczając zmienne izolowane i izolowane podzielne) występuje przyrost wskaźnika rentowności aktywów.

Zastosowane metody preprocessingu to:

- analiza głównych składowych;
- analiza czynnikowa.

Tabela 3. Specyfikacja zmiennych diagnostycznych dla danych w postaci przyrostów wskaźników finansowych

Table 3. Diagnostic variables specification for data in the form of increase finance index

Specyfikacja zmiennych		Zestaw zmiennych diagnostycznych
Metoda Nowaka	wszystkie	$\Delta x_1, \Delta x_{20}$
	wszystkie + rodzaj branży	$\Delta x_1, \Delta x_{20}, x_{49}$
Metoda Hellwiga	wszystkie	$\Delta x_1, \Delta x_5, \Delta x_6, \Delta x_8, \Delta x_9, \Delta x_{15}, \Delta x_{14}, \Delta x_{16}, \Delta x_{18}, \Delta x_{21}, \Delta x_{22}, \Delta x_{24}$
	wszystkie podzielne	$\Delta x_1, \Delta x_5, \Delta x_6, \Delta x_8, \Delta x_{16}, \Delta x_{21}, \Delta x_{18}, \Delta x_{22}, \Delta x_{24}$
	centralne	$\Delta x_1, \Delta x_5, \Delta x_6, \Delta x_8, \Delta x_9, \Delta x_{15}, \Delta x_{16}, \Delta x_{21}$
	centralne podzielne	$\Delta x_1, \Delta x_5, \Delta x_6, \Delta x_8, \Delta x_{16}, \Delta x_{21}$
	izolowane	$\Delta x_{14}, \Delta x_{18}, \Delta x_{22}, \Delta x_{24}$
	izolowane podzielne	$\Delta x_{18}, \Delta x_{22}, \Delta x_{24}$
Algorytm genetyczny		$\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_6, \Delta x_7, \Delta x_8, \Delta x_9, \Delta x_{10}, \Delta x_{13}, \Delta x_{18}, \Delta x_{21}, x_{49}$

Źródło: Opracowanie własne.

Tabela 4. Specyfikacja zmiennych diagnostycznych otrzymanych po preprocessingu danych w formie wskaźników i ich przyrostów

Table 4. Diagnostic variables specification obtained out pre-processing for data in the form of finance index and theirs increase

Specyfikacja zmiennych	Liczba zmiennych		Procent objaśnianej zmienności pierwotnego zestawu danych	
	wskaźniki	przyrosty	wskaźniki	przyrosty
Analiza czynnikowa	7	10	-	-
Analiza głównych składowych	2	2	96,02%	94,85%
	3	5	98,77%	99,71%
	8	8	99,8%	100%

Źródło: Opracowanie własne.

Należy zauważyć, iż każdy z zaproponowanych zestawów zmiennych objaśnia w ponad 94,85% zmienność pierwotnego zestawu danych. Przedstawiono w tabelach 2, 3 i 4 zestawy zmiennych diagnostycznych wykorzystywane w dalszych analizach.

3 Sztuczne sieci neuronowe wykorzystywane w eksperymentach

Sztuczne sieci neuronowe są pojęciem ogólnym obejmującym różne rodzaje sieci uczone³ przy użyciu różnych algorytmów. Sieci neuronowe pełnią różnorodne funkcje i mają rozmaite formy. Idea sieci neuronowych jest bardzo uniwersalna i można ją wykorzystywać do rozwiązań różnego typu problemów, ale oczywiście różne wtedy będzie podejście do rozwiązywania zadania.

Najczęściej stosowanym typem sztucznej sieci neuronowej jest *perceptron wielowarstwowy* poddany procesowi nauczania nadzorowanego przy pomocy algorytmu propagacji wstecznej. Topologia powierzchni funkcji błędu perceptronu jest multimodalna i nie najlepiej poznana. Proces uczenia często „utyka” w minimum lokalnym i wobec braku zmian wartości błędu, jest przerywany. Algorytm propagacji wstecznej i inne stosowane algorytmy uczenia sieci neuronowych są oparte na klasycznych metodach optymalizacji. Metody te bazują na zasadzie twardej selekcji, zakładającej tworzenie nowych punktów bazowych do dalszych poszukiwań wyłącznie na bazie najlepszych z dotychczasowych.

W wielu pracach z zakresu sztucznych sieci neuronowych eksperymentalnie uzasadniono, iż najlepsze rezultaty otrzymuje się w przypadku perceptronu wielowarstwowego jeżeli funkcja aktywacji pomiędzy warstwą wejściową, a ukrytą jest funkcją sigmoidalną⁴.

Sieci realizujące regresję uogólnioną (GRNN – *Generalized Regression Neural Networks*) składają się z czterech warstw:

- a) warstwy wejściowej, którą stanowią neurony liniowe;
- b) pierwszej warstwy ukrytej będącej warstwą radialną – neurony tej warstwy reprezentują skupienia występujące w danych, zaś wartości wyjściowe tych neuronów zmniejszają się wraz ze zwiększaniem się odległości pomiędzy wektorem wejściowym, a wektorem wag - liczba neuronów w tej warstwie jest określona przez liczbę skupień w danych;
- c) drugiej warstwy ukrytej, tzw. warstwy regresyjnej składającej się z dwóch typów

³ Nie wszystkie sieci neuronowe muszą podlegać procesowi uczenia, przykładem może tu być sieć Hopfielda i sieci o radialnych funkcjach bazowych.

⁴ Porównaj Rehkugler H., Schmidt von Rhein A. [4], s. 29.

neuronów: typu A (jest ich tyle ile neuronów w warstwie wyjściowej, każdy neuron tego typu sumuje iloczyny wejść neuronów radialnych i odpowiednich wag) oraz jednego neuronu typu B (neuron tego typu sumuje wartości funkcji gęstości prawdopodobieństwa dla każdego elementu zbioru uczącego, a właściwie wyjść neuronów radialnych);

d) warstwy wyjściowej, która służy do obliczania ilorazu wartości obliczonych przez neuron klasy A i neuron klasy B⁵.

Sieci o radialnych funkcjach bazowych (RBF) składają się z trzech warstw:

a) *wejściowej*, którą stanowią neurony liniowe;

b) *ukrytej*, którą stanowią neurony radialne o radialnej funkcji potencjału postsynaptycznego (PSP) i expotencjalnej funkcji aktywacji;

c) *wyjściowej* reprezentowanej przez neurony liniowe.

Uczenie sieci RBF odbywa się w trzech kolejnych etapach:

a) pierwszy to określenie centrów radialnych, czyli dobór centrów funkcji bazowych poprzez określenie wartości wag dla każdego neuronu radialnego;

b) drugi to określenie odchyleń radialnych, czyli dobór szerokości funkcji aktywacji-parametr określający kształt funkcji aktywacji przechowywany jest jako wartość progowa neuronu radialnego;

c) ostatni to określenie wag neuronu warstwy wyjściowej.

4 Wyniki eksperymentów numerycznych

Celem badań była klasyfikacja klientów banku za pomocą sztucznych sieci neuronowych. Eksperymenty zostały przeprowadzone dla 37 zestawów zmiennych diagnostycznych, przedstawionych w tablicach 2-4. Jakość klasyfikacji oceniono na podstawie błędów⁶, których wartości dla 20-elementowego zbioru testującego zamieszczono w tabeli 5. Podano w niej również symbole zmiennych statystycznie istotnych.

⁵ Porównaj R. Tadeusiewicz, P. Lula [8], s. 40-42.

⁶ Definicję mierników jakości klasyfikacji przedstawiono w pracy I. Staniec [7].

Tabela 5. Porównanie błędów klasyfikacji (w %) wyznaczonych dla różnych zestawów zmiennych diagnostycznych i wybranych sieci neuronowych
 Table 5. Comparison classification errors (in %) calculated for different diagnostic variables and selected artificial neural networks

Specyfikacja zmiennych	Błędy	MLP		RBF		GRNN	
		Wskaźniki	Przyrosty	Wskaźniki	Przyrosty	Wskaźniki	Przyrosty
Metoda Nowaka	E	10	50	0	55	10	50
	E1	0	0	0	80	0	100
	E2	20	10	0	30	20	0
Zestaw zmiennych		x25, x40, x42	$\Delta x1, \Delta x20$	x25, x40, x42	$\Delta x1, \Delta x20$	x25, x40	$\Delta x1, \Delta x20$
Metoda Hellwiga	E	10	35	10	35	5	40
	E1	10	10	10	50	0	10
	E2	10	60	10	20	10	70
Zestaw zmiennych		centralne	centralne podzielne	wszystkie	izolowane	wszystkie podzielne	centralne
AGS	E	35	40	40	25	45	25
	E1	60	80	70	20	90	40
	E2	10	0	10	30	0	10
ACZ	E	30	45	30	40	40	50
	E1	0	0	30	40	40	100
	E2	60	90	30	40	40	0
AG	E	20	40	35	55	40	60
	E1	40	50	10	60	0	60
	E2	0	30	60	50	80	60

Zródło: Opracowanie własne.
 Uwaga: W tabeli przedstawiono najmniejsze błędy klasyfikacji dla próby testowej. W wierszu pod błędami podano zestaw zmiennych dla których uzyskano taki wynik.

5 Wnioski końcowe

Na podstawie przeprowadzonej analizy empirycznej można sądzić, że sieci neuronowe takie jak: sieci o radialnych funkcjach bazowych i realizujące regresję uogólnioną są przy dychotomicznej klasyfikacji klientów kredytowych na klasy: *wiarygodnych* i *niewiarygodnych* kredytobiorców sprawnym instrumentem.

Ponieważ budowa sieci ma wpływ na kompleksowości uzyskanych wyników, więc przedstawiona klasyfikacja przedsiębiorstw jest tylko przykładowa. Przy sieci neuronowej nie obowiązują restrykcyjne akceptacje dotyczące zastosowanych zmiennych jak przy metodach statystycznych więc stosować można sieć neuronową bardziej uniwersalną. Wyniki uzyskane w przypadku sieci dla wszystkich zestawów zmiennych wejściowych są znacznie gorsze niż dla wybranych zmiennych, szczególnie dla sieci o radialnych funkcjach bazowych i sieci realizujących regresję uogólnioną. Zatem przy użyciu tych sieci do klasyfikacji konieczny jest preprocessing zmiennych wejściowych.

Nakład czasu, żeby znaleźć nadające się ustawienie parametrów dla sieci neuronowej, jest tak samo wysoki, jak nakład czasu, żeby przeprowadzić wstępną analizę zmiennych diagnostycznych.

Ważnym punktem krytyki sieci neuronowych jest również fakt, że znalezienie rozstrzygnięć dla sieci neuronowych nie jest proste i w wielu przypadkach nie od razu słuszne. Sieć neuronowa, która w tym badaniu najlepiej klasyfikuje jest jednak relatywnie prosta⁷.

W przypadku każdej z metod klasyfikacji najlepsze wyniki uzyskano dla zmiennych diagnostycznych dobranych metodą Nowaka, tzn. dla wskaźników: x_{25} (ROA w okresie t), x_{40} (wskaźnik ryzyka aktywów w okresie t) i x_{42} (wskaźnik pokrycia majątku trwałego kapitałem stałym w okresie t). Wprowadzenie jako jednej ze zmiennych wejściowych zmiennej jakościowej transponowanej do postaci ilościowej - rodzaju branży powodowało za każdym razem pogorszenie wyników klasyfikacji.

Wyniki analiz empirycznych przemawiają na niekorzyść metod preprocessingu zbioru danych.

Sieć o radialnej funkcji bazowej ze zmiennymi wejściowymi w postaci wskaźników x_{25} , x_{40} i x_{42} rozpoznała 100% przedsiębiorstw poprawnie. Natomiast sieć realizująca regresję uogólnioną ze zmiennymi wejściowymi w postaci wszystkich podzielnych zmiennych Hellwiga zakwalifikowała poprawnie 95% przedsiębiorstw podobnie jak logistyczna funkcja dyskryminacji ze zmiennymi wejściowymi w postaci wskaźników: ROA w okresie t (x_{25}), wskaźnik ryzyka aktywów w okresie t (x_{40}) i wskaźnik pokrycia majątku trwałego kapitałem stałym w okresie t (x_{42}). Należy zauważyć, iż (prawie) dla każdej z metod klasyfikacji uzyskano najniższe błędy dla różnych zestawów zmiennych wejściowych.

Literatura

1. Findeisen W., Szymanowski W., Wierzbicki A.: *Teoria i metody obliczeniowe optymalizacji*, PWN, Warszawa 1977.
2. Gwiazda T. D.: *Algorytmy genetyczne. Zastosowanie w finansach*, Wyższa Szkoła Przedsiębiorczości i Zarządzania im. L. Koźmińskiego, Warszawa 1998.

⁷ Jest to sieć RBF o strukturze 3/14/1.

3. Podolec B, Zając K.: *Ekonometryczne metody ustalania regionów konsumpcji*. PWE Warszawa 1978.
4. Rehkugler H., Schmidt-von Rhein A.: *Kreditwürdigkeitsanalyse und Prognose für Privatkundenkredite Mittels Statischer Methoden und Künstlicher Neuronaler Netze. Eine empirisch - vergleichende Studie*. Otto Friedrich Universität, Bamberg, discussion paper 93, maszynopis 1993.
5. Schumann M., Lohrbach T., Bahrs P.: *Versuche Kreditwürdigkeitsprognose mit Künstlichen Neuronalen Netzen*. Zeszyt nr 2, Göttingen 1992.
6. Shumsky S. A., Yarovoy A. V.: *Self-Organizing Atlas of Russian Banks*, w: *Visual Explorations in Finance with Self-Organizing Maps*, pod red. G. Deboeck, T. Kohonen. Springer-Verlang, London 1998.
7. Staniec Iwona: *Sztuczne sieci neuronowe, a analiza dyskryminacyjna w klasyfikacji wniosków kredytowych*. Zeszyty Naukowe Wydziału Mechanicznego nr 26 „Poliptymalizacja i komputerowe wspomaganie projektowania Mielno '99”, Politechnika Koszalińska, s. 285-292.
8. Tadeusiewicz R., Lula P.: *Statistica Neural Networks*, materiał kursowy. StatSoft Polska, Kraków 1999.
9. Walker R. F., Haasdijk E. W., Gerrets M. C.: *Credit Evaluation Using a Genetic Algorithm*, w: *Intelligent Systems for Finance and Business*. Wiley 1995.
10. Witkowska D.: *Sztuczne sieci neuronowe w analizach ekonomicznych*. KBN, Łódź 1999.

Streszczenie

W celu minimalizowania ryzyka, standaryzowania i obiektywizowania procesu związanego z udzieleniem kredytu, banki na świecie doskonalały metody oceny zdolności kredytowej wykorzystując nowości naukowe, techniczne oraz własne doświadczenia wynikające z wieloletniej praktyki. W ostatnim okresie okazało się, że sztuczne sieci neuronowe mogą spełnić te funkcje.

DYCHOTOMUS CLASSIFICATION BORROWER WITH USE ARTIFICIAL NEURAL NETWORKS

Summary

In order to minimise risk, standardisation and objectivity work connected with giving credit, banks all over the world improve the methods of assessment in the credit granting procedure. They take advantage of news in science, technology and their own experience. Lately it has turned out that neural networks can perform this function.

Referat recenzowany

OPTYMALIZACJA ROZDZIAŁU ZASOBÓW LOKALOWYCH W WARUNKACH ELASTYCZNEGO STUDIOWANIA

1 Wprowadzenie

Praktyczne zagadnienia układania zajęć na wyższych uczelniach są złożonymi problemami harmonogramowania zawierającymi trzy rodzaje danych: zbiór terminów, zbiór zajęć z określonymi wymaganiami oraz zestaw zasobów. Zbiór terminów ma strukturę dwu- lub trój-poziomową zawierającą godziny, dni i tygodnie. Do zbioru zasobów wymaganych przez zajęcia należą nauczyciele, sale oraz studenci. Zasoby mają określoną dostępność w ciągu godzin, dni i tygodni.

W literaturze poświęconej harmonogramowaniu zajęć istnieje wiele prac opisujących liczne metody i algorytmy układania planu zajęć dla różnorodnych, szczególnych typów problemów, np. prace [2], [3], [4], [5]. W stosunku do rzeczywistych problemów układania zajęć rozważane modele są jednak na ogół zbyt uproszczone i nie uwzględniają wielu ważnych uwarunkowań praktycznych.

Do rozwiązywania szczególnie złożonych problemów układania zajęć, w celu zmniejszenia złożoności problemów decyzyjnych jest stosowany strukturalny, hierarchiczny schemat układania zajęć, opisany w pracy [9]. W jego ramach jest dokonywana agregacja występujących w problemie elementów, tj. zasobów, zajęć i terminów. W szczególności, agregacja zasobów polega na rekurencyjnej budowie wielostopniowej struktury kategorii zasobów o tych samych właściwościach ze względu na możliwość ich wykorzystania przez określone zajęcia. W jej wyniku otrzymujemy zasoby wyższego poziomu (agregaty), w skład których wchodzi szereg „identycznych” zasobów niższego poziomu.

Po ułożeniu harmonogramu zajęć na danym poziomie agregacji problemu, jest konieczne przeprowadzenie dezagregacji rozwiązania. Jednym z zasadniczych elementów dezagregacji rozwiązania jest uszczegółowienie alokacji zasobów lokalowych wymaganych do prowadzenia zajęć. Wyniki dotychczasowych badań na ten temat zawiera praca [8].

Punktem wyjścia jest założenie o znajomości dopuszczalnego, w sensie zagregowanych wymagań zasobowych, rozkładu zajęć. Dla każdego zajęcia umieszczonego w rozkładzie jest znany zbiór terminów, w których jest ono realizowane, jego zagregowane wymagania dotyczące zasobów lokalowych oraz liczba i liczebności uczestniczących w nim grup studenckich. Każde zajęcie wymaga dokładnie jednej sali. Tym samym, dla każdej grupy studenckiej, każdego dnia, jest znany zestaw oraz kolejność realizowanych przez nią zajęć. Wynika z tego, że dla każdej pary zajęć odbywających się w dwóch kolejnych terminach jest znana liczba przechodzących studentów. W ogólnym przypadku z zajęcia na zajęcia może nastąpić przejście tylko części uczestniczących w nich

* Instytut Automatyki i Informatyki Stosowanej Politechniki Warszawskiej, ul. Nowowiejska 15/19,
00-665 Warszawa, e-mail: cszwed@ia.pw.edu.pl, etoczyłowski@ia.pw.edu.pl

studentów i każde z zajęć może wymagać sali innej kategorii. Horyzont planowania stanowi dzień składający się z ustalonej liczby godzin. Każdy zagregowany zasób lokalowy reprezentuje zbiór zasobów lokalowych szczegółowych – zbiór indywidualnych sal. Każda sala ma ustaloną pojemność wyrażoną liczbą dostępnych miejsc lub liczbą tzw. *modułów*, odpowiadających kilku lub kilkunastu miejscom.

Sale mogą znajdować się w różnych odległościach od siebie – mogą być położone w różnych budynkach, na różnych piętrach. Dla każdej pary sal jest znany jednostkowy koszt przejścia studentów, który zależy np. od odległości pomiędzy salami lub od czasu potrzebnego na przejście z jednej sali do drugiej.

Celem dezagregacji zasobów lokalowych jest alokacja do zajęć sal o wymaganych pojemnościach, przy minimalizacji sumy kosztów przejść studentów pomiędzy salami przydzielonymi do zajęć. Każda sala musi być przydzielona na cały czas trwania zajęcia.

W pracy przedstawiono analizę zagadnienia w przypadku tzw. *przejsć jednoetapowych*, dla zajęć odbywających się w dokładnie dwóch kolejnych terminach. Sformułowano model matematyczny problemu, zaproponowano metodę jego rozwiązywania oraz przedstawiono wyniki eksperymentów obliczeniowych.

2 Przejścia pomiędzy zajęciami realizowanymi przez dokładnie te same grupy studentów

W przypadku, gdy w dwóch odbywających się kolejno po sobie zajęciach, pomiędzy którymi następuje przepływ studentów, biorą udział dokładnie ci sami studenci (grupa lub grupy studenckie) problem można znacznie uprościć. Ze względu na to, że grupy studenckie uczestniczące w obu zajęciach mają jednakowe liczebności zajęcia te muszą wymagać sal tych samych kategorii. Na wstępie zostaną wprowadzone pewne pojęcia pomocnicze.

Dwa zajęcia odbywające się w dwóch kolejnych terminach nazywamy *szeregowo stycznymi* jeśli uczestniczą w nich dokładnie ci sami i tylko ci studenci. *Łłańcuchem elementarnym* zajęć nazywamy ciąg zajęć szeregowo stycznych, wymagających sal wchodzących w skład tego samego zasobu zagregowanego. W szczególnym przypadku łańcuchem elementarnym zajęć jest jedno zajęcia.

Lemat 1. Jeśli suma wymagań zasobu k w każdym terminie t nie przekracza ogólnej dostępności tego zasobu Q_t , to istnieje przydział sal do grup studenckich nie wymagający przechodzenia od sali do sali tej samej kategorii, tzn. każdy łańcuch elementarny może mieć przydzieloną dokładnie jedną salę.

Dowód. Rozważmy terminy $t, t+1$. Niech S_{t+1} oznacza zbiór łańcuchów elementarnych rozpoczynających się w terminie $t+1$, E_t zbiór łańcuchów kończących się w terminie t , K_t zbiór łańcuchów kontynuowanych w terminach $t, t+1$. Z założenia $E_t + K_t \leq Q_t$ oraz $K_t + S_{t+1} \leq Q_{t+1}$. Dla dowolnego przydziału sal do K_t istnieje dokładnie ten sam przydział sal w terminie $t+1$, ponieważ sale wymagane do przydziału do S_{t+1} w terminie $t+1$ są zwalniane przez E_t w terminie t . Taki przydział jest zawsze dopuszczalny.

Z powyższego lematu wynika, że elementarny łańcuch zajęć można zastąpić agregatem zajęć, które wchodzą w skład tego łańcucha i do których przydziela się sale globalnie na

cały okres trwania łańcucha. W ten sposób ulega redukcji liczba przejść grup studenckich pomiędzy zajęciami a sam problem przejść dotyczy jedynie par zajęć angażujących grupy uczestników, które częściowo się różnią. Natomiast dla tej samej grupy studenckiej przejścia występują pomiędzy zajęciami wchodzącymi w skład różnych łańcuchów elementarnych tej grupy, wymagających sal różnych kategorii. W pewnych szczególnych przypadkach kilka łańcuchów elementarnych kończących się w terminie t i wymagających tej samej bądź różnych kategorii sal, trzeba skojarzyć z jednym łańcuchem rozpoczynającym się w terminie $t+1$ i wymagającym tej samej kategorii sal.

3 Przejścia pomiędzy zajęciami realizowanymi przez różniące się grupy studentów

Problem analizowany w tym rozdziale dotyczy sytuacji, w której w odbywających się kolejno po sobie zajęciach, pomiędzy którymi następuje przepływ studentów, biorą udział różniące się grupy studentów. W takim przypadku zajęcia mogą wymagać sal tych samych kategorii, gdy np. część studentów z wykładu wymagającego sali audytorijnej przechodzi (razem z innymi studentami) na inny wykład również wymagający sali audytorijnej. Możliwa jest również inna sytuacja, w której zajęcia wymagają sal różnych kategorii, gdy np. wszyscy studenci z ćwiczeń wymagających sali o małej pojemności przechodzą (razem z innymi studentami) na wykład wymagający audytorium. Na wstępie zostaną wprowadzone pewne pojęcia pomocnicze.

Zbiory zajęć odbywające się w dwóch kolejnych terminach nazywamy *stycznymi*, jeśli każdy z nich obsługuje dokładnie ten sam podzbiór grup studenckich. Pojedyncze zajęcie może równocześnie wchodzić w skład różnych zbiorów zajęć stycznych. Należy to rozumieć w taki sposób, że część grup studenckich tego zajęcia należy do podzbioru grup studenckich obsługiwanych przez pewien zbiór zajęć stycznych, a inna część grup studenckich zajęcia może należeć do podzbioru grup studenckich obsługiwanych przez inny zbiór zajęć stycznych.

Definicja 1. Stykiem zajęć nazywamy dwa zbiory zajęć stycznych.

Zajęcie nazywamy *kompletnym* w zbiorze grup studenckich, jeśli wszystkie grupy studenckie wymagane przez to zajęcie zawierają się w tym zbiorze grup studenckich. Styk zajęć nazywamy *zupełnym*, jeśli każde z zajęć należące do każdego ze zbiorów zajęć tego styku jest kompletne w podzbiorze grup studenckich obsługiwanych przez ten styk. Styk zajęć nazywamy *minimalnym*, jeśli jest on zupełny i nie można go podzielić na styki zupełne realizowane w podzbiorach grup studenckich o mniejszej mocy.

Definicja 2. Elementarnym stykiem zajęć nazywamy styk, w którym przynajmniej jeden ze zbiorów zajęć składa się z dokładnie jednego zajęcia, kompletnego w zbiorze grup studenckich realizujących to zajęcie.

Sposób podziału zupełnych styków zajęć na styki elementarne jest obecnie przedmiotem badań.

Tak więc w problemie analizowanym w tym rozdziale można wyodrębnić wszystkie minimalne styki zajęć pomiędzy dwoma kolejnymi terminami, a w ramach każdego minimalnego styku zajęć wszystkie elementarne styki zajęć. W takim przypadku pierwotne zadanie przydziału sal do zajęć, pomiędzy którymi w dwóch kolejnych terminach następuje przepływ studentów, można przekształcić do zadania przydziału sal do zajęć należących do elementarnych styków zajęć. Zaletą takiego przekształcenia jest możli-

wość modelowania problemu w postaci dobrze znanego z literatury zadania programowania sieciowego z przepływem wielotowarowym [1].

Każda sala przydzielona w elementarnym styku do zajęcia musi być wypełniona w całości przez studentów uczestniczących dokładnie w jednym zajęciu. Dotyczy to również sal i zajęć, które jednocześnie należą do różnych elementarnych styków zajęć. Powyższe wymagania są określane jako tzw. *warunki strukturalne wypełnienia sal*.

4 Model matematyczny problemu

Model matematyczny problemu przydziału sal do styków elementarnych (model PSSE) został opracowany w postaci zadania programowania sieciowego z przepływem wielotowarowym.

Wprowadzamy następujące indeksy oraz zbiory indeksów: m – indeks elementarnego styku zajęć; i, j – indeksy indywidualnych sal; k, l – indeksy zasobów zagregowanych; R_k, R_l – zbiory sal, których reprezentantami są, odpowiednio: k -ty oraz l -ty zasób zagregowany; R_{i-}^m – zbiór sal, do których może nastąpić przepływ studentów z sali i w styku m ; R_{-j}^m – zbiór sal, z których może nastąpić przepływ studentów do sali j w styku m ; M_{k-} – zbiór elementarnych styków zajęć wymagających k -tego zasobu zagregowanego w terminie t ; M_{-l} – zbiór elementarnych styków zajęć wymagających l -tego zasobu zagregowanego w terminie $t+1$. W ogólnym przypadku, w zależności od rodzaju i liczebności zajęć, pomiędzy którymi następuje przepływ studentów, w skład zbiorów R_{i-}^m oraz R_{-j}^m mogą równocześnie wchodzić różne kategorie sal.

Definiujemy następujące zmienne:

x_{ij}^m – przepływ studentów z sali i do sali j w elementarnym styku zajęć m (w modułach);

x_{i-}^m – odpływ studentów z sali i po terminie t w elementarnym styku zajęć m (w modułach);

x_{-j}^m – dopływ studentów do sali j przed terminem $t+1$ w elementarnym styku zajęć m (w modułach).

Wprowadzamy następujące parametry:

c_{ij} – koszt przejścia z sali i do sali j (dla jednego modułu);

p_{k-}^m – wymaganie na zagregowany zasób k -tego rodzaju w elementarnym styku zajęć m w terminie t (w modułach);

p_{-l}^m – wymaganie na zagregowany zasób l -tego rodzaju w elementarnym styku zajęć m w terminie $t+1$ (w modułach);

U_k, U_l – pojemności sal, odpowiednio: kategorii k i kategorii l (wyrażone w modułach);

U_{ij} – maksymalna liczba studentów (wyrażona w modułach), którzy mogą przejść z sali i do sali j , związana z pojemnościami tych sal.

Koszty przejść pomiędzy salami mogą być związane z tym, że sale znajdują się w różnych odległościach od siebie lub z tym, że potrzebny jest pewien czas na pokonanie odległości pomiędzy salami. W ogólnym przypadku należy przyjąć, że c_{ij} jest różne od c_{ji} , oraz $c_{ii} = 0$. Koszty dojścia studentów na pierwsze zajęcie w ciągu dnia oraz koszty wyjścia z ostatniego zajęcia są zerowe. Parametr U_{ij} jest wyznaczany w następujący sposób:

$$U_{ij} = \min(U_k, U_l) \quad \forall i, j: i \in R_k, j \in R_l \quad (1)$$

Postać modelu *PSSE* jest następująca:

$$\min \sum_{i,j,m} c_{ij} x_{ij}^m \quad (2)$$

przy ograniczeniach:

$$\sum_{i \in R_k} x_{i-}^m = p_{k-}^m \quad \forall k, \forall m \in M_{k-} \quad (3)$$

$$x_{i-}^m = \sum_{j \in R_l^m} x_{ij}^m \quad \forall k, \forall m \in M_{k-} \forall i \in R_k \quad (4)$$

$$\sum_{i \in R_l^m} x_{ij}^m = x_{-j}^m \quad \forall l, \forall m \in M_{-l} \forall j \in R_l \quad (5)$$

$$\sum_{j \in R_l} x_{-j}^m = p_{-l}^m \quad \forall l, \forall m \in M_{-l} \quad (6)$$

$$0 \leq \sum_{m \in M_{k-}} x_{i-}^m \leq U_k \quad \forall k, \forall i \in R_k \quad (7)$$

$$0 \leq \sum_{m \in M_{-l}} x_{-j}^m \leq U_l \quad \forall l, \forall j \in R_l \quad (8)$$

$$0 \leq \sum_m x_{ij}^m \leq U_{ij} \quad \forall i, \forall j \quad (9)$$

$$x_{ij}^m, x_{i-}^m, x_{-j}^m \geq 0 \quad \forall i, \forall j, \forall m \quad (10)$$

W funkcji celu jest minimalizowana suma kosztów związanych z przejściami studentów pomiędzy salami. Ograniczenia (3) zapewniają spełnienie wymagań na przydział odpowiedniej liczby modułów do elementarnych styków zajęć w terminie t . Ograniczenia (4) zapewniają, że odpływ studentów z sali przydzielonej do elementarnego styku zajęć w terminie t będzie równy sumie przepływów studentów z tej sali do wszystkich innych sal, do których jest to możliwe. Ograniczenia (5) zapewniają, że dopływ studentów do sali przydzielonej do elementarnego styku zajęć w terminie $t+1$ będzie równy sumie przepływów studentów do tej sali ze wszystkich innych sal, z których jest to możliwe. Ograniczenia (6) zapewniają spełnienie wymagań na przydział odpowiedniej liczby modułów do elementarnych styków zajęć w terminie $t+1$.

Ograniczenia (7) zapewniają, że odpływ studentów z każdej sali po terminie t , we wszystkich elementarnych stykach zajęć wymagających tej sali, nie przekroczy jej pojemności. Ograniczenia (8) zapewniają, że dopływ studentów do każdej sali w terminie $t+1$, we wszystkich elementarnych stykach zajęć wymagających tej sali, nie przekroczy jej pojemności. Ograniczenia (9) zapewniają, że przepływ z sali i do sali j we wszystkich elementarnych stykach zajęć nie przekroczy pojemności mniejszej z tych dwóch sal.

5 Metoda rozwiązywania problemu

W ogólnym przypadku rozwiązanie modelu *PSSE* może nie być rozwiązaniem dopuszczalnym problemu przydziału sal do elementarnych styków zajęć, ponieważ nie będą spełnione warunki strukturalne wypełnienia sal. Dokładną analizę tego zagadnienia zawiera praca [6].

W pewnych przypadkach, dla niektórych styków elementarnych, niedopuszczalność rozwiązania problemu pierwotnego można usunąć dokonując przekształcenia rozwiązania modelu *PSSE*. Przy spełnieniu pewnych warunków takie przekształcenie nie powoduje utraty optymalności rozwiązania. To zagadnienie zostanie omówione poniżej.

Na wstępie zostaną wprowadzone pojęcia pomocnicze. Mówimy, że wymagania zasobowe dwóch styków elementarnych *częściowo pokrywają się* dla pary zasobów zagregowanych (k, l) , jeśli w terminie t oba styki wymagają zasobu zagregowanego kategorii k , a w terminie $t+1$ zasobu zagregowanego kategorii l .

Lemat 2. Dowolna kombinacja modułów przydzielonych po rozwiązaniu modelu *PSSE* do dwóch styków elementarnych w odpowiedzi na ich wymagania zasobowe częściowo pokrywające się, nie zmienia wartości funkcji celu modelu *PSSE*.

Dowód. Zamiana modułów pomiędzy stykami elementarnymi, których dotyczy powyższy lemat nie zmienia wielkości przepływów po łukach łączących sale indywidualne przydzielone do zajęć w terminach t i $t+1$ w rozwiązaniu zadania *PSSE*, ponieważ moduły można zestawiać w kombinacje tylko spośród tych z nich, które zostały przydzielone do rozważanych styków elementarnych w odpowiedzi na ich wymagania zasobowe częściowo pokrywające się. Tak więc wartość funkcji celu również nie ulega zmianie, gdyż koszty przepływów c_{ij} po łukach łączących sale indywidualne przydzielone do zajęć w terminach t i $t+1$ nie zależą od tego, do których styków elementarnych te sale zostały przydzielone.

Z powyższego lematu wynika, że dokonując cyklicznych zamian pomiędzy modułami indywidualnych sal przydzielonych równocześnie do różnych styków elementarnych można doprowadzić do sytuacji, w której każda z tych sal będzie przydzielona do dokładnie jednego styku elementarnego. Sale pomiędzy którymi są dokonywane cykliczne zamiany przepływów, zarówno przed, jak i po dokonaniu zamian, są w całości wykorzystywane przez dokładnie jeden styk elementarny. Szerszy opis tego zagadnienia można znaleźć w pracy [6].

Wystąpienie opisanej powyżej sytuacji zależy od rodzaju rozwiązania modelu *PSSE*. Cykliczne zamiany przepływów są możliwe jedynie pomiędzy łukami łączącymi indywidualne sale przydzielone do różnych styków elementarnych w odpowiedzi na ich wymagania zasobowe częściowo pokrywające się. Dla danej sali można doprowadzić

do spełnienia warunków strukturalnych jej wypełnienia, jeśli wystarczająca liczba modułów tej sali została przydzielona do styków elementarnych pomiędzy którymi są możliwe cykliczne zamiany przepływów, w odpowiedzi na ich wymagania zasobowe częściowo pokrywające się. Przez wystarczającą liczbę modułów należy rozumieć taką ich liczbę, która umożliwia spełnienie wymagań zasobowych zajęcia, do którego sala została przydzielona.

W ogólnym przypadku rozwiązanie zadania *PSSE* może być niedopuszczalne nawet po dokonaniu opisanych powyżej przekształceń. Zawsze dostarcza ono jednak informacji o dolnym oszacowaniu wartości funkcji celu całego problemu. W celu znalezienia rozwiązania dopuszczalnego, w wyniku strukturalnej analizy problemu przydziału sal do styków elementarnych, został opracowany algorytm przybliżony bazujący na rozwiązaniu zadania *PSSE*. Algorytm ten jest obecnie przedmiotem badań. Jego szczegółowy opis znajduje się w pracy [7].

Przeprowadzone dotychczas eksperymenty obliczeniowe dotyczyły oceny możliwości rozwiązywania modelu *PSSE*. Przyjęto założenie, że w 5 budynkach znajduje się 50 sal o różnych, wyrażonych w modułach, pojemnościach. Pojemności sal były zmieniane w taki sposób, aby łączna liczba modułów w kolejnych eksperymentach zmieniała się od 100 do 200. W każdym zadaniu testowym harmonogram zajęć oraz koszty przejść studentów pomiędzy salami położonymi w różnych budynkach były generowane losowo. Obliczenia wykonano przy pomocy komercyjnego pakietu CPLEX. Łącznie przeprowadzono 30 eksperymentów. W każdym przypadku udało się uzyskać rozwiązanie optymalne. Czas obliczeń zmieniał się od kilkunastu sekund w przypadku 100 modułów do ok. 3 min. w przypadku 200 modułów.

6 Podsumowanie

W referacie podjęto próbę rozwiązania problemu przydziału sal do zajęć, z uwzględnieniem przejść studentów pomiędzy salami. Jest to jeden z trudniejszych problemów pojawiających się w kontekście harmonogramowania zajęć. W wyniku przeprowadzonych badań zaproponowano model matematyczny problemu dla przypadku przejść studentów pomiędzy zajęciami realizowanymi w dokładnie dwóch kolejnych terminach.

Opracowano efektywny model optymalizacyjny będący relaksacją problemu pełnego. W pewnych przypadkach można usunąć przyczyny niedopuszczalności rozwiązań modelu zrelaksowanego bez utraty optymalności wyników. Jednak w ogólnym przypadku nie zawsze jest to możliwe. W związku z tym, w celu znajdowania rozwiązań dopuszczalnych został opracowany algorytm przybliżony, przyjmujący za punkt startu rozwiązanie zrelaksowane. Obecnie algorytm ten znajduje się w fazie badań. Prowadzone są również prace dotyczące uogólnienia zaproponowanego podejścia na przypadek przejść pomiędzy zajęciami realizowanymi w więcej niż dwóch kolejnych terminach.

Sformułowanie problemu oraz wyniki uzyskane w referacie dotyczą zagadnień o charakterze nowatorskim i umożliwiają uwzględnienie ważnych aspektów problemu przydziału sal do zajęć związanych z lokalizacją sal w różnych odległościach od siebie, np. w wielu budynkach. Rezultaty pracy mogą być wykorzystywane zarówno w ramach hierarchicznego schematu rozwiązywania złożonych problemów planowania zajęć, jak też dla innych praktycznych problemów decyzyjnych dotyczących układania rozkładów zajęć.

Literatura

1. Ahuja R.K., Magnanti T.L., Orlin J.B.: *Network Flows. Theory, Algorithms and Applications*. Prentice Hall, New Jersey, 1993 r.
2. Burke E., Varley D.: *Space Allocation: An Analysis of Higher Education Requirements*. Lecture Notes in Computer Science 1408, 1998 r., s. 20-33.
3. Gosselin K., Truchon M.: *Allocation of Classrooms by Linear Programming*. Journal of the Operational Research Society 37, 1986 r., s. 561-569.
4. McDiarmid C.J.H.: *The Solution of a Timetabling Problem*. Journal Institute Mathematics Applications 9, 1972 r., s. 23-34.
5. Mulvey J.A.: *A classroom/time assignment model*. European Journal of Operational Research 9, 1982 r., North Holland, s. 64-70.
6. Szwed C.: *Metody dezagregacji zasobów w harmonogramowaniu zajęć elastycznego studiowania*. Rozprawa doktorska, Wydział Elektroniki i Technik Informacyjnych Politechniki Warszawskiej, 1999 r.
7. Szwed C., Toczyłowski E.: *Analiza problemu przydziału sal do zajęć z uwzględnieniem przejść studentów pomiędzy salami*. Raport Instytutu Automatyki i Informatyki Stosowanej Politechniki Warszawskiej nr 99-23, 1999 r.
8. Szwed C., Toczyłowski E.: *Dezagregacja zasobów lokalowych w harmonogramowaniu zajęć*. T. Trzaskalik (red.) „Metody i zastosowania badań operacyjnych”, Wyd. AE w Katowicach, 1998 r., s. 211-220.
9. Toczyłowski E.: *Analiza strukturalna problemów harmonogramowania zajęć elastycznego studiowania*. Raport Instytutu Automatyki i Informatyki Stosowanej Politechniki Warszawskiej nr 94-18, 1994 r.

Streszczenie

W referacie opisano problem alokacji sal do zajęć w kontekście, opisanego w literaturze, strukturalnego hierarchicznego schematu układania zajęć elastycznego studiowania. Analizy problemu dokonano dla przypadku, w którym uwzględnia się przejścia studentów pomiędzy salami (czasy lub koszty przejść). W pracy przedstawiono model matematyczny problemu, metodę jego rozwiązywania oraz wyniki eksperymentów obliczeniowych.

OPTIMIZATION OF SPACE ALLOCATION IN THE COURSE TIMETABLING PROBLEM

Summary

In this contribution, the problem of allocation rooms to lectures, in the context of solving difficult timetabling problems via structural hierarchical approach, is described. We consider an optimization problem, in which the costs of student transfer between rooms are taken into account. Mathematical formulation and the solution method proposed for this problem are given. The results of the experiments are also described.

Referat recenzowany

POMIAR ZMIENNEJ WYJŚCIOWEJ W UKŁADACH REGULACJI Z OBIEKTAMI CIĄGLYMI

1. Postawienie problemu

W automatyce **przemysłowej**¹ obiekty sterowania są najczęściej układami dynamicznymi ciągłymi w przestrzeni (często nazywane są układami o stałych rozłożonych), to znaczy takimi, że chwilowe wartości zmiennych stanu zależą od miejsca pomiaru. Przykładem może być regulacja temperatury powietrza w pomieszczeniu (natomiast wyjątkiem może być regulacja poziomu cieczy w zbiorniku). Obiekt regulacji jest więc układem dynamicznym o nieskończenie wielu stopniach swobody.

Od układu regulacji automatycznej (URA) żąda się, aby zadanie regulacji było realizowane w **każdym** punkcie przestrzeni². Wymagałoby to nieskończenie wielkiej liczby czujników pomiarowych i nieskończenie wielkiej liczby urządzeń wykonawczych. W rzeczywistych układach przestrzeni musi być zdyskretyzowana. Jeśli – ze względu na koszty aparatury – ograniczymy się do układu SISO³, powstają następujące problemy:

1. Jaka wielkość ma być wielkością regulowaną: czy największa chwilowa wartość wielkości regulowanej w całej przestrzeni, czy najmniejsza, czy średnia po przestrzeni, średnia po czasie i po przestrzeni itp.; jest to więc problem **definicji** wielkości regulowanej;
2. Jak modelować mierzoną wielkość, tzn. jaka jest zależność $y_m(t) = f(y(t), t)$, gdzie $y_m(t)$ jest sygnałem wejściowym czujnika pomiarowego a $y(t)$ jest wielkością regulowaną;
3. W którym miejscu przestrzeni należy ulokować czujnik, np. termometr;
4. W którym miejscu przestrzeni należy ulokować urządzenie wykonawcze, np. grzejnik.

Ta czwarta kwestia nie będzie tu dyskutowana, choć z punktu widzenia projektowania URA jest również ważna.

Problem pomiaru w URA jest dlatego tak ważny, ponieważ sprzężenie zwrotne nie powoduje zmniejszenia błędów oraz opóźnień wprowadzonych przez tor pomiarowy, który zatem musi być starannie zaprojektowany i zrealizowany.

Celem artykułu jest pokazanie wagi definiowania i realizacji toru pomiarowego w przemysłowych URA z obiektami ciągłymi w przestrzeni.

2. Zadanie układu regulacji

Z punktu widzenia teorii sterowania można wyróżnić *zadanie przestawiania* i *zadanie śledzenia*. W obszarze przemysłowych URA, w pierwszym przypadku oznacza to, że ma być dobrze realizowany program sterowania wejście – wyjście, zaś w drugim, że

* Politechnika Koszalińska, Wydział Mechaniczny, Katedra Systemów Sterowania

¹ w odróżnieniu na przykład od serwomechanizmów

² technolog procesu zazwyczaj żąda, aby w każdej chwili 1) wartość średnia była blisko wartości zadanej, oraz 2) aby rozstęp skrajnych wartości wielkości regulowanej w obiekcie był ograniczony

³ Single Input Single Output

mają być skutecznie tłumione zakłócenia. Zatem w przemysłowych układach sterowania obydwą zadania muszą być dobrze realizowane. Na ogół istnieje między nimi sprzeczność i konieczny jest kompromis projektowy.

3. Możliwe rozwiązania układu

W przypadku układu ciągłego możliwe są następujące rozwiązania układowe.

1. Dyskretyzacja przestrzeni na m podobszarów, w każdym osobny czujnik i osobne urządzenie wykonawcze, czyli układ MIMO ⁴;
2. Kilka czujników i jedno urządzenie wykonawcze, na przykład w układzie z kompensacją lub w układzie kaskadowym: układ SIMO;
3. Jeden czujnik pomiarowy i jedno urządzenie wykonawcze: układ SISO.

Układ pierwszy jest najskuteczniejszy ale bardzo drogi; układ drugi – gdy preferowane jest zadanie śledzenia (za zakłóceniami, czyli skuteczne ich tłumienie). Natomiast układ trzeci jest najprostszy i najtańszy, i jest skuteczny dla realizacji zadania przestawiania.

Lokalizacja czujników zależy od ich przeznaczenia. W przypadku śledzenia za zmiennymi zakłóceniami, czujnik powinien być „jak najbliżej” ⁵ głównego zakłócenia.

Przykład

Ograniczymy się tu do przestrzeni 2D, ze względu na mniejszą złożoność obliczeniową ⁶.

Jest dana płyta o jednostkowej grubości (Rys. 1), z prostokątnym otworem, ogrzewana niemal punktowo przez łukową krawędź nr 9 z dolnego prawego rogu czynnikiem grzewczym o temperaturze $T_{st}(t)$. Pozostałe krawędzie są izolowane termicznie, z wyjątkiem lewej krawędzi nr 8, przez którą ciepło może być wymieniane z otoczeniem o temperaturze $T_z(t)$, i ta temperatura jest zakłóceniem. Zakładamy, że nie ma wymiany ciepła w kierunku prostopadłym do płaszczyzny rysunku, problem jest więc dwu-wymiarowy. Wymiary płyty i własności termiczne materiału są dane.

