

**Politechnika Śląska
Wydział Mechaniczny Technologiczny
Katedra Podstaw Konstrukcji Maszyn**

Modele odwrotne i modelowanie diagnostyczne

**Wojciech CHOLEWA
Tomasz ROGALA**

Wydanie 2

Gliwice 2008

Recenzenci

prof. dr hab. inż. Jan Kiciński - Instytut Maszyn Przepływowych PAN
prof. dr hab. Wojciech Moczulski - Politechnika Śląska

Redaktor zeszytów

Wojciech Cholewa

Redaktor techniczny

Marek Wyleżoł

Projekt okładki

Wojciech Cholewa, Marek Wyleżoł

BibTeX

```
@BOOK{, title = {Modele odwrotne i modelowanie diagnostyczne}, publisher =  
{Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn}, year = {2008},  
author = {Cholewa, W. and Rogala, T.}, volume = {136}, series = {Zeszyty},  
address = {Gliwice}, edition = {2} }
```

ISBN 978-83-60759-11-0

Wydawca

Katedra Podstaw Konstrukcji Maszyn
Wydział Mechaniczny Technologiczny
Politechnika Śląska
ul. Konarskiego 18a, 44-100 Gliwice
tel. (032) 237-14-67, fax (032) 237-13-60
<https://kpk.m.polsl.pl>

Druk i oprawa (z dostarczonych materiałów):
Drukarnia cyfrowa D&D sp. z o.o. ul. Moniuszki 6, 44-100 Gliwice

Spis treści

1 Wprowadzenie	5
Bibliografia	6
2 Pojęcia podstawowe	7
2.1 Stan obiektu	7
2.1.1 Symptom	7
2.2 Eksperyment diagnostyczny	8
2.3 Model danych	8
2.3.1 Opis zadania poszukiwania modelu danych	9
2.3.2 Postać modelu	10
2.3.3 Globalny model danych	11
2.3.4 Lokalny model danych	12
2.3.5 Wybrane przykłady modeli interpolacyjnych	13
2.4 Model diagnostyczny	17
2.4.1 Relacja diagnostyczna	18
2.4.2 Użyteczne cechy sygnałów	19
2.4.3 Model odwrotny	20
2.4.4 Model wielowarstwowy	22
2.5 Sieć stwierżeń	23
2.5.1 Treść i wartość stwierdzenia	23
2.5.2 Słownik treści stwierżeń	23
2.5.3 Ziarnistość danych	23
2.5.4 Sieć przekonań	24
Bibliografia	25
3 Model BNBM	27
3.1 Diagnostyka symptomowa i diagnostyka wsparta modelami	27
3.2 Struktura modelu BNBM	28
3.3 Identyfikacja modelu BNBM	30
Bibliografia	30
4 Klasyfikatory	31
4.1 Klasyfikacja w procesie diagnozowania	31
4.2 Klasyfikatory wieloklasowe	32
4.2.1 Klasyfikator bayesowski	33
4.2.2 Liniowa funkcja dyskryminacyjna	34
4.2.3 Klasyfikator Parzena - probabilistyczne sieci neuronowe	35
4.2.4 Klasyfikator k-najbliższych sąsiadów knn	36
4.2.5 Klasyfikatory neuronowe	37
4.2.6 Klasyfikator SVM	38
4.3 Klasyfikatory jednoklasowe	39

4.3.1	Klasyfikator K-środków	39
4.3.2	Jednoklasowy klasyfikator najbliższego sąsiada	40
4.3.3	Jednoklasowy klasyfikator SVM	40
	Bibliografia	40
5	Symulator wirnika wielopodporowego	43
5.1	Prosty model wirnika z łożyskiem hydrodynamicznym	44
5.1.1	Siły hydrodynamiczne	44
5.1.2	Model	49
5.1.3	Weryfikacja poprawności modelu	50
5.2	Model wirnika wielopodporowego	51
5.2.1	Założenia	52
5.2.2	Obciążenia	54
5.2.3	Model podpory	54
5.2.4	Równania ruchu	55
5.2.5	Symulator MLH i MSR	56
	Bibliografia	57
6	Przykład zastosowania proponowanej metody	59
6.1	Dane uczące	59
6.2	Proces dekompozycji i parametryzacja ziarnistości zmiennych dodatkowych	61
6.3	Cechy użyteczne	62
6.4	Klasyfikatory jednoklasowe (BNBM - stopień I)	63
6.5	Sieć przekonań (BNBM - stopień III)	66
6.6	Optymalizacja ewolucyjna modelu BNBM	68
6.7	Uzgadnianie wartości zmiennych dodatkowych (BNBM - stopień II)	69
6.8	Wnioski	72
	Bibliografia	72

Rozdział 1

Wprowadzenie

Wojciech CHOLEWA

Głównym zadaniem diagnostyki technicznej jest rozpoznawanie stanu obiektu (którym może być również proces) na podstawie dostępnych o nim informacji. Stosowane metody takiego rozpoznawania można podzielić ogólnie na dwie klasy [1.6]:

- diagnostyka wsparta modelami,
- diagnostyka symptomowa.

Istotą *diagnostyki wspartej modelami* jest analizowanie różnic (tzw. residuów) występujących pomiędzy wynikami obserwacji obiektu i wynikami stosowania modelu dostrojonego do tego obiektu oraz umożliwiającego symulowanie działania obiektu o zadanym stanie technicznym. O adekwatności możliwych do uzyskania diagnoz decyduje w pierwszym rzędzie jakość stosowanego modelu. Metody te znalazły szerokie zastosowanie zwłaszcza w diagnostyce układów sterujących oraz w diagnostyce procesów [1.5]. W szczególności można je stosować tam gdzie określenie odpowiedniego zdeterminowanego modelu obiektu jest możliwe.

Diagnostyka symptomowa bazuje na założeniu, że skutkiem zmiany stanu obiektu powinno być powstanie objawów świadczących o nim. Większość stosowanych metod przyjmuje, że obserwowane są różne procesy resztkowe [1.1], których cechy są nośnikami informacji o stanie obiektu. Związki zachodzące pomiędzy tymi cechami i stanem obiektu formułowane są w postaci relacji diagnostycznych. Relacje te pozyskiwane na podstawie opinii i doświadczenia specjalistów oraz w wyniku badań diagnostycznych są niestety często niepewne i niedokładne. Diagnostyka symptomowa jest stosowana głównie tam, gdzie określenie wyczerpującego modelu obiektu jest trudne lub niemożliwe co powoduje, że bezpośrednie zastosowanie diagnostyki wspartej modelami nie jest możliwe.

Niniejsza monografia dotyczy zagadnień związanych z budowaniem modeli dla potrzeb diagnostyki technicznej z zastosowaniem metod i technik sztucznej inteligencji. Szczególną uwagę zwrócono na *odwracanie modeli* i wprowadzone w [1.4] pojęcie *diagnostycznych modeli odwrotnych*. Uwzględniono wyniki prowadzonych w Katedrze Podstaw Konstrukcji Maszyn Politechniki Śląskiej badań dotyczących zarówno metod odwracania modeli [1.2], [1.3], [1.9], jak i badań dotyczących wrażliwości cech diagnostycznych [1.8], wielomodeli diagnostycznych [1.10] i metod identyfikacji stanu w układach rozproszonych [1.7]. Celem badań opisywanych w tej pracy było sprawdzenie możliwości równoczesnego zastosowania metod wnioskowania, opracowanych wcześniej dla potrzeb diagnostyki wspartej modelowo oraz metod wnioskowania opracowanych dla potrzeb diagnostyki symptomowej.

Zaproponowano sposób postępowania umożliwiający równoczesne stosowanie metod występujących w obu wskazanych klasach procesów rozpoznawania stanu obiektu. Pokazano, że procesy wnioskowania w diagnostyce wspartej modelami oraz procesy wnioskowania w diagnostyce symptomowej mogą być rozpatrywane w sposób jednolity. Wskazano miejsca modeli odwrotnych w tych procesach oraz omówiono zagadnienia związane z identyfikacją takich modeli.

W monografii opisano wyniki badań wykonanych częściowo jako fragmenty projektu badawczego 4 T07B 05 28, finansowanego ze środków budżetowych na naukę w latach 2005–2008.

Bibliografia

- [1.1] Cempel Cz. *Podstawy wibroakustycznej diagnostyki maszyn*. WNT, Warszawa, 1982.
- [1.2] Cholewa W., Kiciński J., redaktorzy. *Diagnostyka techniczna. Odwrotne modele diagnostyczne*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [1.3] Cholewa W., Kiciński J., redaktorzy. *Diagnostyka techniczna. Metody odwracania nieliniowych modeli obiektów*, wolumen 120 serii Zeszyty. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2001.
- [1.4] Cholewa W., White M.F. Inverse modelling in rotordynamics for identification of unbalance distribution. *Machine Vibration*, (2):157–167, 1993.
- [1.5] Korbicz J., Kościelny J.M., Kowalczyk Z., Cholewa W., redaktorzy. *Diagnostyka procesów. Modele, metody sztucznej inteligencji, zastosowania*. WNT, Warszawa, 2002.
- [1.6] Natke H.G., Cempel Cz. *Model-Aided Diagnosis of Mechanical Systems*. Springer, Berlin, 1997.
- [1.7] Rzydzik S. *Identyfikacja stanu w układach rozproszonych*, wolumen 135 serii Zeszyty. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2007.
- [1.8] Sławik D. *Metody badania wrażliwości cech sygnałów diagnostycznych*, wolumen 126 serii Zeszyty. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, 2006.
- [1.9] Urbanek G. *Zastosowanie algorytmów ewolucyjnych w identyfikacji odwrotnych modeli diagnostycznych*, wolumen 132 serii Zeszyty. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.
- [1.10] Wojtusik J. *Wielomodele diagnostyczne maszyn wirnikowych*, wolumen 128 serii Zeszyty. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.

Rozdział 2

Pojęcia podstawowe

Wojciech CHOLEWA

Diagnostyka techniczna jest stale rozwijającą się dyscypliną. Różnorodne jej zastosowania, realizowane przez niezależnie działające zespoły, są przyczyną braku spójności używanej terminologii, zarówno w literaturze światowej jak i w publikacjach wydawanych w języku polskim. Zestaw zaleceń terminologicznych został zaproponowany między innymi przez *Komitet Techniczny SAFEPROCESS* międzynarodowej organizacji IFAC. Propozycje bazujące na tych zaleceniach i uwzględniające w szczególności diagnostykę procesów, zamieszczono w [2.10]. Obszerny zbiór stosowanych pojęć i ich definicji, zestawionych w układzie encyklopedycznym, zamieszczono w [2.26] .

Wybrane pojęcia omówiono w kolejnych podrozdziałach.

2.1. Stan obiektu

Pojęcie *stan* jest rozpatrywane przez wielu autorów. Jego miejsce w diagnostyce technicznej pokazano w [2.2]. W [2.3] podjęto próbę zdefiniowania pojęć *stan maszyny* oraz *obraz stanu maszyny*. Zestawienie cytatów różnych definicji tego pojęcia można znaleźć w [2.26]. Ciekawą dyskusję pojęcia *stan techniczny* zamieszczono w [2.19].

Rezygnując z nadmiernych formalizmów i zakładając, że interesujące nas własności obiektu opisane są jego cechami nazywanymi dalej *cechami stanu* można przyjąć, że aktualny stan obiektu jest wektorem wartości jego cech stanu. Tak rozumiany stan może być rozpatrywany z różnym stopniem szczegółowości. Należy w tym miejscu wyraźnie zaznaczyć, że założenie o możliwości rozważania cech stanu i ich wartości nie oznacza, iż cechy te są obserwowalne i nie oznacza, iż ich wartości można zmierzyć lub bezpośrednio wyznaczyć. W większości zastosowań nie występuje potrzeba rozpatrywania dokładnych wartości cech stanu. Wystarcza ograniczenie do klas stanu (np. silnie zużyta opona samochodowa) bez konieczności identyfikacji pełnego wektora cech stanu.

Oznacza to, że wprowadzone pojęcie *cecha stanu* służy jedynie do przyjęcia zrozumiałej i posiadającej prostą interpretację definicji pojęcia *stan*. Wartości tych cech najczęściej nie są identyfikowane bezpośrednio.

2.1.1. Symptom

Wprowadzenie ogólnej definicji pojęcia *symptom* jest zadaniem trudnym. Czynione często (jawnie lub domyślnie) założenie, że symptomem jest jakaś cecha sygnału diagnostycznego lub, że jest to miara będąca odwzorowaniem postępującego zużycia badanego obiektu może prowadzić do nieporozumień.

Formalizując to pojęcie należy przyjąć, że symptom jest objawem czegoś. W zakresie diagnostyki technicznej symptom jest objawem występowania jakiegoś stanu lub stanów obiektu. Użycie w tym

miejsu liczby mnogiej (stanów) uzasadnione jest tym, że różne stany mogą objawiać się takimi samymi, nierozróżnialnymi symptomami. Oznacza to, że słowo *symptom* powinno być rozumiane jako objaw występowania klasy lub klas stanu ponieważ, jak wspomniano wcześniej, w większości przypadków wystarcza rozpoznawanie stanu z dokładnością do klasy stanu, bez potrzeby rozpoznawania wartości cech stanu.

Wygodnym ujednoczeniem rozważań dotyczących różnych możliwych postaci symptomów jest przyjęcie założenia, że symptomem określonej klasy stanów jest przynależność chwilowych wartości cech sygnałów diagnostycznych do odpowiedniej *klasy wartości cech sygnałów*.

2.2. Eksperyment diagnostyczny

W diagnostyce technicznej od wielu lat znane są dwie ogólne klasy badań, pozwalających na identyfikację relacji diagnostycznych [2.2]:

- *eksperymenty czynne*, polegające na zamierzonym wprowadzaniu niesprawności (zadanych stanów) w badanym obiekcie, w celu wykrycia zewnętrznych objawów (symptomów) występowania tych niesprawności,
- *eksperymenty bierne*, polegające na obserwowaniu obiektu, którego stan zmienia się niezależnie od eksperymentatora, tzn. bez możliwości zamierzonego modyfikowania jego stanów technicznych, w celu zgromadzenia odpowiednich par *stan-symptom*.

Wymienione klasy badań pozwalają na identyfikację relacji występujących między stanem technicznym i jego symptomami. Eksperymenty bierne wymagają dużej liczby obiektów lub bardzo długiego czasu badań. Eksperymentów czynnych nie można stosować na rzeczywistych, dużych obiektach technicznych (np. turbozespoły), ze względu na wysoki koszt i niedopuszczalny stopień ryzyka.

W ostatnim okresie nastąpił intensywny rozwój metod numerycznego modelowania obiektów technicznych i związanych z nimi procesów. Wynikiem modelowania są modele przyczynowo skutkowe, które pozwalają na wyznaczenie skutków działania obiektu (np. drgań) dla założonego stanu technicznego oraz dla zadanych cech konstrukcyjnych obiektu i warunków jego działania. Rozwinięto techniki strojenia modeli numerycznych, czyli techniki ich dopasowywania do wyników obserwacji działania obiektów rzeczywistych. Wysoka jakość odpowiednio strojonych modeli pozwala na ich stosowanie jako źródeł danych w procesie pozyskiwania wiedzy diagnostycznej. Eksperymenty czynne dotyczą badań prowadzonych na obiektach rzeczywistych. Szczególną klasę badań stanowią

- *eksperymenty symulacyjne*, czyli eksperymenty czynne wykonywane na modelach obiektów. Modele te przekształcają zadane przyczyny (stany) w odpowiednie ich skutki (symptomy).

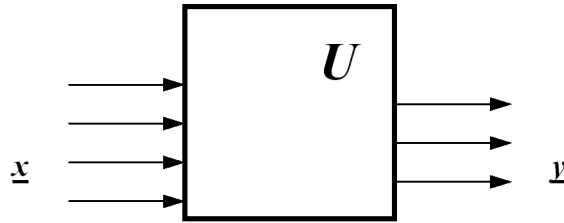
Prowadzenie takich badań wymaga jednak

- odpowiedniej wiedzy dotyczącej parametrycznego doboru potencjalnych defektów, w sposób umożliwiający otrzymanie określonego symptomu,
- opisu i odpowiedniej analizy prowadzonej przez osobę interpretującą wyniki badań symulacyjnych, w celu przekształcania tych wyników w reguły diagnostyczne.

2.3. Model danych

W rozdziale tym, stanowiącym poprawioną i rozszerzoną wersję fragmentów [2.7] oraz [2.8], zwrócono uwagę na modele diagnostyczne identyfikowane na podstawie dostępnych wyników obserwacji lub symulacji działania obiektu, czyli na modele rozpatrywane jako modele danych.

Wybierając omawiane algorytmy szczególną uwagę zwrócono na modele lokalne. Celowo pominięto liczną klasę algorytmów związanych z identyfikacją różnych postaci sieci neuronalnych, ponieważ na podstawie przeprowadzonych prób ich zastosowania [2.13], [2.11] mimo uzyskania odpowiedniej

Rys. 2.1: Obserwowany wirtualny, statyczny układ U

jakości identyfikowanego modelu, stwierdzono brak ich odpowiedniej przydatności do rozwiązywania omawianych zadań diagnostyki technicznej. Głównymi niedogodnościami modeli globalnych bazujących na sieciach neuronalnych są między innymi braki możliwości interpretowania wartości ich parametrów i braki możliwości oceniania wpływu przykładów uczących na postać modelu oraz utrudnienia występujące podczas prób rozszerzania zbioru uwzględnianych przykładów uczących [2.6].

2.3.1. Opis zadania poszukiwania modelu danych

Zadanie identyfikacji modelu danych jest zadaniem bardzo ogólnym. Zadanie takie może występować w różnych dziedzinach i dotyczyć może różnych zastosowań praktycznych. Podejmowane mogą być próby definiowania modeli jakościowych. W niniejszej pracy rozpatrywane są głównie modele ilościowe. Rozpatrujemy zadanie, którego celem jest zbudowanie numerycznego modelu jakiegoś wirtualnego, statycznego (tzn. nie zmieniającego się w czasie) układu U (rys. 2.1), który każdemu zbiorowi wartości wejść \underline{x} przyporządkowuje odpowiedni zbiór wartości wyjść \underline{y} .