Pierwotnym zadaniem regulacji jest utrzymanie całej płyty w zadanej temperaturze $T_{zad}(t)$, w obecności zakłóceń.

Jako wielkość sterującą przyjęto temperaturę $T_{st}(t)$.

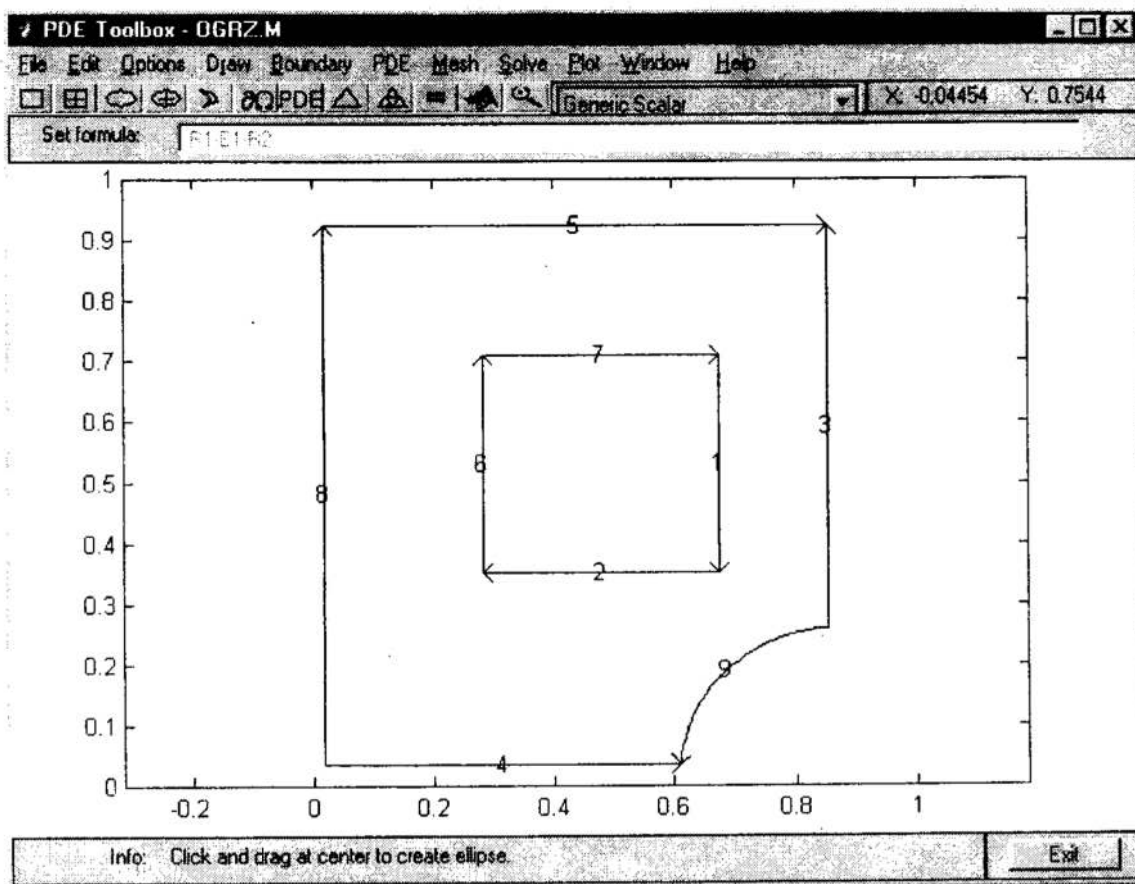
Zadanie 1. Zdefiniowanie wielkości regulowanej

Zadanie regulacji utrzymania całej płyty w zadanej temperaturze jest nierealizowalne, i musi być przeformułowane: jako wielkość regulowaną można przyjąć na przykład wartość średnią dla wybranych kilku punktów, albo różnicę między chwilową wartością maksymalną i minimalną, wariancję odchylenia od wartości zadanej itp. Tu przyjmiemy temperaturę pewnego jednego wybranego punktu płyty, którą uznamy jako reprezentatywną dla całej płyty. Założmy, że to będzie taki punkt w którym temperatura jest w każdej chwili czasu równa temperaturze średniej zdyskretyzowanej płyty. Tam zostanie umieszczony czujnik temperatury. Zadaniem regulacji będzie utrzymanie najmniejszej wartości odchyłki regulacji.

⁴ Multi Input Multi Output

⁵ w sensie szybkości reakcji

⁶ i możliwość wykorzystania narzędzia Matlab/PDE Toolbox; dla układów 3D konieczny byłby inny pakiet, np. FLUENT



Rys. 1. Obiekt w przykładzie: dwuwymiarowa płyta (z ponumerowanymi krawędziami) – edytor Matlab/PDE Tool

Fig. 1. Object in the example: two-dimensional plate (with numbered edges) – Matlab/PDE Tool Editor

Zadanie 2. Znalezienie punktu pomiarowego

Można by to zrobić analitycznie, gdyby istniał model matematyczny procesu wymiany ciepła w rozpatrywanej płycie. To jest możliwe, jeśli są to równania różniczkowe (tutaj cząstkowe), liniowe. Określenie punktu „średniego” można także uzyskać w drodze żmudnych pomiarów i identyfikacji, tym liczniejszych im bardziej układ jest nieliniowy.

Formalnie jest to zadanie optymalizacji: na dyskretnym zbiorze punktów należy znaleźć taki punkt C , którego temperatura $T_c(t)$ jest najmniej różna od temperatury średniej $T_{sr}(t)$ dla wszystkich punktów, dla określonego przedziału czasu $[0, t]$.

$$C = i : \left\{ \int_0^t (T_i - T_{sr}) dt \rightarrow \min_i \right\} \quad i = 1, \dots, m - \text{numery punktów}$$

Tu wykorzystamy symulację komputerową, oraz oszacowanie.

Do symulacji wykorzystamy Matlab/PDE Toolbox [3]. Korzystając z możliwości tego pakietu:

- 1) rysujemy obiekt (zapisany w plik *ogrz.m*), jak na rysunku 1;
- 2) definiujemy zadanie: Options: Heat Transfer;

- 3) określamy ograniczenia na krawędziach⁷ (opcja Boundary);
- 4) deklarujemy własności materiałowe w PDE Specification: Parabolic (analiza stanów nieustalonych)⁸ z parametrami: $c = 1$; $a = 0$; $f = 10$; $d = 18$;
- 5) dyskretyzacja przestrzeni: Mesh, Initialize Mesh;
- 6) wprowadzamy parametry symulacji: Solve Parameters: czas symulacji 1s, wartości początkowe temperatury $u(t_0) = 40^\circ\text{C}$; dopuszczalne błędy obliczeń symulacyjnych $0,001^\circ\text{C}$;
- 7) Uruchomienie symulacji: Solve PDE.

Otrzymany rozkład temperatury pokazuje rysunek 2. Dla rysunku zadeklarowano 16 poziomów wartości temperatury (opcja *Plot – Parameters – Contour plot levels*), zatem izotermy na rysunku są co 5°C , od 20°C do 100°C .

- 8) Wyniki symulacji zapisujemy do Workspace (Solve – Export Solution) w postaci macierzy o nazwie *u*: wiersze odpowiadają numerom węzłów siatki, kolumny odpowiadają kolejnym chwilom czasu.

Dla przykładu, polecenie `size(u)` ujawnia, że jest 151 węzłów i 101 chwil.

- 9) Polecenie `sum(u(:, 101))/151` wylicza wartość średnią temperatury w całej płycie dla chwili 101; wynosi ona 46.3794°C . Z rysunku widać, że jest to izoterma piąta licząc od najniższej, odpowiadającej 20°C (w tym przypadku piąta izoterma o wartości temperatury 45°C przebiega przez prawy górny róg płyty). Zatem na tej izotermie powinien być zlokalizowany czujnik temperatury, wytwarzający sygnał wyjściowy układu⁹.
- 10) Określenie własności dynamicznych: dla układu liniowego to może być transmitancja (od sterowania do wyjścia: tutaj: od *Tst* do *Tsr*. Obieramy jakiś punkt (węzeł siatki) na izotermie „średniej”, np. węzeł nr 123 (patrz rysunek 3)

Polecenie Matlaba `plot(u(123, :))` rysuje odpowiedź czasową w wybranym punkcie płyty (Rys. 4).

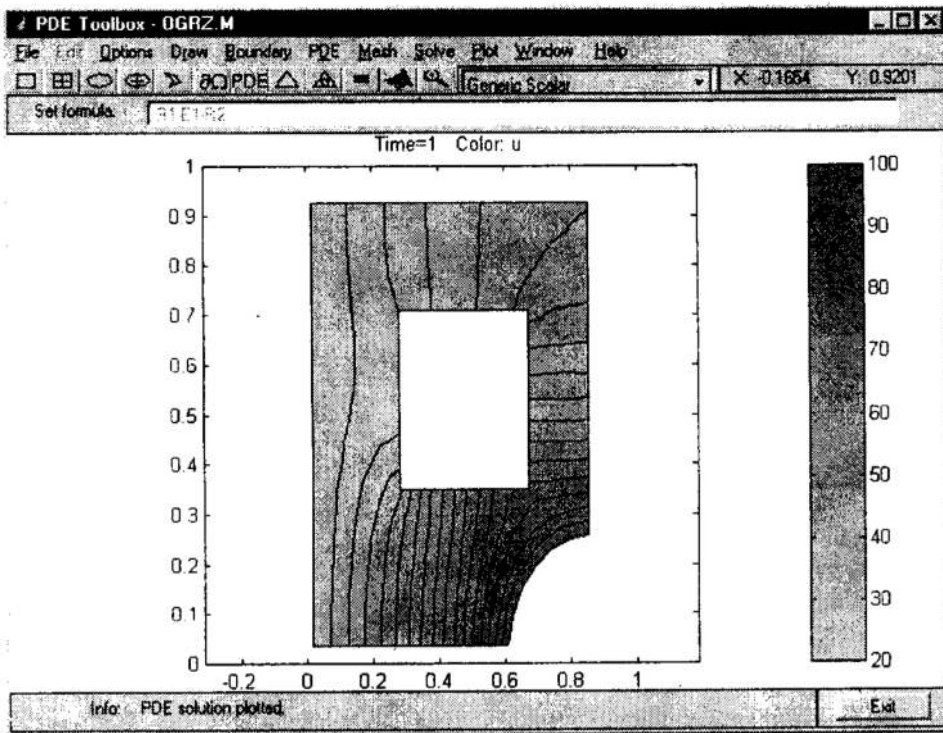
⁷ wymuszenia temperaturowe definiuje się jako ograniczenia Dirichleta, z parametrami $h = 1$ oraz $r = \text{Temp}$; zaś izolację jako ograniczenie Neumanna z parametrami $g = 0$, $q = 0$

⁸ cząstkowe równanie różniczkowe paraboliczne przepływu ciepła w stanach nieustalonych jest:

$$d \cdot \frac{dT}{dt} = \text{div}(c \cdot \text{grad}(T)) - a \cdot T + f$$

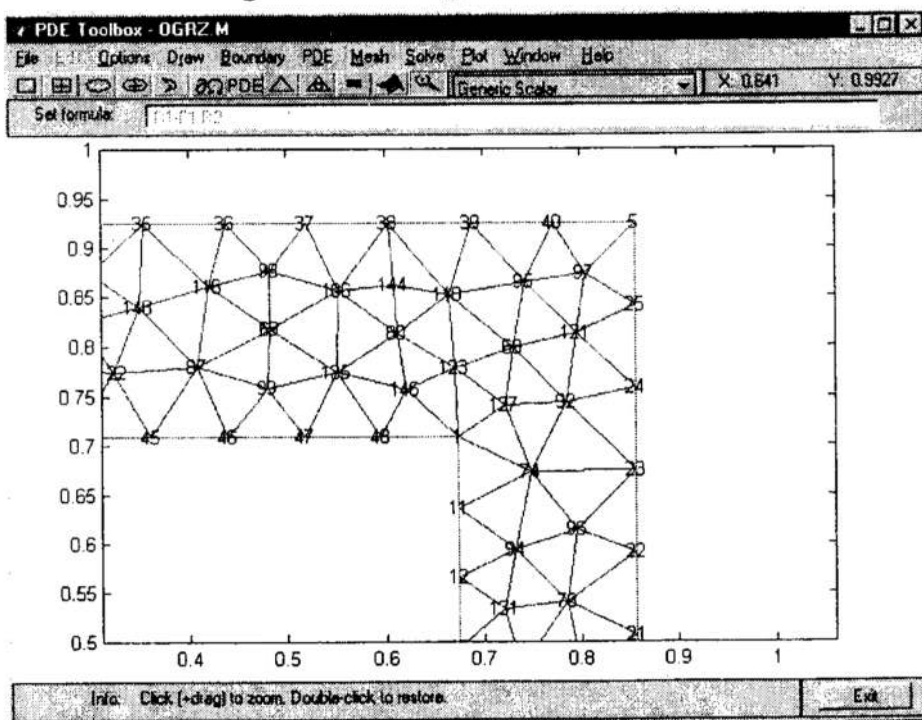
gdzie: d jest iloczynem ciepła właściwego $\text{m}^2/\text{s}^2\text{deg}$ i gęstości kg/m^3 ; c jest współczynnikiem przewodzenia ciepła $\text{kg}\cdot\text{m}/\text{s}^3\text{deg}$; a jest współczynnikiem konwekcyjnego wnikania ciepła na 1 metr długości $\text{kg}/\text{s}^3\text{mdeg}$; oraz f jest strumieniem ciepła wydzielanego wewnątrz ciała na 1 m^3 jego objętości $\text{kg}/\text{s}^3\text{m}$

⁹ jeśli układ dynamiczny jest liniowy (jak w tym przykładzie), położenie izotermy odpowiadającej średniej temperaturze nie zmienia się w czasie gdy układ jest w stanach nieustalonych, zatem czujnik umieszczony na tej izotermie będzie zawsze pokazywał temperaturę średnią całej płyty



Rys. 2. Rozkład temperatury po czasie 1s, dla temp. początkowej 40 °C , dla wymuszeń $T_{st}=100\text{ K}$ $T_z=20\text{ K}$

Fig. 2. Temperature distribution after 1 sec, for starting temperature 40 deg, for $T_{st}=100\text{ deg}$ and $T_z=20\text{ deg}$



Rys. 3. Powiększony fragment płyty pokazujący siatkę elementów skończonych

Fig. 3. Magnified part of the plate, with the mesh of finite elements

Widać, że nie został osiągnięty stan ustalony. Trzeba powtórzyć symulację dla znacznie dłuższego czasu; okazuje się że wystarczający jest czas 15s (Rys. 5).

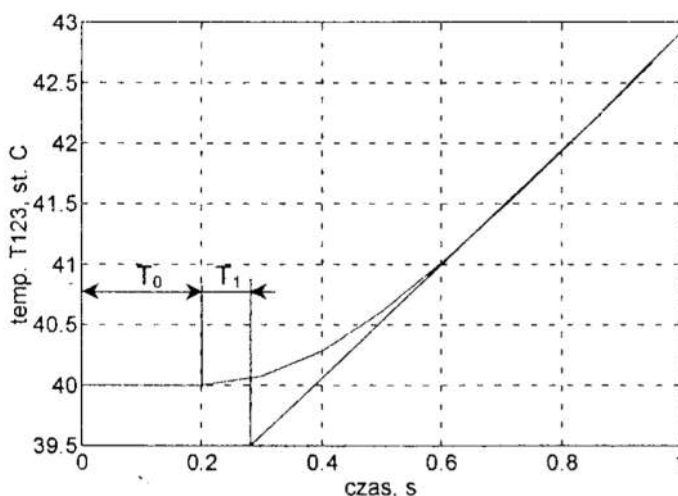
Można teraz, na podstawie przebiegu czasowego możemy zaproponować model matematyczny wielkości mierzonej, czyli zależność:

$$T_{123}(t) = f(T_{st}(t), t)$$

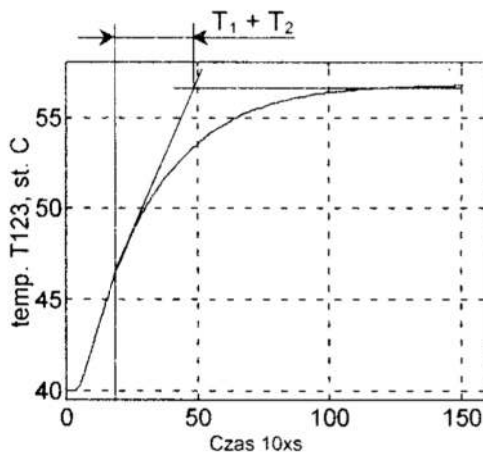
w postaci pewnej aproksymacji. Z rysunków widać, że w tym przypadku niezłym przybliżeniem może być liniowy człon drugiego rzędu z opóźnieniem, o transmitancji:

$$\frac{T_{123}(s)}{T_{st}(s)} = \frac{k}{(T_1s + 1) \cdot (T_2s + 1)} \cdot e^{-T_0s}$$

Na podstawie rysunku 4 i 5 można określić przybliżone wartości parametrów: $T_0 = 0,2s$, $T_1 = 0,1s$ i $T_2 = 6,0 - 2,5 - 0,3 = 3,2s$ oraz $k = 57/40 = 1,4250$, gdzie $40^\circ C$ jest temperaturą wymuszenia, a 57 wartością ustaloną dla punktu (węzła siatki) $i = 123$.



Rys. 4. Odpowiedź czasowa dla węzła nr 123 (początkowy fragment)
Fig. 4. Step response for the nod nr 123 (a starting phase)

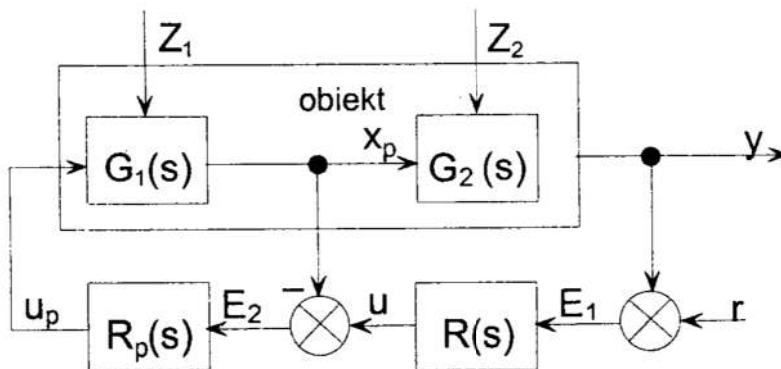


Rys. 5. Odpowiedź czasowa dla węzła nr 123
Fig. 5. Step response for the nod nr 123

4. Tłumienie zakłóceń przez ich pomiar

W jednopętlowym URA zakłócenia przechodzą przez obiekt i ich wpływ manifestuje się przez odchyłki wielkości regulowanej, czyli dopiero na wyjściu z obiektu. Są tłumione dzięki sprzężeniu zwrotnemu od wielkości wyjściowej. Jeśli to jest niewystarczające, są dwie możliwości zwiększenia skuteczności tłumienia: 1) przez pomiar zakłócenia i kompensację (struktura *feed-forward*), lub 2) jeśli nie można zmierzyć zakłócenia lub jest to bardzo trudne i kosztowne (np. jest to wartość opałowa gazu, wilgotność ziarna czy stężenie mieszanki gazowej itp.) możemy zmierzyć wpływ tego zakłócenia na proces w obiekcie nie na wyjściu z obiektu lecz po przejściu sygnału zakłócenia przez część obiektu, i wykorzystanie tego sygnału pomiarowego $v(t)$ w układzie kaskadowym [1] jak na rysunku 6. Ten dodatkowy czujnik pomiarowy powinien być umieszczony jak najbliżej sygnału zakłócenia, w sensie opóźnienia ale i błędu statycznego.

W tym przypadku – dla poprawnej syntezy regulatora kaskadowego – potrzebna jest znajomość charakterystyki przepływu sygnału w tej części obiektu: $x_p(t) = f(u(t), z(t))$, gdzie $x_p(t)$, $u(t)$ i $z(t)$ są odpowiednio: pośrednim sygnałem pomiarowym, sygnałem sterującym i zakłóceniem. Z reguły będziemy tę zależność linearyzować i wówczas zadaniem będzie określenie transmitancji czyli stosunku transformat sygnałów: $V(s)/U(s)$ oraz $X_p(s)/Z(s)$, gdzie s jest operatorem *Laplace'a*. Jest to trudne dla obiektów ciągłych, ale może być wykonane za pomocą symulacji komputerowej, analogicznie jak wyżej pokazano. W naszym przykładzie byłaby to transmitancja $X_p(s)/Tz(s)$, gdzie $Tz(s)$ jest temperaturą otoczenia na lewej krawędzi, które jest tu zakłóceniem.



Rys. 6. Kaskadowy układ regulacji: G_1 – pierwsza część obiektu, G_2 – część obiektu w której występuje znaczne opóźnienie, x_p – pomocnicza wielkość pomiarowa, z_1, z_2 – zakłócenia, u – wielkość sterująca, R_p – regulator pomocniczy

Fig. 6. Cascade control system: G_1 – the first segment of the object, G_2 – part of an object with a substantial delay, x_p – aid measured variable, z_1, z_2 – perturbations, u – control variable, R_p – aid regulator

5. Uwagi końcowe

Jeśli zakłócenia są niemierzalne, to warto zbadać jak można wykorzystać ideę regulatora zmiennych stanu od wybranych zmiennych mierzalnych w obiekcie ciągłym: jak dobrać te wielkości, jakie nastawiać wzmocnienia regulatora. To prowadzi do *regulatora neuronowego*, w którym wagi odpowiadają tym wzmocnieniom. Wagi można

dobierać analitycznie – jeśli znany będzie algorytm – lub metodą uczenia sieci, jeśli będą znane wzorce zachowań regulatora. Te wzorce można uzyskać przez symulację komputerową.

Proponowana symulacja procesów przejściowych wymaga skutecznych modeli matematycznych oraz efektywnych narzędzi komputerowych, bez których modele nie mają praktycznego znaczenia. Są one odmienne dla ciał stałych, dla płynów i dla ciał sypkich.

Literatura

1. Kurman K. J.: *Teoria regulacji: podstawy, analiza, projektowanie*. WNT Warszawa 1975.
2. Staniszewski B.: *Termodynamika*. PWN, Warszawa 1986.
3. Tarnowski W., Bartkiewicz S.: *Modelowanie matematyczne i symulacja komputerowa dynamicznych procesów ciągłych*. Wydawn. Politechniki Koszalińskiej, Koszalin 1998.

Streszczenie

Sformułowano problem definiowania reprezentatywnej wielkości mierzonej w przemysłowych układach automatyki gdzie obiektem jest ciągły w przestrzeni układ dynamiczny (tzw. układ o stałych rozłożonych). W szczególności przedyskutowano sposób określania lokalizacji czujnika pomiarowego. Rozważono wpływ zakłóceń. Dla zakłóceń niemierzalnych wskazano na możliwość wykorzystania regulatora neuronowego.

OUTCOME VARIABLE MEASUREMENT PROBLEM IN INDUSTRIAL CONTROL SYSTEMS WITH THE SPACE CONTINUOUS OBJECTS

Summary

A problem of an adequate definition of a representative measured variable is discussed, in the case of continuous dynamic systems (with distributed parameters). First, a way of how to specify the localisation of a sensor is proposed. Secondly, the influence of perturbations is discussed, and a neural controller is proposed for instances where non-measurable perturbations are present.

Referat recenzowany

ERGONOMICZNE WSPOMAGANIE PROCESÓW PROJEKTOWANIA

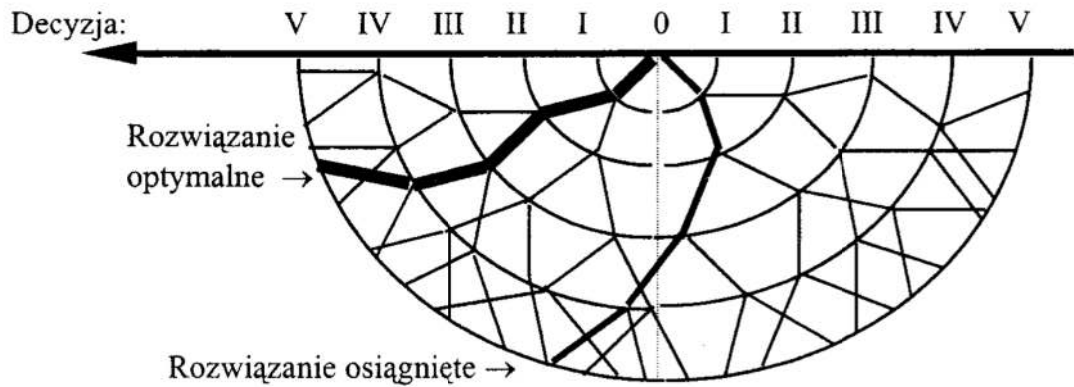
1. Decyzje podejmowane w procesie projektowania

Podejmowanie decyzji jest jednym z najważniejszych zagadnień procesu projektowania. Znane są obecnie różnorodne podejścia mające na celu wyjaśnienie mechanizmów sterujących procesami podejmowania decyzji. W tym celu wykorzystuje się teorie z obszarów: psychologii (teoria decyzji racjonalnych oraz teoria uwzględniająca psychiczne cechy decydenta), cybernetyki, prakseologii i ekonomiki. Wiele problemów dotyczących procesów podejmowania decyzji nie zostało jeszcze w pełni rozpoznanych i nie znalazło naukowego wyjaśnienia, np. niejednoznaczność przyporządkowywania określonych decyzji do określonych danych wejściowych, kształtowanie się preferencji wyborów i pojęcia użyteczności, sprzeczności logiczne i brak konsekwencji przy dokonywaniu wyborów.

Wewnętrzna struktura procesów podejmowania decyzji jest przedmiotem intensywnych badań. H.A.Simon twierdzi, że „...procesy myślowe człowieka łączące się z rozwiązywaniem problemów są sterowane przez program który powoduje, iż niezliczone proste procesy informacyjne - lub jak kto woli - procesy realizujące działania na symbolach, zostają uporządkowane w pewien złożony ciąg. Struktura tego ciągu ulega zmianom, gdyż adaptuje się ona zarówno do otoczenia, w którym dany proces zachodzi, jak i do wskazań wynikających z analizy tego otoczenia prowadzonej w trakcie rozwiązywania zadania” [6, s.106]. Technologia obróbki informacji w umyśle człowieka, zgodnie z tą teorią, nie różni się zasadniczo od działania komputera. Konsekwencją tego spostrzeżenia jest przewidywanie możliwości stworzenia takich programów komputerowych i takich maszyn, które mogłyby wspomagać lub nawet symulować procesy myślowe człowieka.

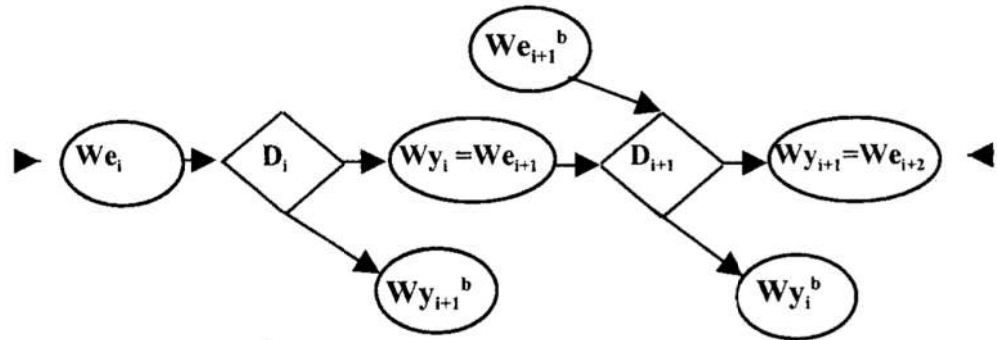
W procesie projektowania koncepcje nowych rozwiązań pojawiają się w wyniku realizacji określonych ciągów decyzji projektowych. Decyzje te dotyczą problemów o różnym stopniu ogólności i złożoności oraz o różnym znaczeniu dla ostatecznego wyniku projektowania. Podjęcie decyzji projektowej oznacza dokonanie **wyboru** określonej koncepcji rozwiązania problemu, na podstawie intuicyjnego lub logicznie czy wręcz matematycznie uzasadnionego przekonania o wyższości tej koncepcji nad innymi. Każda poprzednia decyzja wpływa na wynik następnej i w ten sposób decyzje ogólne (o znaczeniu strategicznym dla rozwiązywanego problemu), podejmowane na początku procesu twórczego, mają zasadniczy wpływ na ostateczne rozwiązanie. Mechanizm ten ilustruje rysunek 1, gdzie dla przejrzystości przedstawiono sytuacje decyzyjne wyłącznie jako dwuwariantowe (alternatywne).

* Politechnika Poznańska, Instytut Inżynierii Zarządzania,
60-965 Poznań, ul. Strzelecka 11, E-mail: Edwin.Tytyk@put.poznan.pl



Rys. 1. Ilustracja mechanizmu konsekwencji w podejmowaniu decyzji alternatywnych
 Fig. 1. Illustration of mechanism of consequence in alternative decision making

Każda decyzja jest podejmowana na podstawie pewnego zbioru informacji wejściowych i może być traktowana jako określony sposób ich transformacji. W ten sposób wejściowe informacje We_i podlegają przetworzeniu w akcie decyzyjnym D_i i uzyskują postać wynikową Wy_i , która z kolei (w całości lub części) może być informacją wejściową We_{i+1} w następnym akcie decyzyjnym (rys. 2).



Rys. 2. Schemat procesu decyzyjnego:

D - decyzja; We - informacje wejściowe; Wy - informacje wyjściowe (wynikowe)

Fig. 2. The scheme of decision making process:

D - decision; We - input informations; Wy - output informations

Decyzje w procesie projektowania podejmowane są najczęściej przy niepełnej i niepewnej informacji. Z tego powodu projektant musi brakującą część informacji uzupełnić własnym doświadczeniem i intuicją. W literaturze [4, ss.103-116] można znaleźć oszacowanie, że co najmniej 75% ilości potrzebnych informacji projektant musi wytworzyć bądź wyszukać samodzielnie, ekstrapolując i przetwarzając dostępne dane oraz przyjmując odpowiednie założenia. W związku z tym procesy decyzyjne w projektowaniu obarczone są zawsze pewnym ryzykiem. Rozterki człowieka podejmującego decyzje opisywał już przed wiekami Arystoteles (384–322 p.n.e.): „Niejednokrotnie trudno jest rozstrzygnąć, co kosztem czego wybrać należy i co w imię czego wytrzymać, a jeszcze trudniej jest wytrwać w przekonaniu, do którego się doszło” [2, s.75].

2. Ergonomiczne kryteria projektowe

Decyzje projektowe podejmowane są na podstawie określonego zbioru informacji wejściowych. Ten zasób informacji można nazwać zbiorem kryteriów decyzyjnych, lub dokładniej – kryteriów projektowych. Można przyjąć za T. Pszczółowskim, że kryterium jest to: „znak wyróżniający coś od czegoś innego, miernik, sprawdzian, wzgląd brany pod uwagę przy wyborze, hierarchizacji, ocenianiu, szacowaniu i tym podobnych działaniach opierających się na pewnych założeniach narzuconych sprawcy lub przyjętych dowolnie przez sprawcę” [5]. Kryteria są zatem merytoryczną i metodologiczną podstawą podejmowanych decyzji. Dostarczają one projektantowi niezbędnych informacji i kierują jego myśli w stronę rozwiązań o pożądanym cechach. Ergonomiczne kryteria projektowe będą się zatem przyczyniać do powstawania projektów o wysokiej ergonomicznej jakości.

Przytoczona wyżej definicja wskazuje, że kryteria są podstawą podejmowania określonych decyzji związanych z ocenianiem, klasyfikacją, wyborem i podobnymi działaniami dokonywanymi na pewnych obiektach. Jeśli działania te dotyczą obiektów fizycznie istniejących, adekwatność tego rozumienia kryteriów nie budzi wątpliwości, gdyż zazwyczaj możliwe jest zebranie wielu potrzebnych informacji o tym obiekcie. Decydują one o poprawności, rzetelności i wiarygodności działań oceniających.

W procesach projektowania taka komfortowa sytuacja nie występuje nigdy. Obiekt oceniany znajduje się „in statu nascendi” („w chwili poczęcia”) i wiele istotnych informacji o nim po prostu nie istnieje. Pomimo, że podejmowanie decyzji projektowych jest również związane z procesami oceny i klasyfikacji, to jednak procesy te dotyczą obiektów innego rodzaju: utworów myślowych nie do końca sprecyzowanych, niepełnych i takich, których skutki działania nie mogą być dokładnie przewidziane. Stąd wynikają formalne i merytoryczne różnice między kryteriami decyzyjnymi o charakterach: diagnostycznym i projektowym (lub prognostycznym). Różnice te dobrze ilustruje następujący przykład.

Dopuszczalny równoważny poziom ciśnienia akustycznego na stanowisku roboczym wynosi 85 dB_A (wg PN-94/N-01307) – tak brzmi kryterium diagnostyczne. Jak tę informację mógłby wykorzystać inżynier projektujący np. przekładnię zębatą obrabiarki? Z pewnością nie w sposób bezpośredni. Wiadomo jednak, że podczas pracy przekładnia hałasuje i że poziom tego hałasu oraz jego częstotliwość zależą od określonych parametrów konstrukcyjnych i eksploatacyjnych: prędkości obrotowej kół oraz ich średnic, kształtu zębów i modułu zazębienia, warunków smarowania i chłodzenia, jakości wykonania, rodzaju zastosowanych materiałów itp. Projektant potrzebuje informacji, że przy prędkości obrotowej „n”, module „m”, liczbach zębów „z₁” i „z₂”, dokładności wykonania „d”, stalowe koła zębate z zębami prostymi, pracujące na pólсуcho, obciążone momentem obrotowym „M₀” emitują hałas o poziomie „L”. Dobierając odpowiednio parametry konstrukcyjne, z których tylko część jest zdeterminowana przez wcześniej ustalone wymagania eksploatacyjne, projektant może unikać „zakazanych” (z uwagi na nadmierny hałas) kombinacji parametrów.

Z powyższego przykładu wynika wniosek, że kryteria ergonomiczne o charakterze diagnostycznym znajdują się na innej płaszczyźnie niż tok rozumowania projektanta i konieczne jest np. związanie tych kryteriów z parametrami konstrukcyjnymi, aby możliwe było wykorzystanie w procesie projektowania zawartych w nich informacji. Faktem wspierającym powyższą tezę są olbrzymie trudności, jakie towarzyszą próbom wykorzystania znanych ergonomicznych list kontrolnych do wspomagania i usprawniania

procesów decyzyjnych w projektowaniu systemów człowiek - obiekt techniczny. Dla potrzeb projektowania opracowano odrębne „listy pytań naprowadzających”, mające służyć jako pomoc przy ergonomicznej ocenie i wyborze wariantów rozwiązań projektowych oraz weryfikowaniu gotowej dokumentacji technicznej i badaniach prototypów lub modeli. Prace te nie zyskały jednak szerokiego poparcia i popularności, być może dlatego, że w ślad za nimi nie dokonano istotnych zmian metodologicznych w strukturach procesów projektowania.

Przeznaczenie kryteriów ergonomicznych do wspomagania działań projektowych powoduje, że ich forma, treść i kolejność występowania muszą być odpowiednie do specyfiki, treści i kolejności decyzji podejmowanych w procesie projektowania. Ze względu na dominację procesów heurystycznych, szczególną wagę ma możliwość ukierunkowania myśli projektanta w stronę rozwiązań ergonomicznie poprawnych poprzez zastosowanie kryteriów decyzyjnych mających postać reguł heurystycznych. W projektowaniu szczegółowym, konstrukcyjnym - w przypadku, gdy zadanie projektowe da się sformalizować (co umożliwia zastosowanie procedur komputerowych) - kryteria ergonomiczne mogą mieć charakter imperatywny: „zrób tak a tak”. Dobrym przykładem są tu komputerowe programy projektowania parametrów przestrzeni pracy człowieka, choć głównym przeznaczeniem tych procedur jest raczej symulacja przestrzennych relacji w systemie człowiek - maszyna. Przede wszystkim jednak kryteria ergonomiczne dostarczają projektantowi wiedzy z zakresu humanizacji techniki, wiedzy potrzebnej dla podjęcia właściwej decyzji w konkretnym kroku procesu projektowego. Ta cecha decyduje o doniosłości roli kryteriów ergonomicznych w projektowaniu, gdyż wyklucza możliwość pojawienia się błędnej informacji „na wejściu” sytuacji decyzyjnej. Informacja nieprawdziwa zawsze prowadzi do przyjęcia błędnych rozwiązań. Informacja prawdziwa, lecz niepełna w takim stopniu, który uniemożliwia podjęcie decyzji, wywołuje „głód informacji” i zmusza projektanta do uzupełnienia braków aż do poziomu uzyskania akceptowalnego ryzyka przy podejmowaniu decyzji.

Kolejność zagadnień rozwiązywanych w projektowaniu ergonomicznym określa jednocześnie strukturę zbioru ergonomicznych kryteriów projektowych. Przedstawiono ją w tabeli 1. Jest to struktura uporządkowana i hierarchiczna. Uporządkowanie zbioru dotyczy treści kryteriów projektowych oraz ich kolejności i odzwierciedla kolejność zagadnień rozwiązywanych podczas projektowania systemu człowiek - obiekt techniczny. Uhierarchizowanie kryteriów wyraża się występowaniem kryteriów mających różne poziomy szczegółowości. Kryteria o charakterze ogólnym (strategicznym) mają zastosowanie podczas formułowania ergonomiczno - techniczno - ekonomicznych założeń do projektu (tzw. faza ZETE) i podczas tworzenia wstępnych koncepcji rozwiązań elementów systemu (projektowanie wstępne). Kryteria szczegółowe przeznaczone są dla fazy projektowania szczegółowego (konstruowania elementów obiektu technicznego) i spełniają rolę „wspomagania ergonomicznego” dla konstruktorów przy podejmowaniu szczegółowych decyzji projektowych.

Od treści kryteriów oraz przyjętego sposobu ich wartościowania zależy decyzja o wyborze określonego, częściowego lub ostatecznego rozwiązania, a więc również końcowy efekt procesu projektowania. Podczas analizy sekwencji kroków projektowych zwraca uwagę fakt, że wszędzie występuje ta sama czynność logiczna, polegająca na dokonywaniu oceny poprzedniego działania przy pomocy ustalonego zbioru kryteriów. Różni autorzy nazywają tę czynność różnie, np. „ocena”, „wybór”, „decyzja”.

Tabela 1. Ergonomiczne kryteria projektowe
Table 1. The ergonomic design criteria

E 1. PROCES PRACY	
1.1. Treść pracy	1.1.1. Stopień autonomii stanowiska 1.1.2. Zadania człowieka 1.1.3. Zadania maszyny (obiektu technicznego) 1.1.4. Optymalizacja fizycznego obciążenia człowieka 1.1.5. Optymalizacja psychicznego obciążenia człowieka 1.1.6. Wymagane kwalifikacje i możliwości personelu
1.2. Metody pracy	1.2.1. Struktura ruchów roboczych 1.2.2. Struktura pracy koncepcyjnej 1.2.3. Tempo i rytm pracy 1.2.4. Przerwy wypoczynkowe
1.3. Funkcje wyposażenia technologicznego	1.3.1. Funkcje maszyny podstawowej 1.3.2. Funkcje pomocy warsztatowych 1.3.3. Środki transportu materiałów 1.3.4. Funkcje wyposażenia pomocniczego
E 2. PRZESTRZEN PRACY	
2.1. Uwarunkowania pozycji ciała	2.1.1. Zakres zmienności pozycji ciała 2.1.2. Wysokość manipulacyjna 2.1.3. Siedzisko, oparcie, podłokietniki, podnóżek 2.1.4. Podest
2.2. Przestrzeń czynności ruchowych i pracy wzroku	2.2.1. Przestrzeń pracy rąk i nóg 2.2.2. Rozmieszczenie wyposażenia technologicznego 2.2.3. Przejścia i dojścia 2.2.4. Przestrzeń pracy wzroku i warunki widoczności
2.3. Architektura obiektu technicznego	2.3.1. Przestrzenna alokacja zespołów funkcjonalnych maszyny 2.3.2. Wstępny kształt zespołu scalającego (korpusu, ramy itp.) 2.3.3. Rozmieszczenie materiałów i pomocy warsztatowych 2.3.4. Składowanie i przemieszczanie materiałów 2.3.5. Estetyka form i barw obiektu technicznego
E 3. ELEMENTY INFORMACYJNE, SYGNALIZACYJNE I STEROWNICZE	
3.1. Elementy informacyjne i sygnalizacyjne	3.1.1. Dobór nośników informacji 3.1.2. Dobór typów urządzeń sygnalizacyjnych i informacyjnych 3.1.3. Rozmieszczenie urządzeń sygnalizacyjnych i informacyjnych
3.2. Elementy sterownicze	3.2.1. Dobór typów urządzeń sterowniczych do rodzaju funkcji 3.2.2. Oznaczenie elementów sterowniczych i powiązanie ich z elementami sygnalizacyjnymi i sterowniczymi 3.2.3. Rozmieszczenie elementów sterowniczych
E 4. ŚRODOWISKO FIZYCZNO - CHEMICZNO - BIOLOGICZNE	
4.1. Źródła zanieczyszczeń powietrza	4.1.1. Zmniejszenie emisji i/lub czasu narażenia 4.1.2. Hermetyzacja źródeł 4.1.3. Instalacje ochronne
4.2. Źródła zakłóceń mikroklimatu	4.2.1. Eliminacja lub osłabienie zakłóceń 4.2.2. Hermetyzacja lub ekranowanie źródeł zakłóceń 4.2.3. Instalacje ochronne
4.3. Źródła drgań, wibracji i hałasu	4.3.1. Zmniejszenie mocy źródeł lub czasu narażenia 4.3.2. Izolowanie lub ekranowanie źródeł 4.3.3. Ochrona indywidualna człowieka
4.4. Źródła szkodliwego promieniowania	4.4.1. Zmniejszenie mocy źródeł lub czasu narażenia 4.4.2. Izolowanie źródeł 4.4.3. Ochrona indywidualna człowieka
4.5. Źródła światła	4.5.1. Dobór mocy źródeł i barwy światła 4.5.2. Rozmieszczenie źródeł oświetlenia miejscowego i ogólnego
4.6. Źródła szkodliwych substancji	4.6.1. Eliminacja lub hermetyzacja źródeł 4.6.2. Neutralizacja szkodliwych substancji 4.6.3. Ochrona indywidualna człowieka

Trudna do przecenienia rola kryteriów decyzyjnych uzasadnia twierdzenie, że „sztuka projektowania jest sztuką godzenia” [1, s.7]. Niewiele procedur projektowania technicznego uwzględnia wagę problemu oceny. Autorzy najczęściej nie precyzują kryteriów oceny, jeśli nie dotyczą one parametrów technicznych. Zazwyczaj mówi się o konieczności ustalenia zbioru kryteriów, segregowaniu ich, przypisywaniu rang lub współczynników preferencji, opracowaniu metody oceny wielokryterialnej itp. A. Góralski zauważa „...godny ubolewania fakt przywiązywania większej wagi do sposobów otrzymywania rozwiązań, a mniejszej - do sposobów ich oceny” [3, s.94]. Podkreślając, że jest to zagadnienie o wielkiej doniosłości, autor wyjaśnia, że jest ono spowodowane swoistą specjalizacją: jedni rozwiązują zadania, a inni je oceniają i decydują o ich wykorzystaniu lub odrzuceniu (gdyż dysponują pieniędzmi na inwestycje). W literaturze można spotkać jedynie bardzo ogólnikowe traktowanie problemu kryteriów oceny, zazwyczaj w kontekście omawiania procedur optymalizacji matematycznej lub w postaci zestawiania rodzajów i typów kryteriów.

3. Struktura procesu projektowania ergonomicznego

Pomiędzy strukturą projektowania technicznego a strukturą procesu projektowania ergonomicznego istnieją znaczne różnice. Proces projektowania technicznego opisywany jest zazwyczaj w postaci sieciowej lub liniowej sekwencji kroków. Charakterystyczne jest tam określenie przedmiotu projektowania jako tworu wyłącznie technicznego. Ma to doniosłe znaczenie dla przebiegu procesu projektowania i określone konsekwencje dla jego wyniku. Racje humanocentryczne są tu wyraźnie usytuowane na dalszym planie, poza racjami technicznymi oraz ekonomicznymi i mają postać zewnętrznych ograniczeń, w niewielkim stopniu modyfikujących przyjęte wcześniej rozwiązania konstruktorskie.

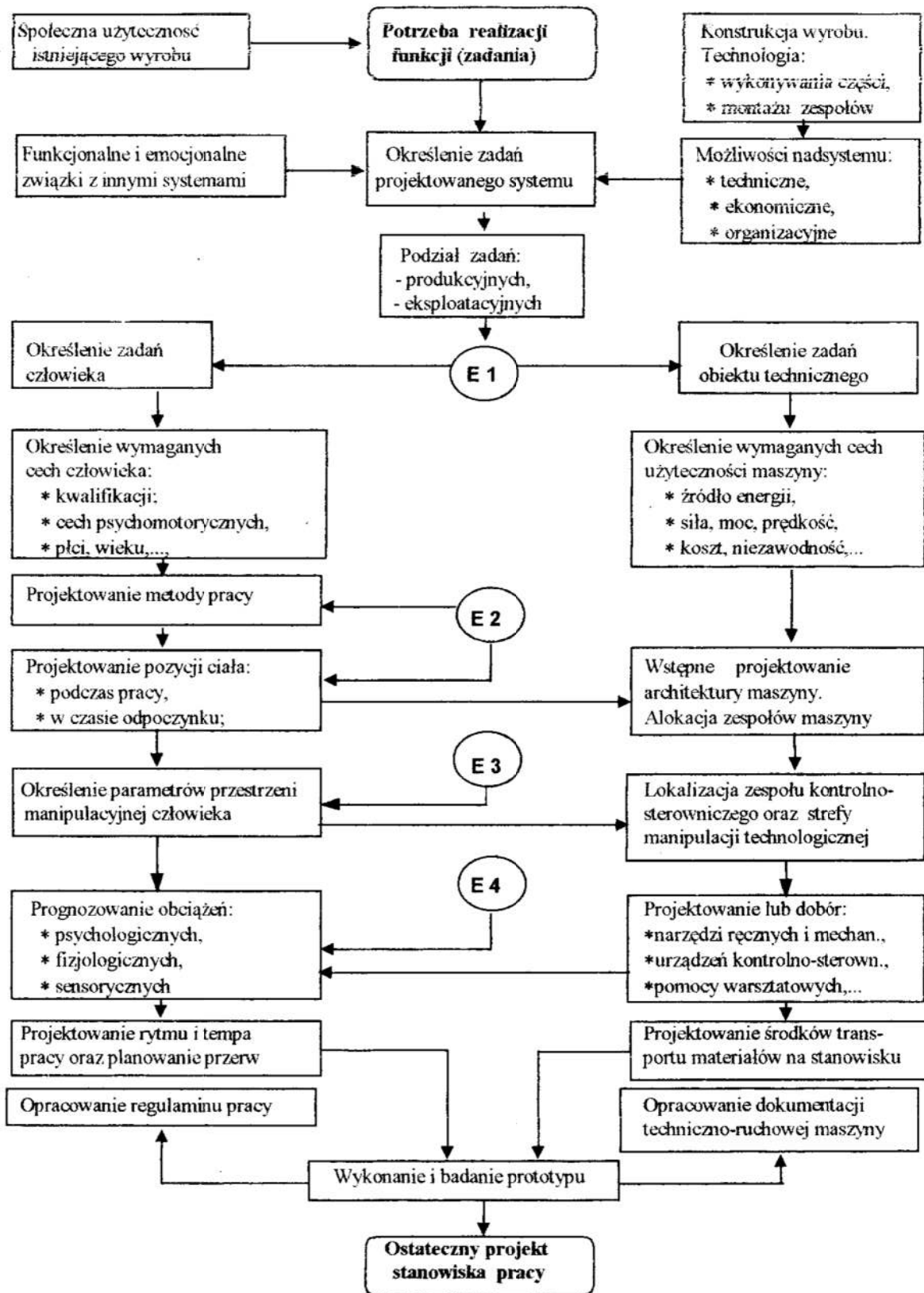
Jednym z głównych założeń metodologicznych projektowania ergonomicznego jest wbudowanie odpowiednio dobranych kryteriów projektowych do każdego kroku procesu projektowania w taki sposób, aby kryteria te były integralną częścią procedury.

Procedura projektowania ergonomicznego może być przedstawiona w postaci algorytmu będącego typową hybrydą: obejmującą obszar pomiędzy dwoma skrajnymi procedurami: ściśle deterministyczną i całkowicie heurystyczną (rys. 3).

Jest to sytuacja, gdy projektant może skorzystać z pewnej liczby gotowych (fizycznych lub teoretycznych) wzorców rozwiązań (np. silniki elektryczne, przekładnie, elementy hydrauliki, komputerowe programy obliczeń inżynierskich lub analiz antropometrycznych), lecz wykorzystuje je do stworzenia nowego podsystemu technicznego. Procedura taka jest często opisywana przez różnych autorów w postaci algorytmów logicznych (stąd jej nazwa), lecz najczęściej dotyczy ona tradycyjnego ujęcia procesu projektowania. Algorytm projektowania ergonomicznego musi być zatem znacznie szerszy i bogatszy niż tradycyjne algorytmy projektowania technicznego, gdyż dotyczy znacznie bardziej skomplikowanego systemu, będącego przedmiotem projektowania.

Charakterystyczna i istotna jest kolejność zagadnień ergonomicznych (kryteriów strategicznych), branych pod uwagę w procesie projektowania systemu:

- 1) proces pracy (E1),
- 2) przestrzeń pracy (E2),
- 3) elementy informacyjne, sygnalizacyjne i sterownicze (E3),
- 4) środowisko fizyczne, chemiczne i biologiczne (E4).



Rys. 3. Struktura procesu projektowania ergonomicznego
 Fig. 3. The structure of ergonomic design process

Powyższe zagadnienia, przetworzone do postaci ergonomicznych kryteriów decyzyjnych, są integralnym składnikiem struktury projektowania ergonomicznego.

Literatura

1. Archer B.L.: *Systematyczna metoda projektowania przemysłowego*. Wyd. Instytut Wzornictwa Przemysłowego, seria: Biblioteka Wzornictwa, nr 7, Warszawa, 1987
2. Arystoteles: *Etyka nikomachejska*. PWN, Warszawa, 1982
3. Góralski A.: *Twórcze rozwiązywanie zadań*. PWN, Warszawa, 1980
4. Miller D.: *Projektowanie metodyczne*. WNT, Warszawa, 1987
5. Pszczołowski T.: *Mała encyklopedia prakseologii i teorii organizacji*. Ossolineum, 1978
6. Simon H.A.: *Podjęcie decyzji kierowniczych. Nowe nurty*. PWE, W-wa, 1982
7. Singleton W.T.: *Man – machine systems*. Penguin Press, 1974
8. Schmidtke H. (Ed.): *Ergonomic data for equipment design*. Plenum Press, 1984
9. Tytyk E.: *Metodologia projektowania ergonomicznego w budowie maszyn*. Wyd. Politechniki Poznańskiej, seria: Rozprawy, nr 252, Poznań, 1991

Streszczenie

W referacie przedstawiono koncepcję włączenia wymagań ergonomicznych do procesów projektowania systemów: człowiek – obiekt techniczny. Takie ujęcie procesu projektowania na tyle różni się od ujęć tradycyjnych, że usprawiedliwione jest przyjęcie nazwy: projektowanie ergonomiczne. Ma ono na celu uzyskanie optymalnego dostosowania projektowanego obiektu technicznego do wymagań człowieka jako użytkownika tego obiektu. Realizacja procesu projektowania ergonomicznego wymaga od projektanta określonej wiedzy z zakresu ergonomii. Wiedzę tę, w odpowiednio przygotowanej formie, mogą dostarczyć tzw. ergonomiczne kryteria projektowe, wspomagające procesy podejmowania decyzji i kierunkujące myśli projektanta w stronę rozwiązań humanizujących technikę.

ERGONOMIC AIDING OF DESIGN PROCESSES

Summary

In the paper is shown the conception of including of ergonomic demands to design processes of man–machines systems. That approach to design process is differ from traditional approaches so much, if is possible to use of name: ergonomic design. Its aim is the achieving of an optimal fitting of the designing technical object to the human demands of user of the object. Designer has to have the definite knowledge from area ergonomics, making of the ergonomic design process. That knowledge, in adequate prepared form, can deliver so called ergonomic design criteria, aiding making decisions processes and directing a designer thoughts to the solutions which are humanize of technology.

WSPOMAGANIE DECYZJI PLANOWANIA DOBOWEGO NA HURTOWYM RYNKU ENERGII ELEKTRYCZNEJ

1. Wstęp

Zagadnienia planowania przez operatora systemu pracy jednostek wytwórczych w warunkach funkcjonującego rynku energii elektrycznej są szczególnie trudnymi i złożonymi zadaniami optymalizacji wielokryterialnej. Często oprócz minimalizacji kosztów realizacji zapotrzebowania na energię elektryczną ważne jest także uwzględnienie dodatkowych aspektów działania systemu elektroenergetycznego, w tym m.in. wymagań bezpieczeństwa systemu elektroenergetycznego, zapewnienia odpowiednich poziomów rezerw, realizacji kontraktów długoterminowych, itp.

Systemy ułatwiające proces planowania powinny umożliwiać z jednej strony automatyczne generowanie rozwiązań, ale z drugiej strony muszą pozwalać na realizację celów i zamierzeń operatora systemu. Ze względu na zakres dostarczanych informacji oraz różnice w wymaganiach stawianych operatorowi, narzędzia planowania pracy jednostek wytwórczych można podzielić na trzy klasy: edytory ręcznego układania planów pracy jednostek, algorytmy optymalizacyjne wyznaczające gotowe plany pracy jednostek oraz systemy wspomaganie decyzji z algorytmami wyznaczania propozycji rozwiązań [7].

W ostatnich latach wraz ze zmieniającymi się wymaganiami dotyczącymi funkcjonowania rynku energii elektrycznej w Polsce pojawiła się konieczność modyfikacji istniejących algorytmów optymalizacji kosztowej [5] i uwzględnienie także innych kryteriów [6]. Specyfika i trudności w pogodzeniu ich z modelem minimalizacji kosztów determinowała potrzebę zaprojektowania i zbudowania systemu wspomaganie decyzji operatora planowania [8]. System taki z jednej strony pozwala na automatyczne generowanie propozycji rozwiązań w sensie minimalnokosztowym, a z drugiej umożliwia operatorowi zmiany otrzymanego harmonogramu tak, aby rozwiązanie końcowe odpowiadało preferencjom użytkownika ze względu na wszystkie kryteria jakości.

Prezentowany w pracy system wspomaganie decyzji operatora został opracowany na potrzeby fazy treningowej wprowadzanego w Polsce rynku energii elektrycznej i jest on obecnie rozwijany i rozbudowywany nie tylko o nowe moduły algorytmiczne, ale także o nowe moduły analizy i prezentacji danych. W niniejszym referacie oprócz

* Instytut Automatyki i Informatyki Stosowanej Politechniki Warszawskiej
ul. Nowowiejska 15/19, 00-665 Warszawa
adresy email: awalczak@ia.pw.edu.pl, toczyłowski@ia.pw.edu.pl

krótkiej prezentacji systemu przedstawione zostaną także kierunki jego dalszego rozwoju.

2. System wspomagania decyzji operatora w zakresie planowania pracy jednostek wytwórczych na ofertowym rynku energii elektrycznej

Charakterystyka potrzeb operatora planowania

Głównymi zadaniami Operatora Krajowego Systemu Elektroenergetycznego jest bilansowanie mocy oraz zapewnienie ciągłości dostaw oraz bezpieczeństwa pracy całego systemu, w tym uwzględnianie szerokiego zbioru ograniczeń systemowych. W zależności od modelu rynku, zakres kompetencji i odpowiedzialności Operatora może być bardzo różny. W szczególności może pokrywać także konieczność realizacji zawartych kontraktów oraz uwzględniania nie tylko ograniczeń sieciowych, ale również ograniczeń technicznych jednostek wytwórczych oraz ograniczeń elektrownianych (np. SOREE [1]). Wymaga to podejmowania szeregu złożonych decyzji, a dodatkowe kryteria jakości nie dają się w sposób jednoznaczny zapisać w postaci modelu optymalizacji jednokryterialnej.

Dla potrzeb planowania pracy jednostek wytwórczych zaprojektowano specjalizowany algorytm poszukiwania optymalnych rozwiązań oraz opracowano system wspomagania decyzji umożliwiający operatorowi zmianę i dostosowanie otrzymywanych rozwiązań do aktualnych wymagań operatora. Dzięki temu zapewniono nie tylko niski koszt harmonogramu, ale również możliwość uwzględnienia innych kryteriów jakości, jak m.in. odpowiedniego poziomu rezerw mocy czy stopnia realizacji kontraktów długoterminowych.

System wspomagania decyzji operatora planowania

System wspomagania decyzji operatora planowania zawiera reprezentację wiedzy o sytuacji decyzyjnej w postaci modeli analitycznych, odpowiednie algorytmy umożliwiające korzystanie z tych modeli, moduły umożliwiające korzystanie z bazy danych oraz wygodny interfejs użytkownika. Wspomaga on cały proces dochodzenia do właściwej decyzji i pozwala on na wypracowywanie rozwiązań i ich ocenę za pomocą odpowiednio określonych kilku kryteriów, w tym kryterium kosztowego.

System rozpoczyna działanie od pobrania danych wejściowych dotyczących ofert produkcyjnych i cenowych jednostek wytwórczych, prognozy zapotrzebowania mocy Krajowego Systemu Elektroenergetycznego wraz z określonymi poziomami rezerw oraz poziomów generacji zewnętrznej (np. eksport, import, generacja hydroelektrowni, itp.). Następnie operator uruchamia algorytm optymalizacyjny, który wyznacza propozycję harmonogramu. Mając wiedzę dotyczącą pożądaných warunków pracy jednostek wytwórczych, może on wielokrotnie weryfikować ewentualne wskaźniki dotyczące jakości harmonogramu wyznaczonego przez algorytm. System wspomagania decyzji pozwala na dokonywanie zmian w harmonogramie oraz ich ocenę względem określonego jednego kryterium lub grupy kryteriów. Proces ten jest powtarzany iteracyjnie aż do uzyskania satysfakcjonującego rozwiązania. W trakcie tego procesu możliwe jest porównywanie poszczególnych wersji oraz powrót do wybranej wersji harmonogramu. Rozwiązanie końcowe jest zgodne z preferencjami operatora i może być przyjęte do realizacji.

Polioptymalizacja

Automatyczne algorytmy optymalizacyjne stosują zwykle jedno kryterium dotyczące minimalizacji sumarycznych kosztów zakupu energii na rynku w okresie doby. W praktyce operator nie zawsze może kierować się tylko kosztami, a przykładem jest potrzeba spełnienia ograniczeń technicznych i systemowych, w tym uwzględnienia odpowiednich poziomów rezerw. Uwzględnienie tych wymagań wiąże się ze zwiększeniem kosztów i może być trudne do interpretacji w sensie kryteriów ekonomicznych. Innym przykładem jest wymóg realizacji kontraktów długoterminowych, których uwzględnienie nie daje się w prosty sposób opisać w modelu kosztowym.

System wspomaganie decyzji umożliwia uwzględnianie więcej niż jednego kryterium. Do tego niezbędne jest dostarczanie przez operatora dodatkowej *informacji o preferencjach*. Preferencje operatora są wykorzystywane do generacji propozycji harmonogramu. W kolejnej iteracji operator może zmienić swoje preferencje dotyczące np. zwiększenia kosztów i podwyższenia wymagań w zakresie bezpieczeństwa pracy systemu, a następnie stwierdzić, jak to wpłynie na harmonogram i poszczególne wskaźniki.

Oprócz definicji kryteriów, operator może również dokonać pewnego rodzaju analizy parametrycznej ograniczeń. System automatycznie zapamiętuje historię procesu interakcji. Zastosowanie do oceny harmonogramu większej liczby kryteriów zwiększa elastyczność wyboru harmonogramów a interaktywna praca operatora z systemem zwiększa jego zaufanie do otrzymanego rozwiązania.

Stworzony system wspomaganie decyzji pozwala obecnie na ocenę harmonogramu pod kątem kilku zdefiniowanych kryteriów. Są to:

- *Koszt całkowity*. Jest to całkowity koszt zakupu energii w ciągu doby, uwzględniający dane ofertowe jednostek, w tym koszty rozruchu.
- *Zapas rezerwy górnej*. Jest to zapas górnej rezerwy mocy w „najgorszej” godzinie harmonogramu. Przez zapas górnej rezerwy mocy rozumieć należy różnicę między sumą mocy maksymalnych, którą mogą generować pracujące jednostki, a prognozą zapotrzebowania powiększoną o wymagany górny poziom rezerw.
- *Zapas rezerwy dolnej*. Jest to zapas dolnej rezerwy mocy w „najgorszej” godzinie harmonogramu. Przez zapas dolnej rezerwy mocy rozumieć należy różnicę między prognozą zapotrzebowania zmniejszoną o wymagany dolny poziom rezerw a sumą mocy minimalnych, którą mogą generować pracujące jednostki.
- *Energia kontraktowa*. Jest to procentowa ilość energii wytwarzanej przez jednostki realizujące zobowiązania kontraktowe w stosunku do energii całkowitej.
- *Zmiana elastyczności*. Współczynnik ten ma za zadanie oceniać wpływ wyboru pracujących jednostek kontraktowych na poprawę lub pogorszenie się wymagań dotyczących zakresów regulacyjności pozostałych jednostek.

Model rynku i cechy rozwiązań

Realizacja funkcji bilansowania wymaga złożonych metod planowania i podejmowania złożonych decyzji. Modele matematyczne wykorzystywane na potrzeby planowania dobowego pracy jednostek wytwórczych na rynku energii elektrycznej (np. rynku SOREE) są często modelami o dużej liczbie zmiennych, złożonych zależnościach pomiędzy nimi oraz trudnych do uwzględnienia ograniczeń [3]. Nierzadko charakteryzują się

nieliniową funkcją celu np. uwzględniającą istnienie ceny krańcowej. W trakcie planowania brany jest pod uwagę szereg ograniczeń, w tym także tych wynikających z potrzeby uwzględnienia wymaganych poziomów zapotrzebowania i rezerw, zmieniających w czasie zakresów regulacyjności jednostek i ich parametrów technicznych oraz kilku charakterystyk rozruchu dla każdej jednostki. Oferta cenowa jednostki składa się z zadeklarowanego kosztu rozruchu oraz nawet kilku cen za wytwarzanie energii w poszczególnych przedziałach czasu po uruchomieniu.

Rozmiar pełnego zadania optymalizacji dyskretnej, mierzony liczbą zmiennych i ograniczeń jest bardzo duży [2]. Powoduje to, że rozwiązanie problemu doboru jednostek w sposób gwarantujący całkowitą optymalność w zadowalającym czasie okazało się niemożliwe, mimo wykorzystania jak najlepszych metod i technik optymalizacji dyskretnej, w tym efektywnych pakietów komercyjnych CPLEX i ILOG [3]. Możliwe natomiast okazało się uzyskiwanie bardzo dobrych rozwiązań dopuszczalnych w krótkim czasie za pomocą przybliżonego, specjalizowanego algorytmu optymalizacyjnego [6].

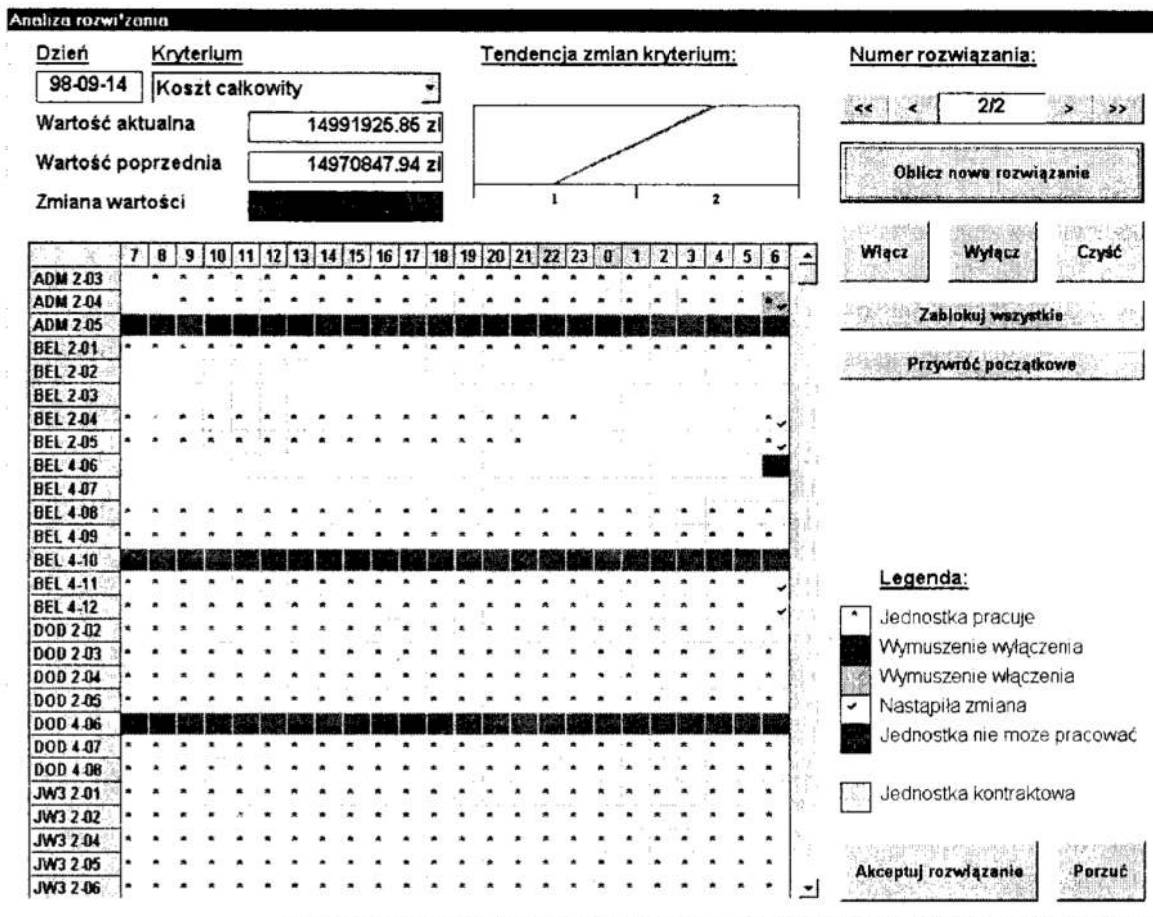
Przyjęcie proponowanego modelu matematycznego [5] z kosztową funkcją celu, wymaganiami dotyczącymi zakresów rezerw mocy, ograniczeniami dotyczącymi zakresów regulacyjności jednostek i czasów rozruchów powoduje, że nie zawsze jednostka, która zadeklarowała cenę niższą od ceny krańcowej w danej godzinie będzie przyjęta do pracy. W szczególnych przypadkach tańsza jednostka o niewielkim zakresie przedziału generacji może zostać wypchnięta z planu przez nieco droższą jednostkę o szerszym zakresie regulacyjności, jeżeli jest to wymagane dla zapewnienia pożądanego poziomu rezerw o tej godzinie [4]. Konieczność uwzględnienia nawet tylko jednego z warunków chwilowych w danej godzinie, np. wymuszonego przez ofertę odstawienia jednej z pracujących jednostek, może w pewnych warunkach spowodować cały łańcuch zależności, w wyniku których może nastąpić uruchomienie i odstawienie wielu jednostek w dłuższym czasie.

Ze względu na postać problemu, w algorytmie podczas planowania uruchamiania jednostek są preferowane te jednostki, które nie tylko zgłaszają niską cenę ofertową, ale również niskie koszty rozruchu, duży zakres oferowanej generacji i stałe w czasie parametry pracy. Negatywne cechy poszczególnych rozwiązań cząstkowych mogą zostać złagodzone lub nawet wyeliminowane poprzez umożliwienie operatorowi modyfikacji harmonogramów dzięki zastosowaniu systemu wspomagania decyzji.

Algorytm wyznaczania propozycji rozwiązań

Propozycje planów pracy jednostek są wyznaczane przez zaprojektowany na potrzeby tego rynku specjalizowany, przybliżony algorytm optymalizacyjny [5]. Algorytm ten ma na celu szybkie znalezienie dopuszczalnych i bliskich optymalnemu rozwiązaniu problemu. Są one generowane automatycznie przy uwzględnieniu preferencji operatora i przy ustalonych przez niego parametrach. Stanowią one podstawę kolejnej iteracji modyfikacji rozwiązań i poszukiwania rozwiązania akceptowanego przez operatora.

Algorytm rozpoczyna działanie od przeprowadzenia analizy kształtu krzywej prognozy popytu na moc. Na tej podstawie horyzont planowania jest dzielony na etapy, w których krzywa prognozy popytu zachowuje cechy monotoniczności. Działanie procedury poszukiwań rozwiązań jest dalej zróżnicowane w zależności od rodzaju etapów.



Rys. 1. Przykładowe okno systemu wspomaganie decyzji operatora planowania jednostek wytwórczych

Fig. 2. Example window of the unit commitment decision support system

W dalszej części algorytmu jest realizowana tzw. *regularyzacja* parametrów problemu, a następnie jest wykonywany złożony, dynamiczny algorytm poszukiwań zestawu jednostek aktualnie uruchamianych przy narastającej krzywej popytu oraz algorytm odstawiania jednostek przy malejącej krzywej popytu. Wynikowy algorytm jest stosunkowo złożony i realizuje niejawną przeglądnę możliwych rozwiązań na horyzoncie złożonym z wielu godzin. Jego celem jest otrzymanie harmonogramu jak najtańszego z punktu widzenia sumarycznego kosztu w całym okresie planowania i związanego zarówno z kosztem uruchamiania jednostek jak i bezpośrednim kosztem zakupu energii dostarczonej przez jednostki już uruchomione.

3. Wykorzystanie systemu wspomaganie decyzji operatora planowania

Narzędzie budowy harmonogramów

Systemy wspomaganie decyzji łącząc proste edytory planów pracy i złożone algorytmy optymalizacji skupiają zalety i funkcjonalność obu rozwiązań. Możliwe jest zdefiniowanie lub zmiana kryteriów wyznaczania rozwiązań oraz iteracyjne poszukiwanie najlepszego możliwego rozwiązania końcowego. Ciężar znajdowania rozwiązania jest podzielony pomiędzy operatora i algorytmy systemu. Operator ma możliwość zmiany automatycznie wyznaczanych rozwiązań, przy czym jest

informowany o ewentualnych konsekwencjach tych działań. W celu znalezienia rozwiązania najbardziej satysfakcjonującego operatora pomocne mogą być narzędzia przeszukiwania przestrzeni rozwiązań, porównywania rozwiązań cząstkowych oraz informacje dotyczące wskaźników jakości wyliczane na bieżąco w trakcie wprowadzania zmian [8]. Dzięki tym cechom system wspomaganie operatora planowania zapewnia z jednej strony automatyczne znalezienie rozwiązania zbliżonego do optymalnego, a z drugiej pozwala na modyfikacje tak znalezionej rozwiązania, aby ostatecznie w pełni odpowiadało ono preferencjom operatora.

Analiza wielokryterialna

Istotnym aspektem wykorzystania systemów wspomaganie decyzji jest także analiza wielokryterialna otrzymanych rozwiązań. Możliwe jest bowiem określenie wartości kilku kryteriów dla znalezionych rozwiązań i wybór takiego, które najbardziej odpowiada preferencjom operatora.

Zdefiniowanie dodatkowych kryteriów i wskaźników, w tym także tych istotnych dla określania dopuszczalności rozwiązania ze względu na niektóre ograniczenia sieciowe i elektrowniane, jest dla operatora zobligowanego do ich uwzględnienia szczególnie istotne i potrzebne. Możliwe jest przecież wykorzystanie zalet i możliwości systemu wspomaganie decyzji, w tym jego możliwości prezentacyjnych, do przygotowania harmonogramów pracy jednostek wytwórczych najlepiej odpowiadających znanym ograniczeniom sieciowym i elektrownianym oraz zapewniających należyty poziom bezpieczeństwa dostaw oraz bezpieczeństwa samego systemu elektroenergetycznego.

Oprócz wymienionych wyżej podstawowych kryteriów oceny harmonogramów interesujące dla operatora są także wskaźniki dodatkowe, takie jak np. liczba uruchomień i odstawień jednostek w poszczególnych godzinach, moc wynikająca z uruchomień w poszczególnych godzinach, czasy nieprzerwanej pracy, czasy odstawienia poszczególnych jednostek, procentowo określony dla każdej jednostki punkt pracy w stosunku do jej zakresu regulacyjności, itp. Każde z powyższych kryteriów może być określone poprzez wartość minimalną, średnią i maksymalną.

Ograniczenie przestrzeni rozwiązań

Przy tak złożonych zadaniach optymalizacji dyskretnej rozmiar przestrzeni rozwiązań jest jednym z najistotniejszych czynników warunkujących nie tylko czas poszukiwania rozwiązania optymalnego, ale nawet możliwość znalezienia w dostępnym czasie takiego rozwiązania w ogóle. Fakt ten nabiera szczególnego znaczenia, jeśli wziąć pod uwagę, że dla celów pełnej analizy wielokryterialnej konieczne może być przeprowadzenie kilku lub kilkunastu serii obliczeń w ograniczonym czasie.

Systemy wspomaganie decyzji dają operatorowi możliwość ograniczenia przestrzeni rozwiązań, tak zakresu jednostek, jak i horyzontu planowania. Przykładowo w przypadku zaimplementowanego specjalizowanego algorytmu harmonogramowania redukcja liczby jednostek o połowę zmniejsza czas potrzebny na znalezienie rozwiązania o 50% (większa część tego czasu przypada na wczytanie i zapis danych). Jednakże przy porównaniu czasów działania samego algorytmu optymalizacyjnego różnice te są znacznie większe.

Dotychczasowe doświadczenia z badaniem możliwości wykorzystania pakietu optymalizacyjnego CPLEX dowiodły, że ograniczenie rozmiaru zadania powoduje znaczne skrócenie czasu obliczeń [3]. Dzięki temu dla ograniczonego zadania planowania (zmniejszonego zestawu jednostek lub zmniejszonego horyzontu planowania) możliwe staje się rozwiązanie go w sposób optymalny.

Algorytmy generowania propozycji rozwiązań

Rozwiązania generowane automatycznie przez specjalizowane algorytmy stanowią dla operatora propozycje i punkt wyjścia do znalezienia rozwiązania docelowego. Zdecydowanie korzystnym jest umożliwienie operatorowi przeglądu kilku takich propozycji rozwiązań. Mogą to być rozwiązania znalezione różnymi metodami, za pomocą różnych algorytmów heurystycznych, metod przybliżonych oraz dzięki rozwiązaniu niektórych zadań i podproblemów optymalizacji.

Operator jest zainteresowany znajomością kilku „konkurencyjnych” rozwiązań o jak najmniejszym koszcie całkowitym, ale także rozwiązań optymalnych z punktu widzenia poszczególnych kryteriów. Jeśli pozwalają na to postać i rozmiar wykorzystywanych zadań programowania całkowitoliczbowego mieszanego, rozwiązania optymalizujące poszczególne wskaźniki jakości mogą być znalezione w sposób dokładny (np. po wykorzystaniu wspomnianych już pakietów). Stosowane są również metody przybliżone (np. wymieniony wcześniej algorytm), czy też inne „alternatywne” metody poszukiwania rozwiązań jak np. algorytm symulowanego wyżarzania dla poprawy rozwiązań.

4. Podsumowanie

Przedstawiony w referacie system potwierdził swoje zalety, łącząc szybkość i dokładność rozwiązań algorytmów optymalizacyjnych z elastycznością ręcznie uzyskiwanych harmonogramów. Dodatkowo wielokryterialna analiza problemu pozwoliła na dostosowanie wyników do bieżących potrzeb operatora, np. umożliwiła zwiększenie udziału energii kontraktowej wymuszone wolniejszym niż się spodziewano tempem realizacji kontraktów długoterminowych.

Ważną zaletą tego systemu jest także możliwość jego użycia do wyznaczania planów pracy uwzględniających nie tylko ograniczenia techniczne jednostek, ale także niektóre ograniczenia elektrowniane i sieciowe. Jest to ogromnie ważne w aspekcie ciągłości dostaw oraz bezpieczeństwa pracy systemu elektroenergetycznego. Zmniejszenie przestrzeni rozwiązań umożliwia także wykorzystanie komercyjnych pakietów do rozwiązania podproblemów optymalizacji.

Literatura

- [1] Energoprojekt Consulting S.A.: *Podstawowe zasady systemowego rynku ofertowego (poolu) - część I i II - wersja 5.0 przyjęta dla celów fazy treningowej przez Zespół Negocjacyjny Umowy SOREE*. PSE Warszawa 1997.
- [2] Sikorski T., Toczyłowski E.: *Badania symulacyjne właściwości rozwiązań problemu doboru jednostek wytwórczych na rynku ofertowym - wyniki badań i wnioski*. Raport IAiS PW nr: 98-3, Warszawa 1998.

- [3] Toczyłowski E., Ogryczak W., Pieńkosz K., Sikorski T., Szwed C., Walczak A., Zadora K., Nowak W.: *Analiza przydatności oprogramowania komercyjnego CPLEX i ILOG do rozwiązywania problemów optymalizacji dyskretnej dla modeli planowania pracy jednostek wytwórczych bez ograniczeń i z ograniczeniami systemowymi (ROPLAN)*. Centrum Badawcze Energetyki i Ochrony Środowiska PW, Warszawa 1998.
- [4] Toczyłowski E., Walczak A.: *Analiza zagadnień planowania generacji w scentralizowanych formach rynku energii*. Raport IAiIS PW nr: 99-18, Warszawa 1999.
- [5] Toczyłowski E., Walczak A.: *Algorytm harmonogramowania pracy jednostek wytwórczych dla modelu rynku energii elektrycznej*. Materiały konferencyjne XIII Krajowej Konferencji Automatyki 1999.
- [6] Toczyłowski E., Walczak A., Kaleta M.: *Algorytm harmonogramowania pracy jednostek wytwórczych w systemie elektroenergetycznym*. Rocznik Zeszyty Naukowe Politechniki Śląskiej 2000 (materiały konferencyjne XII Krajowej Konferencji Automatyzacji Procesów Dyskretnych, Zakopane 2000) – w druku.
- [7] Walczak A., Toczyłowski E., Granat J., Fleszar K.: *Wspomaganie decyzji dobowego planowania pracy jednostek wytwórczych na rynku energii elektrycznej*. Materiały konferencji "Optymalizacja w Energetyce", Jachranka k/o W-wy 1999.
- [8] Walczak A. i inni: *Instrukcja obsługi programu RO-KCO+ v.5.0*. Uczelniane Centrum Badawcze Energetyki i Ochrony Środowiska PW, Warszawa 1999.

Streszczenie

W referacie omówiono system wspomagania decyzji operatora przy dobowym planowaniu pracy jednostek wytwórczych na rynku energii elektrycznej. System wspomaga operatora w procesie wyznaczania dopuszczalnego harmonogramu pracy jednostek spełniającego niektóre ograniczenia techniczne i systemowe. Celem systemu jest również analiza rozwiązań zadań optymalizacji wielokryterialnej, aby ostateczna postać rozwiązania jak najlepiej odzwierciedlała preferencje operatora. W referacie przedstawiono zastosowanie systemu wspomagania decyzji z wbudowanym algorytmem wyznaczania propozycji rozwiązań zaimplementowane dla fazy treningowej Systemowego Ofertowego Rynku Energii Elektrycznej w Polsce (SOREE).

UNIT – COMMITMENT DECISION SUPPORT SYSTEM ON ENERGY WHOLESALE MARKET

Summary

The paper presents unit-commitment Decision Support System with optimization algorithms for the energy wholesale market. DSS combines advantages of advanced optimization algorithms and multicriteria analysis methods. DSS can be also used for constraint analysis of the optimization problems. By reducing the problem size it gives a possibility of using some optimization tools and methods. The system was implemented for training phase of SOREE in Poland.

Referat recenzowany

AUTOMATYZACJA PASTERYZACJI I STERYLIZACJI KONSERW W AUTOKLAWACH POZIOMYCH

1. Wprowadzenie

Wydaje się, że warunkiem zachowania konkurencyjności wyrobów przemysłu spożywczego, w szczególności przetwórczego, w warunkach otwarcia rynku krajowego dla producentów z Unii Europejskiej, jak również w kontekście niezbędnej ekspansji poza rynek krajowy, jest nie tylko samo osiąganie wysokiej jakości przetworów, lecz również zdolność do wyczerpującego udokumentowania tej jakości i to w sposób nie budzący wątpliwości, a więc najlepiej taki, jaki przewidują wymagania norm serii ISO 9000. Pośród wspomnianych wymagań można wyróżnić takie, które odnoszą się do szeroko rozumianej sfery organizacyjnej oraz takie, które dotyczą działań stanowiących poszczególne części procesu wytwarzania. Te ostatnie obejmują m.in. takie operacje, jak sterowanie i nadzór nad pracą maszyn i urządzeń, rejestrowanie historii procesu oraz raportowanie produkcji.

Spełnienie tych wymagań nie jest praktycznie możliwe bez automatyzacji i komputeryzacji procesów wytwarzania. Aby funkcje dokumentowania procesu mogły być realizowane przez systemy komputerowe niezbędne jest jednak zintegrowanie środków produkcji, a więc maszyn, urządzeń, czujników pomiarowych itp. z systemem komputerowym, co sprowadza się do konieczności zautomatyzowania procesu wytwarzania.

W pracy przedstawiono automatyzację procesu pasteryzacji i sterylizacji konserw w naczyniach hermetycznych, mającą na celu sprostanie, przedstawionym powyżej, wymaganiom „procesowym” ujętym w normach serii ISO 9000.

2. Technologia pasteryzacji i sterylizacji żywności w autoklawach poziomych nieobrotowych

Jak wiadomo, konserwowanie żywności, mające na celu umożliwienie jej długotrwałego przechowywania, wymaga zabiegów zmierzających do jej utrwalenia. Utrwalanie to polega na eliminacji drobnoustrojów, które, gdyby nie zostały usunięte, mogłyby doprowadzić do szybkiego zepsucia się zakonserwowanej partii żywności. Proces utrwalania musi być przeprowadzony w taki sposób, aby można było zagwarantować bezpieczeństwo produktu dla konsumentów w całym okresie gwarancyjnym.

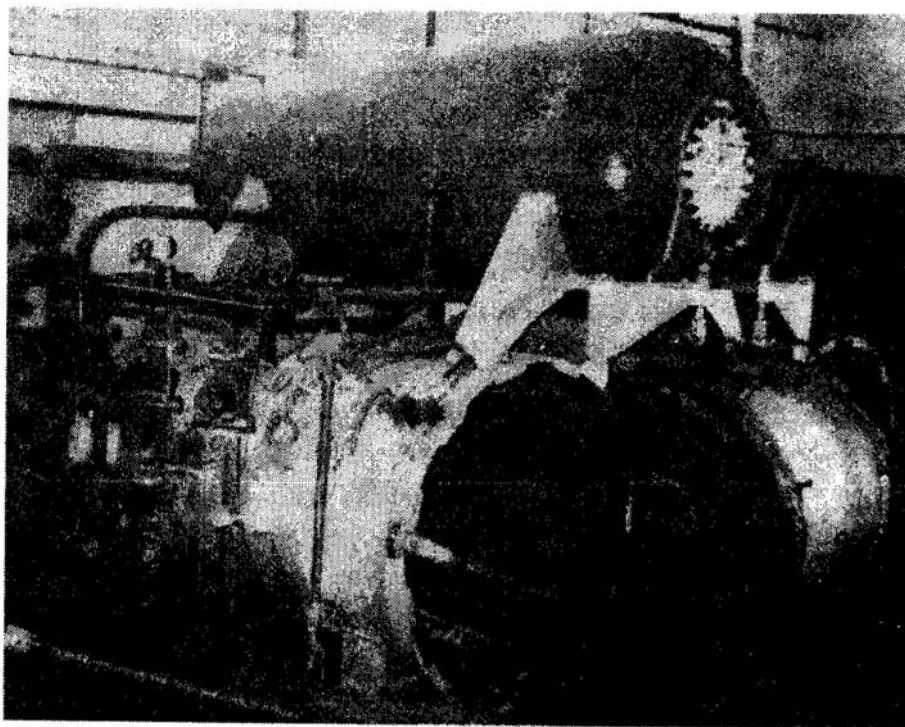
Podstawowym, i jednocześnie najszerzej stosowanym, sposobem utrwalania żywności, zwłaszcza przetworów mięsnych, drobiowych oraz owocowo-warzywnych, jest sterylizacja termiczna przeprowadzana w naczyniach hermetycznych, polegająca na podgrzaniu zanurzonych w wodzie konserw do odpowiedniej temperatury (w przypadku pasteryzacji nieco poniżej 100°C, a w przypadku sterylizacji w granicach 115–121°C)

* *Politechnika Poznańska, Katedra Automatyki, Robotyki i Informatyki, ul. Piotrowo 3A 60-965 Poznań,
E-Mail: :jw@ar-kari.put.poznan.pl.*

i przetrzymaniu ich w tej temperaturze przez określony czas, tak aby uzyskały one technologicznie właściwą dla nich wartość sterylizacyjną.

Najczęściej stosowanymi w Polsce urządzeniami do pasteryzacji bądź sterylizacji konserw są autoklawy poziome, nieobrotowe, takie jak przedstawiony na rysunku 1 autoklaw WAA-06, znajdujący się w zakładach przetwórstwa spożywczego w Pudliszkach. Autoklawy tego typu są obsługiwane w sposób ręczny – nieautomatyczny. Wyposażone są one tylko w termometry rtęciowe oraz manometry. W związku z powyższym proces sterylizacji lub pasteryzacji przeprowadzany jest w oparciu o te dwa mierniki, których wskazania są wykorzystywane przez obsługę techniczną w trakcie nadzorowania i obsługi wspomnianych procesów.

Przeprowadzanie procesu sterylizacji w sposób nieautomatyczny nie gwarantuje otrzymania prawidłowo utrwalonych konserw. Z licznych pomiarów i badań, dokonywanych w praktyce przemysłowej wynika, iż prowadzenie obróbki termicznej przetworów w taki sposób nie uwzględnia szeregu czynników, które mogą wpłynąć na niejednakowe, niepełne, a więc niedostateczne utrwalenie lub zbyt mocne ogrzanie i niepotrzebne, nadmierne wyjałowienie konserw. Trudno też w tych warunkach zapewnić pełne udokumentowanie danych procesowych, niezbędnych dla nadzoru zachowania jakości.



Rys. 1. Autoklaw WAA-06

Fig. 1. The WAA-06 autoclave

Ponieważ większość używanych autoklawów nie jest autoklawami obrotowymi, w ich wnętrzu, w trakcie podgrzewania wody, występuje nierównomierny przestrzenny rozkład temperatur, wynikający z różnego oddalenia od podgrzewacza: różnice pomiędzy

temperaturą górnej i dolnej warstwy wody sięgają 5°C. Ponadto, do nierównomiernego ogrzania poszczególnych partii przetworów może przyczyniać się:

- niejednakowa temperatura początkowa wody użytej w procesie sterylizacji, co jest zrozumiałe, wzięwszy pod uwagę, że niezbędna objętość wody wynosi 3000 l.;
- niejednakowa temperatura początkową poszczególnych partii przetworów; zauważmy, że zróżnicowanie tej temperatury w kolejnych cyklach pasteryzacyjnych lub sterylizacyjnych może sięgać nawet 30°C, co wynika, m.in. ze zróżnicowanego składu surowcowego tych samych przetworów, a co za tym idzie z ich zróżnicowanej pojemności cieplnej;
- zakłócenia w dostawie energii cieplnej do autoklawu i brak informacji, jak dalece wpływają one na stopień dogrzania konserw.

Innymi przyczynami nieudanego procesu mogą być też:

- zróżnicowanie w zakażeniu początkowym surowca;
- subiektywne prowadzenie procesów ogrzewania i chłodzenia; niewłaściwy proces chłodzenia może być przyczyną powstawania bombażu technicznego, prowadzącego do odkształceń wieczek słoików i całych puszek, w wyniku czego może nastąpić nadwężenie zakładek puszek – jeśli zakładka zostanie uszkodzona to produkt nie nadaje się już do sprzedaży;
- nierównomierny stopień schłodzenia konserw w poszczególnych częściach autoklawu lub pasteryzatora.

Wszystkie te czynniki mogą powodować określone straty finansowe, surowcowe, energetyczne, straty wody pitnej, a także mogą nieść zagrożenia dla zdrowia konsumentów, jeśliby budząca zastrzeżenia partia przetworów znalazła się na rynku. Ażeby zredukować prawdopodobieństwo tego ostatniego, w obliczu ograniczonej możliwości zagwarantowania 100%-owej pewności poprawności procesów sterylizacji bądź pasteryzacji, przeprowadzone są bardzo kosztowne, pracochłonne i długotrwałe oceny trwałościowe, bakteriologiczne, oraz fizyko-chemiczne zakonserwowanej żywności. Nadmienmy, że kontrola skuteczności utrwalenia dokonywana jest poprzez przeprowadzenie tzw. testu inkubacyjnego. Polega on na przetrzymaniu utrwalonej partii żywności przez 7 dni w temperaturze 37° C i sprawdzeniu czy w okresie tym nie rozwinęły się drobnoustroje. Kontrole tego typu nie wystarczają jednak do uzyskania certyfikatu jakości ISO9000. Potrzebna jest do tego automatyzacja procesu, która zapewni dokładne spełnienie warunków technologicznych oraz pozwoli na rejestrowanie parametrów procesu, a tym samym na ich dostępność w każdej chwili procesu.

3. Automatyzacja procesu

Dokładne spełnienie wymagań technologicznych procesu sterylizacji lub pasteryzacji możliwe jest tylko w oparciu o sterowanie automatyczne. Zauważmy, że proces taki musi być prowadzony z zachowaniem określonego programu zmian temperatury oraz ciśnienia, co jest bardzo trudne do wykonania w warunkach sterowania ręcznego, gdyż trzeba uwzględnić tutaj cały szereg warunków mających wpływ na proces, które zostały omówione w poprzednim punkcie. Wymaga to od personelu obsługującego proces dużego doświadczenia i praktyki.

Automatyzację procesu postanowiliśmy oprzeć o wykorzystanie sterownika swobodnie programowalnego PLC *Modicon Micro 612 00* oraz systemu oprogramowania typu

MMI/SCADA (ang. Man–Machine Interface/Supervisory Control and Data Acquisition) o nazwie *InTouch* z pakietu *Factory Suite 2000* firmy *Wonderware*, służącego do tworzenia aplikacji sterowania nadrzędnego i akwizycji danych z graficznym interfejsem operatora.

Wykorzystanie sterownika PLC umożliwia zastosowanie regulacji programowej, zgodnie z wymogami technologii: Bardzo ważnym wymaganiem technologicznym automatyzowanego procesu jest stopniowe i równomierne podnoszenie temperatury wody w zbiorniku, zanim osiągnie ona temperaturę procesu właściwego. Czas podgrzewania jest określony przez technologa i nie powinien być ani krótszy ani dłuższy. Jednocześnie należy zagwarantować równomierny wzrost temperatury oraz stałe ciśnienie, również określone przez technologa. Ażeby zdać sobie sprawę, jak trudne jest to zadanie w warunkach sterowania ręcznego zwróćmy uwagę, że temperatura początkowa wody jest na ogół różna, różne są pojemności cieplne sterylizowanych przetworów oraz ich temperatura początkowa, ciśnienie w zbiorniku jest zależne od temperatury, wreszcie mogą występować zakłócenia w dopływie energii cieplnej do autoklawu.