Zakładamy, że struktura i działanie układu U nie są znane. W celu określenia działania tego układu jest on obserwowany przez zewnętrznego obserwatora. Dla obserwatora tego dostępne są, jako wynik obserwacji w chwili t , następujące zbiory wartości:

- zbiór wartości wyjść

$$\{y_1(t), y_2(t), \dots, y_H(t)\} \quad (2.1)$$

- zbiór wybranych wartości wejść

$$\{x_1(t), x_2(t), \dots, x_M(t)\} \quad (2.2)$$

takie że dla każdej chwili czasu t

$$y_h(t) \in Y_h ; h = 1, 2, \dots, H \quad (2.3)$$

$$x_m(t) \in X_m ; m = 1, 2, \dots, M \quad (2.4)$$

gdzie

$$Y_1, Y_2, \dots, Y_H \quad (2.5)$$

$$X_1, X_2, \dots, X_M \quad (2.6)$$

są odpowiednio przestrzeniami wartości wyjść i wejść. Ogólna liczba wejść oraz wartości pozostałych, nie objętych obserwacją, wejść

$$\{x_{M+1}, x_{M+2}, \dots\} \quad (2.7)$$

nie są znane. Wartości wejść (2.7) mogą być traktowane jako zakłócenia. Wynikiem obserwacji prowadzonych w chwilach

$$t_1, t_2, \dots, t_L \quad (2.8)$$

są zbiory wartości:

$$\begin{aligned} \underline{w}_1 &= \{x_{1,1}, x_{1,2}, \dots, x_{1,M}, y_{1,1}, y_{1,2}, \dots, y_{1,H}\} \\ \underline{w}_2 &= \{x_{2,1}, x_{2,2}, \dots, x_{2,M}, y_{2,1}, y_{2,2}, \dots, y_{2,H}\} \\ &\dots \\ \underline{w}_L &= \{x_{L,1}, x_{L,2}, \dots, x_{L,M}, y_{L,1}, y_{L,2}, \dots, y_{L,H}\} \end{aligned} \quad (2.9)$$

Zbiory wartości wyjść (2.1) oraz zbiory wartości wejść (2.2) mogą być zapisywane w postaci macierzy

$$\underline{y}(t) = [y_1(t), y_2(t), \dots, y_H(t)] \quad (2.10)$$

$$\underline{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)] \quad (2.11)$$

Macierze (2.10) (2.11) mogą być interpretowane jako współrzędne punktów przestrzeni wielowymiarowych. Kolejne współrzędne (osie) takiej przestrzeni odpowiadają rozpatrywanym wejściom i wyjściom (atrybutom) badanego obiektu. Wartości współrzędnych (liczby) są wartościami wejść i wyjść (wartościami atrybutów).

Poszukiwany model ma umożliwiać wyznaczanie całego lub części zbioru wartości wyjść (2.1) dla zadanego zbioru wartości wejść (2.2). Należy podkreślić, iż z założenia o braku znajomości struktury i działania układu U wynika brak możliwości uwzględniania informacji o tej strukturze lub tym działaniu podczas wyboru postaci modelu opisującego układ U . Postać tego modelu powinna być dobierana odpowiednio do postaci zbioru danych (2.9).

Przyjęty opis zadania pozwala na zastosowanie omawianych dalej algorytmów do wyznaczania modeli odwrotnych (reguł diagnostycznych) zarówno na podstawie wyników badań symulacyjnych, jak i na podstawie wyników obserwacji obiektu prowadzonych w warunkach biernego eksperymentu diagnostycznego. Możliwość taka wynika z faktu, iż wyniki obu rodzajów badań mogą być zapisywane w postaci zbioru danych uczących (2.9).

2.3.2. Postać modelu

Literatura dotycząca metod wyznaczania modeli jest bardzo obszerna. Wybór klasy modelu może być prowadzony na podstawie następujących kryteriów:

- rodzaj zależności stanowiących elementy modelu, które mogą być identyfikowane na podstawie zbiorów wartości (2.9):
 - *modele odwzorowań* zawierające funkcje odwzorowujące zbiory wartości wejść (2.2) w zbiory wartości wyjść (2.1),
 - *modele relacji* zawierające relacje określające podzbiory, odpowiadających sobie dyskretnych wartości wejść (2.2) i wartości wyjść (2.1), w iloczynie kartezjańskim zbiorów wartości wejść i wyjść,
 - zakres ważności modelu:
 - *modele globalne* opisujące dane uczące (2.9) w pełnej przestrzeni ich wartości
- $$X_1 \times X_2 \times \dots \times X_M \times Y_1 \times Y_2 \times \dots \times Y_H, \quad (2.12)$$
- *modele lokalne* opisujące dane (2.9) w wybranym obszarze przestrzeni (2.12),
 - istota działania modelu:
 - *modele aproksymacyjne*, zastępujące (z zadaną dokładnością) zbiór wartości (2.9),

- *modele interpolacyjne*, wymagające umieszczenia odpowiedniego podzbioru danych (2.9) w definicji modelu,
- stopień złożoności modelu:
 - *modele uogólniające*, w których liczba parametrów niezbędnych do zapisania modelu jest znacznie mniejsza od ogólnej liczby danych zapisanych w (2.9),
 - *modele pełne*, w których liczba parametrów niezbędnych do zapisania modelu jest porównywalna z liczbą danych zapisanych w (2.9), lub przekracza tę liczbę,
- jakość rozwiązania:
 - *modele dokładne* pozwalające na wyznaczanie rozwiązań dokładnych,
 - *modele przybliżone* pozwalające na wyznaczanie wyłącznie rozwiązań przybliżonych lub rozmytych,
- sposób reprezentacji danych (2.1) i (2.2)
 - *modele ilościowe* dla danych zapisanych w postaci liczb,
 - *modele jakościowe* dla danych reprezentowanych w sposób przybliżony,
 - *modele rozmyte* dla danych reprezentowanych w postaci liczb lub zbiorów rozmytych,
 - *modele statystyczne* dla danych reprezentowanych w postaci łącznych rozkładów gęstości prawdopodobieństw ich wartości,
 - *modele przekonań* dla danych reprezentowanych w postaci rozkładów przekonań o ich wartościach, gdzie modele przekonań są odpowiednikami modeli statystycznych, w których miary prawdopodobieństw zastąpiono miarami subiektywnych przekonań,
 - *klasyfikatory* umożliwiające wyznaczenie danych wyjściowych (2.1) reprezentowanych w postaci zbiorów, grup (skupień) lub klas.

2.3.3. Globalny model danych

Typowym przykładem globalnych modeli danych są modele aproksymacyjne, których wyznaczanie polega na identyfikacji, na podstawie zbioru wyników (2.9), funkcji przekształcającej z *odpowiednią dokładnością* wartości wejść (2.2) w wartości wyjść (2.1):

$$\{\hat{y}_1(t), \hat{y}_2(t), \dots, \hat{y}_H(t)\} = f(x_1(t), x_2(t), \dots, x_M(t)) \quad (2.13)$$

Funkcja $f(\cdot)$ ma umożliwić wyznaczanie wartości wyjść dla zadanego zbioru wartości wejść. Ze względu na taki cel postępowania zakłada się często (jawnie lub domyślnie), że wartości wyjść są skutkiem (następstwem) wartości wejść. Oznacza to, że wartość każdego wyjścia y_h w (2.1) zależy wyłącznie od zbioru wartości wejść (2.2) i do jej wyznaczenia nie są potrzebne wartości pozostałych wyjść w (2.1). Pozwala to na uproszczenie zadania (2.13) i przyjęcie iż poszukiwana jest funkcja odwzorowująca zbiory wartości wejść (2.2) w odpowiednio dokładne przybliżenie wartości jednego wybranego wyjścia

$$\hat{y}_h(t) = f_h(x_1(t), x_2(t), \dots, x_M(t)) \quad (2.14)$$

Rozwiązaniem zadania (2.13) będzie odpowiedni zbiór niezależnych rozwiązań zadań (2.14)

$$f = \{f_1, f_2, \dots, f_H\} \quad (2.15)$$

Funkcja (2.15) jest modelem zbioru danych (2.9). Funkcja ta, czyli model aproksymacyjny może występować w różnych postaciach [2.5]. W szczególności mogą to być modele dokładne lub przybliżone, liniowe, nieliniowe, jednostronne, częściowo jednostronne [2.4] itp.

Stosowanie modeli aproksymacyjnych wymaga przyjęcia założenia, że jedną z własności danych jest to iż poszukiwane funkcje (2.14) istnieją. O tym, czy założenie to jest spełnione czy też nie decydują dane (2.9).

2.3.4. Lokalny model danych

Przykładem lokalnych modeli danych są modele interpolacyjne, które pozwalają na wyznaczanie przybliżonej wartości $\hat{y}(\underline{x})$ funkcji $y(\underline{x})$ dla dowolnego argumentu \underline{x} , na podstawie znanych przykładów wartości tej funkcji

$$\{y(\underline{x}_p) : \underline{x}_p \in X_P\} \quad (2.16)$$

określonych dla skończonego zbioru przykładów wartości argumentów

$$X_P = \{\underline{x}_p\}. \quad (2.17)$$

Najczęściej zakłada się, że wartości wyznaczonej funkcji interpolującej $\hat{y}(\underline{x})$ odpowiadają (są równe) znanym wartościom funkcji $y(\underline{x})$ w punktach odpowiadających elementom zbioru przykładów wartości argumentów X_P

$$\forall \underline{x}_p \in X_P : \hat{y}(\underline{x}_p) = y(\underline{x}_p) \quad (2.18)$$

Pierwotny liczonej grupy algorytmów interpolacyjnych został zaproponowany w pracy [2.23]. Nieznane wartości funkcji wyznaczone są na podstawie danych wzorcowych następująco

$$\hat{y}(\underline{x}) = \frac{\sum_{i=1}^P w_i(\underline{x}) y(\underline{x}_i)}{\sum_{i=1}^P w_i(\underline{x})} \quad (2.19)$$

gdzie funkcja wagi w_i powinna uwzględniać odległość pomiędzy rozpatrywanym zbiorem \underline{x} i przykładami wzorcowymi \underline{x}_i definiowana może być na przykład w postaci

$$w_i(\underline{x}) = \|\underline{x} - \underline{x}_i\|^\mu \quad \text{dla } \mu > 0 \quad (2.20)$$

Zapisując (2.19) w równoważnej postaci [2.21]

$$\hat{y}(\underline{x}) = \frac{y(\underline{x}_j) + \sum_{i=1; i \neq j}^P \frac{w_i(\underline{x}) y(\underline{x}_i)}{w_j(\underline{x})}}{1 + \sum_{i=1; i \neq j}^P \frac{w_i(\underline{x})}{w_j(\underline{x})}}; \quad \text{dla } j = 1, 2, \dots, P \quad (2.21)$$

możemy zauważyć, że funkcja wagi (2.20) posiadająca własność

$$w_j(\underline{x}_j) = \infty \quad \text{dla } j = 1, 2, \dots, P \quad (2.22)$$

pozwała na spełnienie warunku

$$\hat{y}(\underline{x}_j) = y(\underline{x}_j) \quad \text{dla } j = 1, 2, \dots, P \quad (2.23)$$

Pokazana metoda postępowania jest skuteczna i prowadzi do stabilnych rozwiązań wtedy, gdy poszukiwane są modele dokładne. Wiadomo jednak, że skutkiem braku znajomości pełnego zbioru wejść modelu jest istnienie zakłóceń oraz potrzeba rozpatrywania modeli przybliżonych. Uzasadnia to rezygnację z uwzględniania, charakterystycznego dla zadań interpolacji, założenia przyjmującego, iż wartości wyznaczonej funkcji interpolującej są równe znanym wartościom funkcji (2.16).

Poszukiwane będą wartości funkcji $y(\underline{x})$ na podstawie znanych wartości tej funkcji, określonych dla skończonego zbioru wartości argumentów $\{\underline{x}\}$. W zadaniach dotyczących interpolacji często zakłada się, że wartości wyznaczonej funkcji interpolującej odpowiadają (są równe) znanym wartościom funkcji dla zbioru argumentów wzorcowych x . Uwzględniając postulaty, uzasadniające potrzebę poszukiwania modeli uogólniających, rezygnujemy jednak z tego założenia. Oznacza to rezygnację z

wymagania, iż wykres funkcji interpolującej ma przechodzić przez węzły interpolacji. Rozwiązywanie tak rozumianego zadania interpolacji wymaga przyjęcia jednego z następujących dodatkowych wymagań:

- określone zostanie kryterium pozwalające na ocenę dokładności lub jakości przybliżenia,
- zadana zostanie klasa funkcji aproksymujących.

Ze względu na potrzeby rozpatrywanych zadań identyfikacji modeli opisane zostaną wybrane algorytmy interpolacji w przestrzeniach wielowymiarowych. Klasyczne algorytmy interpolacji, przeznaczone głównie do stosowania w przestrzeniach o małej liczbie wymiarów, opisywane są w dostępnej literaturze (np. [2.22]).

Lokalne modele danych posiadają szereg zalet. W szczególności pozwalają one na uzyskiwanie wysokiej dokładności w tych obszarach przestrzeni stanów gdzie jest to potrzebne - czyli tam, gdzie poszukiwana jest wartość wyjściowa modelu. Skuteczne stosowanie modeli identyfikowanych na podstawie wyników badań (czynnych eksperymentów) symulacyjnych, wymaga odpowiedniego sterowania tymi badaniami. Celem takiego sterowania jest optymalizacja wyboru warunków działania symulowanego obiektu, gdzie kryterium optymalizacji jest uzyskiwana lokalna jakość wyznaczanego modelu.

2.3.5. Wybrane przykłady modeli interpolacyjnych

Algorytm W11

Wartości funkcji $\underline{y}(\underline{x})$, dla dowolnego \underline{x} , wyznaczone są jako średnie ważone znanych wartości funkcji

$$\underline{y}(\underline{x}) = \frac{\sum_{\underline{x}_i \in \varepsilon(\underline{x})} v(\underline{x}, \underline{x}_i) \underline{y}(\underline{x}_i)}{\sum_{\underline{x}_i \in \varepsilon(\underline{x})} v(\underline{x}, \underline{x}_i)} \quad (2.24)$$

gdzie $\varepsilon(\underline{x})$ jest ograniczonym lub nieograniczonym otoczeniem punktu (elementu) \underline{x} w przestrzeni wartości argumentów funkcji $\underline{y}(\underline{x})$ oraz gdzie funkcja $v(\underline{x}, \underline{x}_i)$ posiada tę własność, iż jej wartości bezwzględne są „duże” w bezpośrednim otoczeniu elementu \underline{x} oraz „małe” (dążące do zera) dla elementów oddalonych znacznie od \underline{x} . Powoduje to, że na wartość $\underline{y}(\underline{x})$ w (2.24) wpływać będą w pierwszym rzędzie elementy bezpośrednio sąsiadujące z \underline{x} . Zakładając, że elementy \underline{x} są elementami przestrzeni metrycznej, w której zdefiniowano miarę odległości $d(\underline{x}, \underline{x}_i)$, możemy rozpatrywać $v(\underline{x}, \underline{x}_i)$ jako $v(d(\underline{x}, \underline{x}_i))$. Przykładem takiej funkcji jest

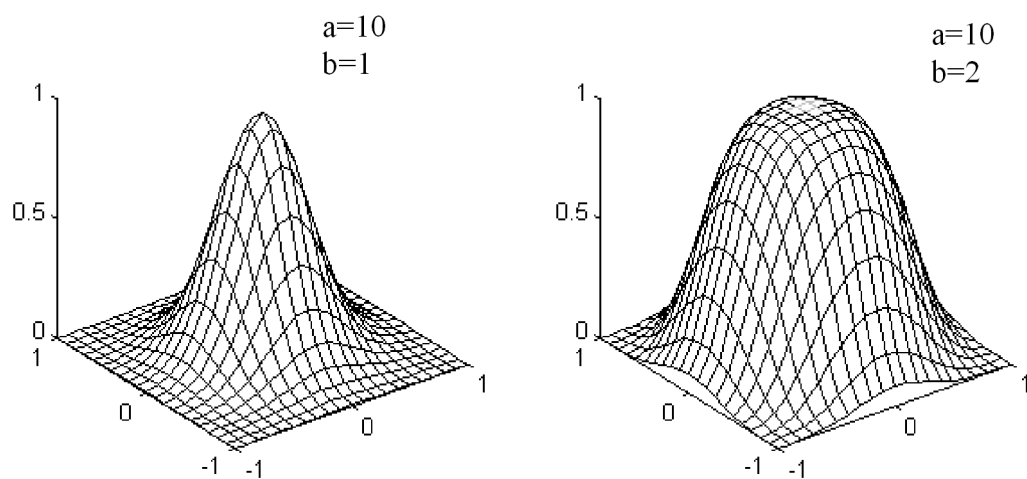
$$v(d(\underline{x}, \underline{x}_i)) = \exp(-a d^b(\underline{x}, \underline{x}_i)) \quad (2.25)$$

Niedogodnością algorytmu W11 jest potrzeba odpowiedniego skalowania wszystkich wartości \underline{x} . Skalowanie to ma (może mieć) znaczny wpływ na wartości odległości występujących w (2.25).