Po zalaniu zbiornika i rozpoczęciu fazy podgrzewania procesu sterownik będzie wyliczał wartość zadaną temperatury według następującej zależności:

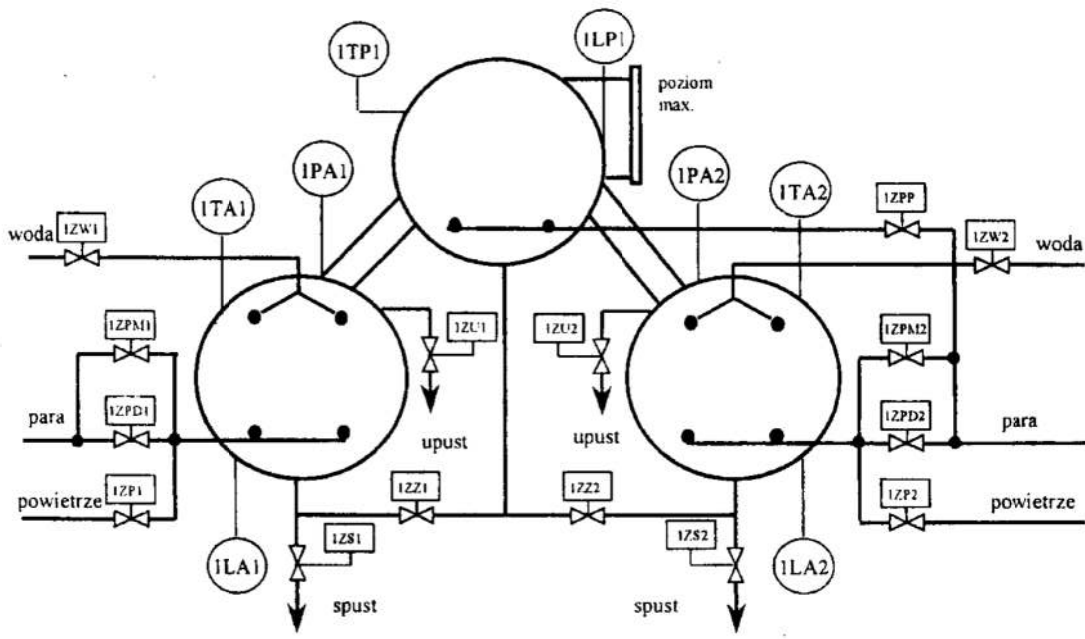
$$T_z(t) = T_0 + \frac{T_p - T_0}{t_p} t \quad (1)$$

gdzie: $T_z(t)$ – wartość zadana temperatury wody w zbiorniku w chwili t ;
 T_0 – wartość początkowa temperatury wody po zalaniu zbiornika;
 T_p – temperatura wody określona dla procesu właściwego;
 t_p – wymagany czas podgrzewania do temperatury procesu właściwego;
 t – czas (bieżąca chwila czasu).

W każdej chwili próbkowania sterownik porównuje temperaturę wody w zbiorniku z wartością wyliczoną według zależności (1). W przypadku gdy rzeczywista temperatura jest mniejsza od temperatury zadanej sterownik otwiera zawór parowy mały 1ZPM1 (porównaj schemat z rysunku 2), dopuszczając parę do autoklawu, co powoduje wzrost temperatury. Jeśli w ciągu 10 sekund od tej operacji temperatura nie przekroczy temperatury zadanej, zostaje otwarty zawór parowy duży (1ZPD1). Zawory parowe mały i duży zostają otwarte do momentu osiągnięcia przez autoklaw temperatury zadanej.

Przez cały czas podgrzewania ciśnienie jest utrzymywane na stałym poziomie i jest stabilizowane przez dwa zawory. W przypadku kiedy ciśnienie spada o 2 kPa poniżej wartości zadanej otwiera się zawór powietrza 1ZP1, natomiast gdy ciśnienie przekroczy wartość zadaną, sterownik otwiera zawór upustowy 1ZU1. Warunkiem zakończenia podgrzewania jest osiągnięcie zadanej temperatury i upłynięcie czasu przeznaczanego na grzanie.

Po osiągnięciu temperatury T_p w zadanym czasie t_p , rozpoczyna się etap procesu właściwego (sterylizacji lub pasteryzacji). W tej fazie procesu temperatura utrzymywana jest na stałym poziomie T_p (temperatura zadana procesu właściwego) poprzez dopuszczanie pary do autoklawu. Ciśnienie w trakcie procesu właściwego utrzymywane jest na poziomie wartości zadanej, identycznie jak podczas podgrzewania.



Rys. 2. Schemat technologiczny autoklawu WAA-06
 Fig. 2. The technological scheme of the WAA-06 autoclave

Etap ten kończy się w momencie kiedy upływa czas zadany procesowi właściwego. Po zakończeniu fazy procesu właściwego następuje faza chłodzenia. Podobnie, jak w fazie podgrzewania, temperatura musi zmieniać się w tym etapie w sposób równomierny przez czas określony przez technologa. W związku z powyższym w etapie chłodzenia temperatura zadana wyliczana jest w sterowniku w oparciu o następującą zależność:

$$T_z(t) = T_p - \frac{T_p - T_k}{t_{ch}} t \quad (2)$$

gdzie: T_k – zadana temperatura końcowa po zakończeniu procesu chłodzenia;
 t_{ch} – czas chłodzenia określony technologią procesu;
 $T_z(t)$, T_p oraz t – jak we wzorze (1).

Gdy temperatura wody w autoklawie jest wyższa od wartości zadanej, sterownik wystawia otwarcie zaworu wody zimnej IZWZ1. Woda gorąca z autoklawu upuszczana jest do podgrzewacza pod warunkiem, że poziom wody w podgrzewaczu jest niższy niż poziom elektrody sygnalizującej poziom maksymalny. W przeciwnym razie woda upuszczana jest do kanału. Spust wody do kanału oraz do podgrzewacza uzależniony jest od poziomu wody w autoklawie. Gdy poziom wody spadnie poniżej wartości zadanej sterownik zamyka zawory spustowe.

Wartość zadana ciśnienia w okresie chłodzenia przyjmuje cztery wartości. W momencie rozpoczęcia chłodzenia zostaje podniesiona o 20% w stosunku do wartości zadanej ciśnienia w procesie właściwym i w podgrzewaniu. Następnie, po każdym z czterech równych odstępach czasu, jest obniżana o wartość progową równą 20%. Utrzymywanie ciśnienia odbywa się identycznie jak w fazie podgrzewania oraz procesie właściwym: sterowane są odpowiednio zawory upustowe bądź dopuszczający sprężone powietrze. Warunkiem zakończenia procesu właściwego jest uzyskanie w autoklawie odpowied-

niej temperatury (zadanej w programie) oraz upływanie zadanego czasu chłodzenia. Po etapie chłodzenia następuje spust końcowy trwający 5 min. W etapie tym otwarte są zawór spustowy wody do kanału oraz zawór upustowy. Po zakończeniu czasu spustu końcowego następuje zakończenie procesu. Przykładowe przebiegi zmian temperatury oraz ciśnienia w trakcie trwania procesu zostały przedstawione na Rys. 4, na którym można łatwo zauważyć, że układ sterowania dokładnie odtwarza programy zadane zmienności temperatury i ciśnienia w autoklawie.

4. Wizualizacja procesu

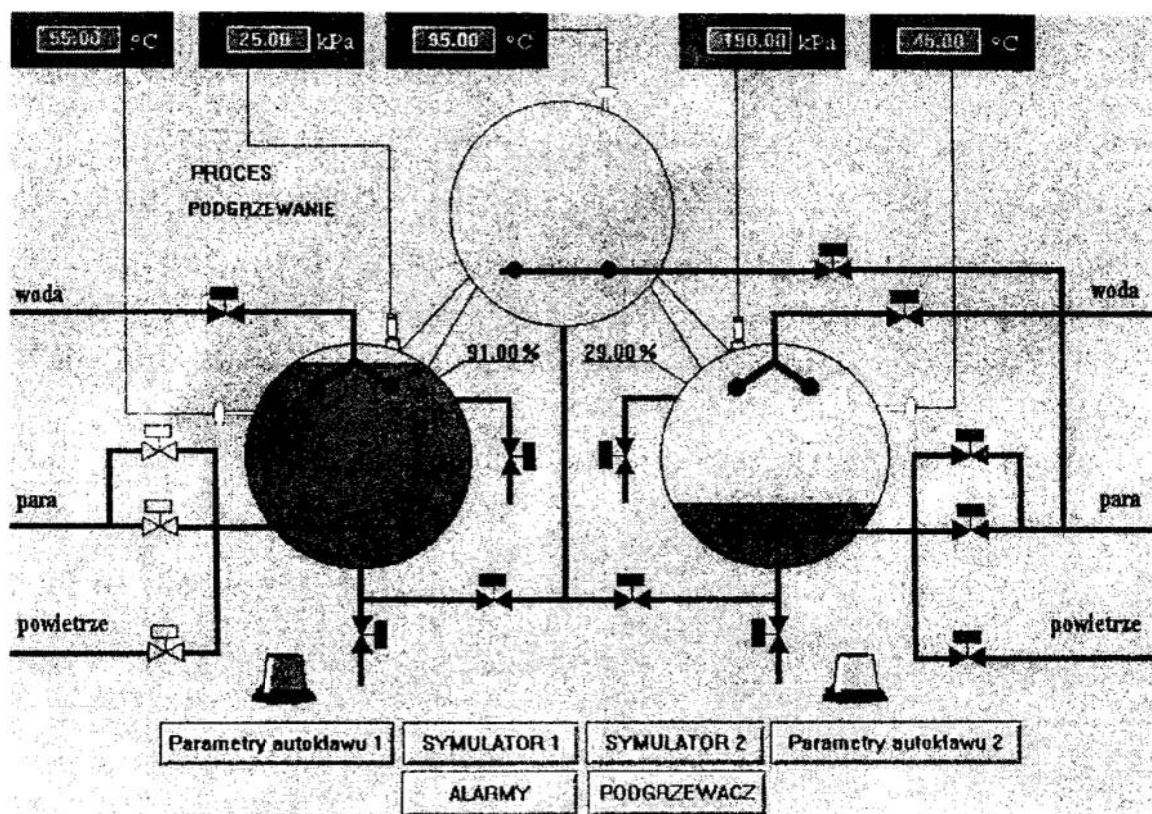
Jak zaznaczono w poprzednim punkcie, do wizualizacji procesu użyto oprogramowania *InTouch*, działającego pod kontrolą systemu operacyjnego *Windows'95*. Wspomniany system *InTouch* firmy *Wonderware* jest w pełni 32-bitowym, obiektowo zorientowanym, zintegrowanym środowiskiem programowym do tworzenia i użytkowania aplikacji sterowania nadrzędnego i akwizycji danych obejmującym:

- szybką i niezawodną komunikację ze sprzętem automatyki, w tym ze sterownikami PLC, co jest możliwe dzięki obszernej bibliotece driverów z wbudowanymi funkcjami diagnostyki komunikacji;
- zbieranie danych oraz wizualizowanie procesu w czasie rzeczywistym; funkcje wizualizacyjne oparte są o wykorzystanie kolorowej, animowanej grafiki;
- bezpośrednią obserwację i analizę trendów bieżących i historycznych przebiegu procesu oraz obsługę receptur, wartości zadanych i parametrów procesu, w tym również (np. w przypadku dużych obiektów) z centralnej sterowni;
- generowanie komunikatów informacyjnych i alarmowych, w tym sygnalizację i rejestrację alarmów;
- rejestrację zmiennych procesowych i generowanie raportów, w tym wykonywanie obliczeń i analiz wraz z możliwością wymiany danych z arkuszami kalkulacyjnymi i relacyjnymi bazami danych, jak również współpracę z systemami bilansowania i rozliczania produkcji.

Tworzenie aplikacji o pożądanej funkcjonalności polega w systemie *InTouch* na osadzeniu wymaganych obiektów na poszczególnych warstwach aplikacji. Obiekty te zapewniają dostęp do danych procesowych, tworzą graficzny interfejs operatora i definiują wszystkie niezbędne procesy systemu.

Aplikacja wizualizacyjna stworzona dla wizualizacji rozpatrywanego gniazda autoklawowego składa się z siedmiu następujących tablic synoptycznych:

- główna tablica synoptyczna, prezentująca gniazdo autoklawowe (została przedstawiona rysunku. 3.);
- tablice parametrów autoklawu 1 oraz parametrów autoklawu 2;
- wykresy temperatury odpowiednio autoklawu 1 i autoklawu 2 oraz
- wykresy ciśnienia odpowiednio autoklawu 1 i autoklawu 2;
- podgrzewacz;
- alarmy;
- symulatory autoklawów 1 i 2 – symulatory te służą do testowania programu sterownika w trybie off-line, zanim zostanie on uruchomiony na obiekcie w trybie on-line.



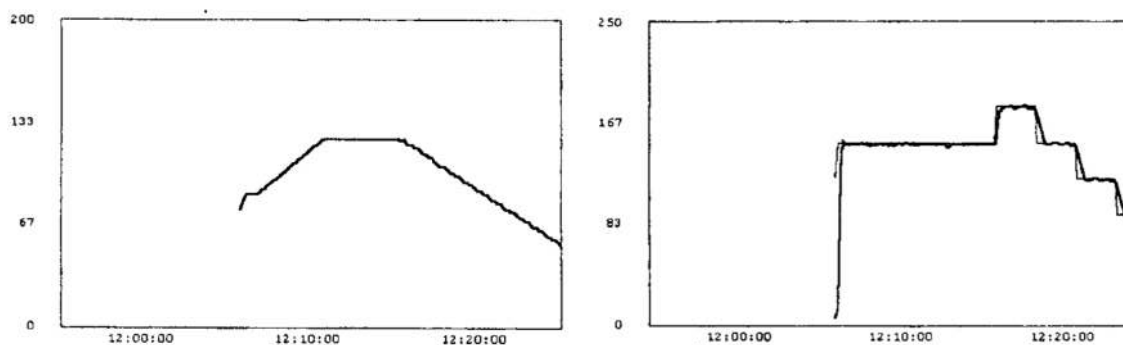
Rys. 3. Ekran synoptyczny gniazda autoklawów
 Fig. 3. The synoptic screen of the autoclave cell

Tablica główna (rys. 3.) uruchamiana jest automatycznie po uruchomieniu programu. Tablica ta umożliwia wejście do sześciu pozostałych ekranów.

Na głównym ekranie synoptycznym możemy obserwować schemat technologiczny gniazda autoklawów ze stanami zaworów autoklawów pierwszego i drugiego. Zawór otwarty sygnalizowany jest kolorem zielonym natomiast zawór zamknięty ma kolor czerwony. Również kolory rurociągów zmieniają się, np. rurociąg parowy w momencie gdy przepływa przez niego para ma kolor czerwony, natomiast gdy zawór parowy jest zamknięty ma kolor czarny. Poziom wody w autoklawie sygnalizowany jest poprzez wypełnienie konturu autoklawu barwą, oraz w sposób cyfrowy, w górnej części autoklawu, gdzie wyświetlana jest jego wartość. W górnej części okna znajdują się również wskaźniki temperatury i ciśnienia w obu autoklawach oraz w podgrzewaczu. Kliknięcie myszą na wskaźnik powoduje wyświetlenie wykresu wskazanej wielkości. Przykłady wykresów ciśnienia i temperatury przedstawione są na rysunku 4. Wykresy te są archiwizowane, co pozwala w każdej chwili na sprawdzenie warunków, w jakich dokonywana była sterylizacja danej partii przetworów. Dotyczy to również wszelkich zdarzeń oraz sytuacji alarmowych.

W dolnej części ekranu głównego znajdują się przyciski, który otwierają okna symulatorów, pozwalających na testowanie programu sterownika w trybie off-line.

Pogram wizualizacyjny komunikuje się ze sterownikiem Modicon Mikro poprzez łącze RS-232 za pomocą protokołu komunikacyjnego MODBUS.



Rys. 4. Przebiegi temperatury w $^{\circ}\text{C}$ (po lewej) i ciśnienia w kPa (po prawej) w czasie
 Fig. 4. The time trends of the temperature in $^{\circ}\text{C}$ (left) and pressure in kPa (right)

5. Wnioski końcowe

Przedstawiony sposób automatyzacji procesów sterylizacji żywności daje możliwość dokładnego spełnienia wymaganych warunków technologicznych. Ponadto zastosowanie systemu MMI/SCADA umożliwia pełne monitorowanie i archiwizowanie przebiegu procesu dla każdej partii sterylizowanych przetworów, pozwalając na pełne odtworzenie tego przebiegu, a tym samym pełną identyfikowalność wyrobów. Obydwa te czynniki stanowią podstawę spełnienia wymagań norm jakościowych serii ISO 9000.

Literatura

1. Janiszewski M.: *Sterowanie procesem pasteryzacji i sterylizacji żywności w autoklawach poziomych*. Praca dyplomowa (niepublikowana) 1999.
2. Warczyński J.: *Zautomatyzowane systemy wytwarzania*. Raport z badań (praca niepublikowana), 1998.

Streszczenie

W pracy przedstawiono sposób automatyzacji procesu pasteryzacji i sterylizacji przetworów spożywczych w autoklawach poziomych nieobrotowych, które są najczęściej używane w Polsce. Automatyzacja ta oparta jest o wykorzystanie sterownika PLC oraz systemu zintegrowanego oprogramowania typu MMI/SCADA. Przedstawiony sposób automatyzacji gniazda autoklawowego pozwala na spełnienie wymagań procesowych norm jakościowych serii ISO 9000.

ON AUTOMATING THE PASTEURIZATION AND STERILIZATION PROCESSES IN HORIZONTAL AUTOCLAVES

Summary

The paper describes an automating solution for the processes of pasteurization and sterilization of the canned food in horizontal not rotating autoclaves which are commonly used in Poland. For this purpose a PLC controller and a MMI/SCADA system of integrated programs has been used. The presented automation solution for the concerned autoclave cell allows to comply with process demands of the quality norms of the ISO 9000 series.

Referat recenzowany

DYCHOTOMICZNA KLASYFIKACJA KREDYTOBIORCÓW ZA POMOCĄ WYBRANYCH METOD KLASYFIKACJI

1 Wprowadzenie

Każda działalność gospodarcza obarczona jest pewnym ryzykiem rozumianym jako prawdopodobieństwo wystąpienia negatywnego lub chociażby niezgodnego z oczekiwaniami zjawiska (sytuacji). Dążenie do minimalizacji ryzyka powoduje, że przed podjęciem decyzji przeprowadza się różnego rodzaju analizy, których celem jest zarówno ocena bieżącej sytuacji, jak i przewidywanie przyszłych zdarzeń, które będą miały istotny wpływ na spodziewane efekty podejmowanych działań gospodarczych.

Szybki rozwój rynków finansowych i transformacja polskiej gospodarki spowodowały, że problematyka ta nie jest już tylko domeną wąskiego grona specjalistów (menedżerów, bankowców, finansistów, czy księgowych), ale dotyczy praktycznie każdego.

Wzrost znaczenia pojęcia ryzyka w finansach powoduje stosowanie nowoczesnych, często bardzo zaawansowanych metod ilościowych. Należą do nich statystyczne metody klasyfikacji (tj. metody taksonomiczne i analiza dyskryminacyjna), metody prognozowania (modele ekonometryczne i szeregi czasowych) oraz sztuczne sieci neuronowe.

Celem niniejszego opracowania jest przedstawienie wyników eksperymentów przeprowadzonych za pomocą wybranych metod statystycznych i sztucznych sieci neuronowych, których celem jest ocena wiarygodności kredytowej przedsiębiorstw, mierzonych decyzjami dotyczącymi przyznania (lub nie) kredytu.

Przeprowadzone badania nie są reprezentatywne, bowiem z uwagi na problemy z pozyskaniem danych w badaniach zastosowano próbę wybraną metodą łatwości dostępu¹.

Wyniki przeprowadzonych badań zostały porównane z rzeczywistymi decyzjami kredytowymi, podjętymi przez ekspertów bankowych ocen zdolności kredytowej. Oznacza to, że w pracy przyjęto założenie, iż decyzje podjęte przez bank są decyzjami poprawnymi-wzorcowymi. Przyjęcie tego założenia spowodowane było brakiem informacji na temat spłat kredytów i nie zostało zweryfikowane empirycznie. Stąd błędy klasyfikacji wyznaczone zostały w stosunku do decyzji ekspertów kredytowych, a nie odzwierciedlają stanu faktycznego².

2 Charakterystyka danych

W badaniach wykorzystano dane pochodzące z wniosków kredytowych 115 przedsiębiorstw ubiegających się o przyznanie kredytu w dwóch bankach łódzkich (75 pierwszych z jednego banku i 40 z drugiego). Na podstawie oceny przeprowadzonej przez inspektorów kredytowych wszystkie rozpatrywane wnioski kredytowe podzielono na dwie klasy tj.

* Zakład Metod Ilościowych Instytutu Zarządzania Politechniki Łódzkiej, dwitkowska.pk1.oiz@oiz.p.lodz.pl.

** Zakład Metod Ilościowych Instytutu Zarządzania Politechniki Łódzkiej, istan@ck-sg.p.lodz.pl.

¹ Dane pochodziły z różnych filii tego samego banku (regionu Łodzi) i dotyczyły tego samego rodzaju kredytu.

² Znany jest fakt, iż część tych decyzji jest nietrafna. Banki posiadają pewien procent trudnych kredytów w Polskim sektorze bankowym było to w roku 1998 około 9,5% akcji kredytowej. Potwierdza to również analiza przeprowadzona w pracy D. Witkowska [1999].

takie, na podstawie których udzielono kredytu (89 firm) i takie, na podstawie których odmówiono przyznania kredytu (26 firm). Wykorzystując dane pochodzących z wniosków kredytowych skonstruowano mierniki opisujące każde z analizowanych przedsiębiorstw za pomocą dziewięciu charakterystyk liczbowych, którymi są:

- a) zysk brutto za okres sprawozdawczy,
- b) zysk netto za okres sprawozdawczy,
- c) średni miesięczny zysk netto,
- d) miesięczna rata spłaty wnioskowanego kredytu,
- e) wielkość maksymalnych miesięcznych odsetek od wnioskowanego kredytu,
- f) rata spłaty wraz z odsetkami,
- g) inne obciążenia miesięczne (np. spłata innych zobowiązań),
- h) pozostały średni zysk miesięczny po odliczeniu rat i odsetek,
- i) decyzja banku odnośnie do przyznania kredytu: udzielenie kredytu -1, odmowa przyznania kredytu - 0.

Wyniki klasyfikacji generowano w oparciu o 8-elementowy zbiór charakterystyk dla każdego z obiektów (przedsiębiorstw), ponieważ większego nie udało się autorkom wyznaczyć na podstawie posiadanej informacji.

3 Metody klasyfikacji

Zadanie oceny wypłacalności klienta instytucji finansowej można sformułować w postaci zadania klasyfikacji, przez które rozumie się określenie reguły decyzyjnej, umożliwiającej rozpoznanie klienta jako należącego do określonej klasy ryzyka kredytowego. Zadania klasyfikacji można rozwiązać za pomocą wielu metod, z których dla potrzeb naszych badań wybraliśmy:

- metody bezwzorcowe, takie jak: metoda Warda, metodę k -średnich, sieć Kohonena;
- metody wzorcowe, tj. perceptron wielowarstwowy.

Metoda Warda należy do metod aglomeracyjnych w których wstępnym założeniem jest przypuszczenie, że każdy obiekt badania O_i tworzy początkowo jedną klasę K_i , gdzie $i=1,2,\dots,n$.

Algorytm tej procedury klasyfikacji jest przedstawiony poniżej.

1. Tworzy się n -skupień, czyli każdy obiekt badania stanowi jedną klasę.
2. W macierzy odległości szuka się pary klas najbardziej podobnych (najmniej odległych od siebie). Założmy, że będą to klasy K_i oraz K_j .
3. Redukuje się liczbę klas o jeden, łącząc klasy K_i oraz K_j w nową klasę.
4. Przekształca się odległości stosownie do metody między połączonymi klasami K_i i K_j oraz pozostałymi klasami.
5. Powtarza się kroki 1–3 do chwili, gdy wszystkie obiekty znajdą się w jednej klasie.

W przypadku metod aglomeracyjnych gdy liczba klas jest znana i wynosi l , należy proces grupowania przerwać po $n-l$ etapach³.

Idea metody k średnich⁴ została stworzona przez T. Daleniusa, który przedstawił iteracyjną procedurę podziału populacji na l grup, tak aby zminimalizować wielkość wewnątrzgrupowej wariancji.

³ *Metody te znalazły szerokie odzwierciedlenie w wielu pakietach statystycznych m.in. w Statistice (porównaj Statistica™ PL [1997], s. 3171).*

⁴ *Z reguły przyjmuje się, że liczba grup w populacji jest równa k i stąd nazwa tej metody. Bliższe szczegóły dotyczące metody k średnich można znaleźć w pracy C. F. Banfielda i L. C. Bassila [1977].*

Sztuczne sieci neuronowe są pojęciem ogólnym obejmującym różne rodzaje sieci trenowane⁵ przy użyciu różnych algorytmów. Sieci neuronowe pełnią różnorodne funkcje i mają rozmaite formy. Idea sieci neuronowych jest bardzo uniwersalna i można ją wykorzystywać do rozważań różnego typu problemów, ale oczywiście różne wtedy będzie podejście do rozwiązywania zadania.

Literatura sieci neuronowych rozróżnia pomiędzy nadzorowanym i nienadzorowanym uczeniem. W nadzorowanym uczeniu, celem jest przewidzenie jednej lub więcej zmiennych wyjściowych spośród jednej lub więcej zmiennych wejściowych. Nadzór polega na wprowadzeniu wartości wzorcowych w treningu. Nadzorowane uczenie porównuje się zazwyczaj do funkcji regresji lub analizy dyskryminacyjnej.

Najczęściej stosowanym typem sztucznej sieci neuronowej jest *perceptron wielowarstwowy* poddany procesowi nauczania nadzorowanego przy pomocy algorytmu propagacji wstecznej. Topologia powierzchni funkcji błędu perceptronu jest multimodalna i nie została najlepiej poznana. Proces uczenia często „utyka” w minimum lokalnym i wobec braku zmian wartości błędu, jest przerywany. Algorytm propagacji wstecznej i inne stosowane algorytmy uczenia sieci neuronowych są oparte na klasycznych metodach optymalizacji⁶. Metody te bazują na zasadzie twardej selekcji, zakładającej tworzenie nowych punktów bazowych do dalszych poszukiwań wyłącznie na bazie najlepszych z dotychczasowych.

Samooorganizująca się sieć Kohonen została opracowana w latach 1979–1982 i wykorzystuje proces nienadzorowanego uczenia. Struktura sieci jest prosta. Sieć składa się z dwu warstw – jednowymiarowej warstwy wejściowej, oraz z dwuwymiarowej (zorganizowanej jako siatka neuronów) tzw. warstwy współzawodnictwa (ang. *competitive layer*). Każdy neuron warstwy wejściowej jest połączony z każdym neuronem warstwy wyjściowej.

Na początku procesu uczenia wagi neuronów w warstwie Kohonena zostają losowo wybrane, przy czym długość wektora wag jest znormalizowana do jedynki. Wielkość sygnału wyjściowego zależy głównie od kąta pomiędzy wektorem wejściowym, a wektorem wag. Jeżeli kąt między wektorem wejściowym, a wektorem wag jest mały to wagi neuronu i jego sąsiadów podlegają uczeniu. Jeśli kąt między wektorem wejściowym, a wektorem wag jest duży to wagi neuronu są zerowane.

Istotą sieci Kohonena jest podział obserwacji na klasy reprezentowane przez poszczególne elementy przetwarzające. Wartości wag neuronów reprezentują uśrednione wektory wejściowe, które sieć przypisała do jednej klasy.

4 Mierniki jakości klasyfikacji

Ogólny błąd klasyfikacji zdefiniowany jest w postaci:

$$E = \frac{N}{n} = \frac{\sum_{s=1}^l N_s}{n} \quad (1)$$

⁵ Nie wszystkie sieci neuronowe muszą podlegać procesowi uczenia, przykładem może tu być sieć Hopfielda i sieci o radialnych funkcjach bazowych.

⁶ Porównaj prace: W. Findeisen, W. Szymanowski, A. Wierzbicki [1977].

gdzie:

N_s – jest liczbą błędnie zaklasyfikowanych obiektów do grupy typologicznej K_s ,
a więc rozpoznanych jako elementy klasy K_s , podczas gdy w rzeczywistości należą one do grupy K_q ($q=1,2,\dots,l, s \neq q$);

$N = \sum_{s=1}^l N_s$ oznacza liczbę wszystkich niepoprawnie zakwalifikowanych obiektów;

n – jest liczbą wszystkich obiektów objętych klasyfikacją.

Ogólny błąd klasyfikacji (1) określa jaka część rozpatrywanych w procedurze klasyfikacyjnej obiektów została niepoprawnie rozpoznana.

Dokonując klasyfikacji dychotomicznej wyróżnia się dodatkowo błędy pierwszego i drugiego rodzaju. Z błędem pierwszego rodzaju mamy do czynienia wtedy, gdy odrzucony przez bank klient zostanie zaklasyfikowany jako *wiarygodny*. Opierając się wówczas wyłącznie na takiej decyzji bank narażony zostałby na ryzyko nieterminowej spłaty udzielonego kredytu wraz z odsetkami.

$$E_1 = \frac{N_1}{Z} \times 100\% \quad (2)$$

gdzie:

N_1 – jest liczbą negatywnie rozpatrzonych przez bank wniosków kredytowych zaklasyfikowanych za pomocą metody grupowania jako wnioski, na podstawie których należy udzielić kredytu;

Z – jest liczbą wszystkich negatywnie rozpatrzonych przez bank wniosków kredytowych znajdujących się w zbiorze uczącym (testującym).

Błąd drugiego rodzaju powstaje wówczas, gdy zaakceptowany przez bank kredyt zostanie rozpoznany jako *niewiarygodny*. Nie przydzielenie kredytu, w przypadku posiadania zdolności kredytowej przez wnioskodawcę powoduje utratę korzyści dla banku z tytułu możliwych do uzyskania dochodów w postaci odsetek od kredytu.

$$E_2 = \frac{N_2}{D} \times 100\% \quad (3)$$

gdzie:

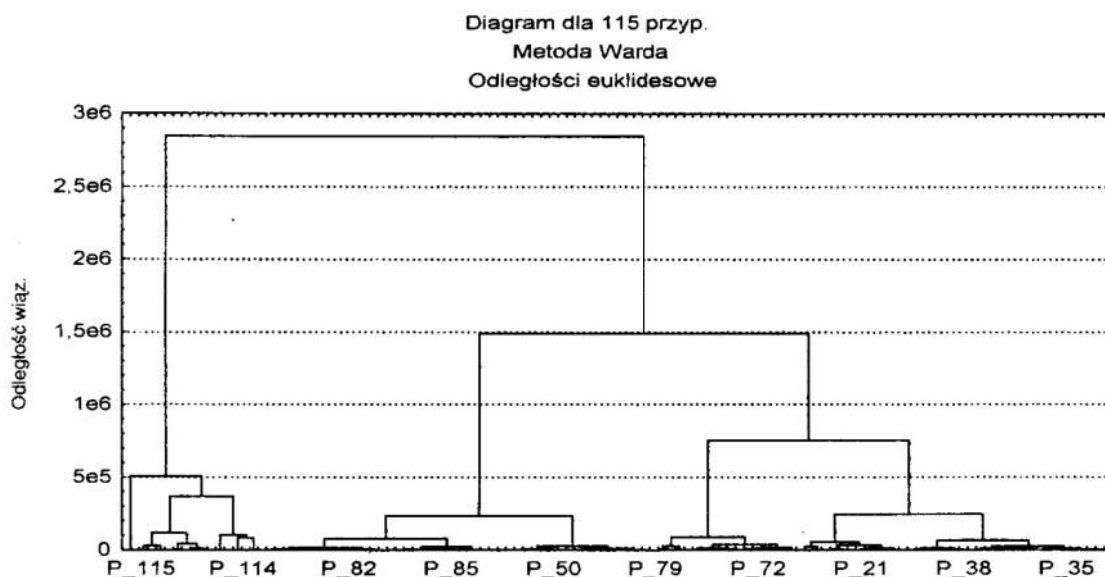
N_2 – jest liczbą pozytywnie rozpatrzonych przez bank wniosków kredytowych zaklasyfikowanych przez niego jako wnioski odrzucone;

D – jest liczbą wszystkich pozytywnie rozpatrzonych przez bank wniosków kredytowych w zbiorze uczącym (testującym).

5 Wyniki eksperymentów

W przypadku klasyfikacji bezwzorcowej następuje klasyfikacja do założonej liczby klas, które nie są jednak precyzyjnie zdefiniowane. Przyjęto założenie, że rozkład utworzonych grup odpowiada strukturze próby. Zatem bardziej liczna grupa odpowiada przedsiębiorstwom, którym przyznano kredyt, a mniej liczna tym, którym nie przyznano kredytu.

Wykorzystując metodę Warda do klasyfikacji określono jako miarę odległości pomiędzy skupieniami odległość Euklidesa. Wyniki klasyfikacji przedstawiono na rysunku 1.



Rys. 1. Wyniki klasyfikacji obiektów przy użyciu metody Warda

Fig. 1. Result classification object by use Ward method

W tabeli 1 przedstawiono numery przedsiębiorstw, które zostały zakwalifikowane przez metodę Warda do dwu klas. Oznaczenia w główce tabeli: klasa 1 obejmuje przedsiębiorstwa, którym można przyznać kredyt, a klasa 2, którym kredytu nie można przyznać. Natomiast oznaczenia w boczku tablicy oznaczają stan faktyczny.

Tablica 1. Obiekty zakwalifikowane do klas otrzymanych metodą Warda

Table 1. Object to classify by select class obtain Ward method

Rzeczywiste zakwalifikowanie	Obiekty skupienia 1 {klasa 1}	Obiekty skupienia 2 {klasa 2}
1	2 -5, 7 - 8, 11 - 16, 18, 19, 22, 25, 27, 31, 33 - 36, 38, 42, 44, 48, 49, 51, 53, 55, 57, 59 - 62, 64 - 66, 68, 70 -72, 74 -76, 78 -81, 82, 85 -91, 93 -97, 100, 101, 106, 108 - 111, 113	9, 20, 24, 26, 29, 39, 46, 56, 58, 63, 69, 77, 81, 83, 84, 99, 102, 104, 114, 115
0	1, 6, 10, 17; 21, 23, 28, 30, 32, 37, 40, 41, 43, 45, 47, 50, 52, 54, 67, 73, 92, 98, 103, 105, 107, 112	

Źródło: Obliczenia własne.

Klasyfikując metodą Warda otrzymaliśmy w pierwszej grupie 95 obiektów, a w drugiej 20. Z tego wynika, że ogólny błąd decyzji dla całej populacji tzn. próby treningowej i testowej wynosi $E = \frac{26 + 20}{115} = 40\%$. Błąd I-ego rodzaju wynosi

$$E_1 = \frac{26}{26} = 100\% . \text{ Błąd II-ego rodzaju wynosi } E_2 = \frac{20}{89} = 22,5\% .$$

W tabeli 2 przedstawiono wyniki klasyfikacji przy użyciu metody k-średnich do każdej z dwu grup. W klasie 1 znalazło się 99 obiektów, a w klasie 2 – 16.

Tablica 2. Obiekty zakwalifikowane do poszczególnych klas otrzymanych metodą k -średnich

Table 2. Object to classify by select class obtain k -means method

Rzeczywistość	Obiekty skupienia 1 {klasa 1}	Obiekty skupienia 2 {klasa 2}
1	3 -5, 7 - 8, 11 - 16, 18, 19, 22, 25 -27, 31, 33 - 36, 38, 42 , 44, 48, 49, 51, 53, 55 - 57, 59 - 66, 68, 70 - 72, 74 -76, 78 -82, 85 -91, 93 -97, 100 -102, 106, 108 - 111, 113	2, 9, 20, 24, 29, 39, 46, 58, 69, 77, 83, 84, 99, 104, 114, 115
0	1, 6, 10, 17; 21, 23, 28, 30, 32, 37, 40, 41, 43, 45, 47, 50, 52, 54, 67, 73, 92, 98, 103, 105, 107, 112	

Zródło: Obliczenia własne.

Klasyfikując obserwacje przy użyciu metody k -średnich tak grupowano obserwacje do poszczególnej grupy, aby zmaksymalizować odległość pomiędzy nimi. Ogólny błąd decyzji dla całej populacji tzn. próby treningowej i testowej w tym przypadku wynosi

$$E = \frac{16 + 26}{115} = 36,5\% . \text{ Błąd I-ego rodzaju wynosi } E_1 = \frac{26}{26} = 100\% , \text{ a błąd II-ego}$$

$$\text{rodzaju - } E_2 = \frac{16}{89} = 18\% .$$

W wielu pracach z zakresu sztucznych sieci neuronowych eksperymentalnie uzasadniono, iż najlepsze rezultaty otrzymuje się w przypadku perceptronu wielowarstwowego jeżeli funkcja aktywacji pomiędzy warstwą wejściową, a ukrytą jest funkcją sigmoidalną⁷.

W eksperymentach numerycznych wykorzystano różne topologie perceptronu wielowarstwowego. Dokonując klasyfikacji, zbiór wektorów charakteryzujących poszczególne przedsiębiorstwa ponumerowano za pomocą generatora liczb pseudolosowych, a cały zbiór podzielono na dwa rozłączne podzbiory tj. trenujący, zawierający 75 lub 85 elementów oraz zbiór testujący, w którym znalazło się 40 lub 30 firm. W próbie testowej znalazły się dane pochodzące z drugiego banku.

Wyniki klasyfikacji dychotomicznej przy użyciu różnych topologii perceptronu wielowarstwowego przedstawiono w tabeli 3.

Najniższy ogólny błąd decyzji dla całej populacji tzn. próby treningowej i testowej jaki udało się uzyskać dla perceptronu wielowarstwowego wynosi $E = \frac{3}{115} = 2,6\% .$

Dla niego błąd I-ego rodzaju wynosi $E_1 = \frac{0}{26} = 0\% ,$ a błąd II-ego rodzaju -

$$E_2 = \frac{3}{89} = 3,4\% .$$

⁷ Porównaj Rehkugler H., Poddig Th. [1992], s. 419.

Tablica 3. Eksperymenty dla ośmiu neuronów wejściowych przy użyciu MLP (wartości błędów podano w %)

Table 3. Experiment for eight input neuron by use MLP

Topologia sieci	Liczba iteracji	Błędy decyzji			Błędy dla zb. trenującego			Błędy dla zb. testującego		
		E	E ₁	E ₂	E	E ₁	E ₂	E	E ₁	E ₂
8-3-1	10 tys.	4,4	0	5,6	1,33	0	1,82 (57)	10,0	0	11,76 (76, 82, 87, 96)
8-5-1	10 tys.	2,6	0	3,4	0	0	0	7,50	0	8,82 (76, 87, 96)
8-5-1	15 tys.	2,6	0	3,4	0	0	0	7,50	0	8,82 (76, 87, 96)
8-5-1	25 tys.	2,6	0	3,4	0	0	0	7,50	0	8,82 (76, 87, 96)
8-3-1	10 tys.	5,2	3,8	5,6	4,71	5,00 (43)	4,62 (57, 76, 82)	6,67	0	8,33 (87,96)

Źródło: Obliczenia własne.

Najwyższy ogólny błąd decyzji dla całej populacji tzn. próby treningowej i testowej uzyskano dla perceptronu wielowarstwowego o topologii 8-3-1 (trenowanego w oparciu o większą próbę treningową – włączono do niej dodatkowo 10 przedsiębiorstw ubiegających się o kredyt w drugim banku) i wynosi on $E = \frac{6}{115} = 5,2\%$. Dla niego błąd I-ego

rodzaju wynosi $E_1 = \frac{1}{26} = 3,8\%$, a błąd II-ego rodzaju – $E_2 = \frac{5}{89} = 5,6\%$.

Wyniki klasyfikacji przy użyciu sieci Kohonena przedstawiono w tabeli 4.

Tablica 4. Obiekty zakwalifikowane do klas otrzymanych przy użyciu sieci Kohonena

Table 4. Object to classify by obtain class Kohonen network

Rzeczywistość	Obiekty skupienia 1 {klasa 1}	Obiekty skupienia 2 {klasa 2}
1	3 -5, 7 -9, 11 -16, 18, -20, 22, 24 -27, 31, 33 -36, 38, 42, 44, 48, 49, 51, 53, 55 -66, 68 -72, 74 -76, 78 -91, 93 -97, 99 -102, 106, 108 -111, 113	2, 29, 39, 46, 77, 104, 114, 115
0	1, 6, 10, 17; 21, 23, 28, 30, 32, 37, 40, 41, 43, 45, 47, 50, 52, 54, 67, 73, 92, 98, 103, 105, 107, 112	

Źródło: Obliczenia własne.

Z wyników klasyfikacji wynika, że ogólny błąd decyzji dla całej populacji tzn. próby treningowej i testowej wynosi $E = \frac{26+8}{115} = 29,6\%$. Błąd I-ego rodzaju wynosi

$E_1 = \frac{26}{26} = 100\%$, a błąd II-ego rodzaju wynosi $E_2 = \frac{8}{89} = 9\%$.

6 Podsumowanie

Metody klasyfikacji bez wzorca takie jak metoda k-średnich, metoda Warda i sieć Kohonena dużo gorzej rozpoznają elementy obu klas niż metody klasyfikacji wzorcowej

takie jak perceptron wielowarstwowy. Jak widać z powyższych analiz empirycznych ogólny błąd decyzji może wynosić:

- dla perceptronu wielowarstwowego 2,6%,
- dla metody Warda 40%,
- dla metody k-średnich 36,5%,
- dla sieci Kohonena 29,6%.

Należy również zauważyć, że dla wszystkich metod bezwzorcowych błąd II-ego rodzaju wynosi 100%.

Wyniki analiz empirycznych przemawiają na niekorzyść metod bezwzorcowych. Jest to zgodnie z panującą w literaturze opinią, że metody bezwzorcowe powinny być wykorzystywane wyłącznie w przypadku klasyfikacji do więcej niż dwie klasy.

Literatura

1. Banfielda C. F. i Bassila L. C.: *a Transfer Algorithm for Non-hierarchical Classification Algorithm AS113*, „Applied Statistics”. 1977, t. 26.
2. Findeisen W., Szymanowski W., Wierzbicki A.: *Teoria i metody obliczeniowe optymalizacji*. PWN, Warszawa 1977.
3. Rehkugler H., Poddig Th.: *Neuronale Netze im Bankbetrieb*. Die Bank 7/92, s. 413–419.
4. *Statistica™ PL*: tom 3, StatSoft 1997.
5. Witkowska D.: *Sztuczne sieci neuronowe w analizach ekonomicznych*. KBN Łódź 1999.

Streszczenie

Celem referatu jest porównanie wyników dychotomicznej klasyfikacji wniosków kredytowych przy użyciu metod statystycznych i sztucznych sieci neuronowych. W zaprezentowanych eksperymentach numerycznych wykorzystano dane pochodzące z dwu różnych banków łódzkich. w analizach zastosowano metody takie jak: Warda, k-średnich, sieć Kohonena i perceptron wielowarstwowy. W wyniku klasyfikacji dychotomicznej otrzymano lepsze wyniki dla metod wzorcowych.

DICHOTOMOUS CLASSIFICATION OF BORROWERS USING SELECTED METHODS

Summary

In the paper we compare the results of borrower classification obtained by statistical methods and neural networks. In our experiments data from two regional banks regarding 115 enterprises were used. Among statistical methods we chose Ward and k-means methods. Feedforward neural networks were trained applying supervised training (back propagation algorithm) and unsupervised learning-Kohonen method. Our experiments show that in case of dichotomous classification pattern recognition methods gives much better results.

Referat recenzowany

ZASTOSOWANIE MATROIDÓW W WYBRANYCH ZAGADNIENIACH MODELOWANIA UKŁADÓW MECHANICZNYCH I OPTYMALIZACJI

1 Wstęp

Teoria matroidów jest dziedziną matematyki w której rozważa się zagadnienia z algebry, teorii grafów, teorii algorytmów i optymalizacji [3], [4], [5], [10], [11], [12]. Matroidy były stosowane w wielu zagadnieniach inżynierskich takich jak modelowanie, analiza i synteza układów elektrycznych [1], [2], [13], [14], [15], analiza układów mechanicznych [17], [18] oraz analizy sztywności struktur prętowych, kratownic [7], [8] i innych. Matroid jest zbiorem z określoną na nim strukturą algebraiczną np. niezależności. Istnieje kilka wzajemnie równoważnych definicji opartych na ideach elementów niezależnych, baz lub cykli. Istnieją matroidy ściśle związane z przestrzeniami liniowymi i grafami. Dzięki temu można przy ich pomocy wygodnie przechodzić od jednych struktur algebraicznych do drugich wykorzystując wiele intuicyjnie spodziewanych własności, które w sposób ścisły zostały sformułowane i udowodnione w wybranych monografiach.

Korzyścią z zastosowania matroidów jest: uzyskanie wzorów analitycznych, w których występują macierze grafu będącego modelem układu oraz algorytmizacja zagadnienia, a zatem i komputeryzacja. W literaturze w języku polskim podstawy teorii zostały omówione skrótowo tylko w kilku pracach [1], [10], [18] oraz w nowym wydaniu pracy Wilsona [19].

2 Wybrane problemy optymalizacyjne rozwiązane za pomocą algorytmów opartych na teorii matroidów

Poniżej sformułowanych będzie kilka znanych problemów w teorii optymalizacji, które rozwiązano za pomocą algorytmów opartych na teorii matroidów.