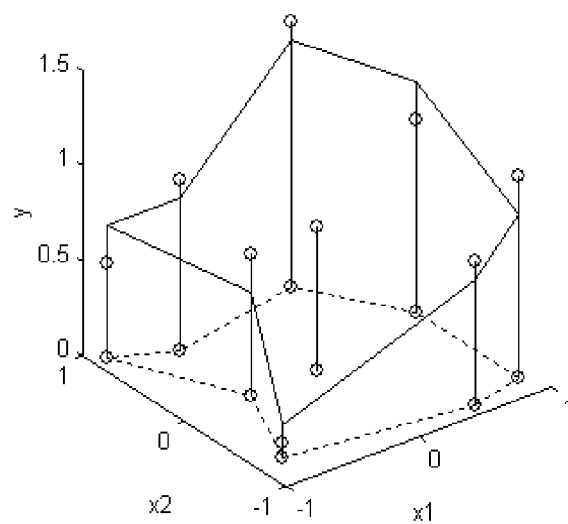
Algorytm W12

Innym sposobem określania modelu interpolacyjnego jest wyznaczanie odpowiedniego lokalnego modelu liniowego, identyfikowanego w bezpośrednim otoczeniu wyróżnionego punktu, dla którego poszukiwany jest ten model (rys. 2.3). Identyfikacja takiego modelu lokalnego następuje na podstawie najbliższych sąsiadów wyróżnionego punktu.

Ze względu na brak sprzężeń zwrotnych w rozpatrywanych modelach, poszukiwanie funkcji $\hat{\underline{y}}(\underline{x})$ stanowiącej przybliżenie funkcji (2.10), można zastąpić poszukiwaniem wyłącznie funkcji jednowymiarowych



Rys. 2.2: Przykłady wykresów funkcji (2.25) w dwuwymiarowym otoczeniu elementu \underline{x}



Rys. 2.3: Lokalny model liniowy

$$\hat{y}_1(\underline{x}), \hat{y}_2(\underline{x}), \dots, \hat{y}_H(\underline{x}) \quad (2.26)$$

Wartości funkcji (2.26) wyznaczone są na podstawie znanych wartości danych wzorcowych. W celu wyznaczenia nieznanej wartości funkcji $\hat{y}_h(\underline{x})$, dla $h=1,2,\dots,H$, oraz dla danego elementu \underline{x} , wykonywane są kolejne kroki algorytmu W12.

1. W zbiorze danych wzorcowych wyznaczany jest najbliższy sąsiad \underline{x}_0 elementu (punktu) \underline{x} , dla którego wartość funkcji $y_h(\underline{x}_0)$ jest znana.
2. W zbiorze danych wzorcowych wyznaczany jest podzbiór

$$X = \{\underline{x}_1, \dots, \underline{x}_K\} \quad \text{gdzie} \quad \underline{x}_0 \notin X \quad (2.27)$$

zawierający kolejnych ($K \geq 2$) najbliższych sąsiadów elementu (punktu) \underline{x} , dla których wartości funkcji $y_h(\underline{x}_k)$ są znane.

3. Wartości elementów zbioru (2.27) zapisywane są w macierzy prostokątnej \underline{A} , jako wartości określone względem elementu \underline{x}_0 (co umożliwia lokalną linearyzację modelu)

$$\underline{A} = \begin{bmatrix} (\underline{x}_1 - \underline{x}_0)^T \\ \vdots \\ (\underline{x}_K - \underline{x}_0)^T \end{bmatrix} = \begin{bmatrix} x_{1,1} - x_{0,1}, & \dots, & x_{1,M} - x_{0,M} \\ \vdots & \vdots & \vdots \\ x_{K,1} - x_{0,1}, & \dots, & x_{K,M} - x_{0,M} \end{bmatrix} \quad (2.28)$$

4. Wartości funkcji $y_h(\underline{x}_k)$, dla $\underline{x}_k \in X$, zapisywane są w macierzy kolumnowej \underline{b}_h , jako wartości określone względem wartości funkcji $y_h(\underline{x}_0)$

$$\underline{b}_h = \begin{bmatrix} y_h(\underline{x}_1) - y_h(\underline{x}_0) \\ \vdots \\ y_h(\underline{x}_K) - y_h(\underline{x}_0) \end{bmatrix} \quad (2.29)$$

5. Przyjmowany jest następujący model, na podstawie którego wyznaczone są wartości funkcji (2.26)

$$\hat{y}_h(\underline{x}) = (\underline{x} - \underline{x}_0)^T \underline{a}_h + y_h(\underline{x}_0) \quad (2.30)$$

Przyjęcie modelu (2.30) oznacza, iż wykresem poszukiwanej funkcji jest płaszczyzna „najlepiej” dopasowana do punktów wyznaczonych przez dane wzorcowe i przechodząca przez punkt

$$(\underline{x}_0, \underline{y}(\underline{x}_0)) \quad (2.31)$$

6. M -elementowa macierz kolumnowa \underline{a}_h w (2.30) wyznaczana jest jako rozwiązanie układu równań (2.32), otrzymanego przez zapisanie danych (2.27) z uwzględnieniem (2.28), (2.29) i (2.30)

$$\underline{b}_h = \underline{A} \underline{a}_h + \underline{\delta}_h \quad (2.32)$$

7. Rozwiązanie układu (2.32) ma minimalizować funkcję kryterialną dla reszt

$$z_h^2 = \sum_{k=1, \dots, K} (\delta_{h,k})^2 \rightarrow \min \quad (2.33)$$

gdzie zakłada się, że

$$\sum_{k=1, \dots, K} (\delta_{h,k}) = 0 \quad (2.34)$$

8. Poszukiwana wartość funkcji $\hat{y}_h(\underline{x})$ i jej odchylenia wyznaczone są

$$\hat{y}_h(\underline{x}) = (\underline{x} - \underline{x}_0)^T \underline{a}_h + y_h(\underline{x}_0) \pm \frac{\beta z}{\sqrt{K-1}} \quad (2.35)$$

gdzie β jest parametrem określającym względną szerokość symetrycznego przedziału ufności, a z wyznaczone jest wg (2.33) i (2.32).

Układ równań (2.32) może mieć różną postać, zależną od wartości danych. Dla uzyskania odpowiedniej niezawodności opisywanego postępowania nie można zakładać, że rząd macierzy (2.28) w (2.32) jest dostatecznie duży. Stwierdzenie to zmusza do rezygnacji z klasycznych metod rozwiązywania zadania (2.33). Zalecane jest postępowanie polegające na dekompozycji macierzy \underline{A} (2.28), zawierającej K wierszy i M kolumn, wg algorytmu SVD (singular value decomposition) [2.15]

$$\underline{A} = \underline{U} \underline{W} \underline{V}^T \quad (2.36)$$

gdzie

\underline{U} - macierz posiadająca K wierszy i M ortonormalnych kolumn, taka że

$$\underline{U}^T \underline{U} = \underline{1} \quad (2.37)$$

\underline{W} - macierz diagonalna posiadająca M kolumn i wierszy oraz nieujemne elementy diagonalne (singular values),

\underline{V} - macierz kwadratowa posiadająca M wierszy i M ortonormalnych kolumn, taka że

$$\underline{V}^T \underline{V} = \underline{1} \quad (2.38)$$

Zaletą proponowanego postępowania jest to, że dekompozycja (2.36) może być przeprowadzana również wtedy, gdy macierz \underline{A} jest macierzą osobliwą lub źle uwarunkowaną:

- Dla osobliwej macierzy \underline{A} macierz \underline{W} zawierać będzie wyłącznie r niezerowych elementów diagonalnych, gdzie r jest rzędem macierzy \underline{A} . Pozostałe elementy diagonalne będą miały wartości zerowe.
- Dla macierzy \underline{A} źle uwarunkowanej mogą wystąpić elementy diagonalne o bardzo małych wartościach. Zalecane zastąpienie tych elementów wartościami zerowymi jest równoważne transformacji macierzy \underline{A} do stabilnej postaci osobliwej.

Poszukiwana macierz \underline{a}_h w (2.32) wyznaczana jest na podstawie macierzy uzyskanych w wyniku dekompozycji (2.36)

$$\underline{a}_h = \underline{V} \underline{W}^{-1} \underline{U}^T \underline{b}_h \quad (2.39)$$

W celu wyznaczenia \underline{W}^{-1} zakłada się, że iloczyn $\underline{W}^{-1} \underline{W}$ jest macierzą *prawie* jednostkową rzędu r , posiadającą M kolumn i wierszy

$$\underline{W}^{-1} \underline{W} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & 0 & & \\ & & 1 & & & \\ & & & 0 & & \\ 0 & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{matrix} 1 \\ \vdots \\ r \\ r+1 \\ \vdots \\ M \end{matrix} \quad (2.40)$$

Elementy diagonalne macierzy odwrotnej \underline{W}^{-1} wyznaczone są następująco

$$\underline{W}^{-1} \Big|_{m,m} = \begin{cases} 1/w_{m,m} & \text{dla } w_{m,m} > 0 \\ 0 & \text{dla } w_{m,m} = 0 \end{cases} \quad (2.41)$$

Algorytm W13

Algorytm W12 pozwala na skuteczne wyznaczanie poszukiwanych wartości funkcji, wtedy gdy dane wzorcowe są zadane dostatecznie gęsto. Uzyskanie takich danych, zwłaszcza w wielowymiarowych przestrzeniach wartości cech, jest zadaniem trudnym. Dla zwiększenia wpływu bezpośrednich sąsiadów rozpatrywanych elementów wprowadzono dodatkową funkcję oceniającą stopień ważności danych. Funkcja ta może występować w różnych wersjach:

- przyjmuje stałe wartości równe 1

$$v(\underline{x}, \underline{x}_i) = 1 \quad (2.42)$$

- zależy od odległości elementów i przyjmuje wartość maksymalną dla zerowej odległości elementów, na przykład

$$v(\underline{x}, \underline{x}_i) = \begin{cases} \frac{1}{1+d(\underline{x}, \underline{x}_i)} & \text{dla } \underline{x}_i \in \varepsilon(\underline{x}) \\ 0 & \text{dla } \underline{x}_i \notin \varepsilon(\underline{x}) \end{cases} \quad (2.43)$$

Algorytm W13 jest zmodyfikowaną wersją algorytmu W12. Modyfikacja polega na następującej zmianie zależności (2.28) i (2.29)

$$\underline{A} = \begin{bmatrix} (x_{1,1} - x_{0,1}) v(\underline{x}, \underline{x}_1), & \cdots, & (x_{1,M} - x_{0,M}) v(\underline{x}, \underline{x}_1) \\ \vdots & \vdots & \vdots \\ (x_{K,1} - x_{0,1}) v(\underline{x}, \underline{x}_K), & \cdots, & (x_{K,M} - x_{0,M}) v(\underline{x}, \underline{x}_K) \end{bmatrix} \quad (2.44)$$

$$\underline{b}_h = \begin{bmatrix} (y_h(\underline{x}_1) - y_h(\underline{x}_0)) v(\underline{x}, \underline{x}_1) \\ \vdots \\ (y_h(\underline{x}_K) - y_h(\underline{x}_0)) v(\underline{x}, \underline{x}_K) \end{bmatrix} \quad (2.45)$$

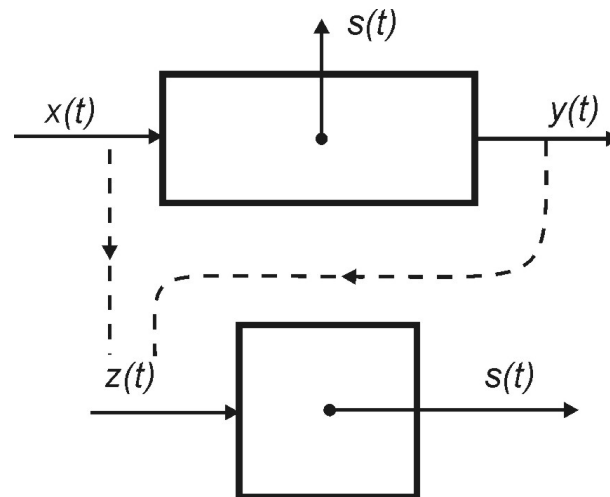
2.4. Model diagnostyczny

Na obecnym etapie rozwoju komputerowych systemów wspomagających diagnostykę maszyn i procesów, wyłoniła się możliwość opracowania skutecznych algorytmów, pozwalających na identyfikację relacji diagnostycznych będących podstawą określania reguł, tworzących bazy wiedzy diagnostycznych systemów doradczych.

Ciekawą klasą modeli obiektów są modele systemowe, których definiowanie wymaga wykonania następujących kroków:

- określana jest umowna granica pomiędzy rozpatrywanym obiektem i jego otoczeniem, umożliwiającą myślowe wyodrębnienie modelowanego obiektu z otoczenia,
- oddziaływanie otoczenia na obiekt reprezentowane jest w postaci wejść obiektu,
- oddziaływanie obiektu na otoczenie reprezentowane jest w postaci wyjść obiektu,
- wprowadza się pojęcie stanu obiektu, który w uproszczeniu można interpretować jako parametr przekształcenia wejść obiektu w wyjścia obiektu.

Schemat tak definiowanego podstawowego modelu systemowego został pokazany w górnej części rys. 2.4. Jak już wspomniano celem badań diagnostycznych jest rozpoznawanie stanu rozpatrywanego obiektu. Stwierdzenie to upoważnia do przekształcenia podstawowego modelu systemowego do postaci pokazanej w dolnej części rys. 2.4. Schemat ten pokazuje model odwzorowujący dane o oddziaływaniach otoczenia na obiekt i obiektu na otoczenie w stan obiektu. Model taki nazywany jest *modelem diagnostycznym*.



Rys. 2.4: Przekształcanie modeli

2.4.1. Relacja diagnostyczna

Jednym z podstawowych pojęć diagnostyki technicznej jest relacja diagnostyczna. *Relacja diagnostyczna* (w tym empiryczna relacja diagnostyczna [2.18]) może być definiowana jako podzbiór R przestrzeni ocen P równej iloczynowi kartezjańskiemu dziedzin S_1, S_2, \dots relewantnych dla rozpatrywanego zadania diagnostycznego. Dla zadania diagnostycznego ograniczonego do jednego obiektu, rozpatrywanego w stałych warunkach działania, minimalna rodzina takich dziedzin obejmuje zbiór stanów S_1 oraz zbiór symptomów S_2 , prowadząc do najprostszej postaci relacji diagnostycznej

$$R \subset P = S_1 \times S_2 \quad (2.46)$$

gdzie R jest zbiorem wszystkich możliwych (lub zaobserwowanych) par (s_1, s_2) określających stan s_1 i odpowiadający mu symptom s_2 .

Badania umożliwiające pozyskiwanie relacji diagnostycznych obejmują:

- badania aktywne, w których możliwe jest modyfikowanie stanu obiektu,
- badania bierne, w których nie jest możliwe celowe modyfikowanie stanu, który można jedynie obserwować,
- badania symulacyjne, odpowiadające badaniom aktywnym, w których obiekt zastąpiony został przez jego model.

W wymienionych klasach modele mogą dotyczyć jednego wskazanego obiektu, zbioru rozpatrywanych, znanych obiektów lub klasy nieznanych obiektów odpowiadających zadanej konstrukcji.

Klasyczne podejście do zadań pozyskiwania relacji diagnostycznych przyjmuje, iż poszukiwany jest model relacji (2.46). Jako szczególnie uprzywilejowaną formę zapisu takiego modelu relacji uznano zbiory reguł występujących w postaci

$$\text{jeżeli przesłanka to konkluzja} \quad (2.47)$$

gdzie w zależności od potrzeb i stopnia złożoności obiektu reguły są regułami przyczynowo-skutkowymi lub regułami diagnostycznymi. W regułach przyczynowo-skutkowych przesłanką jest stan a konkluzją symptom. W regułach diagnostycznych odwrotnie.

Inną, często stosowaną techniką pozyskiwania relacji diagnostycznych jest modelowanie związków zachodzących (lub obserwowanych) pomiędzy dziedzinami omawianych relacji, za pomocą sieci neuronowych. Intensywny rozwój teorii tych sieci oraz dostępne aktualnie oprogramowanie pozwalają na uzyskiwanie zadawalających wyników w większości przypadków.

Wymieniać można liczne przykłady metod postępowania. Ich wspólną cechą jest to, że wynikiem realizowanego procesu identyfikacji są modele globalne uogólniające dane i zastępujące te dane. Pozwala to na znaczne ograniczenie niezbędnych zasobów (obszar pamięci) i przyspieszenie procesu wnioskowania. Wadą takiego postępowania jest globalny charakter rozpatrywanych modeli, pociągający za sobą wysokie koszty (czas obliczeń) ich identyfikacji oraz utrudniający lub ograniczający możliwości ich doskonalenia.

Znaczne zmniejszenie kosztów przechowywania danych oraz ograniczenie czasów dostępu do nich spowodowało wzrost zainteresowania metodami bazującymi na modelach lokalnych wyznaczanych dla wskazanego, ograniczonego obszaru przestrzeni ocen, dopiero wtedy gdy są potrzebne. Odróżnia to je od modeli globalnych wyznaczanych „na zapas” dla wszystkich możliwych obszarów przestrzeni ocen. Stosowanie takich modeli lokalnych jest możliwe wtedy, gdy czasy realizacji procesu wnioskowania przez układ diagnozujący nie są krytyczne.

Ciekawą modyfikacją omawianych metod jest przyjęcie założenia, iż rezygnuje się z wyznaczania globalnych modeli relacji R (2.46) i zastępuje je bezpośrednio gromadzeniem elementów tych relacji. Dysponując wyliczeniem elementów zbioru R można wskazać elementy należące do rozpatrywanego obszaru przestrzeni. Na elementach tych można następnie rozpiąć model lokalny pozwalający na dalsze działania. Model taki może być specyfikowany jawnie lub może występować w postaci ukrytej w systemie diagnozującym.

2.4.2. Użyteczne cechy sygnałów

Efekt ubocznym pozyskiwania wiedzy diagnostycznej na podstawie badań symulacyjnych realizowanych z zastosowaniem modeli numerycznych jest możliwość wyznaczania bardzo licznych zbiorów cech sygnałów. Rozpatrywanie kolejnej nowej cechy nie wymaga kosztownej modyfikacji aparatury i powtarzania badań. Wystarcza uzupełnienie kodu programu wyznaczającego wartości cech na podstawie zarejestrowanych ich przebiegów.