Drzewo grafu o minimalnej wadze

Rozważmy graf obciążony $X = ({}_1X, {}_2X, {}_3X, f)$ gdzie: ${}_1X$ – zbiór wierzchołków grafu X , ${}_2X$ – zbiór krawędzi, a ${}_3X$ relacja incydencji, $f : {}_2X \rightarrow R$; funkcja przyporządkowująca wagi krawędziom grafu [6,16]. Przyjmuje się, następujące oznaczenia: $n = |{}_1X|$, $m = |{}_2X|$, Drzewem grafu nazywamy jego podgraf spójny zawierający $n - 1$ krawędzi. Wagą drzewa nazywamy sumę wag wszystkich krawędzi dane drzewa.

Problem: znaleźć drzewo o minimalnej lub maksymalnej wadze.

Zbiór elementów macierzy

Rozważmy macierz symetryczną A ($k \times k$) o różnych elementach dodatnich, rzeczywistych, przy czym $a_{ii} = 0$.

Problem: znaleźć ciąg elementów macierzy, który każdy należy do innego wiersza i innej kolumny o minimalnej lub maksymalnej sumie.

* Politechnika Śląska w Gliwicach, Katedra Mechaniki Robotów i Maszyn, ul. Konarskiego 18a, 44-100 Gliwice, E-mail: wojnarowski@rmt7.mt.polsl.gliwice.pl

** Politechnika Łódzka Filia w Bielsku-Białej, Katedra Podstaw Budowy Maszyn, ul. Willowa 2, 43-316 Bielsko-B., E-mail: szawislak@aristo.pb.bielsko.pl

Rozwiązanie pierwszego podał Kruskal (1956), a drugiego Boruvka (1926) (za pracą [7]). Poniżej podane będą matroidowe ujęcia obu problemów.

3 Wprowadzenie do teorii matroidów

W literaturze wykazano [2], [7], że następujące definicje matroidu są równoważne. Zakładamy, że matroidy określone są na zbiorze E , który jest skończony i niepusty.

Definicja 1. Matroidem M nazywamy parę (E, J) , gdzie J jest niepustą rodziną podzbiorów zbioru E , nazywanych zbiorami niezależnymi, spełniającą warunki:

(J1) Zbiór pusty należy do J .

(J2) Dowolny podzbiór zbioru niezależnego jest zbiorem niezależnym.

(J3) Jeżeli I_1, I_2 są zbiorami niezależnymi ($I_1, I_2 \in J$) i zawierają odpowiednio k oraz $k+1$ elementów – to istnieje element e należący do I_2 oraz nie należący do I_1 taki, że $I_1 \cup \{e\}$ jest zbiorem niezależnym.

Przykład. Niech A będzie macierzą. Jako zbiór E przyjmujemy zbiór kolumn macierzy A traktowanych jako wektory. Określamy zbiór J jako rodzinę wszystkich liniowo niezależnych podzbiorów kolumn macierzy A . Ta para (E, J) spełnia aksjomaty definicji 1, więc jest matroidem.

Ogólnie jest tu wykorzystana własność przestrzeni liniowej, że dowolny podzbiór niezależnego liniowo podzbioru przestrzeni liniowej jest liniowo niezależny.

Definicja 2. Matroidem M nazywamy parę (E, r) , gdzie r jest funkcją rzędu (lub funkcją rangową), która przyjmuje wartości całkowite nieujemne, jest określona na rodzinie podzbiorów E oraz spełnia następujące warunki:

(R1) $r(\emptyset) = 0$.

(R2) Dla każdego $S \subseteq E$ oraz $e \notin S$, $r(S \cup \{e\}) = r(S) + k$, gdzie $k = 0$ lub $k = 1$.

(R3) Dla każdego $S \subseteq E$ oraz $e_1, e_2 \notin S$ jeżeli

$r(S \cup \{e_1\}) = r(S \cup \{e_2\}) = r(S)$, to $r(S \cup \{e_1, e_2\}) = r(S)$

Przykład. Rozważamy parę (E, J) jako matroid określony za pomocą definicji 1. Dla każdego podzbioru $A \subseteq E$ określamy funkcję rzędu w następujący sposób: $r(A) =$ moc największego zbioru niezależnego zawartego w zbiorze A . Para (E, r) spełnia aksjomaty (R1), (R2) oraz (R3) więc jest matroidem.

Definicja 3. Matroidem nazywamy parę (E, B) , gdzie B jest niepustą rodziną podzbiorów zbioru E , nazywanych bazami, która spełnia następujące warunki:

(B1) żadna baza nie zawiera w charakterze swojego podzbioru właściwego innej bazy;

(B2) jeśli B_1 oraz B_2 są bazami oraz $e \in B_1$, to istnieje element $f \in B_2$ posiadający własność, że $(B_1 - \{e\}) \cup \{f\}$ też jest bazą.

Przykład. Niech $X = ({}_1X, {}_2X, {}_3X)$ będzie dowolnym grafem skończonym. Określamy parę $({}_2X, X_0)$, gdzie X_0 jest zbiorem drzew napinających grafu X . Zbiór X_0 spełnia aksjomaty (B1) oraz (B2), więc para $({}_2X, X_0)$ jest matroidem.

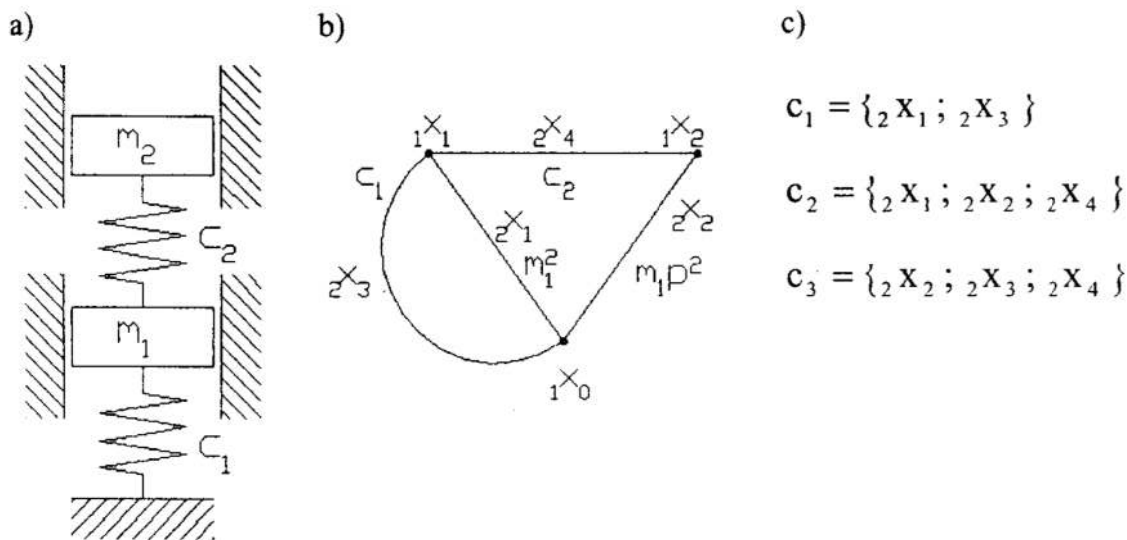
Przykład. Rozważmy E jako skończony podzbiór wektorów z przestrzeni wektorowej V . Jako elementy zbioru B przyjmujemy wszystkie możliwe liniowo niezależne podzbiory zbioru E generujące w przestrzeni V tę samą podprzestrzeń.

Taka para (E, B) tworzy matroid.

Definicja 4. Matroidem nazywamy parę (E, C) , gdzie jest rodziną podzbiorów zbioru E , zwanych cyklami, która spełnia poniższe aksjomaty:

- (C1) żaden element z rodziny C nie jest właściwym podzbiorem innego elementu,
- (C2) niech e_1 oraz e_2 będą różnymi elementami zbioru E , niech C_1, C_2 będą takimi elementami rodziny C , że $e_1 \in C_1 \cap C_2$ oraz $e_2 \in C_1 - C_2$, wtedy istnieje $C_3 \in C$ taki, że $e_2 \in C_3 \subseteq (C_1 \cup C_2) - \{e_1\}$.

Przykład. Rozpatrzmy graf X przedstawiony na rysunku 1. Przez $P(X)$ oznaczmy klasę cykli (konturów) grafu X . Para $({}_2X, P(X))$ jest matroidem.



Rys. 1. Graf X : (a) Schemat układu mechanicznego U , (b) Graf X przyporządkowany układowi U , (c) Rodzina cykli $P(X)$

Fig. 1. Graph X : (a) Scheme of mechanical system U , (b) Graph X assigned to system U , (c) Collection of cycles (polygons) $P(X)$

Definicja 5. Dwa zbiory S oraz T są ortogonalne jeśli $|S \cap T| \neq 1$.

Definicja 6. Niech $M = (E, C)$ będzie matroidem na skończonym zbiorze E oraz Q klasą podzbiorów E , które są ortogonalne do wszystkich elementów klasy C .

Matroid $M^* = (E, C^*)$ gdzie C^* jest klasą minimalnych elementów klasy Q , nazywany matroidem dualnym do matroidu M . Dowód faktu iż C^* spełnia warunki (C1) i (C2) w pracy [2].

Udowodniono następujące własności matroidów dualnych:

(W1) $(M^*)^* = M$,

(W2) Bazy matroidu M są dopełnieniami w zbiorze E baz matroidu M .

(W3) Niech $X = ({}_1X, {}_2X, {}_3X)$ będzie dowolnym grafem skończonym, wtedy

$({}_2X, P(X))^* = ({}_2X, B(X))$, czyli matroidem dualnym do matroidu

cyklicznego $({}_2X, P(X))$ /ang. polygon matroid/ jest matroid odcięć grafu.

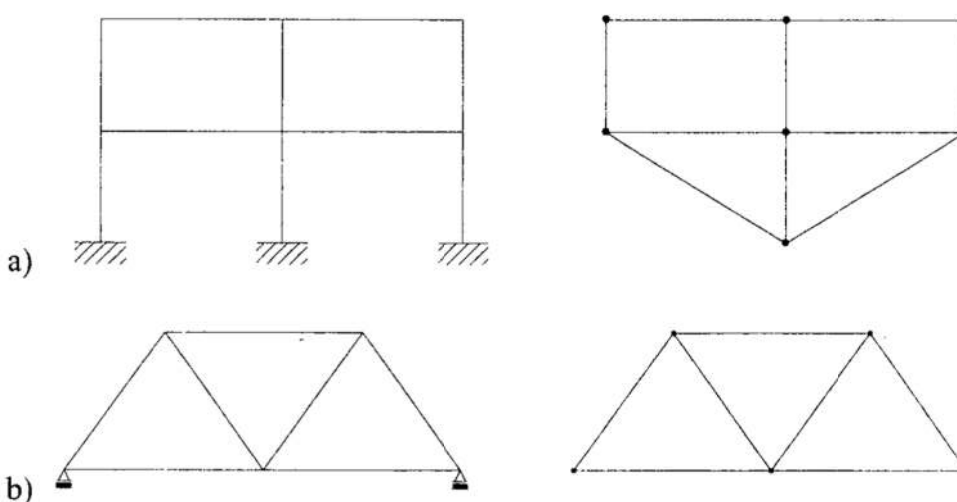
$({}_2X, B(X))$ /ang. bond matroid/.

Wyciągając wnioski z powyższych definicji i przykładów można stwierdzić, że pojęcie matroidu uogólnia pojęcia niezależności, bazy w przestrzeniach liniowych i cyklu oraz odcięcia w grafach. W nowy sposób pokazuje związki między tymi pojęciami. Wprowadza nowe pojęcie dualności w grafach – przeniesione z matroidów dzięki własności (W2). Nie jest ono związane z planarnością grafy, co poprzednio było ograniczeniem – bo nie wszystkie grafy mają graf dualny.

4 Matroidy jako modele układów mechanicznych

Matroidy można przyporządkowywać grafom, a grafami modeluje się (Rys. 1) dyskretne układy mechaniczne [16], [17], [18]. W dwu ostatnich wymienionych pracach przedstawiono sposoby analizy dyskretnych układów mechanicznych w oparciu o pojęcie uogólnionej sieci [2]. Uzyskano nową macierzową metodę analizy tychże układów.

Sposób przyporządkowania grafu strukturze prętowej [7], [8], [9] przedstawiono na rysunku 2.



Rys. 2. Sposób przyporządkowywania grafu strukturze prętowej: (a) Rama, (b) Kratownica
Fig. 2. Method of assigning of graph to a structure (a) Plane frame, (b) Plane truss

Zdefiniowano wielkość określającą stopień statycznej niewyznaczalności dla danej struktury prętowej następującym wzorem [8]:

$$\gamma(X) = a |E(X)| + b |N(X)| + c \gamma_0(X), \quad (1)$$

gdzie:

$|E(X)|$ – liczba krawędzi grafu X , $|N(X)|$ – liczba wierzchołków, a $\gamma_0(X)$ liczba składowych spójności, która zwykle wynosi 1. Wybrane wartości współczynników a , b , c podano w Tabeli 1.

Tabela 1. Wartości współczynników w $\gamma(X)$ dla wybranych struktur
 Table 1. $\gamma(X)$ coefficients for selected truss structures

Rodzaj	a	b	c
Ramy Płaskie	+3	-3	+3
Ramy Przestrzenne	+6	-6	+6
Kratownice Płaskie	+1	-2	+3
Kratownice Przestrzenne	+1	-3	+6

Publikacje w których przedstawiono metody analizy sztywności struktur prętowych za pomocą teorii matroidów omówiono w pracy [7].

Określa się matroid sztywności dla grafu w następujący sposób. Niech X będzie grafem, definiujemy podstawę $N(F_i)$ (ang. support) podzbioru $F_i \subset E(X)$, czyli krawędzi grafu X ; – jako zbiór wierzchołków krawędzi należących do F_i .

Definicja 7. Mówimy, że podzbiór F jest niezależny w $E(X)$ jeśli warunek:

$$|M(F_i)| \leq 2 |N(F_i)| - 3 \quad (2)$$

spełniony jest dla wszystkich podzbiorów F_i zbioru F . Te zbiory niezależne tworzą zbiór F , a zatem para $(E(X), F)$ tworzy matroid oznaczany przez $R(X)$, a nazywany matroidem sztywności.

Definicja 8. Graf X nazywamy sztywnym jeśli

$$r(E(X)) = 2 |N(X)| - 3. \quad (3)$$

Cykl w matroidzie $R(X)$ jest minimalnym podzbiorem zależnym w $E(X)$.

Zatem za pomocą matroidów można sprawdzać sztywność struktur prętowych – szersze rozważania z tego zakresu będą przedmiotem kolejnych publikacji autorów między innymi do czego służy algorytm zachłanny w analizie tychże struktur.

5 Algorytm zachłanny

Algorytm zachłanny można sformułować w sposób ogólny w sposób opisany poniżej. Rozważmy zbiór E ($|E| = n$), funkcję wagową $w: E \rightarrow R_+$, gdzie R_+ – liczby rzeczywiste dodatnie, oraz H rodzinę podzbiorów E . Wagą elementu e nazywamy wartość $w(e)$. Wagę podzbioru $A \subseteq E$ określamy następująco

$$w(A) = \sum_{e \in A} w(e). \quad (4)$$

$e \in A$

Tworzymy pewien zbiór S według zasad:

Krok 1. Uporządkować E według niemalejących (lub nierosnących) wag tzn. $E = \{e_1, e_2, \dots, e_n\}$ przy czym $w(e_1) \geq w(e_2) \geq \dots \geq w(e_n)$;

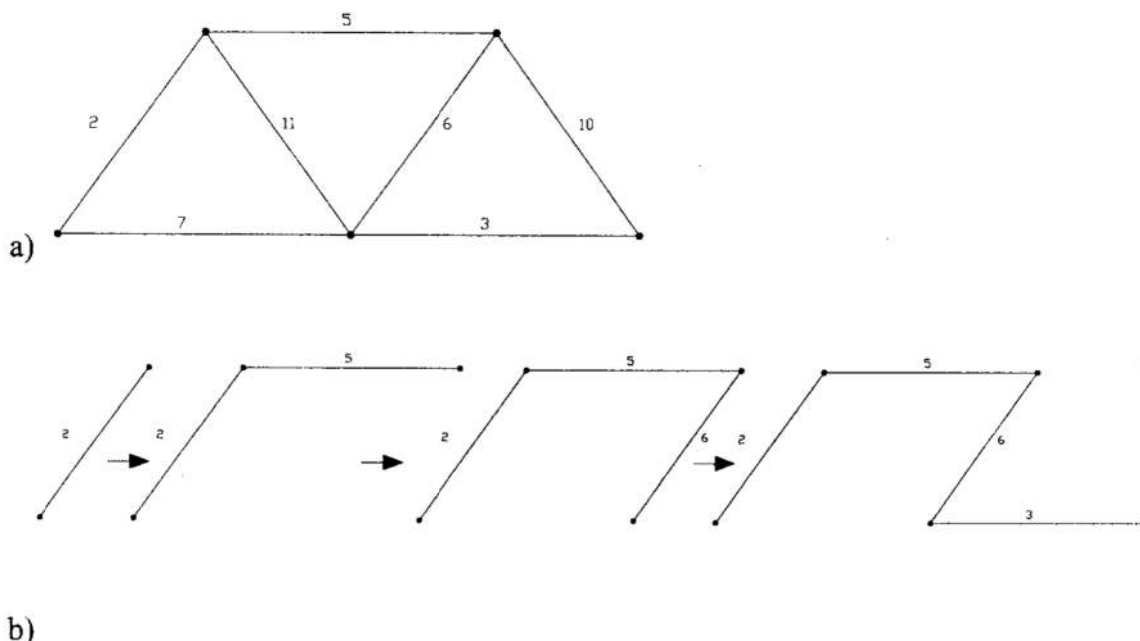
Krok 2. $S := \emptyset$;

Krok 3. Powtarzaj n razy

Krok 4. Jeśli $S \cup \{e_i\} \in H$ to przyjmuje się, że $S := S \cup \{e_i\}$.

Wykazano ([7]) następującą własność: jeśli $M = (E, H)$ jest matroidem, to zbiór S znaleziony przez algorytm zachłanny jest zbiorem niezależnym o największej wadze (lub najmniejszej wadze). Zbiór ten jest równocześnie bazą, gdyż baza w matroidach jest maksymalnym zbiorem niezależnym.

Przykład. Rozpatrzmy graf X , przedstawiony na rysunku 3, którego krawędziom przyporządkowano wagi nieujemne. Baza o minimalnej wadze dla matroidu cyklicznego $M = ({}_2X, P(X))$, którą jest drzewo napinające grafu X o minimalnej wadze – została znaleziona stosując powyższy algorytm. Schematycznie kolejne kroki przedstawiono na rysunku 3.



Rys. 3. Zastosowanie algorytmu zachłannego: (a) graf obciążony, (b) kolejne kroki algorytmu zachłannego, aby uzyskać bazę o minimalnej wadze
Fig. 3. Application of greedy algorithm: (a) weighted graph, (b) consecutive steps of greedy algorithm aiming for obtaining the base of minimal weight

Powyższe rozważania stanowią ilustrację do pierwszego ze sformułowanych na początku problemów. Aby rozwiązać drugi z postawionych na początku problemów trzeba przyporządkować matroid macierzy A ($m \times n$). Oznaczmy jej kolumny przez e_1, e_2, \dots, e_n i rozważamy je jako wektory. Macierzy przyporządkujemy matroid $M(A) = (E, \mathbf{B})$, gdzie E jest zbiorem kolumn, a pewien zbiór kolumn $B \in \mathbf{B}$ wtedy i tylko wtedy gdy kolumny B można rozważać jako zbiór wektorów liniowo niezależnych. Matroid $M(A)$ nazywamy matroidem macierzowym. W macierzy A przedstawiamy teraz kolumny według nierosnących wag. Należy znaleźć niezależny zbiór kolumn o maksymalnej sumie wag. Algorytmy opisano w pracach [10], [12].

Znajdowanie minimalnej bazy cykli ma istotne znaczenie w badaniu sztywności struktur, a omówiona metoda jest szczególnie odpowiednia, co będzie omówione w kolejnych publikacjach autorów.

6 Podsumowanie

Teoria matroidów może być stosowana do opisu różnorodnych problemów inżynierskich. W sensie matematycznym daje eleganckie uogólnienia pojęć z algebry i teorii grafów. Modelowanie układów mechanicznych za pomocą grafów jest powszechnie uznane. Matroidy jako modele układów mechanicznych pozwalają między innymi na algorytmizację analizy oraz umożliwiają zwięzły i poręczny sposób modelowania, uzyskuje się znane wyniki innym sposobem – co pogłębia zrozumienie problemów. Uzyskano

również nowe wyniki – dla których metody grafowe są niewystarczające. Algorytm zachłanny sformułowany został bardzo ogólnie, a jego zmodyfikowane, szczegółowe wersje umożliwiają opracowanie przejrzystych algorytmów wielu problemów kombinatoryki i mają zastosowanie w analizie sztywności struktur prętowych.

Literatura

1. Borowiecki M., Zmyślony B.: *Matroidy i ich zastosowanie do opisu obwodów elektrycznych*. Materiały VIII Ogólnopolskiego Sympozjonu „Metody matematyczne w elektrotechnice”, Pokrzywna k. Głuchołazy, (1979).
2. Bruno J., Weinberg L.: Generalized Networks: Networks Embedded on a Matroid, Part I, Part II, *Networks* 6 (1976) ss. 53–94, 231–272.
3. Bruno J., Weinberg L.: *The principal minors of of a matroid*. Linear Algebra and Appl. 4 (1971), 17 – 54.
4. Edmonds J.: *Minimum portition of a matroid into independent subset*. Journal of Research NBS, Vol. 69B (1965), 67–72.
5. Iri M., Fujishige S.: *Use of Matroid Theory in Operations Research, Circuits and Systems Theory*. Int. J. System Sci., 12, No 1 (1981), ss. 27–54.
6. Jankowski B.: *Grafy. Algorytmys w PASCAL-u*. MIKOM, Warszawa 1998.
7. Kaveh A.: *Matroids in Structural Mechanics*, Computers & Structures, 47, No 1, (1993), ss. 169–174.
8. Kaveh A.: *Graphs, Matroids and Structures*. Proceedings of the First International Conference GRAPHS & MECHANICS, Politechnika Śląska w Gliwicach, Wydział Mechaniczny, Gliwice, (1993), s.91.
9. Kaveh A.: *Optimizing the conditioning of structural flexibility matrices*. Computers and Structures, Vol. 41, No 6 str. 1217–1223 (1991).
10. Lipski W.: *Kombinatoryka dla programistów*. WN–T, Warszawa (1982)
11. Murota K.: *Systems Analysis by Graphs and Matroids*. p. 281, Springer–Verlag, New York (1987).
12. Oxley J.G.: *Matroid Theory*. Oxford (1992)
13. Petersen B.: *Topological Investigation of Linear Active Networks*, Department of Mathematics, Technical University of Denmark, 1980.
14. Recski A.: *Sufficient Conditions for the Unique Solvability of Linear Networks Containing Memoryless 2–Ports*. Circuit Theory and Applications, 8 (1980), ss. 95–103.
15. Weinberg L.: *Matroids, Generalized Networks and Electric Network Synthesis*. J. Comb. Theory, Ser. B, 23, (1977), ss. 106–126.
16. Wojnarowski J.: *Zastosoawanie grafów w analizie drgań układów mechanicznych*. PWN, Wrocław– Warszawa (1981).
17. Wojnarowski J., Zawisłak S.: *Modelling of Mechanical System by Means of Matroids*. Proc. Sec. Inter. Conf. GRAPHS & MECHANICS, Silesian Technical University, Gliwice (1999).
18. Wojnarowski J., Zawisłak S.: *Układ mechaniczny jako uogólniona sieć*. Zeszyty Naukowe Politechniki Śląskiej, Z 53, No 1001 (1989), ss. 87–97.
19. Wilson R. J.: *Wprowadzenie do teorii grafów*. s. 224. PWN, Warszawa, (1998).

Streszczenie

W artykule przedstawiono zastosowanie teorii matroidów do sformułowania wybranych zagadnień optymalizacji i opracowania dla nich algorytmów znajdowania rozwiązania optymalnego. Podano podstawowe pojęcia. Omówiono algorytm zachłanny. Jako przykład podano, że można sformułować problem sztywności struktur kratowych w ujęciu matroidów.

AN APPLICATION OF MATROIDS TO SOME OPTIMIZATION PROBLEMS AND TO MODELLING OF MECHANICAL SYSTEMS

Summary

An application of matroid theory to formulation of selected optimization problems has been presented in the paper. Some basic definitions and properties are enclosed in the paper. The greedy algorithm was described. The exemplary application was mentioned i.e. problem of stiffness of trusses can be formulated in terms of matroids.

Referat recenzowany

PODEJŚCIE GRAFOWE DO UNIKANIA BLOKAD W SYSTEMACH PROCESÓW WSPÓLBIEŻNYCH

1 Wprowadzenie

Współbieżna realizacja zadań jest podstawą działania wielu praktycznie spotykanych systemów czasu rzeczywistego [3]. W pakietowych sieciach komunikacyjnych [4], [5], [9] występują procesy polegające na przesyłaniu danych pomiędzy węzłami sieci. W rozproszonych systemach komputerowych istnieje możliwość realizacji wielu zadań w oparciu o różne komputery [10], [11]. Podobnie w elastycznych systemach produkcyjnych (ESP) możliwa jest współbieżna obróbka elementów zgodnie z zadanymi marszrutami technologicznymi [1], [6], [7], [8]. Przetwarzanie równoległe z wykorzystaniem wspólnych zasobów (maszyn, urządzeń) prowadzi do zwiększenia efektywności funkcjonowania systemu wymaga jednak rozwiązania wielu problemów sterowania, do których zaliczyć należy przede wszystkim problem blokad (impasów) [1], [9], [11].

Stan blokady systemu procesów charakteryzuje się istnieniem zamkniętego łańcucha żądań zasobowych, z których żadne nie może być zrealizowane. W praktyce oznacza to, że żaden proces nie może wykonać kolejnej operacji. Unikanie blokad polega na odrzuceniu tych przydziałów zasobowych, które prowadzą do blokady [11].

W ogólnym przypadku odpowiedź na pytanie czy dany stan rozdziału zasobów jest *bezpieczny* [11], tzn. czy nie prowadzi do blokady, jest problemem NP-zupełnym [5]. W związku z tym większość znanych w literaturze algorytmów unikania impasów to rozwiązania heurystyczne, oparte na warunkach wystarczających do bezblokadowej realizacji zadań. Mogą one odrzucać niektóre, bezpieczne stany rozdziału zasobów i w konsekwencji prowadzić do zmniejszenia efektywności funkcjonowania systemu. Dla określonego systemu procesów rozwiązanie jest tym lepsze im mniej odrzuca poprawnych stanów rozdziału zasobów.

Najbardziej znane algorytmy unikania blokad można podzielić na następujące grupy: rozwiązania wykorzystujące metodę bankiera [11], [13], algorytmy oparte na koncepcji stref operacji (synchronizacji) [1], [14], metody oparte na analizie struktury grafu żądań zasobowych [2], [6], [7], oraz rozwiązania wykorzystujące numerację zasobów [8], [12]. W niniejszej pracy przedstawiono ulepszoną wersję algorytmu opartego na koncepcji minimalnego cyklu w grafie żądań zasobowych. W wyniku uwzględnienia informacji o strukturze cykli elementarnych w grafie żądań opracowano oryginalny algorytm pozwalający zwiększyć rozmiar przestrzeni stanów bezpiecznych w stosunku do rozwiązania klasycznego.

Rozdział drugi zawiera sformułowanie problemu. W rozdziale trzecim przedstawiono algorytm oparty na koncepcji minimalnego cyklu w grafie żądań zasobowych procesów, natomiast w rozdziale czwartym algorytm ulepszony, wykorzystujący informację o strukturze wszystkich cykli grafu. Wnioski końcowe zawarto w rozdziale piątym.

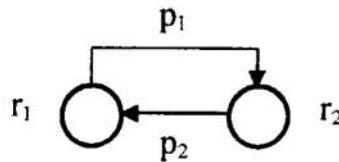
2 Sformułowanie problemu

Przykład ilustracyjny

W celu ilustracji problemu rozpatrzmy następujący system procesów.

* Politechnika Wroclawska, Instytut Cybernetyki Technicznej, ul. Janiszewskiego 11/17, 50-372 Wrocław.
Tel. (071) 320-29-06. E-Mail: wojcik@S104C3S2.ict.pwr.wroc.pl.

Przykład 1. Niech $P=\{p_i \mid i=1,\dots,m\}$ oznacza zbiór procesów, $R=\{r_1,r_2,r_3,r_4\}$ zbiór zasobów. Każdy zasób składa się z jednej jednostki. Procesy o numerach nieparzystych realizują program $T_n=(o_1,o_2,o_3,o_4)$, natomiast procesy o numerach parzystych program $T_p=(o_5,o_6,o_7,o_8)$, gdzie o_k ($k=1,\dots,8$) jest k -tą operacją. Do wykonania każdej operacji niezbędna jest jednostka zasobu $r(o_k) \in R$, gdzie $r(o_1)=r_1$, $r(o_2)=r_2$, $r(o_3)=r_3$, $r(o_4)=r_4$, $r(o_5)=r_3$, $r(o_6)=r_4$, $r(o_7)=r_2$, $r(o_8)=r_1$. Procesy korzystają z zasobów zgodnie z zasadą wzajemnego wykluczania. Proces może uwolnić zasób $r(o_k)$ wtedy i tylko wtedy, gdy dokona rezerwacji zasobu niezbędnego do wykonania następnej operacji. W rozpatrywanym systemie nie wszystkie z możliwych przydziałów zasobowych są poprawne. Załóżmy, że proces p_1 wykonuje operację o_1 , korzystając z zasobu r_1 , natomiast proces p_2 wykonuje operację o_7 , korzystając z zasobu r_2 . W kolejnym kroku pierwszy proces żąda dostępu do zasobu $r(o_2)=r_2$, a drugi proces żąda dostępu do zasobu $r(o_8)=r_1$. Pojawia się cykl żądań zasobowych określony sekwencją $r_1-r_2-r_1$, z których żadne nie może być zrealizowane (rys. 1). System osiąga stan blokady, w którym nie jest możliwa realizacja zadań. Wynika stąd potrzeba syntezy procedur unikania blokad, gwarantujących poprawną realizację programów w rozpatrywanej klasie współbieżnych procesów sekwencyjnych.



Rys. 1. Stan blokady systemu
Fig. 1. A system's deadlock state

Model systemu

W niniejszej pracy wprowadza się następujące definicje. Każdy proces $p_i \in P$ realizuje operacje zgodnie z programem $T_i=(o_1,\dots,o_k,\dots,o_{w_i})$, gdzie o_k jest k -tą operacją. Niech ZT oznacza zbiór różnych programów (niektóre procesy mogą realizować ten sam program), OP zbiór operacji, natomiast $R=\{r_i \mid i=1,\dots,n\} \cup \{r_e\}$ zbiór zasobów wykorzystywanych do realizacji zadań, gdzie r_e jest zasobem wyjściowym żadanym w momencie zakończenia programu. Funkcja pojemności zasobów $c(r)$ określa liczbę dostępnych jednostek zasobu $r \in R$. Przyjmuje się, że do wykonania operacji $o_k \in OP$ niezbędna jest jednostka zasobu $r(o_k) \in R$. Procesy korzystają z zasobów zgodnie z zasadą wzajemnego wykluczania. Proces może uwolnić zasób $r(o_k)$ wtedy i tylko wtedy, gdy dokona rezerwacji zasobu niezbędnego do wykonania następnej operacji. Przyjmuje się, że proces kończący program przechodzi do wykonania operacji wyjściowej o_e takiej, że $r(o_e)=r_e$ & $c(r_e)=\infty$ (zasób r_e ma nieskończoną pojemność).

Stan systemu procesów określa wektor $S=[S(1),\dots,S(k),\dots,S(L)]$, gdzie $S(k)$ jest liczbą procesów realizujących operację $o_k \in OP$, natomiast $L=|OP|$ jest liczbą operacji. *Stan początkowy* definiuje się jako wektor $S_0=[0,\dots,0]$. Każdy przydział zasobu do procesu prowadzi do zmiany stanu. Zbiór stanów osiągalnych ze stanu S_0 nazywa się *przestrzenią stanów* i oznacza SO . Niech $PI \subseteq P$ oznacza zbiór procesów, które korzystają z zasobów w stanie $S \in SO$. Stan jest *bezpieczny* wtedy i tylko wtedy, gdy istnieje sekwencja przydziałów zasobowych prowadząca do stanu S_0 . Stan nazywa się *stanem blokady* wtedy i tylko wtedy, gdy żaden proces $p \in PI$ nie może wykonać kolejnej operacji. Stan S jest określany jako *niebezpieczny* wtedy i tylko wtedy, gdy istnieje proces $p \in PI$, który może wykonać kolejną operację, ale dowolna sekwencja przydziałów zasobowych, dotycząca procesów PI , prowadzi do stanu blokady. Z przedstawionych definicji wynika, że $SO=SB \cup SNB \cup SD$ & $SB \cap SNB \cap SD = \emptyset$, gdzie SB - zbiór stanów bezpiecznych, SNB - zbiór stanów niebezpiecznych, SD - zbiór stanów blokady.

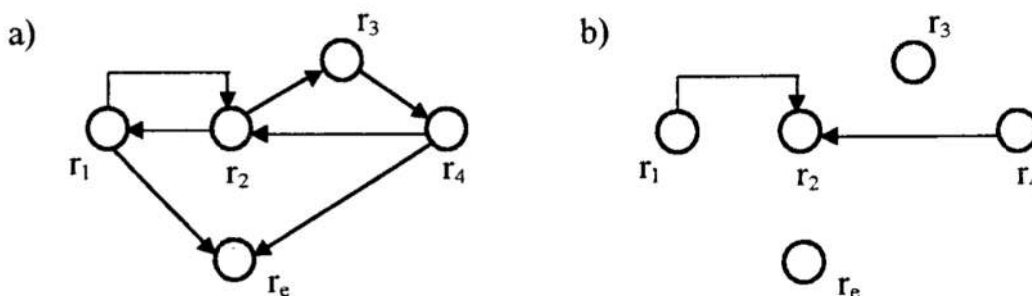
W stanie $S \in SB$ procedura unikania blokad określa, które z możliwych przydziałów zasobowych (żądane zasoby są dostępne) prowadzą do stanów bezpiecznych. Procedura nazywa się *maksymalnie dopuszczającą* wtedy i tylko wtedy, gdy akceptuje wszystkie stany bezpieczne. Jeżeli niektóre z poprawnych stanów rozdziału zasobów mogą być odrzucone przez procedurę unikania blokad, to mówimy wówczas o rozwiązaniu opartym na warunkach wystarczających dla unikania blokad. Algorytm zabezpieczania przed blokadami jest tym lepszy, im mniej odrzuca stanów bezpiecznych.

3 Algorytm minimalnego cyklu

Warunkiem koniecznym dla wystąpienia blokady jest istnienie cyklu w grafie przydziału zasobów [11]. W celu wyeliminowania impasów wystarczy ograniczyć liczbę współbieżnie realizowanych procesów w taki sposób, aby w grafie przydziału nie pojawił się cykl zawierający zasoby, z których żaden nie posiada wolnej jednostki. Punktem wyjścia dla omawianego algorytmu unikania blokad jest znajomość struktury żądań zasobowych zgłaszanych przez procesy podczas realizacji programów.

Struktura żądań zasobowych procesów może być opisana za pomocą grafu żądań zasobowych [2], [6]. Niech $G_i = (R, K_i)$ oznacza graf skierowany opisujący żądania zasobowe występujące w programie T_i , gdzie R oznacza zbiór wierzchołków, natomiast K_i zbiór krawędzi grafu. Krawędź $k_{ab} = (r_a, r_b)$, skierowana od r_a do r_b , należy do $K_i \subset R \times R$ wtedy i tylko wtedy, gdy $r_a = r(o_a)$ & $r_b = r(o_b)$, oraz operacja o_a bezpośrednio poprzedza operację o_b w programie T_i . Graf żądań zasobowych $G = (R, K)$, gdzie $K = \bigcup K_i$ & $i = 1, \dots, |ZT|$ ($|ZT|$ – liczba programów). Graf ten opisuje globalne żądania zasobowe występujące w systemie procesów. Chwilowe żądania zasobowe zgłaszane przez procesy w stanie S można opisać za pomocą grafu stanu $GS = (R, K(S))$, gdzie krawędź $k_{ab} \in K(S)$ wtedy i tylko wtedy, gdy istnieje proces $p \in P$, który w stanie S korzysta z zasobu r_a i do wykonania kolejnej operacji wymaga zasobu r_b [1], [6]. Wprowadzone definicje ilustruje następujący przykład.

Przykład 2. Rozpatrzmy system procesów przedstawiony w Przykładzie 1. Zbiór zasobów $R = \{r_1, r_2, r_3, r_4\}$. Grafy żądań zasobowych związane z programami T_n i T_p mają następującą postać: $G_n = (R, K_n)$ i $G_p = (R, K_p)$, gdzie $K_n = \{(r_1, r_2), (r_2, r_3), (r_3, r_4), (r_4, r_e)\}$ oraz $K_p = \{(r_3, r_4), (r_4, r_2), (r_2, r_1), (r_1, r_e)\}$. Wówczas, graf żądań zasobowych systemu $G = (R, K_n \cup K_p)$ (rys. 2a). Rozpatrzmy stan $S = [1000 0100]$, w którym proces p_1 realizuje operację o_1 , wykorzystując zasób r_1 , natomiast proces p_2 realizuje operację o_6 w oparciu o zasób r_4 . Graf stanu związany ze stanem S ma następującą postać: $GS = (R, K(S))$, gdzie $K(S) = \{(r_1, r_2), (r_4, r_2)\}$ (rys. 2b).



Rys. 2. Struktura żądań zasobowych: a) graf żądań zasobowych; b) graf stanu związany ze stanem $S = [1000 0100]$

Fig. 2. Structure of resource requests: a) resource requests graph; b) state graph associated with state $S = [1000 0100]$

Niech ZL oznacza zbiór wszystkich cykli elementarnych (cykle bez powtórzeń) w grafie żądań zasobowych [5], [6]. Cykl jest określony za pomocą ciągu wierzchołków grafu, z których tylko pierwszy i ostatni są równe (np. $r_1-r_2-r_1$ określa cykl w grafie pokazanym na rys. 2a). Niech WL_i oznacza zbiór wierzchołków grafu należących do cyklu $L_i \in ZL$ (np. dla $L_i=(r_1, r_2, r_1)$ jest $WL_i=\{r_1, r_2\}$). *Pojemność cyklu* L_i jest sumą pojemności zasobów tworzących cykl, tzn. $c(L_i)=\sum_{r \in WL_i} c(r)$. Niech cmL oznacza minimalną spośród pojemności wszystkich cykli występujących w grafie żądań zasobowych. Warunek wystarczający dla bezblokadowej realizacji zadań można sformułować w postaci następującego twierdzenia [1], [2], [7].

Twierdzenie 1. Jeżeli w stanie $S \in SO$ spełniony jest warunek

$$|PI| < cmL, \quad (1)$$

to stan S jest bezpieczny ($|PI|$ – liczba współbieżnie realizowanych procesów).

Z twierdzenia 1 wynika następujący wniosek: w celu uniknięcia impasu należy tak przydzielać zasoby do procesów, aby osiągnęte były tylko stany spełniające warunek (1), tzn. stany, w których liczba procesów w systemie jest mniejsza niż cmL .

Złożoność obliczeniowa algorytmu zależy od złożoności procedur wyznaczania stałych parametrów rozwiązania, realizowanych w trybie off-line, oraz złożoności testu bezpieczeństwa stanu, realizowanego w trybie on-line (warunek sprawdzany przed rezerwacją zasobu). Minimalną spośród pojemności cykli w grafie żądań zasobowych można wyznaczyć za pomocą algorytmu poszukiwania ścieżki o najmniejszej wadze w grafie skierowanym. Złożoność obliczeniowa odpowiedniego algorytmu nie przekracza $O(|R|^3)$ [2], [5], [6], gdzie $|R|$ – liczba wierzchołków grafu (liczba zasobów). Sprawdzenie warunku (1) realizowane jest w jednym kroku, stąd złożoność testu unikania blokad jest rzędu $O(1)$. Sposób działania algorytmu ilustruje następujący przykład.

Przykład 3. W przypadku systemu procesów rozpatrywanego w Przykładzie 1 każdy zasób ma pojemność równą 1. Wówczas, $cmL=2$ (cykl o najmniejszej pojemności $L_1=(r_1, r_2, r_1)$) i warunek (1) przyjmuje postać $|PI| < 2$. Algorytm minimalnego cyklu działa, więc w tym konkretnym przypadku zgodnie z zasadą „wszystko albo nic”, zezwalając na realizację tylko jednego procesu w systemie.

4 Algorytm przesuwania w cyklach

Efektywność algorytmu minimalnego cyklu (mierzoną liczbą akceptowanych stanów bezpiecznych) można zwiększyć poprzez odpowiednie zarządzanie dostępem procesów do zasobów tworzących cykle w grafie żądań zasobowych.

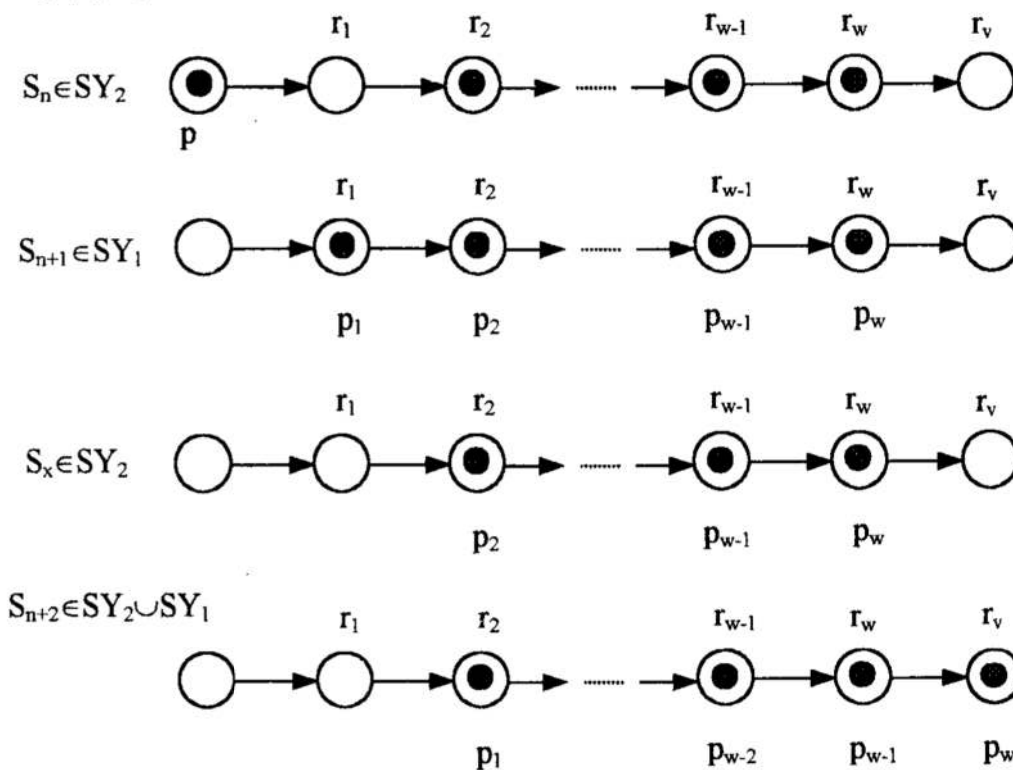
Niech $f(r, S)$ oznacza liczbę wolnych jednostek zasobu $r \in R$ w stanie S , natomiast $P(r, S)$ zbiór procesów, które korzystają z zasobu r w stanie S . Liczba wolnych jednostek zasobowych cyklu $L \in ZL$ w stanie S jest określona zależnością $f(L, S) = c(L) - \sum_{r \in WL} |P(r, S)|$, gdzie $r \in WL$ (WL – zbiór zasobów należących do cyklu L), $c(L)$ – pojemność cyklu L , $|P(r, S)|$ – liczba procesów korzystających z zasobu r w stanie S . Przestrzeń stanów SO może być podzielona na trzy parami rozłączne podzbiory $SO = SY_2 \cup SY_1 \cup SY_0$ & $SY_2 \cap SY_1 \cap SY_0 = \emptyset$, gdzie poszczególne podprzestrzenie są określone następująco:

- $S \in SY_2 \Leftrightarrow (\forall L \in ZL)[f(L, S) \geq 2]$ – zbiór stanów takich, że każdy cykl w grafie żądań zasobowych G posiada co najmniej dwie wolne jednostki zasobowe,
- $S \in SY_1 \Leftrightarrow (\exists L \in ZL)[f(L, S) = 1] \ \& \ (\neg \exists L \in ZL)[f(L, S) = 0]$ – zbiór stanów takich, że istnieje co najmniej jeden cykl w grafie G , posiadający tylko jedną wolną jednostkę zasobową, i nie istnieje cykl, który nie posiada żadnej wolnej jednostki,
- $S \in SY_0 \Leftrightarrow (\exists L \in ZL)[f(L, S) = 0]$ – zbiór stanów takich, że istnieje co najmniej jeden cykl w grafie G nie posiadający ani jednej wolnej jednostki zasobowej.

Można zauważyć, że żaden stan $S \in SY_2 \cup SY_1$ nie jest stanem blokady, gdyż każdy cykl $L \in ZL$ posiada co najmniej jedną wolną jednostkę w stanie S . Zbiór stanów blokady $SD \subset SY_0$. Główna idea algorytmu unikania blokad opiera się na takim sterowaniu przydziałami zasobowymi, aby z każdego stanu $S_n \in SY_2 \cup SY_1$ osiągnąć inny stan $S_{n+1} \in SY_2 \cup SY_1$. Sposób znajdowania ciągu procesów, dla których należy wykonać przydziały zasobowe w celu przeprowadzenia systemu ze stanu S_n do S_{n+1} określa następująca procedura.