Można pokazać, że nadmierne zwiększanie liczby uwzględnianych cech nie zapewnia poprawy jakości wyznaczanego na ich podstawie klasyfikatora i prowadzić może do pogorszenia tej jakości. Fakt ten, czasami pokazywany jako sprzeczny z intuicją, uzasadnia się niekorzystnym wpływem cech nieużytecznych tzn. takich, które nie są związane z rozpatrywanymi klasami. Cechy takie pełnią rolę szumu, pogarszając jakość klasyfikatora. Powoduje to potrzebę redukcji liczby wyznaczonych lub możliwych do wyznaczenia cech, czyli potrzebę wyboru tych cech, które mogą być uznane za cechy użyteczne, przydatne do orzekania o stanie obiektu. Cechy te nazywane są również cechami relewantnymi. Tak określone zadanie jest znane od wielu lat. Dostępne są liczne publikacje pokazujące różne metody wyznaczania cech użytecznych.

Powszechnie stosowanymi metodami ograniczania liczby uwzględnianych cech, są metody skalowania wielowymiarowego. Pozwalają one na

- dokonanie wyboru zadanej liczby cech ze zbioru istniejących cech lub
- wyznaczenie zadanej liczby nowych cech, w wyniku transformacji (najczęściej liniowej) istniejących cech.

O własnościach ograniczonego zbioru cech uznanych za cechy użyteczne, decyduje zastosowane kryterium jakości zbioru cech. Podstawowe metody (np. metoda składowych głównych) bazują na założeniu, iż zachowane powinny być rozproszenia danych i ich główne kierunki lub (np. metoda czynników głównych) zachowane powinny być zależności występujące pomiędzy wartościami cech. Metody te nadają się dobrze do ograniczania liczby cech rozpatrywanych w zadaniach grupowania realizowanych bez nauczyciela. W zadaniach tych zakładany podział danych na klasy nie jest znany i żądanie odpowiedniego zachowania rozproszenia danych jest uzasadnione.

W rozpatrywanym zadaniu budowy klasyfikatora lokalnego występują jednak odmienne wymagania. Wzorcowy podział danych uczących na klasy jest znany. Kryteria optymalizacji wyboru cech powinny uwzględniać ten fakt. Możliwe jest zastosowanie następujących kryteriów:

- różne kryteria bazujące na minimalizowanych średnich wartościach miar rozproszeń wewnętrznych w klasach i maksymalizowanych wartościach miar rozproszeń zewnętrznych pomiędzy klasami,
- kryteria bazujące na wynikowej skuteczności klasyfikatora wyznaczanego dla ograniczonego (w wyniku wyboru) zbioru cech,
- kryteria bazujące na zgodności wyników grupowania (bez nauczyciela) w nowej przestrzeni wyznaczonej przez ograniczony zbiór cech z wynikami klasyfikacji wzorcowej danych uczących.

Dla grupy kryteriów uwzględniających wewnętrzne i zewnętrzne rozproszenia klas znane są liczne efektywne algorytmy ich stosowania. Kryteria te są szczególnie zalecane wtedy gdy oczekuje się klas występujących w postaci spójnych skupień danych. Pozostałe kryteria wymagają iteracyjnego poszukiwania rozwiązania. Rozsądnym kompromisem pomiędzy oczekiwaną trafnością wyboru oraz ograniczeniami dotyczącymi czasu poszukiwania rozwiązania wydaje się postępowanie polegające na tzw. zagłębianiu z nawrotami. Może być ono realizowane cyklicznie w ten sposób, że poszukujemy m kolejnych cech kandydujących do wybrania a następnie poszukujemy n kolejnych cech do pominięcia, gdzie $n < m$. Kryteria te mogą być również skutecznie stosowane w ewolucyjnych metodach optymalizacji zbioru uwzględnianych cech.

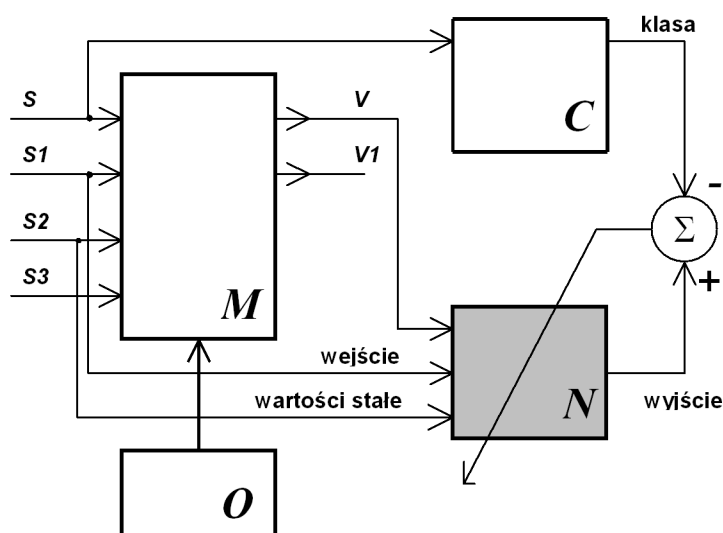
Szczególną uwagę należy zwrócić na ostatnią ze wskazanych grup kryteriów. Należy zauważyć, że zakłada ona, iż możliwe będzie wybranie zbioru cech, które pozwolą na odtworzenie bez nauczyciela podziału danych zgodnego z wzorcowym podziałem uczącym. Znalezienie takich cech pozwoli na uznanie, iż wyznaczony na ich podstawie klasyfikator ma odpowiednie własności uogólniające. Pozwala to równocześnie na ograniczenie możliwości nadmiernego dopasowania klasyfikatora do danych uczących, co może wystąpić podczas stosowania kryteriów bazujących na wynikowej skuteczności klasyfikatora. Spełnienie wskazanych kryteriów ma zapewnić możliwość skutecznego odtwarzania klas stanów na podstawie wartości użytecznych cech sygnałów. Należy zaznaczyć, że omawiane kryteria są sugerowane jedynie jako kryteria przydatne podczas identyfikacji cech użytecznych. Kryteria będące podstawą konstruowania klasyfikatorów występujących jako modele lokalne powinny być dobierane niezależnie.

2.4.3. Model odwrotny

Zaproponowanie skutecznych metod numerycznego odwracania wielowymiarowych, dynamicznych modeli maszyn spowodowało wzrost zainteresowania diagnostyką symulacyjną [2.7, 2.8, 2.9]. Wyjściową fazą badań jest budowanie złożonych, numerycznych modeli symulacyjnych, pozwalających na generowanie sygnałów odpowiadających złożonym stanom technicznym obiektu. Modele te weryfikowane są na obiektach rzeczywistych oraz na stanowiskach laboratoryjnych. Na podstawie wyników badań weryfikacyjnych może być przeprowadzany proces strojenia tych modeli. Polega on najczęściej na dopasowywaniu wartości wybranych parametrów wewnętrznych modelu, tak aby wyniki obliczeń realizowanych na podstawie modelu były zgodne z wynikami obserwacji obiektu rzeczywistego. Na podstawie zweryfikowanych modeli wyznaczane są następnie modele odwrotne [2.12], będące poszukiwanymi relacjami diagnostycznymi, przekształcającymi symptomy diagnostyczne w klasy stanów technicznych. Potrzeba takiego postępowania wynika z braku możliwości bezpośredniego definiowania modeli diagnostycznych oraz braku możliwości analitycznego odwracania wielowymiarowych modeli dynamicznych. Wysoki stopień trudności omawianego zadania związany jest z tym, że wynikiem odwracania modeli opisujących relacje przyczynowo-skutkowe, nie będą w ogólnym przypadku relacje przyczynowo-skutkowe.

We wcześniejszych badaniach [2.12, 2.8] zaproponowano ogólną koncepcję modeli odwrotnych dla potrzeb diagnostyki technicznej. Zakłada się, że znany jest model M (rys. 2.5), który pozwala na wyznaczenie wartości wyjść $\{v, v1\}$ dla zadanego zbioru wartości wejść $\{s, s1, s2, s3\}$. Model ten może być zapisany (na przykład) w postaci relacji empirycznych [2.18] lub programu komputerowego. Z punktu widzenia zastosowań diagnostycznych wymienione wartości wejść i wyjść mogą mieć następujące znaczenie:

- s poszukiwane jako wynik procesu diagnozowania (np. rozkład niewyrównoważenia elementów wirujących, rozosiowanie wałów),
- $s1$ znane wartości, określające warunki działania obiektu (np. wartość prędkości obrotowej, wartość obciążenia),
- $s2$ znane wartości, które dla rozpatrywanego obiektu nie ulegają zmianom (np. wartość cech konstrukcyjnych obiektu),
- $s3$ wartości trudne do oszacowania i pomiaru, które należy przyjmować jako wartości losowe wpływające na ograniczenie dokładności modelu (np. wartość tłumienia drgań przez posadowienie rozpatrywanego obiektu),
- v znane wartości, będące skutkiem działania obiektu i określane w wyniku pomiaru lub odpowiedniej symulacji (np. wartość cech drgań),
- $v1$ nieznane lub pomijane wartości, będące skutkiem działania obiektu.



Rys. 2.5: Wyznaczanie modelu odwrotnego N dla danego modelu M obiektu O

Model M opisuje relacje przyczynowo-skutkowe występujące pomiędzy wartościami wejściowymi $\{s, s1, s2, s3\}$ rozpatrywanymi jako przyczyny oraz wartościami wyjściowymi - skutkami $\{v, v1\}$ rozpatrywanymi jako skutki. W wyniku odwracania modelu M pragniemy (rys. 2.5) uzyskać model odwrotny N , przekształcający znane, określane w wyniku pomiaru lub symulacji wartości $\{v, s1\}$ w poszukiwane wartości $\{s\}$. Należy podkreślić, że z punktu widzenia relacji przyczynowo-skutkowych uwzględniane wartości $\{s, s1, s2\}$ opisują wyłącznie wybrane przyczyny, a uwzględniane wartości $\{v\}$ wybrane ich skutki. Oznacza to, że można przypuszczać, iż dokładny model odwrotny N nie istnieje, ponieważ brak jest podstaw do zakładania, iż rozpatrywany model M jest odwzorowaniem jedno-jednoznacznym. Dla uniknięcia wynikających stąd trudności (powodujących, że zadanie może nie posiadać rozwiązania) ograniczamy dokładność poszukiwanego modelu odwrotnego, zakładając, iż będzie on wyznaczał jedynie klasy wartości parametrów v , a nie dokładne wartości tych parametrów. Klasyfikator definiujący wyznaczane klasy oznaczono literą C (rys. 2.5).

W diagnostyce technicznej maszyn występuje potrzeba określania wartości cech stanu lub klasy stanu na podstawie wyników obserwacji (rys. 2.5 - zmienna v) obiektu opisanego zbiorem jego cech podstawowych (rys. 2.5 - zmienna $s2$) i działającego w znanych warunkach (rys. 2.5 - zmienna $s1$). Wyście modelu odwrotnego N (rys. 2.5) jest zbiorem wartości cech stanu lub identyfikatorów klas stanu, czyli ogólnie zbiorem wartości, które mogą być między innymi wartościami nominalnymi.

Model N (rys. 2.5) rozpatrywany może być jako *czarna skrzynka*, której cechy identyfikowane będą w wyniku procesu jej trenowania lub adaptacji. Liczne badania doświadczalne związane są z

identyfikacją analitycznych modeli obiektów rzeczywistych. Często ogólna postać modelu jest znana i wynika np. z analizy działania badanego obiektu a proces *identyfikacji modelu* sprowadza się do wyznaczenia wartości odpowiedniego zbioru parametrów. Odrębną klasę stanowią zadania, w których postać modelu nie jest znana. Dla podkreślenia iż postać modelu nie jest znana, zadania te określa się jako *poszukiwanie modelu danych*. Rozwiązywanie omawianych zadań wymaga wykonania dwóch kroków:

- wybór postaci modelu,
- identyfikacja modelu (określenie wartości parametrów modelu o wybranej wcześniej postaci).

Jeżeli model N jest znany, to realizacja podstawowego zadania diagnostyki technicznej polega na wyznaczaniu wyjścia tego modelu na podstawie jego wejść. Zakłada się, że poszukiwany model N (rys. 2.5) jest modelem jakiegoś wirtualnego, statycznego (nie zmieniającego się w czasie) układu U , który każdemu zbiorowi wartości wejść \underline{x} przyporządkowuje odpowiedni zbiór wartości wyjść \underline{y} . Sprowadza to zadanie identyfikacji modelu odwrotnego N do ogólnego zadania identyfikacji modelu danych opisanego w rozdziale 2.3.1.

Zaletą proponowanego postępowania jest to, że stosowanie modelu M i wyznaczanie modelu odwrotnego N mogą być traktowane jako procesy rozłączne realizowane w różnym czasie i w różnych środowiskach (co pokazano w [2.12]). Możliwe jest sekwencyjne realizowanie procesów stosowania modelu M i wyznaczania modelu odwrotnego N , pozwalające na iteracyjne doskonalenie wyznaczonego modelu odwrotnego [2.17], [2.24].

2.4.4. Model wielowarstwowy

Zaletą wynikającą ze stosowania modeli numerycznych obiektów jest możliwość generowania licznych zbiorów danych uczących, pozwalających na skuteczne pozyskiwanie wiedzy diagnostycznej. Niestety modele te są równocześnie źródłem niedogodności związanych z

- występowaniem nadmiernej liczby możliwych do wyznaczenia cech sygnałów diagnostycznych,
- występowaniem dużej liczby cech stanu na podstawie których identyfikowana jest klasa stanu.

Wzrost rozmiaru zadania zmusza do rezygnacji z wielu powszechnie stosowanych algorytmów oraz w wysokim stopniu utrudnia intuicyjną weryfikację poprawności otrzymywanych wyników. W celu ograniczenia wskazanych niedogodności przeprowadza się dekompozycję jednego modelu globalnego na zbiór modeli lokalnych.

Dekompozycja modelu globalnego ma na celu ograniczenie liczby uwzględnianych cech wejściowych i wyjściowych. Nie można jej rozpatrywać wyłącznie jako przestrzennej dekompozycji obiektu, polegającej na jego podziale na zespoły i elementy. Zmniejszenie liczby rozpatrywanych cech zostanie uzyskane wyłącznie wtedy, gdy dekompozycja modelu globalnego na modele lokalne będzie połączona z rozprzęgnięciem opisów lokalnych pozwalającym na ich uniezależnienie. Wygodnym sposobem rozprzęgania modeli lokalnych jest wprowadzanie dodatkowych cech występujących w roli zmiennych instrumentalnych, które nazywane są cechami sygnałów dodatkowych. Mogą to być cechy oddziaływań wirtualnych lub ukrytych. Mogą one dotyczyć wielkości fizycznych, które nie są bezpośrednio uwzględniane w definicjach stanu obiektu. Nie wymaga się jednak aby dotyczyły one obserwowalnych oddziaływań rzeczywistych. Dodatkowe cechy modelu lokalnego mogą być wyznaczane metodą prób i błędów np. za pomocą metod skalowania wielowymiarowego zastosowanych do cech związanych z pozostałymi modelami lokalnymi. Poszukiwanie takich cech jest jednak zadaniem trudnym i nie gwarantującym zadawalających wyników, zwłaszcza wtedy gdy nie korzysta się równocześnie z odpowiedniej wiedzy dziedzinowej, a w szczególności wiedzy o istocie działania obiektu. Zagadnienia związane z budowaniem omawianych modeli rozpatrywano w [2.25, 2.24].

Modele lokalne będące wynikiem dekompozycji modelu globalnego stają się elementami mieszanego modelu wielostopniowego. Pierwszym stopniem takiego modelu są modele lokalne z ich wejściami obejmującymi cechy sygnałów diagnostycznych i wyjściami w postaci wprowadzonych cech dodatkowych. Cechy dodatkowe stanowią wejście drugiego stopnia modelu mieszanego.

2.5. Sieć stwierdzeń

Szczególnym rodzajem modeli są sieci stwierdzeń, czyli sieci przekonań w których węzłami są stwierdzenia. Modele takie mogą być stosowane jako modele diagnostyczne.

2.5.1. Treść i wartość stwierdzenia

Najprostszą reprezentacją stwierdzenia jest para uporządkowana składająca się z „*treści stwierdzenia*” i „*wartości stwierdzenia*”. Treść stwierdzenia wskazuje obserwowany fakt lub rozpatrywaną opinię, o uznaniu których orzeka stwierdzenie. Wprowadzając pojęcie „*wartość logiczna stwierdzenia*” lub prościej „*wartość stwierdzenia*”, oznaczające dla zdań logicznych „*prawda*” lub „*fałsz*” zapisywane np. jako wartości „1” i „0”, możemy znacznie rozszerzyć klasę rozpatrywanych stwierdzeń. Przyjęcie, iż wartość stwierdzenia jest liczbą z przedziału $[0; 1]$ pozwala na interpretowanie tej wartości jako stopnia prawdziwości stwierdzenia, stopnia przekonania o słuszności stwierdzenia lub prawdopodobieństwa tego, że rozpatrywane stwierdzenie jest prawdziwe. Pozwala to na realizowanie wnioskowania przybliżonego, wnioskowania o różnym stopniu pewności itp. Stwierdzenia takie mogą tworzyć sieci nazywane sieciami przekonań (ang. belief networks) lub sieciami bayesowskimi. Sieci te spotykają się w ostatnim okresie z coraz większym uznaniem jako skuteczne rozwiązania, przydatne do wnioskowania na podstawie stwierdzeń przybliżonych. Znalazły one zastosowanie w wielu dziedzinach.

Dla potrzeb dynamicznego wnioskowania o stanie, czyli wnioskowania realizowanego dla zmieniających się w czasie przesłanek i konkluzji, wprowadza się stwierdzenia dynamiczne. Najprościej można je rozpatrywać jako stwierdzenia, których wartość jest funkcją czasu. Jedną z metod analizowania takich stwierdzeń jest ich „zamrażanie” w kolejno analizowanych dyskretnych chwilach czasu.