Algorytm wyznaczania sekwencji przydziałów (AWSP)

Z definicji $S_0 \in SY_2$. Rozpatrzmy stan $S_n \in SY_2$. Załóżmy, że w wyniku alokacji zasobu r_1 do procesu $p \in P$ pojawi się stan $S_{n+1} \in SY_1$. Niech $RS = (r_1, r_2, \dots, r_{w-1}, r_w, r_v)$ będzie dowolną ścieżką w grafie stanu GS prowadzącą od r_1 do r_v , dla którego $f(r_v, S_{n+1}) > 0$ (zasób r_v posiada co najmniej jedną wolną jednostkę). Rozpatrzmy dowolny ciąg procesów $(p_1, p_2, \dots, p_{w-1}, p_w)$ takich, że $p_i \in P(r_i, S_{n+1})$ & $i=1, \dots, w$. Można zauważyć, że w wyniku alokacji zasobów zgodnie z sekwencją $AL = (p_w, p_{w-1}, \dots, p_2, p_1)$ otrzymamy stan $S_{n+2} \in SY_1$ lub $S_{n+1} \in SY_2$. Istotnie, proces p_w żąda dostępu do zasobu r_v , który posiada wolną jednostkę. W wyniku przydzielenia zasobu do procesu p_w uwolniona zostanie jednostka zasobu r_w , która stanie się dostępna dla procesu p_{w-1} . Podobną analizę można przeprowadzić dla kolejnych procesów należących do sekwencji AL . Po zrealizowaniu wszystkich żądań zasobowych z listy AL pojawi się stan S_{n+2} , w którym liczba wolnych jednostek zasobów r_2, \dots, r_w nie zmieni się, tj. $f(r_i, S_{n+2}) = f(r_i, S_{n+1})$ dla $i=2, \dots, w$, natomiast $f(r_1, S_{n+2}) = f(r_1, S_{n+1}) + 1$ oraz $f(r_v, S_{n+2}) = f(r_v, S_{n+1}) - 1$. Uwolnienie jednostki zasobu r_1 prowadzi do stanu $S_x \in SY_2$, gdyż stan S_n , który istniał przed zajęciem zasobu r_1 przez proces p należał do zbioru SY_2 . Wynika stąd, że zajęcie jednostki zasobu r_v w stanie S_x może prowadzić, w najgorszym przypadku, do stanu S_{n+2} należącego do zbioru $SY_2 \cup SY_1$ (rys. 3). Sposób działania algorytmu ilustruje kolejny przykład.



Rys. 3. Wyznaczanie sekwencji przydziałów zasobowych
Fig. 3. Finding a sequence of resource allocations

Przykład 4. Graf żądań zasobowych systemu rozpatrywanego w Przykładzie 1 pokazano na rys. 2a. Graf żądań zawiera dwa cykle elementarne: $L_1=(r_1, r_2, r_1)$, o pojemności $c(L_1)=2$, oraz $L_2=(r_2, r_3, r_4, r_2)$, dla którego $c(L_2)=3$. Niech p_1 realizuje program T_n , natomiast p_2 program T_p . Przyjmując, że żądania zasobowe procesów są rozpatrywane w sposób cykliczny otrzymamy następującą sekwencję przydziałów.

1. Wybieramy p_1 . W wyniku zajęcia zasobu r_1 przez p_1 następuje przejście ze stanu S_0 do stanu $S_a=[1000\ 0000]$. Zmianę stanu zapisujemy w postaci $S_0(p_1, r_1)S_a$. Stan $S_a \in SY_1$, gdyż $f(L_1, S_a)=1$. Sekwencja przydziałów $AL=(p_1)$, a więc należy przydzielić zasób dla p_1 .
2. $S_a(p_1, r_2)S_b$, gdzie $S_b=[0100\ 0000]$ & $S_b \in SY_1$ (gdyż $f(L_1, S_b)=1$). Stąd, $AL=(p_1)$.
3. $S_b(p_1, r_3)S_c$, gdzie $S_c=[0010\ 0000]$ & $S_c \in SY_2$ (gdyż $f(L_1, S_c)=2$ & $f(L_2, S_c)=2$).
4. Wybieramy p_2 . Przydział nie jest możliwy, gdyż zasób r_3 jest zajęty przez p_1 . Wybieramy p_1 . Wówczas, $S_c(p_1, r_4)S_d$, gdzie $S_d=[0001\ 0000]$ & $S_d \in SY_2$.
5. Wybieramy p_2 . $S_d(p_2, r_3)S_e$, gdzie $S_e=[0001\ 1000]$ & $S_e \in SY_1$ (gdyż $f(L_2, S_e)=1$). Wówczas, $AL=(p_1, p_2)$.
6. Przydzielamy zasoby dla p_1 i p_2 . $S_e(p_1, r_e)S_f(p_2, r_4)S_g$, gdzie $S_f=[0000\ 1000]$ & $S_g=[0000\ 0100]$ & $S_g \in SY_2$.

Kolejne przydziały zasobowe są wykonywane dla procesu p_2 w sposób analogiczny.

Twierdzenie 2. Jeżeli przydziały zasobowe realizowane są zgodnie z algorytmem AWSP, to osiągnane stany są bezpieczne.

Dowód twierdzenia wynika bezpośrednio z zasady działania AWSP, który określa sposób przydzielania zasobów do procesów gwarantujący przeprowadzenie systemu ze stanu $S_n \in SY_2 \cup SY_1$ do stanu $S_{n+1} \in SY_2 \cup SY_1$. Ponieważ zbiór $SY_2 \cup SY_1$ nie zawiera stanów blokady jest to warunek wystarczający globalnej żywotności systemu.

Złożoność obliczeniowa procedury wyznaczania cykli w grafie żądań zasobowych jest rzędu $O\{(|R|+|K|)(w+1)\}$ [6], gdzie $|R|$ - liczba wierzchołków grafu, $|K|$ - liczba krawędzi grafu, w - liczba cykli w grafie. W ogólnym przypadku złożoność czasowa procedury może być wykładnicza, gdyż graf żądań może zawierać wykładniczą liczbę cykli. Nie ma to jednak wpływu na test unikania blokad, gdyż cykle wyznaczone są w trybie off-line. Zarządzanie dostępem do zasobów, realizowane w trybie on-line, wymaga sprawdzania stanu zapelnienia cykli oraz poszukiwania sekwencji przydziałów zasobowych. Złożoność obliczeniowa odpowiednich procedur nie przekracza $O(w+|P|)$, gdzie w - liczba cykli, $|P|$ - liczba procesów.

5 Podsumowanie

W pracy przedstawiono algorytmy unikania blokad przeznaczone dla klasy współbieżnych procesów sekwencyjnych z wzajemnym wykluczeniem w dostępie do wspólnych zasobów. Klasyczny algorytm oparty na koncepcji minimalnego cyklu w grafie żądań zasobowych [2], [5], [6] prowadzi do rozwiązania czasowo efektywnego, które gwarantuje unikanie blokad w systemach procesów o dowolnej strukturze programów. W wyniku uwzględnienia informacji o strukturze wszystkich cykli w grafie żądań można opracować rozwiązanie, pozwalające zwiększyć liczbę akceptowanych stanów bezpiecznych. Przedstawiony algorytm wymaga jednak wyznaczenia wszystkich cykli w grafie w trybie off-line oraz odpowiedniej procedury sterowania zajętością zasobów, tworzących cykle, realizowanej w trybie on-line.

Przestrzeń stanów systemu procesów rozpatrywanego w Przykładzie 1 można przedstawić w postaci drzewa stanów, które składa się z 11 warstw $SO = \cup SO_i$ & $i=1, \dots, 11$, gdzie zbiory SO_i zawierają następujące stany: $SO_1 = \{S_0 = [0000\ 0000]\}$, $SO_2 = \{S_1 = [1000\ 0000], S_2 = [0000\ 1000]\}$, $SO_3 = \{S_3 = [0100\ 0000], S_4 = [1000\ 1000], S_5 = [0000\ 0100]\}$,

$SO_4 = \{S_6 = [0010\ 0000], S_7 = [1100\ 0000], S_8 = [0100\ 1000], S_9 = [1000\ 0100], S_{10} = [0000\ 1100], S_{11} = [0000\ 0010]\}$, $SO_5 = \{S_{12} = [0001\ 0000], S_{13} = [1010\ 0000], S_{14} = [1100\ 1000], S_{15} = [0100\ 0100], S_{16} = [1000\ 0010], S_{17} = [1000\ 1100], S_{18} = [0000\ 1010], S_{19} = [0000\ 0001]\}$, $SO_6 = \{S_{20} = [0001\ 1000], S_{21} = [1001\ 0000], S_{22} = [0110\ 0000], S_{23} = [1100\ 0100], S_{24} = [0010\ 0100], S_{25} = [0100\ 1100], S_{26} = [1000\ 1010], S_{27} = [0000\ 0110], S_{28} = [0000\ 1001]\}$, $SO_7 = \{S_{29} = [1001\ 1000], S_{30} = [0101\ 0000], S_{31} = [1110\ 0000], S_{32} = [1100\ 1100], S_{33} = [1010\ 0100], S_{34} = [0010\ 0010], S_{35} = [1000\ 0110], S_{36} = [0000\ 1110], S_{37} = [0000\ 0101]\}$, $SO_8 = \{S_{38} = [0101\ 1000], S_{39} = [0011\ 0000], S_{40} = [1101\ 0000], S_{41} = [0110\ 0100], S_{42} = [1010\ 0010], S_{43} = [0001\ 0010], S_{44} = [0010\ 0001], S_{45} = [1000\ 1110], S_{46} = [0000\ 1101], S_{47} = [0000\ 0011]\}$, $SO_9 = \{S_{48} = [1101\ 1000], S_{49} = [1011\ 0000], S_{50} = [1110\ 0100], S_{51} = [1001\ 0010], S_{52} = [0001\ 0001], S_{53} = [0000\ 1011]\}$, $SO_{10} = \{S_{54} = [0111\ 0000], S_{55} = [0001\ 1001], S_{56} = [0000\ 0111]\}$, $SO_{11} = \{S_{57} = [1111\ 0000], S_{58} = [0000\ 1111]\}$.

Przejścia pomiędzy stanami można zapisać za pomocą listy incydencji drzewa stanów, która określa wszystkie stany osiągalne ze stanu o numerze i , gdzie $i=0,1,\dots,58$. W szczególności, zapis $0| \rightarrow 1 \rightarrow 2$ oznacza, że ze stanu S_0 można przejść (w wyniku alokacji określonych zasobów) do stanu S_1 lub S_2 , natomiast zapis $16| \rightarrow d$ oznacza, że stan S_{16} jest stanem blokady. Lista incydencji drzewa stanów ma następującą postać:

$0| \rightarrow 1 \rightarrow 2$, $1| \rightarrow 3 \rightarrow 4$, $2| \rightarrow 4 \rightarrow 5$, $3| \rightarrow 6 \rightarrow 7 \rightarrow 8$, $4| \rightarrow 8 \rightarrow 9$, $5| \rightarrow 9 \rightarrow 10 \rightarrow 11$, $6| \rightarrow 12 \rightarrow 13$, $7| \rightarrow 13 \rightarrow 14$, $8| \rightarrow 14 \rightarrow 15$, $9| \rightarrow 15 \rightarrow 16 \rightarrow 17$, $10| \rightarrow 17 \rightarrow 18$, $11| \rightarrow 16 \rightarrow 18 \rightarrow 19$, $12| \rightarrow 0 \rightarrow 20 \rightarrow 21$, $13| \rightarrow 21 \rightarrow 22$, $14| \rightarrow 23$, $15| \rightarrow 23 \rightarrow 24 \rightarrow 25$, $16| \rightarrow d$, $17| \rightarrow 25$, $18| \rightarrow 26 \rightarrow 27 \rightarrow 28$, $19| \rightarrow 0 \rightarrow 28$, $20| \rightarrow 2 \rightarrow 29$, $21| \rightarrow 1 \rightarrow 29 \rightarrow 30$, $22| \rightarrow 30 \rightarrow 31$, $23| \rightarrow 32 \rightarrow 33$, $24| \rightarrow 33 \rightarrow 34$, $25| \rightarrow d$, $26| \rightarrow 35$, $27| \rightarrow 35 \rightarrow 36 \rightarrow 37$, $28| \rightarrow 2 \rightarrow 37$, $29| \rightarrow 4 \rightarrow 38$, $30| \rightarrow 38 \rightarrow 39 \rightarrow 40$, $31| \rightarrow 40$, $32| \rightarrow d$, $33| \rightarrow 41 \rightarrow 42$, $34| \rightarrow 42 \rightarrow 43 \rightarrow 44$, $35| \rightarrow d$, $36| \rightarrow 45 \rightarrow 46$, $37| \rightarrow 5 \rightarrow 46 \rightarrow 47$, $38| \rightarrow 8 \rightarrow 48$, $39| \rightarrow 6$, $40| \rightarrow 7 \rightarrow 48 \rightarrow 49$, $41| \rightarrow 50$, $42| \rightarrow 51$, $43| \rightarrow 11 \rightarrow 51 \rightarrow 52$, $44| \rightarrow 6 \rightarrow 52$, $45| \rightarrow d$, $46| \rightarrow 10 \rightarrow 53$, $47| \rightarrow 11 \rightarrow 53$, $48| \rightarrow 14$, $49| \rightarrow 13 \rightarrow 54$, $50| \rightarrow d$, $51| \rightarrow 16$, $52| \rightarrow 12 \rightarrow 19 \rightarrow 55$, $53| \rightarrow 18 \rightarrow 56$, $54| \rightarrow 57$, $55| \rightarrow 20 \rightarrow 28$, $56| \rightarrow 27 \rightarrow 58$, $57| \rightarrow 31$, $58| \rightarrow 36$.

Zbiór stanów blokady $SD = \{S_{16}, S_{25}, S_{32}, S_{35}, S_{45}, S_{50}\}$, zbiór stanów niebezpiecznych $SN = \{S_{17}, S_{26}, S_{33}, S_{41}, S_{42}, S_{51}\}$, natomiast zbiór stanów bezpiecznych $SB = SO \setminus (SD \cup SNB)$. Wynika stąd, że $|SD|=6$ & $|SNB|=6$ & $|SB|=|SO|-|SD|-|SNB|=47$. Algorytm minimalnego cyklu akceptuje stany o numerach 0,1,2,3,5,6,11,12,19, natomiast algorytm przesuwania w cyklach akceptuje stany o numerach 0,1,2,3,4,5,6,8,9,10,11,12,13,18,19,20,21,22,24,27,30,34,37,39,43,52. Wynika stąd, że w przypadku analizowanego systemu procesów algorytm minimalnego cyklu akceptuje 9 stanów, natomiast algorytm ulepszony 26 stanów spośród całkowitej liczby 47 stanów bezpiecznych, należących do przestrzeni 59 stanów osiągalnych ze stanu początkowego. Przedstawione rozwiązania oparte są na warunkach wystarczających dla bezblokadowej realizacji zadań i dlatego mogą odrzucać niektóre z poprawnych stanów rozdziału zasobów. Interesujące może być pytanie o średnią liczbę stanów akceptowanych przez algorytmy w przypadku systemów o dowolnej strukturze programów. Odpowiednia analiza może być przedmiotem dalszych prac.

Literatura

1. Banaszak Z., Krogh B.: *Deadlock Avoidance in Flexible Manufacturing Systems with Concurrently Competing Process Flows*. IEEE Trans. Robot. and Automation, 1990, Vol. 6, No. 6, S. 724-734.
2. Banaszak Z., Wójcik R.: *A Certain Solution of Resource Allocation Problem in Pipeline Concurrent Processes*. Proc. of Int. Conf. System Modelling Control 7, Zakopane, Poland, 1993, Vol. 1, S. 55-60.
3. Ben-Ari M.: *Podstawy programowania współbieżnego*. WNT Warszawa 1989.
4. Błażewicz J., Brzeziński J., Gambosi G.: *Time-Stamp Approach to Store-and-Forward Deadlock Prevention*. IEEE Trans. Commun., 1987, COM-35, No. 5, S. 490-495.

5. Błażewicz J.: *Złożoność obliczeniowa problemów kombinatorycznych*. WNT, Warszawa, 1988.
6. Fanti M. P., Maione B., Mascolo S., Turchiano B.: *Event-Based Feedback Control for Deadlock Avoidance in Flexible Production Systems*. IEEE Trans. Robot. and Automation, 1997, Vol. 13, No. 3, S. 347-363.
7. Hasegawa K., Wójcik R., Sugisawa M., Banaszak Z.: *Graphical Synthesis of Deadlock Avoidance Procedures for Flexible Manufacturing Systems*. Proc. of the 1st Int. Conf. on Engineering design and Automation – Advances in Eng. Design and Automation Research I, Bangkok, Thailand, 1997, S. 358-365.
8. Lawley M. A., Reveliotis S. A., Ferreira P. M.: *A Correct and Scalable Deadlock Avoidance Policy for Flexible Manufacturing Systems*. IEEE Trans. Robot. and Automation, 1998, Vol. 14, No. 5, S. 796-809.
9. Merlin P. M., Schweitzer P. J.: *Deadlock Avoidance in Store-and-Forward Networks - I: Store-and-Forward Deadlock*. IEEE Trans. Commun. 1980, COM-28, No. 3, S. 345-356.
10. Raynal M.: *Algorithms for Mutual Exclusion*. MIT Press, Massachusetts, 1986.
11. Silberschatz A., Peterson L.: *Operating Systems Concepts*. Addison-Wesley, Amsterdam, 1983.
12. Wójcik R.: *Podjęcie scentralizowane do unikania blokad w pewnej klasie elastycznych systemów produkcyjnych*. Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wrocławskiej, Nr 95, Seria: Konferencje, Nr 42, V KKR, Tom 1, Wrocław 1996, S. 148-155.
13. Wójcik R.: *Wyznaczanie przestrzeni stanów bezpiecznych w systemach procesów współbieżnych*. Zeszyty Naukowe Wydziału Mechanicznego Politechniki Koszalińskiej, Materiały XVII Ogólnopolskiej Konferencji Polioptymalizacja i Komputerowe Wspomaganie Projektowania, Nr 26, Mielno 1999, S. 325-332.
14. Wójcik R., Banaszak Z., Muszyński W.: *Deadlock-Free Workflow Control*. Proc. of Int. DAAAM Symposium, Intelligent Manufacturing & Automation: Past-Present-Future, Vienna Univ. of Tech., Vienna, Austria, 1999, S. 607-609.

Streszczenie

W pracy przedstawiono algorytmy unikania blokad (wyznaczania przestrzeni stanów bezpiecznych) dla klasy współbieżnych procesów sekwencyjnych dzielących wspólne zasoby. Rozwiązania oparto na wykorzystaniu grafu żądań zasobowych procesów. W wyniku uwzględnienia informacji o strukturze cykli grafu opracowano algorytm pozwalający zwiększyć rozmiar przestrzeni stanów bezpiecznych w stosunku do klasycznego algorytmu opartego na koncepcji minimalnego cyklu. Przedstawiono przykłady ilustrujące sposób działania algorytmów.

GRAPH BASED APPROACH TO DEADLOCK AVOIDANCE IN SYSTEMS OF CONCURRENT PROCESSES

Summary

This paper presents algorithms for deadlocks avoidance (for determining a space of safe states) for a class of concurrent sequential processes sharing common resources. The solutions are based on the approach using processes' resource requests graph. As a result of using an information about a structure of cycles of the graph an algorithm accepting more safe states than the classical algorithm based on a concept of the minimal cycle has been designed. Examples illustrating the algorithms are presented.

Referat recenzowany

OPTIMALIZACJA PARAMETRÓW OKRĘTOWYCH UKŁADÓW NAPĘDOWYCH W CELU ROZPRZĘŻENIA DRGAŃ SKRĘTNYCH I WZDŁUŻNYCH

1. Wprowadzenie

Problematyka drgań sprzężonych skrętno-wzdłużnych w okrętowych układach napędowych (OUN) o sztywno łączonej linii wałów, z wolnoobrotowymi dwusuwowymi silnikami diesla, mimo kilkudziesięciu lat prowadzonych badań, jest tematem ciągle aktualnym. O ważności zagadnienia świadczyć może fakt, iż w ciągu ostatnich lat wielu badaczy [1], [2], [7] podkreślało potrzebę przeprowadzania obliczeń sprawdzających dynamiczne własności OUN już na etapie projektowania statku, dzięki czemu można uniknąć kosztownych zmian w konstrukcji zbudowanej już jednostki pływającej. Naturalną konsekwencją takiego postulatu jest wprowadzenie w procesie projektowania statku metody optymalizacji, której jednym z celów jest zmniejszenie drgań sumarycznych skrętnych i wzdłużnych, a w szczególności zminimalizowanie oddziaływania pomiędzy tymi postaciami drgań.

Jedną z cech charakterystycznych efektu sprzężenia się drgań skrętnych i wzdłużnych jest jego silna zależność od relacji pomiędzy wartościami częstości drgań własnych skrętnych i wzdłużnych [6]. Im wartości własne obu postaci drgań są do siebie bardziej zbliżone tym zjawisko sprzężenia jest silniejsze. Własność ta wynika z faktu, iż w przypadku zbliżonych wartości własnych drgań skrętnych i wzdłużnych, przekazanie strumienia energii z jednej postaci drgań do drugiej wskutek sprzężenia wystarcza, aby spowodować istotne drgania drugiej (sprzężonej) postaci. W związku z powyższym, optymalizację mającą na celu rozprężenie drgań skrętnych i wzdłużnych można zdefiniować jako proces, którego celem jest „odsunięcie od siebie” częstości drgań własnych odpowiadających postaci drgań skrętnych i wzdłużnych oraz zmniejszenie sumarycznych drgań wzdłużnych i skrętnych.

Proces optymalizacji w celu zmiany własności dynamicznych OUN prowadzącej do osłabienia sprzężeń drgań skrętnych i wzdłużnych można podzielić na trzy następujące etapy:

- **Etap I** - opracowanie modelu obliczeniowego okrętowego układu napędowego wraz z konstrukcją nośną linii wałów, przy uwzględnieniu [3], [5], [11]:
 - a. współczynników sprzęgających drgania skrętne i wzdłużne wału korbowego,
 - b. współczynników „skośnych” konstrukcji łożyskująco-nośnej linii wałów (współczynników opisujących niezachowawcze własności filmu olejowego,
 - c. tłumienie filmu olejowego oraz oddziaływanie pomiędzy podporami łożyskowymi)
 - d. współczynników sprzęgających drgania skrętne i wzdłużne śruby napędowej,

- Etap II - matematyczne sformułowanie problemu optymalizacji, polegające na ustaleniu zmiennych decyzyjnych x , funkcji celu $F(x)$ oraz warunków ograniczających g_i ,
- Etap III - rozwiązanie problemu sformułowanego w poprzednim podpunkcie, polegające na znalezieniu takiego wektora zmiennych decyzyjnych x , który zapewnia wystarczające „odsunięcie od siebie” częstości postaci drgań własnych skrętnych i wzdłużnych (dla co najmniej trzech najniższych częstości drgań), przy równoczesnym spełnieniu wszystkich nałożonych warunków ograniczających.

2. Zmienne decyzyjne

W zagadnienie optymalizacji parametrów OUN w celu rozprężenia drgań skrętnych i wzdłużnych można stosować wszystkie rodzaje zmiennych decyzyjnych, tzn. zarówno zmiennych wymiarowych, jak i fizycznych oraz konfiguracyjnych. Jako zmienne wymiarowe z powodzeniem można zastosować średnice wałów pośrednich i śrubowego, średnice i szerokości sprzęgieł sztywnych, średnicę i szerokość koła zamachowego czy też obudowy tłumika drgań skrętnych. Zmierzonymi fizycznymi, które w praktyce mogą być wykorzystane, są sztywności podpór łożyskowych i podwójnego dna statku oraz masa i masowy moment bezwładności śruby napędowej. Natomiast jako zmienne konfiguracyjne wygodnie jest przyjąć położenie podpór łożyskowych, liczbę łopatek śruby napędowej i rozkład mas występujących w OUN.

Z powyższego opisu wyraźnie wynika, iż proponowany w niniejszej pracy sposób optymalizacji własności dynamicznych OUN bazuje na modyfikacjach parametrów układu niezwiązanych z konstrukcją silnika spalinowego.

3. Warunki ograniczające

Optymalizacja wymaga wielu warunków ograniczających, dotyczących zarówno bezpośrednio zmiennych decyzyjnych oraz takich, które odnoszą się do własności mechanicznych układu. Warunki ograniczające nałożone na własności mechaniczne OUN dotyczą:

- a) stosunku częstości drgań własnych skrętnych do wzdłużnych,
- b) maksymalnych wartości amplitud drgań sumarycznych skrętnych i wzdłużnych,
- c) maksymalnych wartości amplitud naprężeń w elementach składowych modelu dyskretnego OUN.

Jednym z najistotniejszych warunków optymalizacji w celu rozprężenia drgań skrętnych i wzdłużnych jest to, aby „odsuwając” od siebie najniższe częstości drgań własnych skrętnych i wzdłużnych nie zbliżyć się z tymi częstościami do kolejnych wyższych częstości.

4. Funkcja celu

W celu przeprowadzenia modyfikacji OUN, której rezultatem ma być zmniejszenie oddziaływania pomiędzy drganiami skrętnymi i wzdłużnymi, słuszne jest przyjęcie funkcji celu będącej sumą mas i masowych momentów bezwładności elementów

wirujących OUN oraz masy konstrukcji nośnej linii wałów. Zasadność zastosowania tak sformułowanej funkcji celu jest oczywista, jeśli weźmie się pod uwagę fakt, iż dążymy do odpowiedniego przesunięcia częstości drgań własnych OUN. Jak wiadomo, taką operację można przeprowadzić albo poprzez modyfikację macierzy mas albo poprzez modyfikację macierzy sztywności modelu matematycznego układu, ewentualnie równoległe zmieniając parametry masowe i sztywnościowe OUN.

Funkcję celu najogólniej można zapisać w postaci

$$F(\mathbf{x}) = J_{sum} + m_{sum} + a_1 \cdot (x_1^2 \cdot H_1 + x_1^4 \cdot H_1) + a_2 \cdot (x_2^2 \cdot H_2 + x_2^4 \cdot H_2) + \dots \\ a_j \cdot (x_j^2 \cdot H_j + x_j^4 \cdot H_j) + b_1 \cdot (y_1^2 \cdot l_1 + y_1^4 \cdot l_1) + b_2 \cdot (y_2^2 \cdot l_2 + y_2^4 \cdot l_2) + \dots \\ b_k \cdot (l_k^2 \cdot h_k + l_k^4 \cdot h_k) + m_{st1} + m_{st2} + \dots + m_{stv} + m_{KL-N} + m_{sr} + J_{sr}$$

gdzie:

J_{sum} , m_{sum} - odpowiednio suma masowych momentów bezwładności i mas elementów wirujących oraz mas konstrukcji łożyskująco-nośnej linii wałów OUN,

PARAMETRY SES MODELU DYSKRETNEGO

x_1, x_2, \dots, x_j - promienie powierzchni walcowych sztywnych elementów skończonych występujących w modelu dyskretnym OUN,

H_1, H_2, \dots, H_j - długości tworzących powierzchni walcowych sztywnych elementów skończonych występujących w modelu dyskretnym OUN,

a_1, a_2, \dots, a_j - współczynniki, j - liczba SES, których parametry są modyfikowane.

PARAMETRY EB MODELU DYSKRETNEGO

y_1, y_3, \dots, y_p - promienie przekroju elementów belkowych występujących w modelu dyskretnym OUN,

y_2, y_4, \dots, y_q - długości elementów belkowych występujących w modelu dyskretnym OUN,

b_1, b_2, \dots, b_k - współczynniki, k - liczba EB, których parametry są modyfikowane.

PARAMETRY KONSTRUKCJI ŁOŻYSKUJĄCO - NOŚNEJ

$m_{st1}, m_{st2}, \dots, m_{stv}$ - masa stojaków łożyskowych, v - liczba stojaków, których parametry są modyfikowane.

m_{KL-N} - masa podwójnego dna statku.

W praktyce z przedstawionej funkcji należy wybrać kilka parametrów, które są łatwe do modyfikacji w budowanym OUN mając na uwadze wrażliwość układu na te parametry.

Drugą istotną kwestię stanowi fakt, iż zmieniających długość EB sąsiadujących z elementami sprężysto-tłumiącymi (EST), które modelują łożyska linii wałów, zmieniamy położenie podpór łożyskowych.

Matematyczne sformułowanie problemu optymalizacji sprowadza się do zadania optymalizacji wybranych parametrów okrętowego układu napędowego mającej na celu rozprężenie drgań skrętnych i wzdłużnych, można sformułować następująco: znaleźć taki wektor zmiennych decyzyjnych

$$\mathbf{x} = \text{col} \{x_i\} \quad i=1, 2, \dots, n,$$

dla którego funkcja celu $F(\mathbf{x})$ dąży do minimum, przy warunkach ograniczających $g_i(\mathbf{x})$ nałożonych na:

1. zmienne wymiarowe – średnice i grubości walcowych SES oraz średnice przekroju i długości EB,

$$x_j^d \leq x_j \leq x_j^g \quad (j=1,2..m)$$

gdzie: x_j^d - dolna wartość średnicy lub grubości / długości SES i EB,
 x_j^g - górna wartość średnicy lub grubości / długości SES i EB,

2. częstości własne drgań skrętnych i wzdłużnych ω_{js} i ω_{jw}

$$\Delta\omega_j^d \leq (\omega_{js} - \omega_{jw}) \leq \Delta\omega_j^g \quad (j=1,2..k)$$

gdzie: $\Delta\omega_j^d, \Delta\omega_j^g$ - dolna i górna wartość różnicy częstości własnych drgań skrętnych i drgań wzdłużnych,

3. amplitudy sumarycznych drgań skrętnych $A_{drg\ skr}$ i wzdłużnych $A_{drg\ wzd}$

$$A_{drg\ skr} \leq A_{drg\ skr\ dopuszcz}$$

$$A_{drg\ wzd} \leq A_{drg\ wzd\ dopuszcz}$$

4. amplitudy naprężeń stycznych i normalnych w elementach składowych modelu dyskretnego

$$\tau \leq \tau_{dop} ; \sigma \leq \sigma_{dop}$$

5. Rozwiązanie problemu optymalizacji

Do rozwiązania sformułowanego problemu optymalizacji można zastosować dowolną metodę optymalizacji. W badaniach zaprezentowanych w niniejszej pracy wykorzystano metodę rozszerzonej wewnętrznej funkcji kary, co było podyktowane prostotą metody oraz jej skutecznością działania [10].

Metoda funkcji kary polega na powiększeniu funkcji celu o człon, który stanowi „karę” za wyjście poza dopuszczalny obszar rozwiązań. Nowoutworzona funkcja celu ma postać:

$$\Phi(\mathbf{x}, r_k) = F(\mathbf{x}) + r_k \cdot \sum_{i=1}^m G_i(\mathbf{x})$$

gdzie:

$F(\mathbf{x})$ – funkcja celu zagadnienia z ograniczeniami,

$\sum_{i=1}^m G_i(\mathbf{x})$ – funkcja kary,

r_k – współczynnik kary,

m – liczba warunków ograniczających,

$$G_i(\mathbf{x}) = \begin{cases} \frac{1}{g_i(\mathbf{x})} & \text{dla } g_i(\mathbf{x}) \geq \varepsilon \\ \frac{2\varepsilon - g_i(\mathbf{x})}{\varepsilon^2} & \text{dla } g_i(\mathbf{x}) < \varepsilon \end{cases}$$

przy czym:

ε - parametr przejścia.

6. Analiza wrażliwości OUN na zmianę parametrów

Przed przystąpieniem do procesu optymalizacji odnoszącego się do wybranego OUN należy określić w jaki sposób zmieniają się jego częstotliwości drgań własnych i postacie przy małych zmianach wybranych parametrów masowych lub sztywnościowych układu. Do tego celu można wykorzystać np. metodę perturbacji.

Zakładając, że znana jest macierz spektralna Ω^2 i macierz modalna V układu mechanicznego, przy czym

$$K V = M V \Omega^2,$$

M - globalna macierz bezwładności układu mechanicznego,

K - globalna macierz sztywności układu mechanicznego,

oraz wprowadzając zależność

$$K_* V_* = M_* V_* \Omega_*^2$$

w której:

$$K_* = K + dK \quad M_* = M + dM$$

$$V_* = V + dV \quad \Omega_*^2 = \Omega^2 + d\Omega^2,$$

można uzyskać wyrażenie reprezentujące zmianę częstotliwości drgań własnych w funkcji małych zmian współczynników macierzy mas M i sztywności K

$$d\omega_r^2 = \frac{\mathbf{v}_r^T dK \mathbf{v}_r - \omega_r^2 \mathbf{v}_r^T dM \mathbf{v}_r}{\mathbf{v}_r^T M \mathbf{v}_r}$$

W celu wyznaczenia zmian postaci własnych układu w zależności od zmian jego parametrów masowych i sztywnościowych, należy przeprowadzić obliczenia szacunkowe wg poniższego schematu:

I etap

$$F = V^T dK V - V^T dM V \Omega^2$$

$$L = V^T M V$$

II etap

$$G = L^{-1} F = [g_{jk}]$$

$$\mathbf{H} = [h_{jk}] \quad h_{jk} = \begin{cases} 0 & \omega_j^2 = \omega_k^2 \\ \frac{g_{jk}}{\omega_k^2 - \omega_j^2} & \omega_j^2 \neq \omega_k^2 \end{cases}$$

Korzystając z powyższych zależności zmianę macierzy modalnej V w funkcji niewielkich zmian parametrów masowych i sztywnościowych można wyrazić zależnością:

$$dV = V H$$

7. Przykład optymalizacji parametrów OUN pod kątem rozprężenia drgań skrętnych i wzdłużnych

Optymalizację własności dynamicznych przeprowadzono dla OUN o następujących, podstawowych danych technicznych:

silnik główny SULZER 6 RTA 76 (moc 11500 kW, nominalna prędkość obrotowa 86 obr/min)

długość sumaryczna wałów pośrednich i śrubowego - 22,5m

śrubę napędową z pięcioma skrzydłami (masa 50t, biegunowy moment bezwładności 93280kgm^2)

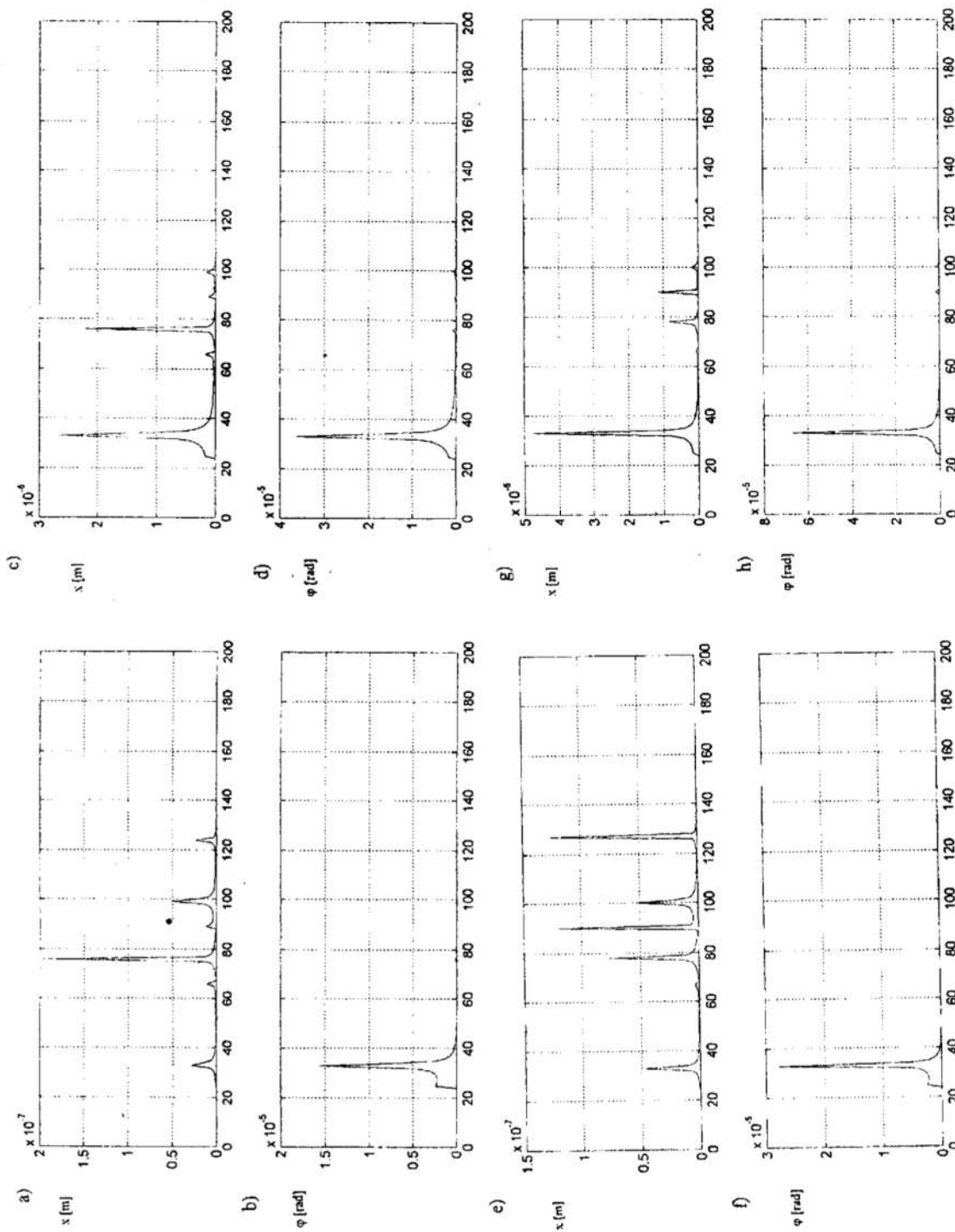
liczba łożysk wałów pośrednich – 2; liczba łożysk wału śrubowego – 1.

Ponieważ wykorzystany w przykładowych obliczeniach OUN jest układem dobrze pracującym na statku, w którym sprzężenia były słabo eksponowane, przed przystąpieniem do optymalizacji przeprowadzone zostały modyfikacje jego parametrów masowo-szywnościowych, w celu zestrojenia częstości własnych drgań skrętnych i wzdłużnych. W tak przygotowanym układzie, po przyłożeniu skrętnych wymuszeń harmonicznym stwierdzono silny efekt sprzęgania się drgań skrętnych i wzdłużnych w obszarze 78-100 rad/s (rys.2e-h). Rozwiązanie zagadnienia własnego dla układu sprzężonego wykazało, że 78 rad/s, 90 rad/s i 100 rad/s, są to wartości własne, dla których postacie własne drgań skrętnych i wzdłużnych są ze sobą sprzężone. Po rozprężeniu układu otrzymano częstość własną drgań wzdłużnych 94 rad/s i częstość własną drgań skrętnych 96 rad/s. Obie częstości przyjęto jako parametry początkowe w optymalizacji parametrów dla rozprężonego OUN. Wartości własne uzyskane w wyniku optymalizacji wyniosły odpowiednio 91 rad/s i 98 rad/s.

Na rysunku 1 przedstawiono amplitudy drgań skrętnych i wzdłużnych swobodnego końca wału korbowego i śruby napędowej wymuszanych momentami skracającymi; po optymalizacji a),b),c),d) i przed optymalizacją e),f),g),h) parametrów masowych OUN. Zmniejszenie amplitud w widmie drgań skrętnych i drgań wzdłużnych było rezultatem zmniejszenia masy tarczy sprzęgła sztywnego oraz zmniejszenia masowego momentu bezwładności obudowy tłumika drgań skrętnych.

8. Wnioski

Zaprezentowana metodyka mająca na celu istotne osłabienie zjawiska sprzęgania się drgań skrętnych i wzdłużnych została zastosowana w odniesieniu do pracującego OUN. W procedurze optymalizacji w przyjętej funkcji celu uwzględniono zmienne decyzyjne, które wyznaczały masę sztywnego sprzęgła kołnierzewego i masowy moment bezwładności obudowy tłumika. Należy podkreślić tu fakt, że funkcję dobrano tak, aby wdrożenie zmian konstrukcji w rzeczywistym OUN w oparciu o wyznaczone parametry, nie stwarzało trudności realizacyjnych. Uzyskane z procesu optymalizacji rezultaty potwierdziły skuteczność proponowanej metody i uzasadniają prowadzenie dalszych szczegółowych badań w tym kierunku. Bardzo istotna jest analiza wrażliwości OUN na zmianę parametrów, której wyniki mogą znacznie ułatwić zastosowanie metody na etapie projektowania.



Rys. 1. Amplitudy drgań czopa wału korbowego po optymalizacji a) wzdłużne i b) skrętne oraz odpowiednio przed optymalizacją e), f). Amplitudy drgań śruby po optymalizacji c) wzdłużne i d) skrętne i odpowiednio przed optymalizacją g), h)

Fig. 1. Vibration amplitudes of the free end of crankshaft after optimising a) axial and b) torsional and appropriate before optimising e), f). Vibration amplitudes of the propeller after optimising c) axial and d) torsional and appropriate before optimising g), h)

Literatura

1. Camisetti C., Molinari R.: *Vertical and axial vibration of a ship propulsion system*. CETENA, Quaderno no 49, pp. 16-28, Genova 1982.
2. Jenzer J.: *Some vibration aspects of modern ship installation*. New Sulzer Diesel Ltd, Switzerland, July 1996.
3. Jenzer J., Welte, Y.: *Coupling effect between torsional and axial vibrations in installations with two-stroke diesel engines*. NSD, May 1991.
4. Kleiber M.: *Komputerowe metody mechaniki ciał stałych*. Praca zbiorowa, Wydawnictwo Naukowe PWN, Warszawa 1995.
5. Kucharski T., Wittbrodt E., Wójcik T.: *A superelement for analyzing the coupling between crankshaft torsional and axial vibration*. Polish Maritime Research, December 1997.
6. van der Linden C. A. M., Hart H. H. Dolfin E. R.: *Torsional-axial vibrations of a ship's propulsion system*. International Shipbuilding Progress, vol. 16, no 173, pp. 16=26, January 1969.
7. Murawski L.: *Drgania wzdłużne układu mechanicznego, linia wałów napędowych – wał korbowy silnika spalinowego*. Praca doktorska, Gdańsk 1996
8. Schiffer W. Dr.: *Calculation of coupled vibrations*. New Sulzer Diesel Ltd, Switzerland, September 1996.
9. Tsuda K., Teon H.: *Theoretical analysis of coupled torsional - axial forced vibration of marine shafting*. Japan Shipb. & Marine Eng., vol. 4, no 5, pp. 15-30, 1969.
10. Wilczyński B.: *Optymalizacja okrętowych układów napędowych*. Praca doktorska, Gdańsk 1977.
11. Kucharski T., Wójcik T.: *Analiza wpływu parametrów konstrukcyjnych na sprzężenie się drgań skrętnych i wzdłużnych okrętowych układów napędowych*. Materiały Konferencji „Drgania i Fał”, Błażejewko 1996.

Streszczenie

Zaprezentowano metodę optymalizacji okrętowego układu napędowego w zakresie rozprężania drgań skrętnych i wzdłużnych. Podano wektor zmiennych decyzyjnych i funkcję celu. Do rozwiązania problemu optymalizacji zastosowano metodę rozszerzonej wewnętrznej funkcji kary. Przykład obliczeniowy wykonano dla pracującego na statku układu napędowego.

OPTIMUM DESIGN OF THE SHIP PROPULSION SYSTEM TO UNCOUPLE TORSIONAL AND AXIAL VIBRATION

Summary

Optimum Design of the ship propulsion system to uncouple torsional and axial vibration is given. The design variable vector and cost function is established. The penalty function method is used. To test the realism of the method the real ship propulsion system design problem is treated.

Referat recenzowany

APLIKACJA WSPOMAGAJĄCA DOBÓR GENERATORÓW DLA SIŁOWNI OKRĘTOWEJ

1 Wprowadzenie

Systemy ekspertowe znajdują coraz szersze zastosowanie w rozwiązywaniu różnorodnych zadań praktycznych. Z analizy procesu projektowania maszynowni statku [1] wynika, że jest to proces, który można wspomóc poprzez system ekspercki z bazą wiedzy. Proces ten charakteryzuje się dużą złożonością wynikającą z następujących powodów:

- duża różnorodność rozwiązań systemu energetycznego,
- układy automatyki spełniają różnorodne funkcje,
- wymaga się od systemu wysokiej niezawodności,
- układy automatyki oferowane przez poszczególne firmy ulegają poważnym zmianom, z powodu rosnących wymagań i szybkiego rozwoju techniki w tym zakresie,
- proces projektowania automatyki jest ściśle powiązany z procesem projektowania całego statku,
- proces projektowania musi uwzględniać wymagania: armatora, towarzystw klasyfikacyjnych i międzynarodowych konwencji oraz możliwości stoczni.

Decyzje podejmowane w trakcie realizacji projektów i przyjmowane rozwiązania bazują na niesformalizowanej wiedzy wynikającej z intuicji i doświadczenia projektantów (ekspertów). Bardzo często proces projektowania sprowadza się do adaptacji istniejących już rozwiązań przy uwzględnieniu wymagań klienta. Stworzenie systemu ekspertowego [2], [3], [4], [5], [6] z bazą wiedzy umożliwiającą tworzenie nowych projektów na podstawie już istniejących, oraz zawierającego informacje na temat możliwych do zastosowania układów automatyki wydaje się być rozwiązaniem umożliwiającym skrócenie czasu wykonywania projektu.

W referacie przedstawiono przebieg procesu projektowania układów automatyki dla siłowni statku. Przedstawiono także ogólne założenia i wymagania dla systemu ekspertowego z bazą wiedzy. Efektem prowadzonych prac ma być stworzenie programu komputerowego wspomagającego projektowanie systemów automatyki siłowni okrętowej. Pierwszym elementem projektowanego systemu jest aplikacja wspomagająca obliczenie bilansu energetycznego statku.

2 Proces projektowania układów automatyki maszynowni statku

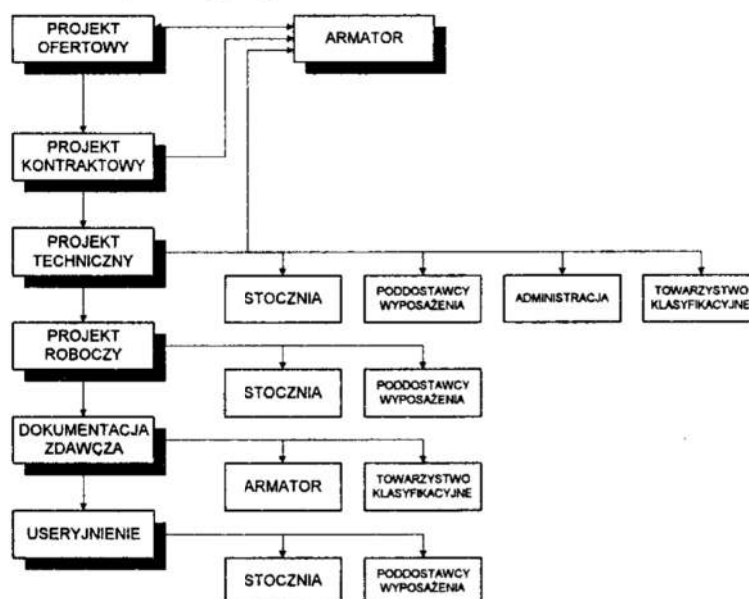
Proces projektowania [1], [2], [3] układów automatyki maszynowni okrętowej można podzielić na następujące etapy (rys. 1):

- Projekt ofertowy. W ramach tego projektu zostaje wykonana dokumentacja ofertowa zgodna z wymaganiami klienta (ogólna koncepcja i zakres automatyki siłowni okrętowej).
- Projekt kontraktowy. W ramach tego projektu następuje dokładna specyfikacja urządzeń i systemów. Określa się między innymi: zakres automatyzacji (z klasą automatyki uzgodnioną z towarzystwem klasyfikacyjnym), możliwe do zastosowania urządzenia i preferowanych producentów.
- Projekt techniczny. W ramach tego projektu opracowywana jest następująca dokumentacja: schematy AKPiA (aparatura kontrolno pomiarowa i automatyka), wykazy AKPiA, rozmieszczenie stanowisk kontrolnych, schematy zasilania urządzeń AKPiA, listy sygnałów WE/WY do sterowników, regulatorów itp., schematy systemów siłowni, schematy obwodów iskrobezpiecznych. Przygotowywane są zapytania ofertowe dla poddostawców. Dokonuje się w nich specyfikacji wymagań odnośnie

* Politechnika Gdańska, Wydział Elektrotechniki i Automatyki, Katedra Automatyki,
ul. Narutowicza 11/12, 80-952 Gdańsk, E-mail: kasiaw@pg.gda.pl

zamawianych u firm trzecich systemów i urządzeń. Opracowywane są wykazy mechanizmów, urządzeń i elementów automatyki dla działu logistyki. Projekt techniczny podlega akceptacji w towarzystwie klasyfikacyjnym jak i u klienta.

- Projekt roboczy. W ramach tego projektu wykonuje się szczegółową dokumentację układów automatyki. Podstawą do wykonania niniejszej dokumentacji są uwagi towarzystwa klasyfikacyjnego, dodatkowe wymagania klienta, dokumentacja techniczna pochodząca od wybranych poddostawców.
- Dokumentacja zdawcza. Po zakończeniu budowy i odbiorze statku przez klienta następuje faza przygotowania dokumentacji techniczno-ruchowej. Dokumentacja ta przedstawia rzeczywistą sytuację na statku. Przekazywana jest ona do klienta oraz do towarzystwa klasyfikacyjnego.



Rys. 1. Fazy projektowania statku
Fig. 1. The phases of the ship design

Bilans energetyczny statku jest obliczany na poziomie projektu ofertowego. Wszystkie urządzenia elektryczne znajdujące się na statku dzielone są na grupy. W każdej grupie znajdują się odbiory, dla których podawane są następujące dane:

- ilość,
- moc znamionowa,
- moc pobrana.

Dla tych danych obliczana jest sumaryczna moc pobrana dla następujących warunków pracy statku:

- jazda w morzu,
- postój w porcie,
- przeładunek w porcie,
- manewry,
- awaria.

Ilość warunków pracy zależy od rodzaju statku i może ulegać zmianie. Następnie dla tak obliczonych mocy dla poszczególnych grup odbiorów i dla poszczególnych warunków pracy statku tworzony jest bilans energetyczny statku, który jest podstawą do doboru ilości i mocy zainstalowanych prądnic.

3 Tworzenie systemu ekspertowego

Opracowanie systemu ekspertowego z bazą wiedzy, przeznaczonego do zastosowania praktycznego, jest procesem żmudnym i pracochłonnym. Właściwa jego realizacja wymaga wykonania następujących zadań:

- określenie zadania,
- opracowanie prototypu systemu,
- testowanie systemu,
- wdrożenie systemu.

Określenie zadania obejmuje między innymi analizę dziedziny, w której tworzony system ma być zastosowany, w tym przypadku wiąże się to z możliwością zdobycia wiedzy w rozważanej dziedzinie oraz co w dzisiejszych czasach jest najważniejsze istnienia zapotrzebowania na dany produkt.

Opracowanie prototypu systemu zawiera w sobie tak naprawdę fizyczną realizację systemu. Należy, więc przede wszystkim zaprojektować interfejs użytkownika (łatwość obsługi systemu i sposobu prezentacji wyników). Zastanowić się nad sposobem zapisywania i przechowywania danych. Należy też podjąć decyzję, co do samego systemu ekspertowego. Należy, więc ustalić sposób uzyskiwania i przedstawiania wiedzy eksperta, metody reprezentacji wiedzy w systemie ekspertowym (reguły, ramy, obiekty, itp.), efektywności działania systemu (metody wnioskowania).

Opracowanie i testowanie prototypu dają podstawę do rozpoczęcia realizacji systemu docelowego. Przebieg postępowania jest taki sam jak w przypadku realizacji prototypu.

Wdrożenie systemu ekspertowego w środowisku użytkownika wiąże się z wieloma problemami (udział użytkownika w pracach wydatnie umożliwia zmniejszenie liczby problemów). Należy, więc przede wszystkim przetestować i przebadać system na tym etapie.

W oparciu o szczegółową analizę zakresów czynności realizowanych na poszczególnych etapach projektowania założono, że w pierwszym etapie powstanie system akwizycji danych, oraz system ekspertowy umożliwiający dobór generatorów dla siłowni okrętowej. W ramach realizacji pierwszej fazy należało:

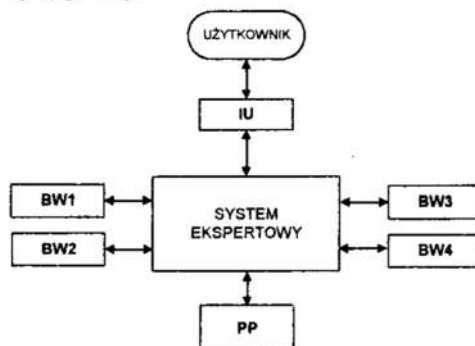
- zebrać informację na temat procesu doboru ilości generatorów w siłowni okrętowej,
 - przeszukać bazy danych dotyczące projektów już wykonanych.
- Założono, że system powinien realizować następujące funkcje:
- w przypadku braku pewnych rozwiązań wśród zgromadzonych projektów system powinien sugerować rozwiązanie,
 - w przypadkach wystąpienia dalszych braków uniemożliwiających podjęcie decyzji lub wystąpienia rozwiązań wariantowych system powinien zwracać się o rozstrzygnięcie do użytkownika,
 - system powinien prezentować rozwiązania częściowe na przykład dla aktualnej grupy odbiorców,
 - w trakcie wnioskowania (szukania rozwiązania) system powinien uwzględniać wymagania towarzystwa klasyfikacyjnego,
 - generowanie wynikowej dokumentacji projektowej w postaci: zestawień i innych dokumentów tekstowych oraz danych (baz danych) do wykorzystania przez inne systemy.

Przeprowadzona wstępna selekcja najbardziej rozpowszechnionych systemów ekspertowych wytypowała spośród wielu dostępnych aplikacji, dwa programy: EXSYS Professional firmy MultiLogic [7] i KAPPA-PC firmy IntelliCorp [8]. Obydwa te systemy pracują w systemie Windows. Po analizie obydwóch systemów wydaje się, że system KAPPA-PC [8] będzie lepiej odpowiadał stawianym wymaganiom. System KAPPA-PC oparty jest o podejście obiektowe, dające bardzo bogate możliwości prezentacji wiedzy. Posiada także różnorodne metody wnioskowania oraz język programowania. Istnieją mechanizmy współpracy z innymi programami, w szczególności z oprogramowaniem baz danych. Istnieje możliwość realizacji połączenia pomiędzy aplikacjami Windows poprzez wykorzystanie DDE.

4 Propozycja realizacji systemu

Ogólna struktura systemu

Realizacja wyżej wymienionych funkcji wymagać będzie złożonej struktury systemu eksperckiego z bazą wiedzy (rys. 2).



Rys. 2. Struktura systemu ekspertowego z bazą wiedzy

Fig. 2. The structure of the expert's system with knowledge base

Przyjęto następujące oznaczenia:

- IU – interfejs użytkownika, pełniący rolę modułu sterującego,
- BW1 – baza wiedzy zawierająca dane z poprzednich projektów,
- BW2 – baza wiedzy zawierająca przepisy towarzystw klasyfikacyjnych i inne wymagania,
- BW3 – baza danych zawierająca informacje o urządzeniach i układach automatyki,
- BW4 – baza wiedzy powstająca na podstawie schematów AKPiA
- PP – programy pomocnicze.

Realizacja problemu

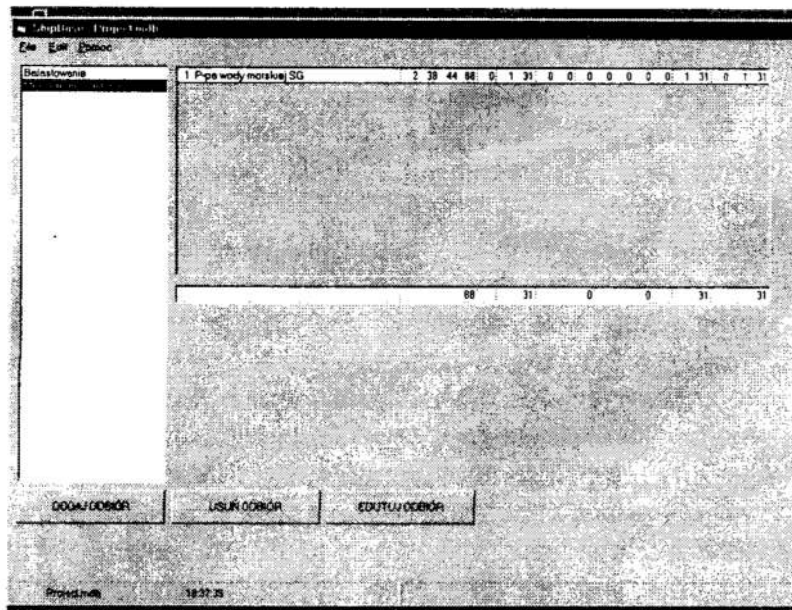
Opracowanie i implementacja ekspertowego systemu wspomagającego dobór ilości generatorów w siłowni okrętowej wymagała przeprowadzenia szeregu prac.

W pierwszej kolejności należało zapoznać się z procesem. Zidentyfikowano specyficzne cechy procesu doboru liczby generatorów. Zdecydowano się na następujące określenie danych: każdy projekt to osobna baza Accessow'a, w której poszczególne tabele odpowiadają poszczególnym grupom odbiorów. Każda tabela zawiera jednakową liczbę pól. Przyjęto, że jeżeli jakiś tryb pracy statku nie występuje to wpisujemy w odpowiadające mu komórki zera. Dodatkowo tworzona jest tabela z bilansem energetycznym. Cały interfejs użytkownika został napisany przy użyciu programu Microsoft Visual Basic 6.0. Przykładowy wygląd ekranu użytkownika jest przedstawiony na rysunku 3. Napisana aplikacja umożliwia:

- edycję utworzonych już projektów, o ile są one zgodne z przyjętymi założeniami,
- tworzenie nowych projektów.

Aby ułatwić jak najbardziej pracę przyszłym użytkownikom, wprowadzono możliwość kopiowania całych projektów, wybranych grup z poszczególnych projektów, czy też pojedynczych odbiorów. W programie dodano procedury sprawdzające poprawność wprowadzanych danych tak, aby ograniczyć do minimum możliwość popełnienia błędu przez użytkownika. Po wpisaniu danych możliwe jest uruchomienie procedury obliczającej bilans energetyczny statku. Istnieje także możliwość wydrukowania interesującego nas projektu. Przykładową stronę gotową do wydruku przedstawiono na rysunku 4. Następnym działaniem będzie dodanie systemu ekspertowego umożliwiającego dobór

ilości generatorów. W tym celu konieczne jest stworzenie bazy danych uwzględniającej oferty firm produkujących generatory.



Rys. 3. Program ShipBase
Fig. 3. ShipBase application

Rys. 4 Bilans energetyczny
Fig. 4 Electric energy balance

5 Zakończenie

Realizacja przedstawionej koncepcji systemu wspomagającego dobór liczby generatorów dla siłowni okrętowej umożliwi znaczne przyspieszenie prac. Utworzenie bazy danych ułatwi gromadzenie doświadczeń, systematyzowanie i porównywanie informacji o już wykonanych projektach. Napisana aplikacja nie wprowadza żadnych ograniczeń, co do dalszej rozbudowy.

Literatura

1. Piotrowski J., Dziworski J., Kozłowska A.: *Analiza procesu projektowania układów automatyki statku, założenia dla systemu z bazą wiedzy*. Politechnika Gdańska, Wydział Elektrotechniki i Automatyki. Gdańsk, 1997.
2. Kowalski Z., Arendt R., Zieliński S., Koleśnikow A.: *Koncepcja systemu z bazą wiedzy wspomagania projektowania układów automatyki okrętowej*. Materiały III Krajowej Konferencji Naukowej „Inżynieria wiedzy i systemy ekspertowe”. Wrocław. 10–12 Czerwca 1997. Oficyna Wydawnicza PW. Wrocław 1997.
3. Kowalski Z., Arendt R., Zieliński S., Koleśnikow A.: *Systemy z bazą wiedzy dla wspomagania projektowania układów automatyki statków*. Materiały Konferencji Naukowo–Technicznej „Automatyzacja – nowości i perspektywy”. Warszawa 1998.
4. Traczyk W.: *Systemy eksperckie – zasady działania i możliwości*. Referat wygłoszony na seminarium naukowym. 28–30.07.1997. Wydział Elektrotechniki i Automatyki PG.
5. Majewski W.: *Kierunki rozwoju systemów komputerowego wspomagania projektowania statków*. XVII sesja Naukowa Okrętowców. Jurata. Maj 1996.
6. Żyłkowski.: *Zintegrowany system komputerowego wspomagania projektowania i produkcji statków*. XVII sesja Naukowa Okrętowców. Jurata, maj 1996.
7. Zielinski S.: *Analiza możliwości systemu EXSYS Professional*. Politechnika Gdańska. Wydział Elektrotechniki i Automatyki. Gdańsk. Listopad 1997.
8. Kosmowski K.: *Analiza możliwości oprogramowania KAPPA-PC w tworzeniu systemów ekspertowych*. Politechnika Gdańska. Wydział Elektrotechniki i Automatyki. Gdańsk. Grudzień 1997.

Streszczenie

Projektowanie układów automatyki siłowni statku należy zaliczyć do zadań bardzo złożonych. Rozwiązywanie tego typu zadań można znacznie uprościć poprzez wykorzystanie systemu ekspertowego z bazą wiedzy. W pracy przedstawiono etapy projektowania układów automatyki siłowni statku oraz przedstawiono pierwsze wyniki realizowanego projektu. Zaprezentowano ogólną koncepcję systemu oraz omówiono planowany sposób jej realizacji.

APPLICATION FOR SUPPORTING SHIP ELECTRIC ENERGY BALANCE

Summary

A design of the automation for ship power system belongs to very complicated problems. Solution of such problems can be much easier when the Expert's system with knowledge-based system is used. The process of ship power system automation design is presented. The results obtained by using the presented system are discussed. The general concept of the expert's system with knowledge-based system is presented and discussed.

Referat recenzowany

ESTYMACJA OPTIMALNYCH PARAMETRÓW ALGORYTMU KOREKCJI OBRAZU W OPARCIU O ODEBRANY SYGNAŁ

1. Wprowadzenie

Przetwarzanie i transmisja sygnałów cyfrowych znajduje coraz powszechniejsze zastosowanie. Podstawowym powodem wykorzystywania tej postaci danych jest swoboda, z jaką można je przekształcać oraz ich duża podatność na kompresję. Jednak dane w postaci skompresowanej, jakie najczęściej poddaje się transmisji, są szczególnie czułe zakłócenia. Konieczne jest zatem stosowanie technik korekcyjnych, spośród których szczególne znaczenie mają rozwiązania klasy maskowania błędów. Zasadniczym problemem przy tego rodzaju korekcji jest dobór parametrów algorytmu maskowania błędów, tak by uzyskać możliwie największą skuteczność operacji.

2. Hybrydowy algorytm maskowania błędów transmisji

Omawiany hybrydowy algorytm maskowania błędów transmisji [4], [5] stanowi efekt prac nad poprawieniem jakości procesu detekcji i maskowania błędów. Zastosowanie techniki hybrydowej wpływa na istotne podniesienie skuteczności operacji.

Hybrydowy algorytm maskowania błędów podzielony jest na pięć etapów:

- wstępna filtracja danych,
- detekcja wzorców błędów,
- detekcja nietypowych współczynników,
- korekcja danych,
- końcowa filtracja danych.

Pierwszym etapem przetwarzania jest filtrowanie, usuwające współczynniki DCT, nie niosące informacji, natomiast mogące być poddane działaniu błędów transmisji. W toku przetwarzania sygnału obrazu, obejmującym dwuwymiarową transformatę DCT oraz kwantyzację współczynników, otrzymuje się kilka do kilkunastu znaczących wartości współczynników DCT, pozostała ich część w praktyce może być przyrównana do zera. Liczba niezerowych współczynników transformaty zależna jest od przyjętego rozmiaru bloku DCT oraz dynamiki danego fragmentu obrazu.

Kolejne dwa etapy mają za zadanie wykrycie prawdopodobnego położenia błędów. Pierwszy z nich bada kolejne fragmenty (bloki) obrazu, wstępnie odtworzonego do postaci przestrzennej, w poszukiwaniu wyróżnialnych wzorców błędów, charakterystycznych dla zastosowanej transformaty. Błąd obecny w danym bloku 8×8 współczynników DCT, po odtworzeniu obrazu (przez dokonanie odwrotnej transformaty współczynników) do postaci bloku 8×8 punktów zakłóca cały blok odtworzonego obrazu. Ponieważ wzorce błędów są niepowtarzalne, detekcja określonego wzorca daje jednoznaczne informacje o tym, który ze współczynników transformaty prawdopodobnie

został przetransmitowany z błędem. Współrzędne tego współczynnika zaznaczane są w macierzy błędów.

Detekcja wzorców błędów opiera się na analizie przebiegu funkcji danych w wektorach danych na obrzeżach analizowanego bloku. Ponieważ dane obrazu praktycznie nigdy nie posiadają przebiegu dokładnie pokrywającego się z wzorcem uzyskanym na drodze matematycznej, konieczne jest kwantowanie tych danych w celu przybliżenia ich przebiegu do funkcji wzorca.

Współczynnik kwantyzacji oznaczono Q . Jest to jeden z dwóch parametrów, których wartości są dobierane tak, by uzyskać najlepszy efekt korekcyjny. Drugi z nich, S , określa czułość analizy na etapie progowania współczynników.

Kolejny etap algorytmu dokonuje detekcji nieprawidłowych wartości danych obrazu w postaci współczynników DCT. Dokonuje się tego na drodze porównania różnicy między wartością badanego współczynnika a średnią wartością odpowiednich współczynników z sąsiednich bloków. Dla każdego z 64 współczynników został ustalony próg różnicy, powyżej którego dana uznawana jest za zakłóconą.

Analiza skuteczności algorytmu wykazała, iż przy korekcji obrazów silnie zakłóconych ($BER > 3\%$) lepszy efekt korekcyjny osiąga się przy zaostreniu kryterium porównawczego. Przyczyną jest fakt, iż algorytm wykrywa błędy na drodze porównywania danych z pewnego otoczenia badanego bloku. W przypadku gdy znaczna ilość danych jest uszkodzona, traci się poprawną informację odniesienia i skuteczność detekcji spada. Korzystne jest wobec tego zmniejszenie dopuszczalnego progu wspomnianej różnicy, co powoduje co prawda możliwość uszkodzenia danych poprawnych, jednak daje znaczną korzyść dzięki znacznemu zwiększeniu wykrywalności rzeczywistych błędów.

Czwartym etapem algorytmu jest korekcja, oparta na macierzy błędów, uzyskanej w poprzednich krokach. Operacja korekcji dokonywana jest na współczynnikach DCT, przez odtworzenie najbardziej prawdopodobnej wartości błędnego współczynnika w operacji uśrednienia wartości odpowiednich współczynników z sąsiednich bloków (na krawędziach obrazu dane są kopiowane w sposób lustrzany). Po odtworzeniu danych uznanych za błędne, dokonywana jest ostateczna odwrotna transformata obrazu.

Ostatni etap obejmuje filtrowanie obrazu w poszukiwaniu bloków o wyróżnialnie dużej różnicy w poziomie jasności względem otoczenia. Koryguje on błędy zakłócające pierwszy współczynnik DCT, tj. składową stałą sygnału, których to błędów nie da się wykryć przez analizę wzorców błędów.

3. Dobór parametrów korekcji

Podstawą do podjęcia decyzji o wartościach omawianych parametrów jest estymacja jakości odebranego obrazu. Do jej określenia proponuje się wykorzystanie powszechnie stosowanej miary jakości sygnału, współczynnika PSNR. Ze względu na niedostępność obrazu wzorcowego w odbiorniku przyjęto, iż jakość badanego obrazu odniesiona będzie do jakości tego samego obrazu, ale poddanego korekcji. Wynikiem operacji będzie informacja o korekcji, jaką dało zastosowanie algorytmu przy nieoptymalnych, lecz znanych parametrach. Pomierzona w ten sposób różnica daje estymatę ilości zakłóceń, jakie uszkodziły badany sygnał. Różnica zerowa oznacza brak poprawy jakości sygnału po zastosowaniu korekcji, co implikuje brak zakłóceń, zaś jej wzrost oznacza wzrost liczby błędów transmisji i pogorszenie jakości obrazu.

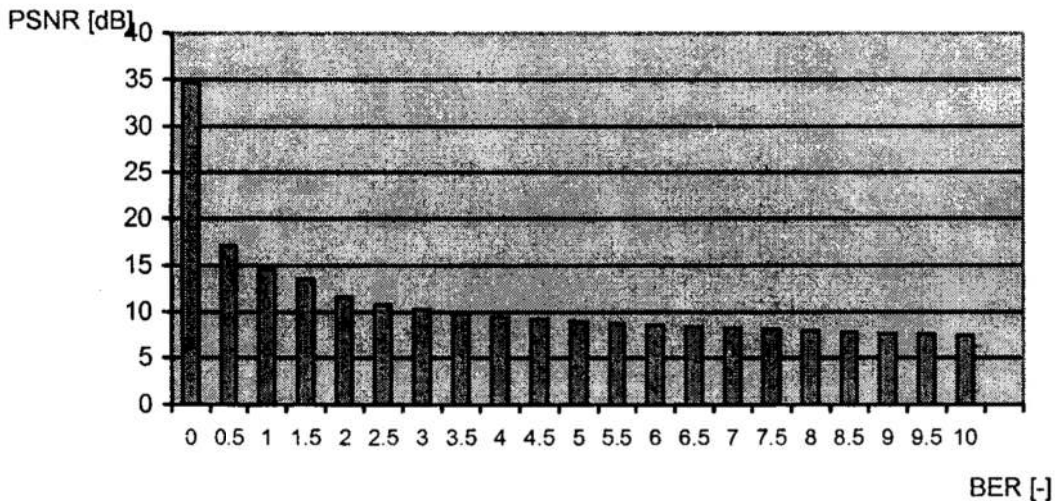
Do określenia „poprawności” obrazu stosuje się powszechnie różnicowy współczynnik PSNR, którego formuła ma postać określoną równaniem (1).

$$PSNR = 10 \log_{10} \frac{x^2 \max N^2}{\sum_{obrazu} (x - x)^2} dB \quad (1)$$

gdzie mamy N^2 elementów obrazu, x i x są odpowiednio oryginalnymi i odtworzonymi elementami obrazu.

Określenie wartości współczynnika wymaga dostępu do danych obrazu oryginalnego. Ze względu na brak możliwości dokonania takiego porównania, do estymacji jakości obrazu zastosowano porównanie jakości odebranego obrazu przed i po korekcji przez analizowany algorytm, dokonanej przy parametrach nieoptymalizowanych. Uzyskane wyniki pozwalają na określenie zależności poprawy jakości obrazu, jaką daje zastosowanie omawianego algorytmu od współczynnika błędów bitowych BER. Jej graficzną postać pokazano na rysunku 1.

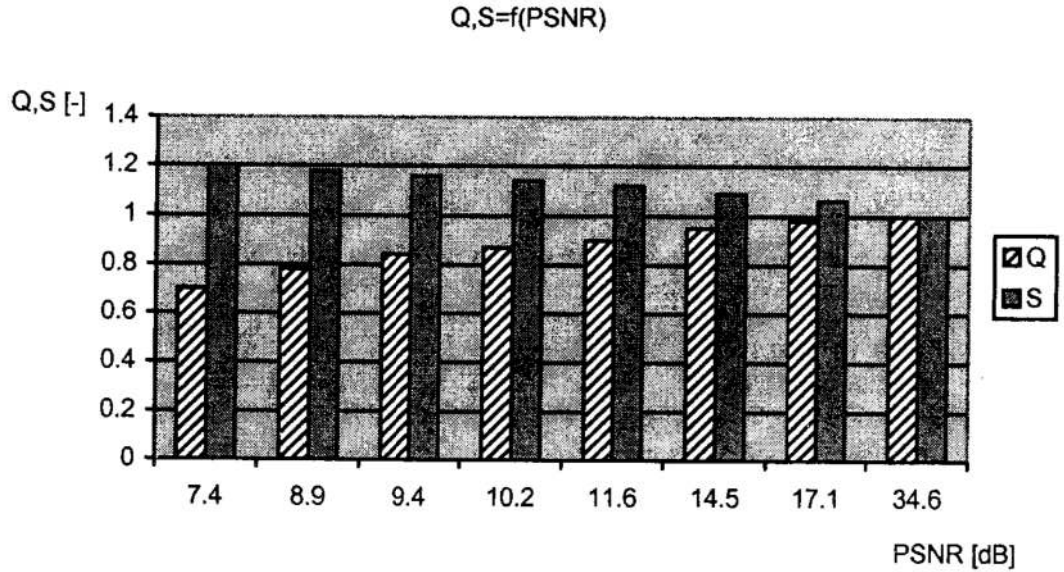
PSNR=f(BER)



Rys. 1. Zależność współczynnika poprawy jakości obrazu PSNR od współczynnika ilości błędów BER

Fig. 1. Relation of image quality improvement coefficient PSNR to BER ratio

Określenie parametrów Q i S , dających najlepsze wyniki korekcji dokonany został na drodze eksperymentu. Przebadano szereg obrazów testowych, symulując różną liczbę błędów transmisji. Tak przetworzone dane poddano korekcji hybrydowym algorytmem, ustalając różne wartości parametrów wewnętrznych Q i S oraz analizując jakość uzyskanego w taki sposób obrazu korygowanego. Uzyskana w ten sposób „tabela poprawek” wykorzystywana jest przez człon decyzyjny algorytmu korekcyjnego, który dla określonego poziomu współczynnika PSNR (stosowanego jako pośrednia miara poziomu zakłóceń sygnału) wybiera wartości parametrów, optymalne z punktu widzenia skuteczności korekcji. Graficzna postać wartości parametrów optymalnych przedstawiona została na rysunku 2.



Rys. 2. Wykres zależności optymalnych wartości parametrów wewnętrznych algorytmu korekcyjnego w funkcji wyniku korekcji wstępnej, mierzonego współczynnikiem PSNR
 Fig. 2. Relation of optimal parameters of error concealment algorithm as a function of pre-correction coefficient PSNR

Bibliografia

1. Woźnicki J.: *Podstawowe techniki przetwarzania obrazu*. Wydawnictwo Komunikacji i Łączności, Warszawa 1996
2. Skarbek W.: *Metody reprezentacji obrazów cyfrowych*. AOW PLJ, Warszawa 1998
3. Clarke R. J.: *Digital Compression of Still Images and Video*. Academic Press London 1995
4. Lynn P.A., Fuerst W.: *Introductory Digital Signal Processing with Computer Applications*. John Wiley and Sons, Chichester 1989
5. Zając W.: *A Hybrid Method of Error Elimination in Digital Image Transmission w Mixed Design of Integrated Circuits and Systems*. Proceedings of the 5th International Conference, Łódź 1998
6. Kisilewicz J., Zając W.: *An improved method of error concealment in digital image transmission*. Proceedings of the 33rd International Conference Modelling and Simulation MOSIS'99, April 99, Volume I, pp. 39-44

Streszczenie

W artykule przedstawiono propozycję rozwiązania problemu doboru parametrów wewnętrznych autorskiego algorytmu zwalczania błędów powstałych w procesie transmisji cyfrowego monochromatycznego obrazu stałego, transformowanego z wykorzystaniem dyskretnej transformaty cosinusowej (ang. Discrete Cosine Transform – DCT). W proponowanym rozwiązaniu wykorzystano estymację poziomu założeń w odebranych sygnale na drodze pośredniej, z wykorzystaniem pomiaru różnicy jakości obrazu przed i po korekcji hybrydowym algorytmem korekcyjnym, uruchomionym z parametrami nieoptymalnymi. Tak określona jakość sygnału jest podstawą do doboru parametrów optymalnych właściwej korekcji sygnału. Ustalenie tych parametrów dokonane zostało eksperymentalnie dla założonej klasy obrazów testowych.

DEFINITION OF THE OPTIMAL PARAMETERS OF AN ALGORITHM FOR CORRECTION OF IMAGES ON THE RECEIVED SIGNAL

Summary

In the paper there is presented a solution to solve a problem of optimisation of error concealment algorithm parameters. The method uses estimation of image quality by evaluation of PSNR ratio in the way of comparing the temporarily corrected image data and non-corrected data received. Estimated image quality is used to select the predefined optimal values of the algorithm internal parameters. Value of the parameters were determined in experimental way as described.

Referat recenzowany

WPROWADZENIE OGRANICZEŃ DO FUNKCJI BŁĘDU SIECI NEURONOWEJ NA PRZYKŁADZIE SYNTEZY REGULATORA NEURONOWEGO WEDŁG SCHEMATU WIDROW'A – NGUYEN'A

1 Wstęp

Rozważmy proces optymalizacji dynamicznej procesu mający na celu syntezę optymalnego regulatora ze względu na przyjęte kryteria oceny jakości regulacji. Jeśli dopuścimy optymalizację funkcji sterowania to rozwiązanie tak postawionego problemu wymaga zastosowania zaawansowanych metod analizy funkcjonalnej, podlega silnym ograniczeniom co do modelu procesu i z reguły jest bardzo pracochłonne [1]. Jeżeli jednak rozpatrywany proces możemy opisać na jednoelementowym zbiorze zmiennych niezależnych oraz dopuścimy dyskretyzację tej zmiennej (czasu), to jako narzędzia do syntezy regulatora możemy użyć procedury uczenia sieci neuronowej znanej w literaturze pod nazwą schematu Widrow'a – Nguyen'a [2, 3, 5, 7]. Zadanie sprowadza się zatem do znalezienia optymalnych wag sieci neuronowej realizującej regulator. Niestety metody optymalizacji stosowane do uczenia sieci nakładają specyficzne wymagania na funkcję błędu i nie pozwalają na wprowadzenie ograniczeń w jawnej postaci [4–6]. W tej pracy autor podejmuje próbę zdefiniowania funkcji błędu pozwalającej ominąć ten problem.

2 Definicja problemu

Na początek przypomnijmy krótko wspomnianą wyżej procedurę uczenia regulatora neuronowego [3]. Jest to procedura dwuetapowa. Pierwszy etap, wstępny, polega na syntezy neuronowego emulatora (modelu) procesu. Mając model procesu możemy przystąpić do właściwej syntezy regulatora. Dzielimy przedział czasu, w którym regulator ma wykonać zadanie, na k równych części $t = \{t_1, \dots, t_k\}$, następnie wykonujemy symulację procesu z regulatorem, z zainicjowanymi losowo wagami, oraz przygotowanym wcześniej emulatorem procesu dla zbioru wybranych P wartości zadanych $y_z = \{y_{z1}, \dots, y_{zP}\}$. Obliczamy sumaryczny błąd kwadratowy dla k -tej chwili czasu:

$$SSE = \sum_{i=1}^P (y_{z_i} - y_k)^2 \quad (1)$$

a następnie obliczamy błąd dla chwil t_{k-1}, \dots, t_1 wykorzystując własność sieci jednokierunkowych pozwalającą na propagację błędu do warstw poprzednich. Tak wyznaczony błąd służy do korekcji wag sieci. Cały proces powtarzamy tak długo jak długo błąd SSE jest większy od zadanej wartości granicznej.

Jak łatwo zauważyć mamy do czynienia z zadaniem programowania nieliniowego bez ograniczeń, w którym zmiennymi decyzyjnymi są wagi sieci pełniące funkcję regulatora. Funkcja celu (1) jest miarą statystyczną, i jedyne czego przy takiej definicji możemy być pewni to to, że średni uchyb sterowania dla wszystkich realizacji w chwili czasu t_k , będzie mniejszy niż pewna wartość graniczna. Nic natomiast nie wiemy o przebiegu funkcji sterowania i odpowiedzi emulatora obiektu dla poszczególnych realizacji, podczas, gdy wymagania jakościowe, np. dla zadania przestawiania, definiujemy zazwyczaj jako graniczne (maksymalne lub minimalne) dopuszczalne wartości kilku wskaźników

* Katedra Systemów Sterowania, Politechnika Koszalińska, E-mail: ciebielm@lew.tu.koszalin.pl.

jakości, będących funkcjonalami określonymi na zbiorze wartości zmiennych stanu $\mathbf{x} = \{x_1, \dots, x_k\}$ i wartości sterowań $\mathbf{u} = \{u_1, \dots, u_k\}$:

$$\begin{aligned} f_i(\mathbf{x}, \mathbf{u}) &\leq g_i, i = 1 : m_m \\ f_i(\mathbf{x}, \mathbf{u}) &\geq g_i, i = m_{m+1} : m \end{aligned} \quad (2)$$

Żądamy ponadto, aby pewna funkcja skalarna, określona na tychże funkcjonalach, a więc pośrednio na zbiorze wag \mathbf{W} sieci realizującej funkcje regulatora, będąca miarą jakości procesu przyjęła wartość ekstremalną (minimalną) w obszarze dopuszczonym ograniczeniami:

$$\min_{\mathbf{W}} F(f_1, \dots, f_m) \quad (3)$$

Tak sformułowana miara jakości wraz z ograniczeniami jest oczywiście deterministyczna.

Jeśli udałooby się zbudować taką funkcję F , która spełniałaby wymagania stawiane przez algorytmy optymalizacyjne stosowane do uczenia sieci neuronowych jednokierunkowych, tzn. ciągłą i posiadającą łatwo wyznaczalną wartość pochodnej, najlepiej z wartości samej funkcji, oraz zawierającą w sobie ograniczenia (2), to po zastosowaniu jej zamiast standardowej funkcji celu w procedurze Widrow'a–Nguyen'a, moglibyśmy mówić o regulatorze quasi – optymalnym ze względu na przyjęte kryteria oceny jakości regulatora.

3 Propozycja rozwiązania

Rozważmy funkcję sigmoidalną (nazywaną też funkcją logistyczną):

$$S_1(x) = \frac{1}{1 + e^{-\beta x}}, \quad (4)$$

która oczywiście spełnia warunki postawione wyżej:

$$[S_1(x)]' = \frac{-\beta e^{-\beta x}}{(1 + e^{-\beta x})^2} = \beta S_1(x)(1 - S_1(x)). \quad (5)$$

Niech $\Omega : [x_1, x_2, \dots, x_{k-1}, x_k]$ będzie zbiorem wartości, które mają spełniać warunek zapisany w jednej z trzech przedstawionych poniżej postaci:

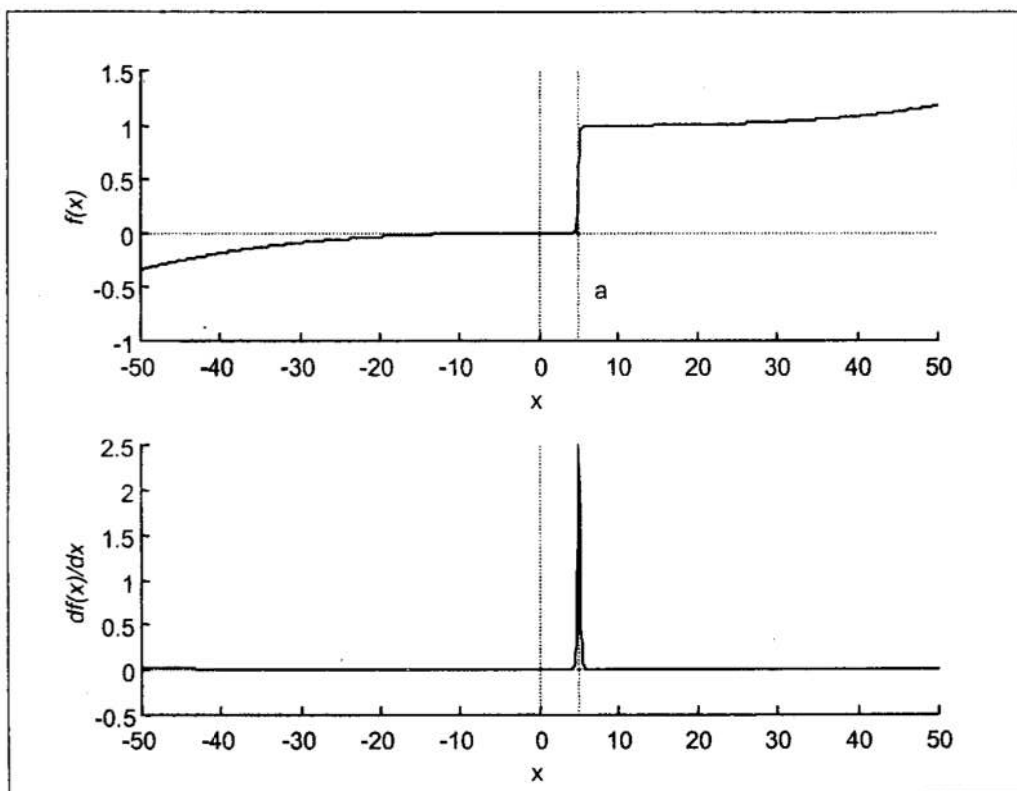
$$\forall_{x \in \Omega} x > a \rightarrow C_I(\Omega, a, \beta, \gamma) = \sum_{i=1}^k [S_1(x_i - a) + \gamma^3 (x_i - a)^3] > \frac{1}{2} \quad (6)$$

$$\forall_{x \in \Omega} x < a \rightarrow C_{II}(\Omega, a, \beta, \gamma) = \sum_{i=1}^k [S_1(a - x_i) + \gamma^3 (a - x_i)^3] > \frac{1}{2} \quad (7)$$

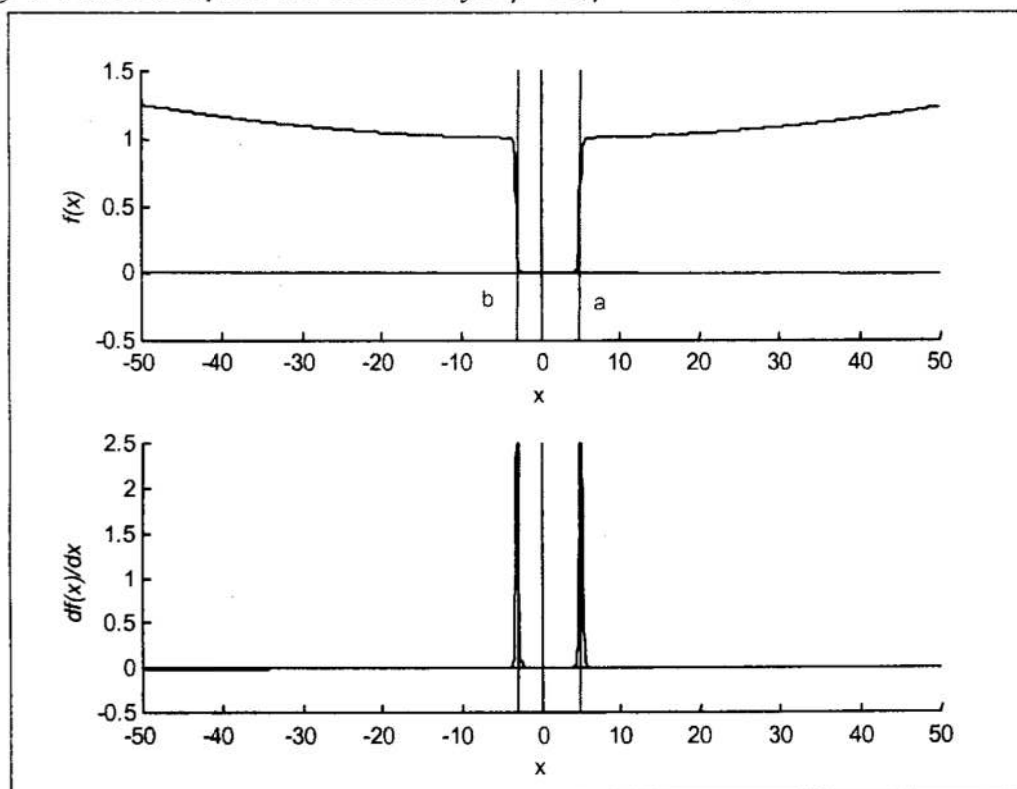
$$\begin{aligned} \forall_{x \in \Omega} b < x < a \rightarrow C_{III}(\Omega, a, b, \beta, \gamma) = \\ = \sum_{i=1}^k \left[S_1(x_i - a) + S_1(b - x_i) + \gamma^2 \left(x_i - \frac{a+b}{2} \right)^2 \right] > \frac{1}{2} \end{aligned} \quad (8)$$

Gdzie γ jest współczynnikiem z przedziału $(10^{-3}, 2 \cdot 10^{-2})$.

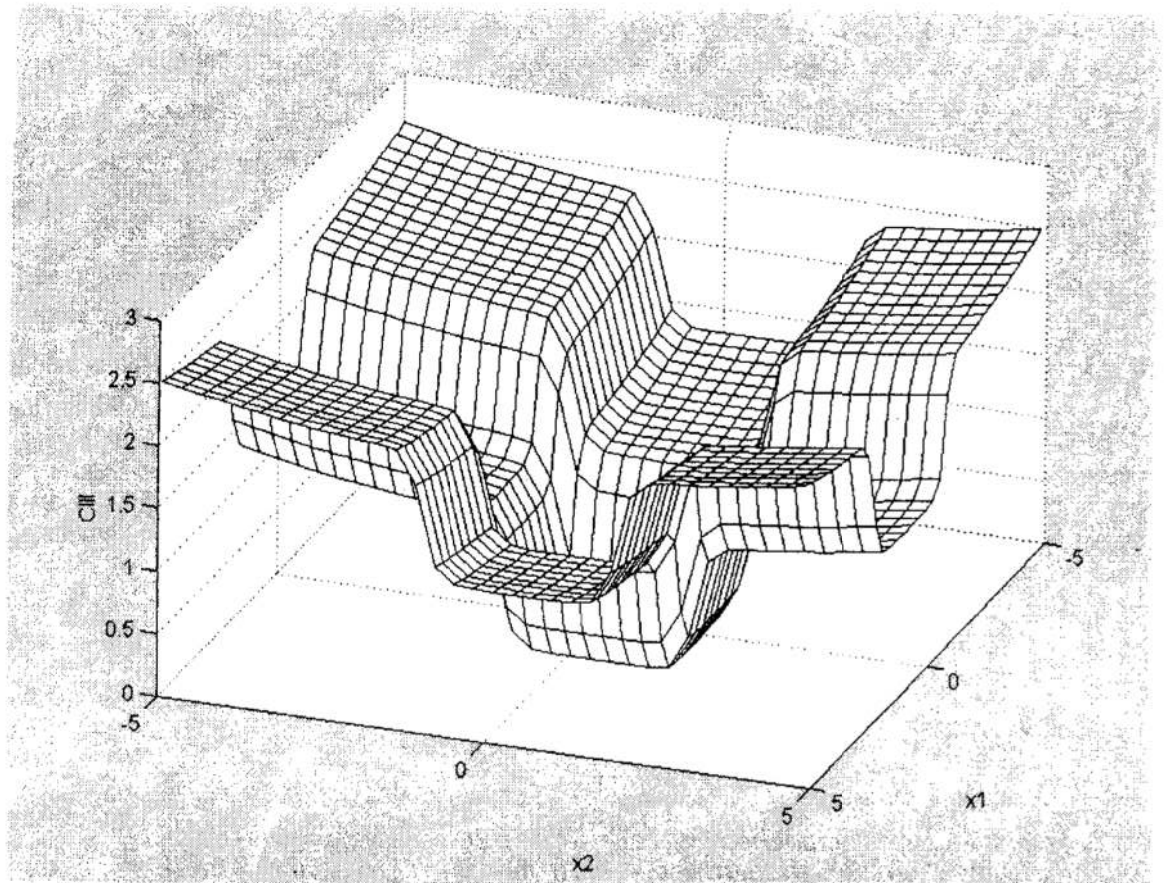
Funkcje C_I (6) i C_{III} (8) oraz ich pochodne w \mathbf{R}' są pokazane na rysunkach 1 i 2. Funkcja C_{II} (7) jest lustrzanym odbiciem funkcji C_I .