2.5.2. Słownik treści stwierdzeń

Bardzo ważnym, świadomie wprowadzanym założeniem związanym ze stosowaniem stwierdzeń, jest rozpatrywanie zamkniętego zbioru stwierdzeń. Założenie to oznacza, że zadaniem systemu wnioskującego jest orzekanie o prawdziwości wybranych stwierdzeń, na podstawie informacji o uznaniu innych stwierdzeń, gdzie wynikiem takiego procesu nie może być żadne nowe, wcześniej nieznanie, stwierdzenie.

Taka własność uwzględnianych stwierdzeń pozwala na szczegółowe opracowanie (specyfikację) zbioru tych stwierdzeń, nazywanego czasami „*słownikiem treści stwierdzeń*”. Każdemu elementowi słownika, czyli każdemu rozpatrywanemu stwierdzeniu, można przyporządkować opis i ilustrację objaśniającą jego treść oraz tłumaczenie ewentualnej przyczyny lub skutków jego uznania lub odrzucenia. Oznacza to, że możliwe jest wygodne zarządzanie zbiorem stwierdzeń z jednoczesną możliwością szczegółowego dokumentowania procesu jego konstruowania i rozwoju.

2.5.3. Ziarnistość danych

Zadanie wnioskowania o stanie obiektu na podstawie zaobserwowanych symptomów jest równoważne wnioskowaniu o wartości logicznej konkluzji, czyli stwierdzeń dotyczących stanu obiektu, na podstawie wartości logicznej przesłanek, czyli innych stwierdzeń dotyczących zaobserwowanych symptomów. Przejście od stanów i symptomów do stwierdzeń wymaga rozwiązania szeregu trudnych zadań dotyczących między innymi racjonalnej dyskretyzacji danych. Trudność tych działań jest związana z tym, że zbyt duża ziarnistość danych wynikająca z przyjętych parametrów dyskretyzacji może decydować o jakości procesu wnioskowania, podczas gdy zbyt mała ziarnistość danych zwiększa koszty procesu wnioskowania i znacząco ogranicza możliwości uogólniania. Dokładniejszą dyskusję zagadnień związanych z ziarnistością danych można znaleźć w ciekawej monografii [2.1].

Oczywiście można w tym miejscu pokazać przykłady systemów diagnostycznych, w których do procesu wnioskowania wprowadza się ciągłe wartości cech obserwowanych sygnałów, gdzie trudno

zbiory tych wartości reprezentować skończonym (mało licznym) zbiorem stwierdzeń. Typowymi przykładami takich systemów są systemy bazujące na sieciach neuronowych.

Zwykle uznajemy, że słuszne jest założenie o oczekiwanej ograniczonej „rozdzielczości” konkluzji w procesie wnioskowania, czyli o potrzebie rozpoznawania klasy stanu a nie wartości cech stanu. Oznacza to, że akceptowana jest ziarnistość wyjścia procesu wnioskowania. Konsekwencją tego jest możliwość wykazania, że ziarniste wejścia tego procesu pozwalają na uzyskanie sprawności (jakości) wnioskowania porównywalnej ze sprawnością takiego procesu z wejściami ciągłymi w ramach, które- go przeprowadzana jest dyskretyzacja niezbędna dla otrzymania ziarnistego wyjścia. Argumentując inaczej można stwierdzić, że skoro dopuszczamy ziarnistość wyjścia procesu to dopuszczalna jest również ziarnistość jego wejścia. Oznacza to, że rozpatrywanie procesu wnioskowania jako procesu realizowanego na stwierdzeniach jest rozwiązaniem wystarczająco ogólnym.

2.5.4. Sieć przekonań

Sieć przekonań, nazywana również siecią bayesowską, stanowi zwartą formę reprezentacji gęstości łącznego dyskretnego rozkładu zbioru zmiennych losowych. Sieć ta bazuje na koncepcji warunkowej niezależności zmiennych umożliwiającej ograniczenie liczby danych niezbędnych do zapisania gęstości łącznego rozkładu dyskretnego. Ograniczenie polega na wyznaczaniu tej gęstości jako iloczynu kilku innych gęstości o znacznie mniejszych rozmiarach. Jedna wielowymiarowa tablica prawdopodobieństw łącznych zostaje zastąpiona zbiorem dwuwymiarowych tablic prawdopodobieństw warunkowych, które pozwalają na wyznaczenie wartości prawdopodobieństwa łącznego, wtedy gdy jest ono potrzebne.

Niech $\underline{Z} = \{Z_1, Z_2, \dots\}$ będzie zbiorem dyskretnych zmiennych (losowych), przyjmujących wartości ze skończonych zbiorów. Siecią przekonań (siecią bayesowską) nad \underline{Z} jest para (D, P) gdzie D jest acyklicznym (nie zawierającym cykli) grafem skierowanym. Węzły tego grafu reprezentują zmienne ze zbioru \underline{Z} , a

$$P = \{p_1(z_1|pa(z_1)), p_2(\dots), \dots\} \quad (2.48)$$

jest zbiorem prawdopodobieństw warunkowych określonych dla kolejnych zmiennych. Zapis $p_i(z_i|pa(z_i))$ oznacza prawdopodobieństwo warunkowe

$$p_i(Z_i = z_i|(pa(Z_i) = pa(z_i))) \quad (2.49)$$

gdzie $pa(Z)$ jest zbiorem rodziców węzła Z . Prawdopodobieństwa warunkowe (2.49) są zapisywane w postaci tablic przyporządkowanych węzłom (zmiennym). Wartości prawdopodobieństw warunkowych mogą być pozyskiwane z różnych źródeł, do których należą znane analityczne lub numeryczne modele obiektów, wyniki eksperymentów czynnych, wyniki eksperymentów biernych, opinie specjalistów.

Podczas tworzenia sieci przekonań można przyjąć założenie, że gałęzie sieci będą dobierane wyłącznie tak, aby reprezentowały relacje przyczynowo-skutkowe. Sieci takie nazywane są przyczynowymi sieciami przekonań. Założenie to nie obowiązuje jednak jako założenie domyślne dla wszystkich sieci. Podkreślić należy, że skierowane gałęzie ogólnej sieci przekonań wyznaczają jedynie uporządkowanie węzłów (węzeł nadrzędny / podrzędny) niezbędne dla prawidłowego reprezentowania łącznego rozkładu zmiennych losowych. Nie wskazują one na istnienie relacji przyczynowo-skutkowych.

W procesie konstruowania sieci przekonań można wyróżnić dwie fazy: określanie węzłów sieci i łączących je gałęzi czyli wyznaczanie postaci sieci oraz ustalanie wartości w tablicach prawdopodobieństw warunkowych, które są przyporządkowane węzłom sieci. Próby praktycznego zastosowania sieci pokazują, że obie fazy związane są z koniecznością rozwiązania szeregu trudnych zadań. Postać sieci i wartości prawdopodobieństw warunkowych można rozpatrywać jako wynik odpowiedniego procesu optymalizacji.

Skuteczne algorytmy wnioskowania z zastosowaniem sieci przekonań omawiane są w licznych pracach, np. [2.20], [2.16], [2.14]. Szczególną grupą trudnych zadań, omawianych w literaturze w niewielkim zakresie, jest weryfikacja i ocenianie jakości (wartościowanie) sieci.

Bibliografia

- [2.1] Bargiela A., Pedrycz W. *Granular Computing. An Introduction*. Kluwer Academic Publishers, Boston, 2003.
- [2.2] Cempel Cz. *Podstawy wibroakustycznej diagnostyki maszyn*. WNT, Warszawa, 1982.
- [2.3] Cholewa W. *Metoda diagnozowania maszyn z zastosowaniem zbiorów rozmytych*, wolumen 764 serii *Zeszyty Naukowe*. Dział Wydawnictw Politechniki Śląskiej, Gliwice, 1983.
- [2.4] Cholewa W. Unilateral diagnostic models. *2nd International Conference on Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques*, wolumen 2, strony 805–815, Senlis, 1995.
- [2.5] Cholewa W. Wybrane klasy modeli danych. *Zeszyty Naukowe Politechniki Śląskiej*, wolumen 121 serii *Mechanika*, strony 62–70, Gliwice, 1995.
- [2.6] Cholewa W. The pros and cons of neural networks. *Computer Integrated Manufacturing*, strony 43–48, Zakopane, 1996.
- [2.7] Cholewa W. Diagnostyka symulacyjna. *Problemy Eksploatacji*, 27(4):13–23, 1997.
- [2.8] Cholewa W., Kiciński J., redaktorzy. *Diagnostyka techniczna. Odwrotne modele diagnostyczne*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [2.9] Cholewa W., Kiciński J., redaktorzy. *Diagnostyka techniczna. Metody odwracania nieliniowych modeli obiektów*, wolumen 120 serii *Zeszyty*. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2001.
- [2.10] Cholewa W., Kościelny J.M. Wprowadzenie. Korbicz J., Kościelny J. M., Kowalczyk Z., Cholewa W., redaktorzy, *Diagnostyka procesów. Modele, metody sztucznej inteligencji, zastosowania*, strony 3–27. WNT, Warszawa, 2002.
- [2.11] Cholewa W., White M. Inverse vibration modelling for diagnosis of unbalance in machinery. *15th International Congress on Acoustics*, strony 503–506, Trondheim, 1995.
- [2.12] Cholewa W., White M.F. Inverse modelling in rotordynamics for identification of unbalance distribution. *Machine Vibration*, (2):157–167, 1993.
- [2.13] Cholewa W., White M.F. *Application of neural networks for inversing of vibration models*. Mechanika z. 116. *Zeszyty Naukowe Politechniki Śląskiej*, Gliwice, 1994.
- [2.14] Cowell R. G., Dawid A. P., Lauritzen S. L., Spiegelhalter D. J. *Probabilistic Networks and Expert Systems*. Springer, New York, 1999.
- [2.15] Golub G.H., Van Loan C.F. *Matrix Computations*. Hopkins University Press, Baltimore, 1989.
- [2.16] Jensen V. J. *Bayesian Networks and Decision Graphs*. Springer, New York, 2002.
- [2.17] Klimek A. *Metody doskonalenia odwrotnych modeli diagnostycznych*, wolumen 134 serii *Zeszyty Naukowe. Mechanika*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1999.
- [2.18] Moczulski W. *Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn*, wolumen 1382 serii *Zeszyty Naukowe*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [2.19] Moczulski W. *Diagnostyka techniczna. Metody pozyskiwania wiedzy*, wolumen 36 serii *Monografie*. Wydawnictwo Politechniki Śląskiej, Gliwice, 2002.
- [2.20] Pearl J. *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.
- [2.21] Powell M.J.D. A review of methods for multivariable interpolation at scattered data points. Duff I.S., Watson G.A., redaktorzy, *The State of the Art in Numerical Analysis*, strony 283–309. Clarendon Press, Oxford, 1997.
- [2.22] Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1992.

- [2.23] Shepard D. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 23rd National Conference ACM*, strony 517–524, Princeton, 1968.
- [2.24] Urbanek G. *Zastosowanie algorytmów ewolucyjnych w identyfikacji odwrotnych modeli diagnostycznych*, wolumen 132 serii *Zeszyty*. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.
- [2.25] Wojtusik J. *Wielomodele diagnostyczne maszyn wirnikowych*, wolumen 128 serii *Zeszyty*. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.
- [2.26] Żółtowski B., Ćwik Z. *Leksykon diagnostyki technicznej*. Wydawnictwo Uczelniane Akademii Techniczno-Rolniczej, Bydgoszcz, 1996.

Rozdział 3

Model BNBM

Wojciech CHOLEWA

W rozdziale tym opisano główne zalety stosowanych powszechnie rodzajów diagnostyki technicznej, tzn. diagnostyki symptomowej i diagnostyki wspartej modelami. Zwrócono uwagę na potrzebę łącznego stosowania metod występujących w obu wymienionych rodzajach diagnostyki. Zaproponowano ogólny model BNBM (belief network based model) umożliwiający połączenie koncepcji, na których bazują oba rodzaje diagnostyki.

3.1. Diagnostyka symptomowa i diagnostyka wsparta modelami

Diagnostyka obiektów technicznych może być realizowana jako

- diagnostyka symptomowa,
- diagnostyka wsparta modelami.

Istotą diagnostyki symptomowej jest wykorzystywanie wiedzy specjalisty¹, umożliwiającej wskazanie symptomów świadczących o występowaniu określonego stanu obiektu. Metody stosowane przez ten rodzaj diagnostyki będą oznaczane jako metody MDS (metody diagnostyki symptomowej). Dotyczą one różnych sposobów porównywania wyników obserwacji z opisami symptomów. Można je rozpatrywać jako zadania klasyfikacji, w których definicjami klas są opisy symptomów wskazujących na występowanie określonego stanu obiektu. Przegląd wybranych klasyfikatorów zamieszczono w rozdziale 4.

Metody MDS mogą być stosowane wtedy gdy znamy definicje symptomów. Źródłami takich definicji byli początkowo specjaliści. Wysoki stopień złożoności rozpatrywanych obiektów i ograniczone możliwości bezpośredniego formułowania opisów symptomów przez specjalistów, były przyczyną poszukiwania metod badań umożliwiających wyznaczenie definicji symptomów. Stosowane mogą być sformalizowane metody pozyskiwania wiedzy [3.2] oraz metody identyfikacji symptomów na podstawie czynnych i biernych eksperymentów diagnostycznych (patrz rozdział 2.4.1).

Przykłady metod identyfikacji modeli diagnostycznych pokazane w rozdziale 2 mogą być rozpatrywane jako dotyczące głównie diagnostyki symptomowej i bazujące na założeniu, że dysponujemy wystarczającym zbiorem danych uczących (2.9), umożliwiających ustalenie postaci modelu i wyznaczenie jego parametrów. Postępowanie takie nie pozwala jednak na bezpośrednie wykorzystanie posiadanej wiedzy związanej z rozpatrywanym obiektem. Charakterystycznym przykładem, ilustrującym występujące ograniczenia, są sieci neuronowe (neuralne). Odpowiednie ich wyznaczenie prowadzi do poprawnych modeli, których główną wadą jest brak możliwości interpretowania wartości

¹ Na przykład wybrane składowe widma drgań mogą być objawami uszkodzenia łożysk tocznych lub zużycia uzębienia w przekładni zębatej.

parametrów modelu oraz brak możliwości bezpośredniego doskonalenia modelu na podstawie posiadanej wiedzy. Skutecznym sposobem celowego wprowadzania posiadanej wiedzy do takich modeli jest działanie pośrednie polegające na uzupełnianiu danych uczących o dane reprezentujące posiadaną przez nas wiedzę. Jest to zadanie trudne, wymagające gruntownej znajomości stosowanych metod identyfikacji modeli i związane z koniecznością wielokrotnego trenowania budowanych modeli.

Postępowaniem alternatywnym do diagnostyki symptomowej jest diagnostyka wsparta modelami. Stosowane przez nią metody będą oznaczane jako metody MDM (metody diagnostyki modelowej). Jej istotą jest porównywanie wyników obserwacji działającego obiektu z wynikami symulowanego jego działania za pomocą posiadanego modelu i przy założeniu, że wejścia do procesu symulacji odpowiadają wejściom obserwowanym dla rozpatrywanego obiektu. Porównanie takie przeprowadzane jest w celu rozpoznania klasy stanu obserwowanego obiektu. Proces wnioskowania o stanie może być realizowany dwoma drogami:

- Każdej klasie stanu obiektu odpowiada inny model i wyniki zastosowania tych modeli w równoległe prowadzonych symulacjach porównywane są z wynikami obserwacji obiektu. Wyniki symulacji najbardziej podobne do wyników obserwacji, wskazują klasę stanu.
- Stosowany jest jeden model reprezentujący poprawnie działający obiekt. Residua, czyli różnice pomiędzy wyjściami tego modelu i odpowiadającymi im wynikami obserwacji stanowią następnie dane wejściowe do procesu wnioskowania, który może być prowadzony z zastosowaniem technik rozpoznawania wzorców.

Istotną zaletą diagnostyki symptomowej jest możliwość jawnego wykorzystania posiadanej wiedzy podczas budowania i stosowania omawianych modeli. W diagnostyce wspartej modelami wiedza ta może być zastosowana do strojenia używanych modeli symulacyjnych. Zarówno metody MDS jak i metody MDM, pozwalają na budowanie skutecznie działających układów diagnozujących. Liczne zalety tych niezależnie stosowanych metod skłoniły do poszukiwania metody ogólnej łączącej diagnostykę wspartą modelami i diagnostykę symptomową. Rozpatrywano możliwość szeregowego stosowania metod charakterystycznych dla obu klas badań diagnostycznych.

Wynikiem poszukiwań jest opisany w kolejnych podrozdziałach model BNBM. Należy zwrócić uwagę, że jest to model heurystyczny, tzn. taki który umożliwia uzyskanie poprawnych wyników w większości zadań ale nie gwarantuje ich uzyskania we wszystkich zadaniach. Z procedur heurystycznych korzystamy z powodzeniem w rozwiązywaniu wielu rutynowych zadań (np. podejmując decyzję o zabraniu lub pozostawieniu w domu parasola).

3.2. Struktura modelu BNBM

Model BNBM (belief network based model) jest modelem wielostopniowym. Model ten zaproponowano w celu umożliwienia bezpośredniego wykorzystania wiedzy dotyczącej rozpatrywanego obiektu lub dziedziny i pochodzącej od specjalisty lub z dowolnego repozytorium wiedzy.

Zakładamy, że dysponujemy danymi uczącymi (2.9), które mogą być rozpatrywane jako dane reprezentujące realizacje nieznanego modelu diagnostycznego. Wyjścia (2.1) tego modelu obejmują

- zmienne, których wartości są poszukiwane jako wynik procesu diagnozowania.

Wejścia (2.2) tego modelu obejmują

- zmienne (ich wartości) określające warunki działania obiektu (np. wartość prędkości obrotowej, wartość obciążenia),
- zmienne, które dla rozpatrywanego obiektu mogą nie ulegać zmianom (np. wartość cech konstrukcyjnych obiektu),
- zmienne, których wartości określają skutki działania obiektu i są określane w wyniku pomiaru lub odpowiedniej symulacji (np. wartość cech drgań).

Poszukiwany model BNBM ma określać przekształcenie

$$\underline{x} \longmapsto \underline{y} \quad (3.1)$$

Dla potrzeb modelu BNBM wprowadza się zbiór dodatkowych zmiennych reprezentujących cechy występujących w rzeczywistości lecz nieobserwowanych lub wirtualnych oddziaływań zachodzących pomiędzy obiektem i otoczeniem lub pomiędzy elementami obiektu

$$\{z_1(t), z_2(t), \dots, z_J(t)\} \quad (3.2)$$

takich że dla każdej chwili czasu t

$$z_j(t) \in Z_j ; j = 1, 2, \dots, J \quad (3.3)$$

gdzie

$$Z_1, Z_2, \dots, Z_J \quad (3.4)$$

są przestrzeniami wartości tych zmiennych (cech nieobserwowanych lub wirtualnych oddziaływań).

Zakłada się, że wartości zmiennych (3.2) mogą być wyznaczone za pomocą innych modeli reprezentujących wiedzę rozpatrywanej dziedziny, na podstawie dostępnych wejść (2.2) i wyjść (2.1) modelu. Ważnym założeniem jest również to, wartości wprowadzonych dodatkowych zmiennych (3.2) mogą być w prosty sposób interpretowane przez specjalistę. Na wartości te można nałożyć liczne ograniczenia wynikające z wiedzy specyficznej dla rozpatrywanego obiektu lub związanej z nim dziedziny. Zaletą modelu BNBM jest to, że dodatkowe zmienne (3.2) pozwalają na bezpośrednie wprowadzenie wiedzy do omawianego modelu.

Przykładem dodatkowych zmiennych (3.2) wprowadzonych do zbioru danych, które mają stanowić podstawę do wyznaczania modelu diagnostycznego umożliwiającego ocenianie poprawności aktualnego osiowania wielopodporowego wirnika na podstawie drgań podpór mogą być reakcje podpór [3.1, 3.4, 3.3]. Reakcje te nie są obserwowane bezpośrednio i w związku z tym nie występują w zbiorze wartości wejść (2.2). Ich wartości mogą być wyznaczone np. podczas gromadzenia danych uczących generowanych w badaniach symulacyjnych. Na wartości tych zmiennych mogą być nakładane dodatkowe ograniczenia wynikające np. z podstawowych warunków równowagi rozpatrywanego wirnika.

Przyjęto następującą trójstopniową strukturę modelu BNBM:

- Pierwszy stopień modelu BNBM jest zbiorem równolegle działających przybliżonych klasyfikatorów jednoklasowych *OCC* (one-class classifier)

$$\{OCC_1, OCC_2, \dots, OCC_K\} \quad (3.5)$$

Klasyfikator jednoklasowy *OCC* może być interpretowany jako definicja zbioru rozmytego. Dla każdego zbioru wartości jego wejść klasyfikator OCC_k wyznacza odpowiadający tym wejściom stopień przynależności do k -tej klasy. Tak wyznaczone stopnie przynależności mogą być interpretowane jako wartości funkcji charakterystycznej odpowiedniego zbioru rozmytego. Wejściami klasyfikatorów OCC_k są wybrane elementy zbioru wszystkich wejść modelu BNBM. Oznacza to, że zbiory wejść klasyfikatorów (3.5) nie muszą być takie same. Wyjściami klasyfikatorów *OCC* są klasy wartości zmiennych dodatkowych (3.2). Stosowanie takich klas wartości (np. *duża wartość reakcji podpory 3*) wymaga podjęcia odpowiednich decyzji dotyczących ziarnistości tych wartości.

- Drugi stopień modelu BNBM przeznaczony jest do uzgadniania wartości zmiennych dodatkowych. Podstawą przeprowadzanego uzgadniania są ograniczenia wynikające z wiedzy rozpatrywanej dziedziny. Stopień ten nie jest stopniem obligatoryjnym. Jeżeli brak jest podstaw do definiowania ograniczeń, to stopień ten może być pominięty.

- Trzeci stopień modelu BNBM jest siecią przekonań BN (belief network). Wejściami tej sieci są wyjścia drugiego stopnia. Wyjściami tej sieci są stopnie przekonania o prawdziwości stwierdzeń reprezentujących wartości (odpowiadających wartościom) wyjść (2.1) modelu BNBM.

W zależności od potrzeb można przyjmować różne zasady łączenia drugiego i trzeciego stopnia modelu, uwzględniające specyficzne własności klasyfikatorów jednoklasowych. Odpowiedzi pozytywne tych klasyfikatorów pozwalają na uznawanie przynależności elementu do klasy bez konieczności interpretowania odpowiedzi negatywnych jako podstaw do uznawania braku przynależności elementu do klasy. Oznacza to, że połączenie drugiej i trzeciej warstwy modelu BNBM można realizować alternatywnie

- Wyjście drugiego stopnia ustala wartość wejściowego węzła sieci na trzecim stopniu.
- Wyjście drugiego stopnia, którego wartość przekracza zadaną wartość progową ustala wartość wejściowego węzła sieci na trzecim stopniu. Wyjście drugiego stopnia, które nie przekracza tej wartości progowej nie pociąga za sobą zmian wartości wejściowego węzła sieci na trzecim stopniu. Wartość tego węzła pozostaje równa wartości domyślnej.

3.3. Identyfikacja modelu BNBM

Identyfikacja modelu BNBM jest procesem iteracyjnym prowadzonym dla odpowiednio przygotowanego zbioru danych uczących i jego podziału na dane trenujące i testujące.

- (A) Pierwszym krokiem procesu identyfikacji modelu BNBM jest parametryzacja ziarnistości zbioru wartości zmiennych dodatkowych (3.2) i przyjęcie początkowych wartości wprowadzonych parametrów.
- (B) Kolejnym krokiem jest niezależne budowanie (identyfikacja) każdego klasyfikatora OCC_k oraz budowanie (identyfikacja) sieci przekonań BN , na podstawie danych trenujących.
- (C) Następnym krokiem jest testowanie otrzymanego modelu polegające na zastosowaniu kolejnych jego stopni dla danych testujących.

Model otrzymany w wyniku opisanych działań może być doskonalony. Optymalizacja modelu może polegać na modyfikowaniu wartości parametrów określających ziarnistość zmiennych dodatkowych. Dogodnym sposobem takiej optymalizacji jest zastosowanie algorytmów ewolucyjnych, które w kolejnych cyklach obejmujących kroki (B) i (C) realizują odpowiednie operacje genetyczne na wartościach rozpatrywanych parametrów.

Bibliografia

- [3.1] Kiciński J., redaktor. *Modelowanie i diagnostyka oddziaływań mechanicznych, i magnetycznych w turbozespołach energetycznych*, Gdańsk, 2005. Wydawnictwo IMP PAN.
- [3.2] Moczulski W. *Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn*, wolumen 1382 serii *Zeszyty Naukowe*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [3.3] Urbanek G. *Zastosowanie algorytmów ewolucyjnych w identyfikacji odwrotnych modeli diagnostycznych*, wolumen 132 serii *Zeszyty*. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.
- [3.4] Wojtusik J. *Wielomodele diagnostyczne maszyn wirnikowych*, wolumen 128 serii *Zeszyty*. Politechnika Śląska, Katedra Podstaw Konstrukcji Maszyn, Gliwice, 2006.

Rozdział 4

Klasyfikatory

Tomasz ROGALA

Lokalizacja oraz identyfikacja uszkodzeń wymaga stosowania odpowiednich metod pozwalających na uzyskanie odpowiedzi, do jakiej klasy stanu należy obserwowany obiekt techniczny? W najprostszym przypadku może być to odpowiedź na pytanie, czy obiekt jest zdalny i spełnia stawiane mu wymagania zgodnie z charakterystyką techniczną lub jest niezdatny i na przykład powinien zostać wyłączony z dalszej eksploatacji. Ocena takiego jak i bardziej szczegółowych zagadnień nazywana jest diagnozowaniem i polega na ocenie zmian stanu lub jego identyfikacji prowadzonej na podstawie obserwacji wyjść badanego obiektu. W diagnostyce można wyróżnić dwie grupy metod diagnozowania [4.12]. Pierwsza z nich to metody wnioskowania, których podstawą jest znajomość przynajmniej częściowej relacji stan-symptom. Druga grupa to metody klasyfikacji, których zadaniem jest przyporządkowanie odpowiednich symptomów właściwym klasom reprezentującym dany stan obiektu lub procesu. Relacje diagnostyczne, w tym ostatnim przypadku, pozyskiwane są w procesie uczenia maszynowego. Tej grupie metod poświęcony jest niniejszy rozdział.

4.1. Klasyfikacja w procesie diagnozowania

Proces diagnozowania z wykorzystaniem metod klasyfikacji jest zadaniem złożonym, a jego odpowiednie przeprowadzenie stanowi o końcowej jakości systemu diagnostycznego. W niniejszej sekcji poruszono niektóre aspekty związane z identyfikacją klasyfikatorów dla potrzeb stosowanych dotychczas systemów wnioskujących, w tym również identyfikacji klasyfikatorów dla potrzeb modelu BNBM.

Proces przyporządkowywania nowych wektorów symptomów odpowiednim klasom stanu poprzedzony jest procesem identyfikacji klasyfikatora i pozyskaniem relacji diagnostycznych. Pozyskanie relacji wymaga odpowiedniego przeprowadzenia eksperymentu symulacyjnego, podczas którego, najczęściej dla zadanych „ukrytych” klas stanów obiektu pozyskiwany jest zbiór przykładów uczących. Wybór relewantnych klas stanów nie zawsze jest zadaniem trywialnym [4.15], [4.8], szczególnie wtedy gdy konieczne jest opracowanie klas typu np. duże reakcje podporowe, małe reakcje podporowe, dopuszczalne reakcje podporowe itp.. W takim przypadku definiowanie klas jest dość często kwestią dyskusyjną. Przydatnym narzędziem może okazać się zastosowanie tzw. klasyfikatorów przybliżonych. Klasyfikatory te, w przeciwieństwie do klasyfikatorów dokładnych, przyporządkowują nowym zbiorom cech stopień przynależności do zadanego zbioru klas. W ten sposób określane jest „podobieństwo” rozpatrywanego przykładu do rozpatrywanych klas wcześniej zdefiniowanych na podstawie identyfikacji klasyfikatorów. Istotną trudnością w przypadku stosowania klasyfikatorów różnego typu jest odpowiednie skalowanie funkcji przynależności, które dla klasyfikatorów parametrycznych może być określone na podstawie prawdopodobieństwa posteriori, natomiast w przypadku klasyfikatorów nieparametrycznych może być zdefiniowane na podstawie np. wartości funkcji potencjalnej.

W badaniach diagnostycznych, często operujemy ograniczonym zbiorem danych uczących. Z tego względu zadaniem trudnym, a wręcz niemożliwym jest określenie typu i rozkładu prawdopodobieństwa dla poszczególnych zbiorów danych uczących, reprezentantów różnych klas stanów. Wymaga to arbitralnego przyjęcia rozkładu dla danych uczących. Konieczne staje się w tym przypadku stosowanie nieparametrycznych metod klasyfikacji, ponieważ nie wymagają one identyfikacji parametrów rozkładów tak jak ma to miejsce w przypadku metod parametrycznych, gdzie konieczna jest wiedza na temat charakteru rozkładu funkcji gęstości prawdopodobieństwa.

Biorąc pod uwagę konieczność doskonalenia klasyfikatorów wykorzystywanych w systemach diagnozujących, a w szczególności w odwrotnych modelach diagnostycznych ważnym aspektem jest zdolność klasyfikatorów do ich iteracyjnego doskonalenia na bazie nowych przykładów uczących. Można tu wymienić klasyfikatory, które są budowane od nowa w przypadku pozyskania nowych przykładów uczących jak i klasyfikatorów „sterowanych” za pomocą danych, ponieważ nie wymagają one prowadzenia ponownego procesu identyfikacji.

Prowadzenie eksperymentu symulacyjnego w celu pozyskania przykładów uczących wymaga przygotowania stosunkowo równomiernych zbiorów przykładów uczących dla każdej z przyjętych klas stanów. Ma to istotne znaczenie w przypadku identyfikacji klasyfikatora wieloklasowego. Uczenie klasyfikatora wieloklasowego na zbiorze przykładów gdzie liczebność jednej z klas jest wyraźnie mniejsza od pozostałych może prowadzić do wyznaczenia niewłaściwej granicy w przestrzeni cech [4.13], a tym samym do pogorszenia jakości klasyfikacji. Przeciwnie do opisanych klasyfikatorów wieloklasowych zachowują się klasyfikatory jednoklasowe, dla których wyznaczana jest jedynie jedna granica określająca przynależność do rozpatrywanej klasy [4.13] co pozwala identyfikować klasyfikator dla dowolnego reprezentatywnego zbioru przykładów uczących w których przykłady są nierównomiernie rozłożone.

Podsumowując, ważnymi aspektami przy wyborze typu klasyfikatorów dla potrzeb modelu BNBM jest :

- zdolność do doskonalenia klasyfikatorów w sposób czynny, a więc w sposób sterowany zewnątrz na podstawie nowych danych co stanowi ich istotną zaletę w przypadku np. systemów on-linowych,
- zdolność do uzyskiwania przybliżonego rozwiązania, np. na zasadzie uzyskania stopnia przynależności do rozpatrywanej klasy,
- zdolność do korzystania z danych niesymetrycznych w których liczebność przykładów uczących należących do rozpatrywanej klasy różni się w stosunku do liczebności pozostałych klas,
- stosowanie w większości przypadków nieparametrycznych metod uczenia klasyfikatorów.

Biorąc pod uwagę omawiane w niniejszej pracy badania dotyczące modelowania diagnostycznego oraz modeli odwrotnych w dalszej części dokonano krótkiego opisu klasyfikatorów jednoklasowych oraz wieloklasowych, które mogą być stosowane w systemach diagnostycznych. W związku z powyższym w dalszej części przedstawiono nieparametryczne (z wyjątkiem klasyfikatora bayesowskiego) klasyfikatory wielo- i jednoklasowe, dla których możliwe jest uzyskanie wyniku w postaci przybliżonej. Opisane klasyfikatory należą zarówno do grupy klasyfikatorów, które mogą być doskonalone w sposób czynny i bierny. Dodatkowo opis został uzupełniony o informacje dotyczące bibliografii, gdzie można znaleźć szczegółowe informacje na temat danego typu klasyfikatora. Dla każdego z przedstawionych w dalszej części klasyfikatorów wskazano ogólnodostępną bibliotekę procedur napisanych w środowisku obliczeń inżynierskich Matlab.

4.2. Klasyfikatory wieloklasowe

Najbardziej konwencjonalne metody klasyfikacji prowadzone są z zastosowaniem klasyfikatorów wieloklasowych. Dają one możliwość opisywania bardzo dużych zbiorów danych uczących, reprezentujących różne klasy stanu obiektu. Poniżej przedstawiono wybrane klasyfikatory wieloklasowe z przeznaczeniem do stosowania w modelowaniu systemów diagnostycznych. Niektóre z wymienionych

klasyfikatorów definiowane są na podstawie estymat funkcji gęstości prawdopodobieństwa i należą do tzw. metod parametrycznych. Przykładem może być parametryczny klasyfikator bayesowski, który przedstawia najbardziej ogólny przypadek klasyfikacji, w którym nie są znane zarówno rozkłady prawdopodobieństw warunkowych jak i a priori.

4.2.1. Klasyfikator bayesowski

Problemy związane z klasyfikacją łączą się nieodłącznie z teorią podejmowania decyzji. Działania te są najczęściej wspierane statystyką bayesowską [4.8] poprzez wprowadzenie rachunku prawdopodobieństwa oraz funkcji kosztów związanych z podejmowanymi decyzjami. Oprócz wspomnianej funkcji kosztów, jedynym koniecznym warunkiem wykorzystania rachunku Bayes'a jest znajomość wszystkich prawdopodobieństw istotnych dla uzyskania rozwiązania [4.15].