Rys. 1. Funkcja C_I oraz jej pochodna dla $\beta=10, \gamma=1.26 \cdot 10^{-2}, a=5$ w przestrzeni \mathbf{R}^1
 Fig. 1. Function C_I and it's derivative for $\beta=10, \gamma=1.26 \cdot 10^{-2}, a=5$ in \mathbf{R}^1



Rys. 2. Funkcja C_{III} oraz jej pochodna dla $\beta=10, \gamma=10^{-2}, a=5, b=3$ w przestrzeni \mathbf{R}^1
 Fig. 2. Function C_{III} and it's derivative for $\beta=10, \gamma=10^{-2}, a=5, b=3$ in \mathbf{R}^1



Rys. 3. Funkcja C_{III} dla $\beta=10, \gamma=10^{-1}, a=2, b=-1$ w przestrzeni \mathbf{R}^2
 Fig. 3. Function C_{III} for $\beta=10, \gamma=10^{-1}, a=2, b=-1$ in \mathbf{R}^2

Spełnienie warunku (7) oznacza, że każda z wartości ze zbioru Ω jest mniejsza od wartości granicznej a , z dokładnością zależną od wartości współczynnika β .

Jeśli wyznaczymy wartość średnią wyrażenia pod znakiem sumy:

$$S_1(x-a) + \gamma(x-a)^3 = \frac{1}{2k}, \quad (9)$$

to rozwiązanie ze względu na x dane jest zależnością:

$$x = a - \frac{1}{\beta} \ln \left[\frac{2m(1 + \gamma(x-a)^3) - 1}{1 - 2m\gamma(x-a)^3} \right] = a - \delta \quad (10)$$

gdzie δ jest dokładnością odwzorowania zbioru wartości dopuszczalnych.

Założmy ponadto, że:

$$\gamma \rightarrow 0^+ \text{ oraz } \eta = \frac{\delta}{a}, \quad (11)$$

gdzie η jest dokładnością względną odwzorowania zbioru wartości dopuszczalnych.

Otrzymamy wtedy :

$$\beta = \eta \cdot a \ln(2k - 1). \quad (12)$$

Zależność ta jest ważna również dla funkcji (8,9).

Przykładowa interpretacja geometryczna funkcji C_{III} jest pokazana na rysunku 3. Tworzy ona w przestrzeni argumentów R^n rodzaj studni odpowiadający zbiorowi rozwiązań dopuszczalnych. Stromość zboczy funkcji błędu powoduje, że algorytmy uczące „chętnie wpadają” do wewnątrz studni i równie niechętnie ją opuszczają.

Jeśli zapiszemy warunki nałożone na funkcjonały w postaci funkcji (7–9), a następnie ponownie zastosujemy funkcję postaci (7) do zbioru Ω_{war} przetworzonych wartości funkcjonałów, to uzyskamy poszukiwaną postać funkcji celu:

$$F_{war} = C_I(\Omega_{war}, \frac{1}{2}, \beta, \gamma) \quad (13)$$

gdzie $\Omega_{war} : [q_1 C_{X_1}(\Omega_1, a_1, \langle b_1 \rangle, \beta_1, \gamma_1), \dots, q_m C_{X_m}(\Omega_m, a_m, \langle b_m \rangle, \beta_m, \gamma_m)]$.

Współczynniki q_i są wagami definiującymi istotność funkcjonałów składowych, C_x oznacza dowolną z trzech przedstawionych postaci warunku (7–9) zapisaną dla zbioru wartości funkcji Ω_i , określonej na dowolnym podzbiórze sumy zbiorów sterowań u i zmiennych stanu x . Współczynniki β_i określamy dla każdego warunku niezależnie z zależności (12).

Jeżeli w wyniku uczenia sieci uzyskamy wartość tak zdefiniowanej funkcji błędu mniejszą niż 0.5 to możemy być pewni, że wszystkie warunki nałożone na funkcjonały składowe są spełnione, poniżej tej wartości funkcji błędu wszystkie składowe są minimalizowane proporcjonalnie do wartości współczynników q_i .

Wyznaczenie wartości pochodnej funkcji celu bezpośrednio z wartości funkcji nie jest możliwe, możemy jednak wyznaczyć wartości pochodnych dla składowych funkcji celu z zależności (2), a następnie zsumować je z odpowiednimi wagami q_i .

Możliwe jest oczywiście zbudowanie funkcji celu o większej liczbie poziomów, na przykład wtedy, gdy grupujemy funkcjonały w podgrupy stanowiące o pewnych odrębnych właściwościach rozpatrywanego procesu. Należy jednak pamiętać o tym, że dołożenie każdego nowego poziomu ogranicza dokładność odwzorowania obszaru rozwiązań dopuszczalnych, ponieważ iloczyn wartości dokładności bezwzględnych δ nie może być mniejszy niż dokładność reprezentacji liczb zmiennopozycyjnych stosowanych w implementacji algorytmu.

Podsumowanie

Przedstawiona w pracy propozycja zapisu funkcji błędu pozwala rozwiązać postawiony na wstępie problem kosztem stosunkowo niewielkiego zwiększenia ilości obliczeń, w porównaniu z algorytmem propagacji wstecznej dla funkcji SSE i jej podobnych. Wydaje się, że można stosować ją również dla innych zagadnień, gdyż zapisanie ograniczeń w funkcji celu pozwala zastosować prostszą metodę optymalizacji bez ograniczeń w porównaniu do optymalizacji wielokryterialnej z ograniczeniami. Może też być – zdaniem autora – użyta do sprawdzania, czy przyjęte ograniczenia generują niepusty zbiór rozwiązań dopuszczalnych – wystarczy bowiem sprawdzić czy funkcja celu w zaproponowanej postaci osiąga wartości mniejsze niż 0.5 dla przyjętych przedziałów zmienności zmiennych decyzyjnych.

Literatura

1. Tarnowski W.: *Optymalizacja statyczna procesów nieliniowych*. Materiały XV Ogólnopolskiej Konferencji Polioptymalizacja i CAD, Mielno 1997
2. Nguyen D.H., Widrow B.: *The Truck Backer-Upper: An example of Self-Learning in Neural Networks*. w: International Joint Conference on Neural Networks, Washington D.C. 1989, pp. II-357 – II-365

3. Nguyen D.H., Widrow B.: *Neural Networks for Self-Learning Control Systems*. Special Issue of Neural Networks in Control Systems vol. 10, no. 3, April 1990 pp.18-23.
4. Hertz J., Krogh A., Palmer R.G.: *Introduction to the Theory of Neural Computation*. Addison –Wesley Publishing Co. Redwood City 1991.
5. Korbicz J., Uciński D., Obuchowicz A.: *Sztuczne sieci neuronowe Podstawy i zastosowania*. Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994
6. Osowski S.: *Sieci Neuronowe w ujęciu algorytmicznym*. WNT, Warszawa 1996
7. Kaźmierczak T.: *Neuronowe układy sterowania obiektami technicznymi*. Praca dyplomowa, Politechnika Zielonogórska, Instytut Robotyki i Inżynierii Oprogramowania, Zielona Góra 1994

Streszczenie

W artykule autor proponuje definicję funkcji celu pozwalającą sprowadzić zadanie optymalizacji wielokryterialnej z ograniczeniami do zadania programowania nieliniowego bez ograniczeń. Poszukiwanej postaci funkcji celu postawiono dodatkowe wymagania co do różniczkowalności, wynikające z pierwotnego zastosowania poszukiwanej funkcji jako definicji funkcji błędu dla sieci neuronowej realizującej regulator w schemacie Widrow'a – Nguyen'a. Postać funkcji uzyskana w wyniku poszukiwań może być – zdaniem autora – stosowana do dowolnych zadań optymalizacji wielokryterialnej, zarówno jako funkcja celu, jak i jako funkcja sprawdzająca realizowalność ograniczeń.

THE CONSTRAINTS INSERTING INTO A NEURAL NETWORK ERROR FUNCTION ON THE EXAMPLE OF NEURAL CONTROLLER SYNTHESIS IN THE WIDROW – NGUYEN SCHEME

Summary

In this paper the author presents some non-linear function that permits to transform a multicriteria optimisation with constraints into the optimisation without any constraints. The demanded function has to comply with any additional conditions, because of primary intends to train a neural controller on the Widrow–Nguyen scheme. Presented function may be used both as an aim function with a build in constraints, and as a test function for constraints.

Referat recenzowany

KOMUNIKATY

AUTOMATYZACJA STEROWANIA MAGAZYNEM ZBOŻOWYM ZAKŁADU MŁYNARSKIEGO

1. Wprowadzenie

Nowe warunki rynkowe, jakie wyłoniły się w ostatnim czasie, będące skutkiem gwałtownego nasilenia się konkurencji, nabierającej zasięgu ogólnoswiatowego, zmuszają przedsiębiorstwa, które chcą sprostać tym warunkom do zastosowania nowych strategii, metod i środków technicznych, pozwalających na obniżanie kosztów wytwarzania oraz osiąganie wysokiej jakości produkcji. Wysoka jakość produkcji powinna być udokumentowana w sposób nie budzący wątpliwości, a więc najlepiej taki, jaki przewidują wymagania norm serii ISO 9000. Posiadanie certyfikatu jakości wg norm serii ISO 9000 staje się coraz częstszym wymogiem w eksporcie, kooperacji z firmami zachodnimi lub przetargach. Ponadto firmy posiadające zaawansowane systemy zapewnienia jakości preferują dostawców z certyfikatami.

W pracy przedstawiono automatyzację podstawowych funkcji magazynu zbożowego zakładu młynarskiego, która ma na celu podniesienie jakości produktu finalnego - mąki, obniżenie kosztów działania magazynu oraz sprostanie tzw. procesowym wymaganiom norm serii ISO 9000 w zakresie pracy magazynu, tzn. takim, które obejmują takie operacje, jak sterowanie i nadzór nad pracą maszyn i urządzeń, rejestrowanie historii procesu oraz raportowanie produkcji.

Wyprodukowanie dobrej mąki wymaga starannego przygotowania surowca, którym jest zboże. Głównymi cechami zbóż chlebowych, decydującymi o ich ocenie technologicznej są: wartość wymiłowa ziarna i wartość wypiekowa mąki otrzymanej z tych ziaren [2].

Wysoka wartość wymiłowa ziarna zapewnia uzyskanie dobrej wydajności surowca (tzw. wysoki wyciąg mąki), przy równoczesnych niskich kosztach przemiału. Dobre właściwości przemiałowe ziarna zależą od takich czynników, jak: wilgotność ziarna, jego struktura, wykształcenie, czystość oraz niska zawartość popiołu [2]. Natomiast, wartość wypiekowa mąki zależy od właściwości fizycznych i chemicznych ziarna i należy rozumieć przez nią zespół tych właściwości, które pozwalają otrzymać ciasto zdolne do dobrego wyrośnięcia oraz maksymalną wydajność gotowego pieczywa.

W mąkach pszennych właściwość ta zależy głównie od jakości i ilości zawartego w mące glutenu, odgrywającego najważniejszą rolę w procesie tworzenia się ciasta. Gluten jest substancją białkową (ok. 80% białka), składającą się głównie z gliadyny i gluteniny, która - dzięki charakterystycznej właściwość pęcznienia w wodzie podczas zarabiania mąki z wodą - tworzy w cieście kleistą i elastyczną masę, zatrzymującą pęcherzyki dwutlenku węgla, powstającego w procesie fermentacji ciasta, co zapewnia mu porowatą, elastyczną i ciągliwą strukturę, umożliwiającą dobre jego wyrośnięcie w procesie wypieku [2].

* Politechnika Poznańska, Katedra Automatyki, Robotyki i Informatyki, ul. Piotrowo 3A 60-965 Poznań,
E-mail: jaroslaw.warczyński@ar-kari.put.poznan.pl., robertpurol@poczta.onet.pl

2. Zadania magazynu zbożowego zakładu młynarskiego

Istnieją trzy główne zadania magazynu: Pierwszym zadaniem, które nie wymaga szerszego komentarza, jest gromadzenie zboża w silosach.

Zadaniem drugim jest właściwa segregacja zboża, co jest niezbędne do wykonania zadania trzeciego, którym jest dostarczanie zboża do działu produkcji mąki. Dostarczanie to odbywa się za pośrednictwem transportera, który napełnia specjalny zbiornik produkcyjny, z którego zboże trafia bezpośrednio do przemiału.

Wyprodukowanie mąki o wysokich parametrach jakościowych wymaga, aby średnia zawartość glutenu w zbożu kierowanym do zbiornika produkcyjnego była zgodna z zapotrzebowaniem działu produkcyjnego oraz, żeby była stała wzdłuż przekroju zbiornika. Oznacza to, że zboże powinno być mieszane w taki sposób, aby w każdej warstwie zboża średnia zawartość glutenu nie odbiegała od pożądanej wartości niezbędnej dla procesu produkcyjnego. Zauważmy, że sprostanie temu zadaniu wymaga takiego gromadzenia zapasów zboża, aby umożliwiała ono ciągłe otrzymywanie mieszanki zboża o zadanej średniej zawartości glutenu. A zatem, drugim ważnym zadaniem magazynu zbożowego, obok gromadzenia zapasów do produkcji, jest segregowanie zboża w sposób, który jest dogodny z punktu widzenia zadania mieszania zboża wysyłanego do zbiornika produkcyjnego.

Należy podkreślić, że skupowanie zboża o tylko określonej zawartości glutenu jest w praktyce niemożliwe, na co składa się szereg przyczyn, takich jak: rozdrobnienie rolnictwa pociągające za sobą odmienne sposoby uprawy, różnice w jakości gleby, zróżnicowanie parku maszynowego dostawców itp.

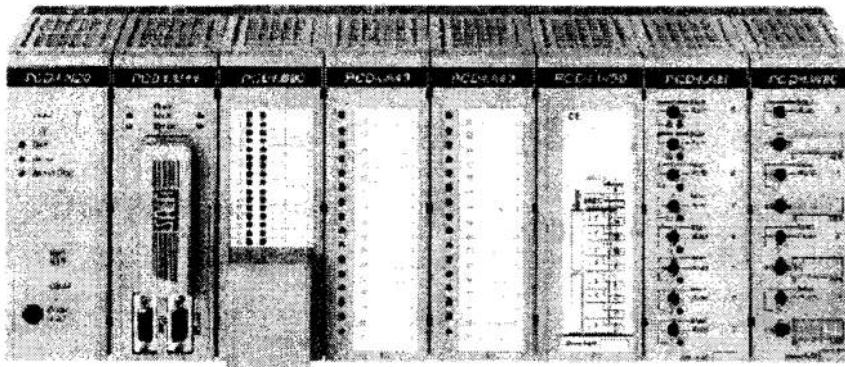
Wywiązanie się z opisanych zadań spada na obsługę magazynu. Opisany zakład młynarski, znajdujący się w miejscowości Poniec, należy do grupy zakładów o średniej wielkości: Jego magazyn składa się z sześciu silosów, a rozpiętość magazynu wynosi kilkadziesiąt metrów. W związku z powyższym, obsługa magazynu musi liczyć 5-6 osób, gdyż wszystkie niemal czynności związane ze wspomnianymi zadaniami wykonywane są ręcznie - w sposób niezautomatyzowany: Przede wszystkim, po ustaleniu przez laboratorium, zawartości glutenu w partii zboża z dostawy należy zanotować wagę oraz gluten danej partii i zdecydować, do którego silosu powinna ona być wysypana. Operator sam musi otworzyć zasuwę wlotową wybranego silosu. Na podstawie notatek operatora przyjmującego zboże do magazynu sporządza się mieszankę zboża, która jest transportowana do zbiornika produkcyjnego. Powinna ona charakteryzować się określoną średnią zawartością glutenu. Ponieważ zboże nie może zbyt długo zalegać w silosie magazynowym, technologicznie wskazane jest sporządzanie mieszanki ze zboża wysypywanego z możliwie największej liczby silosów. Przy sporządzaniu mieszanki należy notować ubytki zboża z poszczególnych silosów, jak również uwzględnić wcześniejsze notatki o rozchodzie ziarna z silosów. Wszystko to musi być dokonywane na bieżąco tak, aby zbiornik produkcyjny, z którego zboże trafia do przemiału był stale napełniony. Oznacza to, że operatorzy magazynu muszą też śledzić prędkość ubywania ziarna z tego zbiornika, żeby - z jednej strony - nie dopuścić do przestoju młynów, a z drugiej strony, aby nie przepełnić zbiornika, w przypadku spowolnienia odbioru mieszanki przez dział produkcyjny. Komplikuje to, oczywiście jeszcze bardziej pracę operatorów magazynu, gdyż nie wystarcza tutaj obliczenie proporcji mieszania zboża z poszczególnych silosów - trzeba jeszcze ustalić prędkość zsypu ziarna. Przyjmuje się,

że prędkość zsypania ziarna jest proporcjonalna do stopnia uchylenia zasuw spustowej. Ponieważ prędkość ta zależy od wilgotności ziarna, konieczne jest, co pewien czas, sprawdzanie wilgotności zboża i korygowanie stopnia otwarcia zasuw.

Biorąc pod uwagę wszystkie omówione czynniki, które trzeba uwzględnić w trakcie kierowania pracą magazynu, łatwo możemy sobie wyobrazić, że przy ręcznym sposobie obsługi nawet tylko zgrubne sprostanie stawianym zadaniom jest niezwykle trudne i nie do utrzymania w sposób ciągły, co ma oczywiście negatywny bezpośredni i pośredni wpływ na jakość produkowanej mąki oraz wydajność zakładu.

3. Automatyzacja pracy magazynu

Aby sprostać wymaganiom dostarczania do zbiornika produkcyjnego mieszanki zboża o zadanej, ustabilizowanej średniej zawartości glutenu, przyjęto, że mieszanie zboża będzie dokonywane w sposób ciągły na transporterze przenoszącym zboże do zbiornika produkcyjnego. Mieszanie to polega na zasypywaniu transportera z silosów we właściwym stosunku ilościowym, tak aby w warstwie zboża pokrywającej transporter po przejściu pod spustami wszystkich silosów, średnia zawartość glutenu była zgodna z zadaną. Wymaga to, oczywiście zapewnienia możliwości dozowania zboża z poszczególnych silosów z określoną wydajnością, która musi podlegać sterowaniu.



Rys. 1. Sterownik PLC SAIA PCD4
Fig. 1. SAIA PCD4 PLC controller

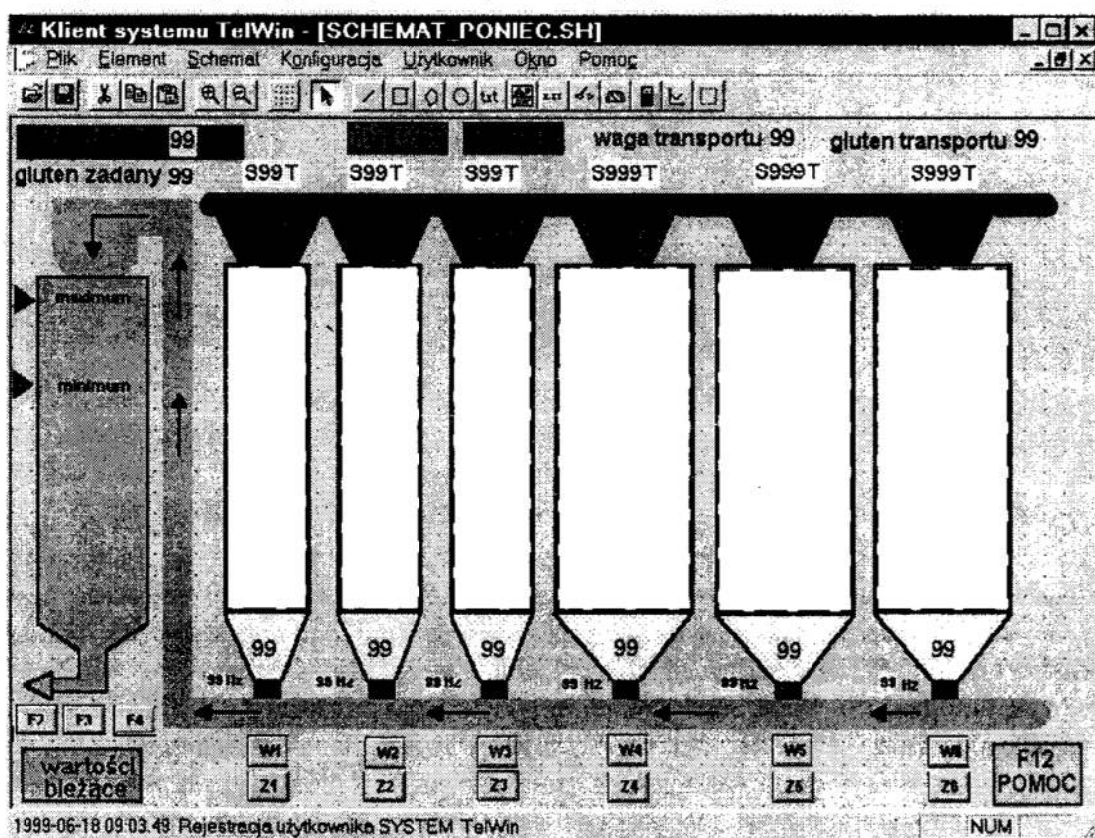
Przyjęta koncepcja automatyzacji opiera się o wykorzystanie sterownika przemysłowego SAIA PCD4 (rys. 1.) oraz systemu sterowania nadrzędnego i wizualizacji typu MMI/SCADA o nazwie *TelWin*, co pozwala na skomputeryzowanie stanowiska operatora. W kluczowych punktach magazynu zostaną wprowadzone napędy ze sterowaniem automatycznym: Zasuw górne silosów byłyby otwierane oraz zamykane poprzez siłowniki elektromagnetyczne, natomiast w miejsce zasuw dolnych zaproponowano układ wirnika z łopatkami zabierającymi, napędzanego poprzez falownik. W takim rozwiązaniu ilość zboża wysypującego się ze zbiornika jest proporcjonalna do obrotów wirnika i nie zależy od wilgotności zboża. Odpada więc konieczność korygowania nastaw, jak miało to miejsce przy ręcznej obsłudze zsypania. Prędkość obrotowa wirnika jest sterowana poprzez falownik. Rozwiązanie takie zapewnia dokładną regulację w szerokim zakresie.

Sterownik wykorzystany w systemie spełnia następujące funkcje:

- Steruje otwieraniem i zamykaniem zasuw wlotowych poszczególnych silosów.
- Steruje falownikami, napędzającymi wirniki z łopatkami zabierającymi.
- Oblicza proporcje mieszania zboża z sześciu zbiorników jednocześnie wg specjalnego algorytmu mieszania.
- Załącza cały system dozowania zboża na sygnał minimalnego poziomu zboża, przychodzący od czujnika ilości zboża w zbiorniku produkcyjnym i wyłącza go na sygnał poziomu maksymalnego, pochodzący z tego samego zbiornika.

Stanowisko operatora zawiera komputer klasy PC wraz z zainstalowanym systemem wizualizacji *TelWin* (rys. 2.), który udostępnia operatorowi następujące usługi:

- Informuje o ilości i jakości zboża w zbiornikach.
- Informuje o stanach zasuw w silosach (otwarte – zamknięte).
- Wykrywa i sygnalizuje stany alarmowe wynikające z różnych powodów np.: uszkodzeń sprzętu, błędów programowych, błędów transmisji, przekroczeń poziomów niektórych parametrów.
- Generuje raporty o operacjach wykonywanych na obiekcie.
- Pozwala operatorowi na sterowanie nadrzędne, nadzorowanie i monitorowanie procesu z jednego miejsca - za pomocą wygodnego interfejsu MMI.



Rys. 2. Ekran synoptyczny magazynu zbożowego

Fig. 2. The synoptic screen of the grain storehouse

Cały układ pozwala na w pełni automatyczną pracę magazynu od przyjęcia zboża do wysłania mieszanki do działu produkcji. Proces przyjęcia zboża do magazynu będzie teraz wyglądał następująco: Operator po otrzymaniu informacji z laboratorium o parametrach próbki dostawy wpisze je, za pośrednictwem komputera, do systemu. System dokona wyboru silosu, do którego należy wsypać zboże i otworzy odpowiednią zasuwę. Następnie, system obliczy proporcje mieszania zboża wg odpowiedniego algorytmu i na sygnał operatora rozpocznie zsymp. Gdy operator wybierze tryb automatyczny pracy systemu, mieszanie i zsymp będzie odbywał się w trybie ciągłym, a jego załączanie i wyłączanie będzie realizowane na podstawie sygnałów informujących o minimalnej bądź maksymalnie dopuszczalnej ilości ziarna w zbiorniku produkcyjnym.

Centralnym zadaniem systemu jest właściwe mieszanie zboża z poszczególnych silosów. W tym zakresie chcielibyśmy przedstawić dwa algorytmy.

Algorytm pierwszy, którego pierwsza wersja została podana w [1], umożliwia automatyczne wyznaczanie proporcji dozowania zboża z sześciu silosów magazynu, zapewniające otrzymanie mieszanki o zadanej średniej zawartości glutenu. W tym celu przyjęto, że każdy z silosów będzie gromadził zboże o trzech kolejnych zawartościach glutenu. Oznacza to, że po oznaczeniu przez laboratorium zawartości glutenu w zbożu z bieżącej dostawy operator wpisuje tę zawartość oraz wagę dostawy w odpowiednie miejsce formularza dostaw w systemie *TelWin*, a system powiadamia operatora, do którego silosu powinno zostać ono wsypane oraz sam otwiera zasuwę wlotową tego silosu, który przeznaczony jest do przechowywania ziarna o danej zawartości glutenu. System oblicza też średnią ważoną zawartość glutenu w zbożu poszczególnych silosów wg następującej zależności:

$$\bar{g}_i = \frac{\sum_j m_j g_j}{\sum_j m_j} \quad (1)$$

gdzie:

- \bar{g}_i – średnia zawartość glutenu w zbożu znajdującym się w silosie i ;
- g_j – zawartość glutenu w zbożu dostawy j skierowanej do silosu i ;
- m_j – waga dostawy j , o zawartości glutenu g_j , skierowanej do silosu i .

Następnie, do pięciu - spośród sześciu silosów magazynu - przypisuje się wagi, które będą uwzględnione w procesie mieszania ziarna. Najczęściej jest tak, że w magazynie znajduje się najmniej zboża o wysokiej zawartości glutenu, w związku z czym silos zawierający to zboże otrzyma najniższą wagę, np. równą 1. Będzie to oznaczało, że do sporządzenia mieszanki zboże to będzie wykorzystywane w małych ilościach. Silos o kolejnej, niższej już zawartości glutenu otrzymuje wyższą wagę itd. Silos z piątą kolejną zawartością glutenu otrzyma więc wagę najwyższą, np. równą 5. Waga szóstego silosu - w malejącej kolejności zawartości glutenu - zostaje obliczona zgodnie z (2):

$$w_6 = \frac{g_z (w_1 + w_2 + w_3 + w_4 + w_5) - (w_1 \bar{g}_1 + w_2 \bar{g}_2 + w_3 \bar{g}_3 + w_4 \bar{g}_4 + w_5 \bar{g}_5)}{\bar{g}_6 - g_z} \quad (2)$$

gdzie:

- \bar{g}_i – średnia zawartość glutenu w zbożu znajdującym się w silosie i ;
- g_z – zadana zawartość glutenu sporządzanej mieszanki;
- w_6 – waga obliczana;
- w_i – waga przypisana do silosu i , $i \neq 6$.

oraz

$$\begin{aligned} \Delta K_i^- \geq 0, \Delta K_i^+ \geq 0, \Delta x_i^- \geq 0, \Delta x_i^+ \geq 0 \\ x_i \geq 0, \quad v_i \geq 0 \\ \sum_i x_i = 1, \quad \sum_k v_k = 1 \\ i = 1, 2, \dots, N \quad (N - \text{liczba silosów}) \end{aligned} \quad (5)$$

gdzie:

- $K_k(\mathbf{x})$ – k -ta funkcja celu;
- x_i – poszukiwany udział zboża z silosu i w mieszance zbóż;
- $\Delta K_k^-, \Delta K_k^+$ – odchyłki ujemne i dodatnie od optymalnej wartości k -tej funkcji celu;
- g_z, w_z, c_z – optymalne wartości poszczególnych funkcji celu, np. zawartość glutenu w mieszance, wilgotność mieszanki, itp.;
- $\bar{g}_i, \bar{w}_i, c_i$ – odpowiednio, wartości średnie glutenu, wilgotności oraz innych parametrów w zbożu silosu i ;
- $x_{i\max}$ – maksymalny dopuszczalny udział w mieszance zboża z silosu i ;
- v_i – waga przypisana i -tej funkcji celu.

Przedstawiony powyżej problem programowania liniowego może być rozwiązany w oparciu o znane algorytmy, np. metodę sympleksów. Zwróćmy uwagę, że minimalizowana jest tutaj suma ważona odchyłek od optymalnych wartości poszczególnych funkcji celu. Wektor wag v pozwala hierarchizować kryteria. W ten sposób, możemy np. spowodować, aby w pierwszym rzędzie minimalizowana była odchyłka średniej zawartości glutenu w mieszance zboża, gdyż jest to najważniejszy parametr, dopiero – niejako w drugiej kolejności – minimalizowana byłaby odchyłka od pożądanej średniej wilgotności mieszanego ziarna oraz odchyłki innych jeszcze parametrów, w szczególności mogą to być np. pożądane udziały zboża z poszczególnych silosów.

Po wyznaczeniu proporcji w mieszance zboża, analogicznie jak poprzednio, są one zamieniane na sterowania dla falowników napędzających łopatkę zabierającą zboże z silosów.

4. Podsumowanie

Przedstawiony sposób automatyzacji pracy magazynu zbożowego umożliwia właściwe dozowanie zboża z poszczególnych silosów w procesie sporządzenia mieszanki o pożądanej średniej zawartości glutenu, przeznaczonej dla działu produkcyjnego. Zastosowanie drugiego spośród przedstawionych algorytmów mieszania, oprócz zapewnienia pożądanej średniej zawartości glutenu w mieszance, która jest najważniejszym parametrem zapewniającym właściwą wartość wypiekową mąki, pozwala uwzględnić również

kryteria odnoszące się do wartości wymiałowej, co umożliwi osiągnięcie wysokiego wyciągu mąki przy niskich kosztach przemiału. Reasumując, daje to możliwość dokładnego spełnienia wymaganych warunków technologicznych. Ponadto zastosowanie systemu MMI/SCADA umożliwia pełne monitorowanie i archiwizowanie przebiegu

procesów w magazynie, pozwalając na pełne odtworzenie tego przebiegu, a tym samym stwarza warunki dla pełnej identyfikowalności wyrobów. Obydwa te czynniki stanowią podstawę spełnienia wymagań norm jakościowych serii ISO 9000.

Warto też zauważyć, że automatyzacja pracy magazynu pozwoli zredukować licznosc jego obsługi, a co za tym idzie obniżyć koszty jego działania.

Literatura

1. Charnes A., Cooper W.W.: *Management models and industrial applications of linear programming*. J. Wiley & Sons, New York 1961.
2. Dłużewski M., Nowotny F., Pałasiński M. i in.: *Chemia i technologia przemysłów rolnych*. PAN, PWRiL, Warszawa 1961.
3. Kulikowski R.: *Sterowanie w wielkich systemach*. WNT, Warszawa 1974.
4. Purol R.: *Sterowanie magazynem zbożowym w miejscowości Poniec z wykorzystaniem sterownika SAIA PCD4*. Praca dyplomowa (niepublikowana) 1999.
5. Warczyński J.: *Wiekokryterialna optymalizacja rozdziału zasobów w systemach typu kompleks operacji*. Praca doktorska, Poznań 1983. (praca niepublikowana).
6. Warczyński J.: *Zautomatyzowane systemy wytwarzania*. Raport z badań (praca niepublikowana), 1998.

Streszczenie

W pracy przedstawiono sposób automatyzacji podstawowych funkcji magazynu zbożowego zakładu młynarskiego. Chodzi tutaj o takie funkcje jak: odbiór zboża z dostawy, sporządzanie mieszanki ziarna o zdanej średniej zawartości glutenu, oraz zapewnienie ciągłości transportu zboża do działu produkcyjnego. Sterowanie wymienionymi funkcjami powierzono sterownikowi PLC. Zastosowano również system MMI/SCADA do nadzoru i sterowania nadrzędnego oraz wizualizacji procesów zachodzących w ramach wspomnianych funkcji.

ON AUTOMATIC CONTROL OF A GRAIN STOREHOUSE OF A FLOUR-MILL WORKSHOP

Summary

An automation solution for a grain storehouse of a flour-mill workshop has been presented in the paper. The solution includes automating main functions of a storehouse such as grain receiving, preparing the grain mixture of the desired mean value of the gluten content, and the assurance of its transportation continuity to the mill workshop. For the purpose of automated controlling the above mentioned functions a PLC controller has been used. Also a MMI/SCADA system of integrated programs for supervising and monitoring of the automated processes has been applied.

Referat recenzowany

DYDAKTYCZNY MODEL SYMULACYJNY PROCESU MONTAŻU SAMOCHODÓW

1 Wprowadzenie

Wydział Montażu samochodów, którego model symulacyjny przedstawiono w niniejszej pracy, jest jednym z wydziałów utworzonego dla celów dydaktycznych modelu fabryki samochodów osobowych. Model ten powstał i nadal jest rozwijany w Zakładzie Robotyki i Automatyzacji Procesów Dyskretnych Instytutu Automatyki Politechniki Śląskiej.

Wydział Montażu składa się z 3 komórek produkcyjnych, które łącznie zawierają 28 centrów roboczych i 38 stanowisk roboczych. Na każdym stanowisku wykonywana jest jedna operacja, która może składać się z jednej lub kilku czynności. Do każdego stanowiska zostali przydzieleni pracownicy (monterzy, kontrolerzy oraz kierowcy) w liczbie równej liczbie czynności, które na danym stanowisku mogą być wykonywane współbieżnie.

2 Opis procesu montażu

Przed Wydziałem Montażu, za Lakiernią, znajduje się bufor, z którego pobierane są pomalowane nadwozia (Rys.1). W przypadku krótkotrwałych awarii linii montażowej w buforze tym gromadzone są nadwozia, dzięki czemu w lakierni nie ma przestojów. Podobnie w przypadku awarii lakierni z bufora tego pobierane są zmagazynowane tam nadwozia, co zabezpiecza linię montażową przed przestojem.



Rys. 1. Rozmieszczenie elementów modelu symulacyjnego Wydziału Montażu Samochodów w systemie Taylor II

Fig. 1. Layout of the Car Assembly Department in Taylor II System

Pierwszą komórką na wydziale montażu jest linia montażu samochodów poruszająca się z taktem 2 min. Na linii montażowej znajdują się 24 stanowiska montażu obsługiwane przez 28 monterów (Rys.2). Komponenty montowane w samochodzie są

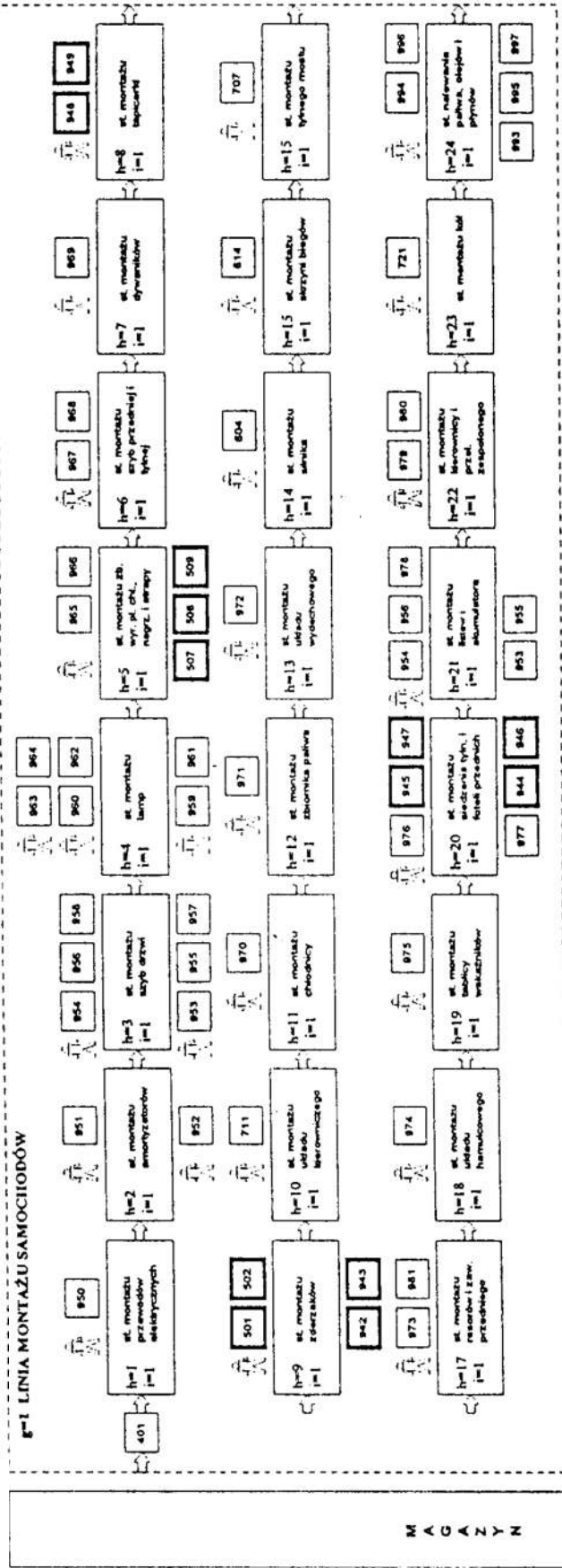
*Zakład Robotyki i Automatyzacji Procesów Dyskretnych Instytut Automatyki Politechniki Śląskiej

e-mail: mزابورowski@zeus.polsl.gliwice.pl

** e-mail: a.pusiewicz@kato.apexim.com.pl

d=8 WYDZIAŁ MONTAŻU SAMOCHODÓW

F=1 LINIA MONTAŻU SAMOCHODÓW



M A G A Z Y N

Rys. 2. Schemat przepływu produkcji przez Wydział Montażu Samochodów
Fig. 2. Material flow diagram through the Car Assembly Department

pobierane z odpowiednich buforów znajdujących się przy stanowiskach. Bufory te na schemacie oznaczono numerami indeksu materiałowego pozycji, która się w nim znajduje. Są one zasilane z magazynu wydziałowego. Na schemacie (Rys. 2) wyróżniono przez pogrubienie bufory komponentów wersyjnych, których montaż uzależniony jest od zamówienia klienta (są to: atrapa, tapicerka, zderzaki, fotele przednie).

Komórką drugą jest **dział kontroli i regulacji zespołów samochodów** (Rys. 2). W dziale tym znajdują się 3 centra robocze. W skład każdego z nich wchodzi po dwa stanowiska, przy czym jest obojętne, na którym stanowisku będzie się odbywać kontrola. Do centrów tych przydzielono po dwóch kontrolerów, z których każdy może dokonywać kontroli na dowolnym stanowisku wchodzącym w skład danego centrum. Z działu kontroli samochody trafiają do komórki **transport samochodów na plac postojowy** (Rys.2). Komórka ta formalnie zawiera jedno centrum robocze składające się z 8 stanowisk, którymi są trasy jazdy samochodem obsługiwane przez 8 kierowców. Transport samochodu może odbywać się dowolną trasą. Jest też obojętne, który z kierowców będzie dokonywał tego transportu.

3 Symulacja w systemie Taylor II

Model Wydziału Montażu utworzono wykorzystując jedynie cztery (machine, buffer, reservoir, aid) z dostępnych dziewięciu elementów. Zdecydowano się na takie posunięcie, aby maksymalnie uprościć i ujednoczyć budowę modelu. Tak więc do budowy wszystkich stanowisk użyto elementów typu machine, do budowy buforów i magazynu wydziałowego użyto elementów typu buffer oraz reservoir, natomiast pracownicy reprezentowani są przez elementy typu aid. Do zgrupowania elementów należących do tych samych komórek produkcyjnych wykorzystano klastry (Rys. 1).

Magazyn wydziałowy zbudowano z 54 elementów typu buffer oraz 5 elementów typu reservoir, oraz napelniono je (stan początkowy symulacji) taką ilością komponentów, aby wystarczyła na 24 godziny symulacji. Uniknięto w ten sposób konieczności użycia elementów typu In/Out. Magazyn ten zasila bufory znajdujące się przy stanowiskach linii montażowej. Bufory te także zapełniono, aby przyspieszyć rozruch symulacji (gdyby tego nie uczyniono, przy każdym starcie bufory te byłyby napelniane z buforów magazynu wydziałowego). Napelnianie buforów przystanowiskowych odbywa się partiami, w momencie gdy zapas spadnie w nich poniżej połowy ich pojemności nominalnej. Bufory oznaczono numerami indeksu materiałowego półproduktu lub materiału w nim znajdującego się.

Na linię montażową (Rys. 2) pomalowane nadwozia pobierane są z bufora 401. Przy opuszczaniu tego bufora nadwoziom nadawane są atrybuty, w których zapisane są opcje cech fakultatywnych montowanego samochodu. Dzięki temu na stanowiskach, na których wybór montowanego komponentu zależy od opcji, możemy przed pobraniem komponentu sprawdzić odpowiednią opcję cechy fakultatywnej (są to stanowiska: 5-atrapa, 8-tapicerka, 9-zderzaki i 20-fotele przednie). Na stanowiskach montażowych zainstalowano ikony dynamiczne (typu „tekst:wartość”) które na bieżąco monitorują przebieg symulacji wyświetlając: nr zamówienia klienta, nr wersji, nr aktualnie wykonywanej czynności w ramach operacji oraz oznaczenia opcji cech fakultatywnych aktualnie montowanego samochodu.

Na linii montażowej czasy trwania czynności zdefiniowano wykorzystując rozkłady dyskretne. Do wyznaczenia taktu linii wykorzystano cyklicznie wywoływane zdarzenia użytkownika (user events). Przed przemieszczeniem się samochodu na następne stanowisko sprawdzana jest zmienna mówiąca, czy na choć jednym stanowisku występuje awaria, dzięki czemu wystąpienie awarii na choć jednym stanowisku zatrzymuje całą linię montażową.

W czasie budowy modelu w bardzo szerokim zakresie wykorzystywano będący silną stroną systemu język TLI (Taylor Language Interface). Wykorzystano go także do napisania uruchamianego okresowo (co 1 min) programu zapisującego aktualne wyniki symulacji do tabeli czasowej, co umożliwi np. zapisanie tej tabeli w pliku i następnie importowanie jej do Excela w celu dalszej analizy.

4 Eksperymenty symulacyjne

Przedstawiony w niniejszej pracy model zaimplementowano w systemie Taylor II i przy jego wykorzystaniu przeprowadzono szereg eksperymentów symulacyjnych. Poprzez odpowiednią definicję średnich czasów międzyawaryjnych stanowisk (MTBF) oraz średnich czasów trwania awarii (MTTR) wprowadzono do symulacji losowo pojawiające się awarie linii montażowej zmniejszające średnią wydajność produkcji z 30 samochodów/godz (takt linii = 2 min) na ok. 24 samochody/godz. Badano reakcje procesu montażu na występowanie tych awarii oraz na nieobecność pracowników (monterów, kontrolerów i kierowców).

Badano między innymi: długości kolejek, czasy oczekiwania na obsługę, stopień wykorzystania stanowisk i buforów. W badaniach tych wykorzystano w szerokim zakresie możliwości systemu Taylor w zakresie wyświetlania różnego typu wykresów, diagramów, histogramów, przebiegów czasowych oraz generowania własnych raportów.

Przewiduje się wykorzystanie modelu w zajęciach dydaktycznych dla studentów specjalności Robotyka na Wydziale Automatyki, Elektroniki i Informatyki Politechniki Śląskiej.

Literatura

1. Pusiewicz A.: *Sterowanie produkcją na wydziale montażu przykładowej fabryki samochodów*. Praca dyplomowa, Instytut Automatyki Politechniki Śląskiej, Gliwice, 1999.
2. *Taylor II for Windows – Simulation software for personal computers*, F&H Simulations B.V., Utrecht, The Netherlands, 1996
3. Zaborowski M.: *Struktura danych o technologii w systemach zarządzania i sterowania produkcją*. Materiały II Ogólnopolskiej Konferencji „Komputerowo Zintegrowane Zarządzanie”, Zakopane 1999, str. 481-490
4. Zaborowski M., Żmuda Z.: *Modele struktur produktów wielowersyjnych*, Materiały II Ogólnopolskiej Konferencji „Komputerowo Zintegrowane Zarządzanie”, Zakopane 1999, str. 499-508

Streszczenie

W pracy przedstawiono dydaktyczny model symulacyjny procesu montażu samochodów osobowych i jego implementację w systemie Taylor II. Modelowany Wydział Montażu obejmuje 38 stanowisk roboczych, a jego produkt – samochód osobowy – może być wytwarzany w 24 wersjach. Przedstawiono także przykładowe eksperymenty symulacyjne.

EDUCATIONAL SIMULATION MODEL OF THE CAR ASSEMBLY PROCESS

Summary

The educational simulation model of the car assembly process and its implementation in Taylor II system are presented in the paper. Modelled Assembly Department includes 38 work stations. The car, which is its product, can be manufactured in 24 versions. The sample simulation experiments are presented too.