Zakładając, że badany jest obiekt techniczny, który opisany jest wektorem cech $\mathbf{x} \in X$ zebranych w wyniku pomiarów oraz ukrytymi zmiennymi klas stanów $c \in C$ klasyfikacja bayesowska polega na sklasyfikowaniu odpowiedniego zbioru tych danych $\mathbf{x} \in X$ do odpowiedniego zbioru klas $c \in C$ za pomocą odpowiedniej reguły decyzyjnej $q: X \rightarrow C$ określonej poprzez klasyfikator bayesowski [4.9]. Zakładając określoną liczbę klas K $C = \{c_1, c_2, c_3, \dots, c_K\}$ przyjmujemy, że poszczególne klasy c są wzajemnie wykluczające się, a prawdopodobieństwo wystąpienia określonej klasy definiuje się jako prawdopodobieństwo wystąpienia klasy c niezależnie od zebranego zbioru danych X . Dla skończonego zbioru klas c obowiązuje warunek [4.15]:

$$\sum_{n=1}^K P(c_n) = 1 \quad (4.1)$$

Każdy obiekt techniczny należący do innej klasy stanu powinien posiadać inny wektor danych wyjściowych \mathbf{x} . Niestety, obiekty należące do tej samej klasy stanu posiadają często różne wektory danych. Statystyka bayesowska uwzględniając ten fakt posługuje się gęstością prawdopodobieństwa warunkowego $p(\mathbf{x}|c_n)$ [4.15]. Funkcja ta oznacza gęstość prawdopodobieństwa warunkowego wystąpienia określonego wektora danych \mathbf{x} z obiektu O pod warunkiem jego przynależności do określonej klasy c_n stanu. Jeżeli jednak obiekt pochodzi z nieznannej klasy to gęstość prawdopodobieństwa wystąpienia określonego wektora danych \mathbf{x} jest ważoną funkcją prawdopodobieństwa a priori [4.15]

$$p(\mathbf{x}) = p(\mathbf{x}|c_n) \sum_{n=1}^K P(c_n) \quad (4.2)$$

natomiast prawdopodobieństwo przynależności danego wektora cech \mathbf{x} do klasy c_n jest zgodnie z twierdzeniem Bayes'a równe [4.15]:

$$P(c_n|\mathbf{z}) = \frac{p(\mathbf{x}|c_n)P(c_n)}{P_{\mathbf{x}}} p(\mathbf{x}|c_n) \quad (4.3)$$

W celu określenia klasyfikatora konieczne jest również określenie funkcji kosztów $W(q|c_n)$, która może przyjmować różną postać [4.9]. Ostatecznie brana jest pod uwagę wartość oczekiwana z prawdopodobieństwa wystąpienia wartości funkcji kosztów W pod warunkiem, że zastosowano określoną regułę decyzyjną q . Funkcję tę często określa się mianem ryzyka bayesowskiego lub ryzyka warunkowego [4.9]:

$$R(q|\mathbf{x}) = E\{W(q|c_n)|\mathbf{x}\} = \sum_{n=1}^K W(q|c_n)P(c_n|\mathbf{z}) \quad (4.4)$$

W wyniku działania algorytmu poszukiwany jest taki element c_n w zbiorze klas C , dla którego funkcja 4.4 przyjmie wartość minimalną.

$$R_{opt} = \operatorname{argmin}_{c_n \in C} R(q|\mathbf{x}) \quad (4.5)$$

Bibliografia Szczegółowe informacje na temat klasyfikacji bayesowskiej można znaleźć w pozycjach [4.8], [4.6], [4.11], [4.15].

Oprogramowanie Przykładowe procedury w środowisku Matlab dotyczące klasyfikatorów bayesowskich można znaleźć w Statistical Pattern Recognition Toolbox for Matlab [4.9] oraz w pakiecie oprogramowania Netlab [4.10] [4.11].

Przykładowo w pakiecie STPRtool zaimplementowano dwie procedury dla których określono funkcje kosztów dotyczących minimalizacji błędnego sklasyfikowania przykładu do określonej klasy, oraz procedurę z możliwością odrzucenia tj. z klasą uwzględniającą brak zakwalifikowania przypadku do którejkolwiek z klas.

4.2.2. Liniowa funkcja dyskryminacyjna

Zadanie klasyfikacji realizowane poprzez klasyfikator bayesowski wymaga znajomości prawdopodobieństw a priori oraz prawdopodobieństw warunkowych. Analogicznie zastosowanie liniowych funkcji dyskryminacyjnych wymaga znajomości postaci tych funkcji w celu określenia parametrów klasyfikatora na podstawie dostępnych wektorów cech. Jest to zatem metoda nieparametryczna.

Zatem reguła przyporządkowująca elementy ze zbioru N wymiarowego wartości cech X do odpowiedniej klasy jest spełniona, gdy odpowiadająca danej klasie funkcja dyskryminacyjna $f_c(x)$ osiąga wartość maksymalną, co można zapisać jako [4.9]:

$$q : X \longrightarrow C = c_1..c_n \quad (4.6)$$

$$c = \operatorname{argmax}_{c \in C} f_c(x) = \operatorname{argmax}_{c \in C} (w_c x + b_c) \quad (4.7)$$

gdzie $x \in X$ jest wektorem wejściowym, $w \in R^N$ jest wektorem wag natomiast b_c jest przesunięciem liniowym dla każdej klasy $c \in C$.

Naturalnym rozszerzeniem tej metody jest stosowanie uogólnionych funkcji dyskryminant, które są stosowane do budowy nieliniowych klasyfikatorów np. kwadratowych funkcji dyskryminacyjnych [4.9]. Wieloklasowy wektor tych funkcji może przyjmować następującą formę:

$$c = (w_c x + b_c) \quad (4.8)$$

gdzie

$$x = [x, x_1^2, x_2^2, \dots, x_1 x_2, x_2 x_3, \dots] \quad (4.9)$$

Wektor parametrów klasyfikatora w jest uzyskiwany w procesie uczenia, w taki sposób aby uzyskać jak najlepszą wydajność klasyfikatora dla danych trenujących. Stosowane są w tym celu zarówno algorytmy nieiteracyjne jak i iteracyjne popularnie stosowane w metodach optymalizacji.

Jednym z najprostszych klasyfikatorów jest tzw. perceptron liniowy. Jest to klasyfikator binarny, którego podstawową cechą jest podział na dwie klasy $C = \{c_1, c_2\}$ [4.15]. Algorytm jest bardzo prosty i wymaga jedynie wstępnego przygotowania danych w taki sposób, że określony wektor danych przykładowych dla klasy pierwszej jest dodatni, natomiast wektor danych dla drugiej klasy jest ujemny. Parametry klasyfikatora określa dwójka wektorów [4.9]:

$$v = [w, b] \quad (4.10)$$

Algorytm perceptronowy w iteracyjny sposób sprawdza dla każdego rekordu danych warunek:

$$(v x_i) > 0 \quad (4.11)$$

dla każdego $i = 1..N$, gdzie x_i jest wektorem przygotowanych danych wejściowych. Jeżeli dane są liniowo separowalne, to wydajność klasyfikatora jest zadowalająca.

Algorytm ten jest dostępny w [4.15], [4.9].

Innym przykładem jest liniowa dyskryminanta Fishera, w przypadku której parametry liniowej funkcji dyskryminacyjnej określane są na podstawie następującego kryterium separowalności:

$$\mathbf{w}' = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w} \mathbf{S}_b \mathbf{w}}{\mathbf{w} \mathbf{S}_w \mathbf{w}} \quad (4.12)$$

gdzie \mathbf{S}_b jest macierzą rozproszenia pomiędzy klasami, \mathbf{S}_w jest wewnętrzną macierzą rozproszenia klasy, \mathbf{w} jest wektorem parametrów kierunkowych liniowej funkcji dyskryminacyjnej natomiast, \mathbf{w}' jest takim wektorem tych parametrów, dla których kryterium separowalności przyjmuje wartość maksymalną. Wyznaczenie współczynnika przesunięcia liniowego, wymaga uwzględnienia dodatkowego warunku np. dla binarnego klasyfikatora spełnienia następującej równości:

$$\mathbf{w} \nu_1 + b = -(\mathbf{w} \nu_2 + b) \quad (4.13)$$

gdzie ν_1, ν_2 są wartościami średnimi dla przykładów należących do jednej i drugiej klasy.

Bibliografia Szczegółowe informacje na funkcji dyskryminacyjnych można znaleźć w pozycjach [4.8], [4.6], [4.1].

Oprogramowanie Matlab z toolboxem Statistics, [4.1], oraz Statistical Pattern Recognition Toolbox for Matlab [4.9].

4.2.3. Klasyfikator Parzena - probabilistyczne sieci neuronowe

Neuronowa wersja klasyfikatora Parzena opiera się na zastosowaniu prostych probabilistycznych sieci neuronowych. Klasyfikacja ta jest nieparametryczną wersją klasyfikatora bayesowskiego, w którym prawdopodobieństwo posteriori jest wyznaczane na podstawie danych z wykorzystaniem okien Parzena [4.8], [4.2]. Technika ta polega na sparametryzowaniu nieznanymi prawdopodobieństw warunkowych lub a priori i estymacji tych parametrów na podstawie danych. Estymacji dokonuje się podając szerokość okna, oraz odpowiednią całkowalną funkcję okna. Neuronowa wersja klasyfikatora Parzena jest budowana w postaci dwuwarstwowej sieci, gdzie liczebność wyjść oraz wejść zależy od wielkości wektora danych oraz liczby klas. Sieć jest trenowana w następujący sposób, w pierwszej kolejności dane wejściowe są normalizowane, a następnie przetwarzane w pierwszej warstwie sieci. Następnie każda próbka danych otrzymuje neuron ze znormalizowaną wartością wag \mathbf{w} . W wyniku iteracyjnego procesu trenowania dla każdego przykładu uczącego przyjmowany jest odpowiedni wektor wag zgodny z tymi danymi. Podczas klasyfikacji dane testujące są poddane iloczynowi z wcześniej uzgodnionymi wagami, a następnie uzyskany wynik jest przetwarzany przez funkcję przejścia o postaci [4.8], [4.2]:

$$f_a = e^{\frac{z-1}{\sigma^2}} \quad (4.14)$$

i otrzymywany jest sygnał aktywacji. Neuron wyjściowy przyporządkowany odpowiedniej klasie sumuje sygnały aktywacji ze wszystkich neuronów pierwszej warstwy wejściowej. Klasa dla której dany przykład testujący jest najbardziej odpowiedni jest wskazana poprzez największą wartość.

Parametrem sterującym jest w tym przypadku wartość σ odpowiadająca szerokości efektywnej okna gaussowskiego.

Bibliografia Szczegółowe informacje na temat okien Parzena oraz probabilistycznych sieci neuronowych można znaleźć w pozycjach [4.8], [4.2], [4.11], [4.1].

Oprogramowanie Matlab z toolboxem Neural networks, [4.1], Pakiet NETLAB [4.10]

4.2.4. Klasyfikator k-najbliższych sąsiadów knn

Klasyfikator k-najbliższych sąsiadów jest klasyfikatorem nieparametrycznym. Interpretacja tej metody klasyfikacji jest bardzo prosta. Klasa $c \in C$ przydzielona do odpowiedniego zbioru przykładów $\mathbf{x} \in X$ uczących na podstawie reguły $q : X \longrightarrow Y$ jest klasą z największą wartością wyznaczoną na podstawie metryki dla zbioru k przykładów najbliższych zbiorowi przykładów uczących \mathbf{x} [4.9].

$$q(\mathbf{x}) = \operatorname{argmax}_{c \in C} d(\mathbf{x}, c) \quad (4.15)$$

Prowadząc klasyfikację w przestrzeni \mathbb{R}^N można stosować różne miary odległości pomiędzy elementami \mathbf{x} i \mathbf{y} tej przestrzeni [4.7]:

- Minkowskiego

$$d_p(x, y) = \frac{1}{l} \sum_{i=1}^l |x[i] - y[i]|^p \quad (4.16)$$

- Euklidesa

$$d(x, y) = \frac{1}{l} \sum_{i=1}^l (x[i] - y[i])^2 \quad (4.17)$$

- Hamminga

$$d(x, y) = \frac{1}{l} \sum_{i=1}^l |x[i] - y[i]| \quad (4.18)$$

- Sebestyena

$$d(x, y) = (x - y)W(x - y)^T \quad (4.19)$$

gdzie W jest diagonalną macierzą wag.

- Mahalanobisa

$$d(x, y) = (x - y)C^{-1}(x - y)^T \quad (4.20)$$

gdzie C^{-1} jest odwrotną macierzą kowariancji.

- Canbera

$$d_p(x, y) = \frac{1}{l} \sum_{i=1}^l \frac{|x[i] - y[i]|}{x[i] + y[i]} \quad (4.21)$$

gdzie C^{-1} jest odwrotną macierzą kowariancji.

Istotną zaletą omawianego algorytmu jest fakt, że klasyfikacja jest "sterowana" danymi, co nie powoduje konieczności ponownego wyznaczania klasyfikatora. Wadą tej metody klasyfikacji jest konieczność wyboru odpowiedniej metryki, która nie zawsze jest optymalna. Znalezienie właściwej jest zadaniem trudnym. W przypadku bardzo dużych zbiorów danych uczących złożoność obliczeniowa jest znaczna.

Bibliografia Szczegółowe informacje na temat klasyfikatorów najbliższego sąsiada, można znaleźć w większości pozycji literaturowych dotyczących problemów klasyfikacji i grupowania m.in.: [4.8], [4.2], [4.11], [4.1].

Oprogramowanie Matlab [4.1], Netlab [4.10], STPRTTool [4.9], PRTTools [4.14] itp..

4.2.5. Klasyfikatory neuronowe

Neuronowy klasyfikator perceptronowy

W rozdziale 4.2.2 przedstawiono perceptron liniowy spełniający rolę klasyfikatora binarnego. Perceptronowa sieć neuronowa jest rozszerzeniem tego klasyfikatora, łącząc wyjścia kilku perceptronów z innym perceptronem. Możliwe jest zatem rozpatrywanie przypadków wieloklasowych, gdzie liczba identyfikowanych klas najczęściej odpowiada liczbie wyjść warstwy wyjściowej sieci neuronowej. W odróżnieniu od perceptrona liniowego jako funkcja aktywacji neuronu wykorzystywana jest funkcja signum, w sieci neuronowej funkcja ta może przyjmować bardziej ogólną postać funkcji ciągłej np. funkcji sigmoidalnej [4.15].

$$f(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + \exp^{-\mathbf{w}^T \mathbf{x}}} \quad (4.22)$$

gdzie \mathbf{w} jest wektorem wag, natomiast \mathbf{x} jest wektorem cech sygnałów. Najczęściej neurony stosowane w sieciach połączone są w tzw. warstwy. Jeżeli wszystkie warstwy neuronowe generują swoje wyjścia na podstawie wyjść wcześniejszych warstw, to sieć taka nazywana jest siecią przednią jeżeli są sprzężenia zwrotne pomiędzy warstwami, to sieci te są nazywane wstecznymi.

Jeżeli sieć posiada n ukrytych warstw to wyjście sieci obliczane jest w następujący sposób:

$$g(\mathbf{x}) = f\left(\sum_{k=1}^n \mathbf{v}_{z,k} f(\mathbf{w}_k^T \mathbf{x} + \mathbf{v}_{z,k+1})\right) \quad (4.23)$$

gdzie \mathbf{w}_k jest wektorem wag pomiędzy warstwą $k - 1$, a neuronem w warstwie k , a \mathbf{v}_k jest wektorem wag pomiędzy warstwą k , a warstwą $k + 1$.

Uczenie sieci polega na wyznaczeniu wartości tych wag oraz wartości stałych (nie przedstawionych w powyższym wzorze). W tym celu wykorzystuje się szereg metod rozwiązywania równań, m.in. metodę najmniejszych kwadratów [4.15].

$$J_{mse} = \frac{1}{2} \sum_{n=1}^{N_s} \sum_{k=1} K(g_k(x_n) - t_{t,n})^2 \quad (4.24)$$

gdzie $t_{n,k}$ przyjmuje określone wartości w zależności, czy dana wejściowa y_n należy do rozpatrywanej klasy, czy też nie. Stosowane w sieciach neuronowych funkcje przejścia powinny być funkcjami różniczkowalnymi, tak aby możliwe było wyznaczanie pochodnych cząstkowych błędu kwadratowego względem wag.

Nie ma jednoznacznych przesłanek dotyczących struktury sieci tj. rozmiarów warstw ukrytych tych sieci. Dobór odpowiedniego klasyfikatora bazującego na sieciach neuronowych wymaga stosowania sieci o różnych strukturach i wyborze na podstawie odpowiednich miar jakości i wydajności sieci właściwej struktury.

Neuronowy klasyfikator radialny

W odróżnieniu od klasyfikatora perceptronowego, neuronowy klasyfikator radialny jest siecią o określonej strukturze. Zazwyczaj składa się z trzech warstw:

- wejściowej o liniowej funkcji przejścia, wagach wejść równych jednej oraz składowych stałych różnych zero i liczbie neuronów równej liczbie wymiarów rozpatrywanej przestrzeni cech,
- ukrytej, z nieliniową radialną funkcją przejścia, wagach wejść równych jeden oraz składowych stałych równych zero,
- wyjściowej, składającej się z jednego neuronu o liniowej funkcji przejścia, składowej stałej równej zero i wagach wyjść wyznaczanych w procesie uczenia sieci.

Dobór liczby neuronów warstwy ukrytej jest wyznaczany w procesie identyfikacji sieci. Neurony są dodawane w sposób iteracyjny, aż do osiągnięcia minimalnego błędu średniokwadratowego, lub osiągnięcia maksymalnej liczby neuronów równej liczbie rozpatrywanych klas.

Bibliografia Szczegółowe informacje na temat klasyfikatorów neuronowych, dotyczących problemów klasyfikacji i grupowania m.in: [4.8], [4.2], [4.11], [4.1]

Oprogramowanie Matlab z pakietem Neural Toolbox [4.1], Netlab [4.10].

4.2.6. Klasyfikator SVM

Metoda wektorów wspierających jest metodą identyfikacji klasyfikatora na podstawie wyznaczania hiperpłaszczyzny separującej elementy należące do różnych klas.

Wieloklasowy klasyfikator SVM wykorzystuje zbiór liniowych funkcji dyskryminujących

$$f_c(\mathbf{x}) = \langle \alpha \mathbf{k}_s + \mathbf{b}_c \rangle \quad (4.25)$$

dla każdej klasy $c \in C$. gdzie α_c jest wektorem wag dla rozpatrywanej klasy. \mathbf{k}_s jest wektorem ocen $\mathbf{k}(s) = [k(x, s_1), k(x, s_2), k(x, s_2), \dots, k(x, s_d)]$ wyznaczonych na podstawie funkcji jądra wyśrodkowanych na podstawie wektorów wspierających $S = [s_1, s_2, \dots, s_d]$, które są zazwyczaj podzbiorem przykładów uczących. Wektor \mathbf{b}_c jest wektorem przesunięcia funkcji dyskryminującej.

Klasyfikacja wieloklasowa polega na spełnieniu takiej reguły,

$$q(\mathbf{x}) = \operatorname{argmax}_{c \in C} f_c(\mathbf{x}) \quad (4.26)$$

która za pomocą zbioru funkcji dyskryminującej przyporządkowuje elementy przestrzeni cech odpowiednim klasom stanu.

W przeciwieństwie do innych metod, metoda SVM poszukuje jednego rozwiązania szczególnego polegającego na wyznaczeniu największego przedziału wokół granicy klas, która nie zawiera żadnych przykładów uczących. W przypadku klas liniowo nieseparowalnych w przestrzeni cech \mathbb{R}^l poszukuje się przestrzeni \mathbb{R}^L o większej liczbie wymiarów ($L > l$), w której zadanie będzie liniowo separowalne. Przestrzeń ta jest określona funkcją jądra, do najczęściej stosowanych należą funkcje [4.13]:

- liniowe,

$$k(x_i, s_j) = x_i^T s_j \quad (4.27)$$

- wielomianowe,

$$k(x_i, s_j) = (x_i^T s_j + \theta)^p \quad (4.28)$$

- radialne.

$$k(x_i, s_j) = \exp\left(-\frac{\|x_i - s_j\|}{\sigma^2}\right) \quad (4.29)$$

Możliwe jest stosowanie wielu strategii wyznaczania hiperpłaszczyzny, np. jeden przeciwko wszystkim, każdy z każdym itp..

Bibliografia Więcej szczegółów na temat klasyfikatorów SVM można znaleźć m.in. w: [4.16], [4.5], [4.8], [4.2], [4.11], [4.1].

Oprogramowanie Matlab svm toolbox [4.1], SVM and Kernel Methods Matlab Toolbox [4.4], SimpleSVM toolbox [4.3] [4.10].

4.3. Klasyfikatory jednoklasowe

Każde zadanie klasyfikacji wieloklasowej można rozłożyć na kilka prostych zadań klasyfikacji dwuklasowej, gdzie konieczna jest znajomość informacji tylko o jednej z tych klas [4.13]. Zadanie to różni się od konwencjonalnego zadania wieloklasowego tym, że rozróżnia się jedynie dwa zbiory przykładów uczących. Pierwszy z nich to zbiór przykładów uczących należących do rozpatrywanej klasy oraz zbiór przykładów uczących nie należących do rozpatrywanej klasy. Wytrenowany na takim zbiorze klasyfikator nazywany jest klasyfikatorem jednoklasowym.

Klasyfikatory jednoklasowe posługują się dwoma wielkościami [4.13], pierwsza z nich to miara odległości $d(\mathbf{x})$ lub prawdopodobieństwa $p(\mathbf{x})$ przynależności danego zbioru przykładów uczących do rozpatrywanej klasy. Zarówno prawdopodobieństwa jak i odległości stosowane są w zależności od wykorzystywanej metody parametrycznej lub nieparametrycznej. Drugą wielkością jest próg ustalony na wcześniej ustalonej miarze odległości lub prawdopodobieństwa.

Reguła przyporządkowania do zbioru klas ma postać, dla miar odległościowych:

$$q(\mathbf{x}) = I(d(\mathbf{x}) < \theta_d) \quad (4.30)$$

gdzie θ_d jest ustalonym progiem odległości.

W większość metod klasyfikacji jednoklasowych w pierwszej kolejności optymalizowana jest miara odległości lub prawdopodobieństwa, a dopiero później ustalany jest próg. W niektórych przypadkach jak również w przypadku metody *SVM* kolejność ta jest odwrotna [4.13].

Można wyróżnić trzy grupy metod rozwiązywania problemu klasyfikacji jednoklasowej [4.13]. Pierwsza z nich polega na estymacji gęstości przykładów uczących i ustawienia progu dla rozkładów gęstości dla tych danych jest to grupa metod należących do tzw metod uczenia parametrycznych. Zadanie to wymaga przygotowania bardzo licznego i odpowiednio przygotowanego zbioru danych uczących co nie zawsze jest możliwe. Druga z metod polega na wyznaczeniu zamkniętej granicy wokół danych uczących należących do rozpatrywanej klasy. Zadanie to nie wymaga znacznej liczby przykładów uczących w porównaniu do pierwszej grupy metod, ale jest czuła na skalowanie danych, wymagając tym samym odpowiedniego przygotowania danych. Trzecia z metod należy do grupy metod rekonstrukcji, które są raczej wykorzystywane do modelowania danych i w dalszej części nie będą rozważane. Ponieważ stwierdzono istotną przydatność metod nieparametrycznych dla potrzeb systemów diagnozujących w dalszej części zostaną przedstawione tylko klasyfikatory jednoklasowe należące do grupy metod nieparametrycznych. Podobnie jak w przypadku klasyfikatorów wieloklasowych podstawowymi informacjami na temat jakości uzyskanych klasyfikatorów są dwie miary błędów [4.13]:

- czyli liczba błędnie sklasyfikowanych obiektów nie należących do rozpatrywanej klasy,
- jak i liczba niesklasyfikowanych obiektów, które należą do rozpatrywanej klasy.

4.3.1. Klasyfikator K-środków

Metoda wyznaczania granic klas z wykorzystaniem tego klasyfikatora polega na pokrywaniu przykładów danych należących do klasy sferami o określonej średnicy. Środek sfery określany jest na podstawie znalezienia minimalnej odległości pomiędzy przykładami uczącymi i środkiem klasy. Podczas dopasowywania, minimalizowany jest błąd [4.13]:

$$\varepsilon_{km} = \max_i (\min_k \|x_i - \mu_k\|^2) \quad (4.31)$$

Metoda jest czuła na występowanie danych, które nie należą do przykładów należących do danej klasy ale są blisko, powodując zwiększanie frakcji przykładów błędnie sklasyfikowanych do rozpatrywanej klasy.

Bibliografia Więcej szczegółów na temat klasyfikatorów k-środków można znaleźć m.in. w: [4.13].

Oprogramowanie Dodatek DD tools do pakietu PRTools [4.14].

4.3.2. Jednoklasowy klasyfikator najbliższego sąsiada

W algorytmie jednoklasowego klasyfikatora najbliższego sąsiada analogicznie jak w przykładzie wieloklasowym ocena prowadzona jest na podstawie gęstości komórki, która zwiększając swą podprzestrzeń w przestrzeni cech obejmuje k przykładów ze zbioru uczącego. Lokalna gęstość klasyfikatora jest oceniana na podstawie zależności [4.13]:

$$p_{NN}(\mathbf{x}) = \frac{k/N}{V_k(\|\mathbf{x} - NN_k^{tr}(\mathbf{x})\|)} \quad (4.32)$$

gdzie NN_k^{tr} przedstawia k najbliższych sąsiadów ze zbioru przykładów uczących \mathbf{x} natomiast V_f jest podprzestrzenią w przestrzeni cech, która te elementy obejmuje.

W klasyfikatorze najbliższego sąsiada przykład testujący jest pozytywnie sklasyfikowany, jeżeli jego lokalna podprzestrzeń jest większa lub równa podprzestrzeni lokalnej pierwszego najbliższego sąsiada na zbiorze trenującym. Można więc zapisać, że:

$$f_{NN^{tr}}(x) = I\left(\frac{V(\|\mathbf{x} - NN^{tr}(\mathbf{x})\|)}{V(\|NN^{tr}(\mathbf{x}) - NN^{tr}(NN^{tr}(\mathbf{z}))\|)} \leq 1\right) \quad (4.33)$$

Porównując do metody k -najbliższych sąsiadów przedstawionej w sekcji 4.2.4 dla omawianej metody $k = 1$. Konstrukcja tego klasyfikatora powoduje, że jeżeli dane są dwa blisko siebie położone punkty należące do rozpatrywanej klasy to pole wokół nich jest zwężone, jeżeli jednak punktów jest mało, to klasyfikator w ich otoczeniu akceptuje o wiele większą podprzestrzeń w przestrzeni cech.

Bibliografia Więcej szczegółów na temat jednoklasowego klasyfikatora jednoklasowego można znaleźć w: [4.13].

Oprogramowanie Dodatek DD Tools do pakietu PRTools [4.14].

4.3.3. Jednoklasowy klasyfikator SVM

Opis tego algorytmu jest podobny do opisu wieloklasowego klasyfikatora SVM. Istotna zmiana polega na tym, że klasyfikator ten prowadzi procedurę optymalizacji na zbiorze przykładów nie należących do rozpatrywanej klasy. Użytkownik dodatkowo podaje jedynie liczbę określającą frakcję przykładów należących do klasy, które mogą zostać błędnie sklasyfikowane. Na podstawie tego drugiego parametru ustalana jest w końcowej fazie próg definiujący klasę. Metoda ta jest bardzo czuła na odpowiednie skalowanie danych w przestrzeni cech, w szczególności przy zastosowaniu radialnej funkcji gaussowskiej jako funkcji jądra.

Bibliografia Więcej szczegółów na temat jednoklasowych klasyfikatorów svm można znaleźć m.in. w: [4.8].

Oprogramowanie Dodatek DD Tools do pakietu PRTools [4.14].

Bibliografia

- [4.1] Matlab manual. <http://www.mathworks.com>.
- [4.2] Parzen probabilistic neural networks. <http://www.mathworks.com/matlabcentral/files/11880>.
- [4.3] Simple svm toolbox. <http://asi.insa-rouen.fr/enseignants/gloosi/simpleSVM.html>.

- [4.4] Svm and kernel methods matlab toolbox. <http://asi.insa-rouen.fr/enseignants/arakotom/toolbox/index.html>.
- [4.5] Abe S. *Series: Advances in pattern recognition, Support vector machines for pattern recognition*. Springer, London, 2005.
- [4.6] Berry M., Browne M. *Lecture notes in data mining*. World Scientific, USA, 2006.
- [4.7] Cholewa W. *Metoda diagnozowania maszyn z zastosowaniem zbiorów rozmytych*. Zeszyty naukowe-Mechanika z.79, Politechnika Śląska, Gliwice, 1983.
- [4.8] Duda R.O., Hart P.E., Stork D.G. *Pattern recognition*. Wiley Interscience, New York, 2001.
- [4.9] Franc V., Hlavac V. *Statistical Pattern Recognition Toolbox for Matlab*. <ftp://cmp.felk.cvut.cz/pub/cmp/articles/Franc-TR-2004-08.pdf>, Praha, 2004.
- [4.10] Nabney I. Netlab software. <http://www.ncrg.aston.ac.uk/netlab/>.
- [4.11] Nabney I. *Netlab. Algorithms for pattern recognition*. Springer Verlag, 2006.
- [4.12] Control reviews in Annual, redaktor. *Model based fault detection and diagnosis -status and applications*, wolumen 29 pp.71-85. Elsevier, 2005.
- [4.13] Tax D.M.J. *One-class classification*. Praca doktorska, Delft Univesity of Technology, 2001.
- [4.14] Heijden van der F., Duin R.P.W., Ridder de D., Tax D.M.J. Prt tools. <http://prtools.org/>.
- [4.15] Heijden van der F., Duin R.P.W., Ridder de D., Tax D.M.J. *Classification, Parameter Estimation and State Estimation*. John Wiley and Sons Ltd, England, 2006.
- [4.16] Vapnik V.N. *Statistical learning theory*. Wiley and Sons, New York, 1998.

Rozdział 5

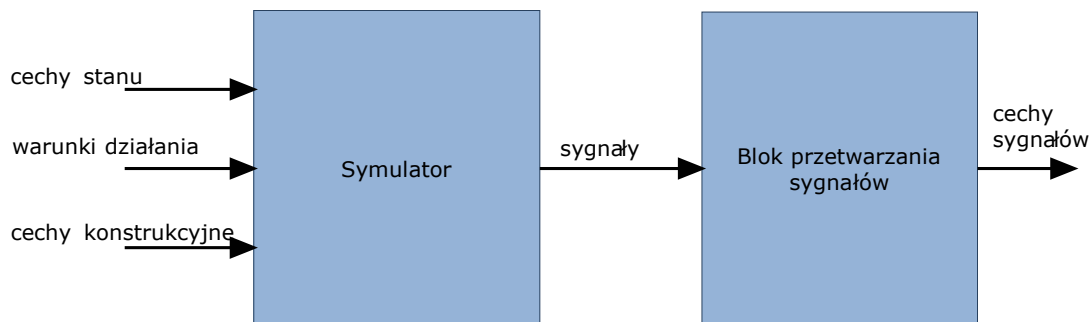
Symulator wirnika wielopodporowego

Tomasz ROGALA

W celu oceny zmian stanu lub identyfikacji stanu przeprowadza się eksperymenty diagnostyczne. Mogą to być zarówno eksperymenty czynne jak i bierne w zależności od dostępności danego środka technicznego. Obserwacje działania środka technicznego są źródłem informacji o jego stanie lub zmianach tego stanu. Niestety, eksperymenty bierne powalają czasami jedynie na obserwacje stanu w otoczeniu jednego punktu pracy. Brak możliwości poznania relacji diagnostycznych w szczególności w przypadku zmian stanu wymusza prowadzenie eksperymentów czynnych. Z drugiej strony one same nie zawsze są możliwe ze względu na ekonomikę działania przedsiębiorstwa w którym znajduje się obiekt techniczny jak np. w energetyce. Jedyną możliwością przeprowadzenia eksperymentu czynnego jest w tym przypadku prowadzenie badań na maszynie należącej do tej samej klasy maszyn co obiekt obserwowany, lub też prowadzenie symulacji w oparciu o model numeryczny w odpowiednim stopniu aproksymujący rzeczywisty obiekt techniczny lub klasę obiektów podobnych. Obecnie takie metody są coraz powszechniej stosowane do czego przyczynił się szybki rozwój technik komputerowych. W diagnostyce maszyn i procesów symulacja stała się podstawą stosunkowo nowej gałęzi nazywanej diagnostyką wspartą modelowo, w tym również jedną z podstawowych metod pozyskiwania wiedzy na potrzeby identyfikacji relacji diagnostycznych stan-symptom. Opracowane na bazie symulacji diagnostyczne modele obiektów nie tylko są wykorzystywane w procesie diagnozowania na podstawie oceny residuów czy identyfikacji zmian parametrów obiektu, ale z powodzeniem mogą być stosowane w procesie trenowania oraz testowania odwrotnych modeli diagnostycznych. W tym właśnie celu został opracowany symulator przedstawiony w niniejszym rozdziale.

Poznanie relacji diagnostycznych na bazie symulacji wymaga odpowiedniego zaplanowania prowadzonych badań symulacyjnych. Przede wszystkim badania takie wymagają określenia punktów pracy obiektu w dyskretniej przestrzeni stanu, jego warunków działania, jak również cech konstrukcyjnych symulowanego obiektu technicznego. Zebrane w ten sposób zbiory przebiegów sygnałów dowolnych wielkości fizycznych elektrycznych, termicznych czy mechanicznych poddawane są dalszym zabiegom począwszy od wstępnego przetwarzania sygnałów, poprzez ich przetwarzanie właściwe do uzyskania cech sygnałów, a ostatecznie do uzyskania użytecznych sygnałów diagnostycznych odpowiadającym omawianym symptomom. Na rysunku 5.1 przedstawiono wejścia i wyjścia bloku symulatora oraz bloku przetwarzania sygnałów.

Największą klasę maszyn pod względem liczby stanowią maszyny zaliczane do klasy maszyn wirnikowych. Uzasadnieniem dla stosowania systemów diagnostycznych dla tej grupy maszyn są przede wszystkim maszyny krytyczne, stanowiące z punktu widzenia bezpieczeństwa lub ekonomii danego przedsiębiorstwa lub nawet gałęzi gospodarki krajowej element strategiczny. Przykładem takich obiektów są turbospoły energetyczne, które są przedmiotem wielu badań związanych z szeroko pojmowaną diagnostyką. Są to obiekty bardzo złożone w których występuje bardzo wiele zjawisk zachodzących procesów [5.7] [5.6]. Układ ten jest narażony na działanie wielu sił, które oddziałują



Rys. 5.1: Schemat pozyskiwania relacji diagnostycznych na bazie badań symulacyjnych

na linię wirników i łożysk wpływają zasadniczo na stan obiektu [5.7]. Można tutaj wyróżnić wiele wymuszeń takich jak na przykład siły aerodynamiczne, mechaniczne, magnetyczne i elektryczne [5.7]. Siły te powodują powstawanie wielu niesprawności takich jak na przykład niestabilna praca łożysk hydrodynamicznych wywołana rozoosiowaniem.

Dla potrzeb badań związanych z identyfikacją oraz testowaniem modelu odwrotnego opracowano symulator wirnika wielopodporowego z łożyskami hydrodynamicznymi, będącego reprezentantem szerokiej klasy powszechnie spotykanych maszyn wirnikowych. W opracowanym modelu można wyodrębnić następujące podukłady:

- podukład opisujący linię wirników,
- podukład opisujący linię łożysk,
- podukład związany z dynamicznymi własnościami podpór.

5.1. Prosty model wirnika z łożyskiem hydrodynamicznym

Model prostego wirnika z łożyskiem hydrodynamicznym jest rozbudowanym podukładem wyodrębnionym z systemu wirnika wielopodporowego. Model ten został wykonany w celu weryfikacji poprawności działania części składowych modelu, w którym model łożyska hydrodynamicznego stanowi jeden z najbardziej złożonych modeli cząstkowych. W niniejszej sekcji rozpatrywany jest model płaski w którym przedstawiono działanie pojedynczej podpory z łożyskiem hydrodynamicznym.

5.1.1. Siły hydrodynamiczne

Niniejszy model został opracowany na podstawie dostępnej wiedzy analitycznej na temat charakteru rozkładu ciśnień hydrodynamicznych i nieliniowych reakcji hydrodynamicznych wywołanych w podporach łożyskowych. W modelu tym wykorzystano opis rozkładu ciśnienia filmu olejowego według rozwiązania równania Reynolds'a z rozwiązaniem szczególnym dla łożysk krótkich znanym jako rozwiązanie Ocvirk'a (przyjęcie rozwiązania Ocvirka nie pozwala na rozpatrywanie drgań poosiowych koniecznych do identyfikacji przekoszeń panwi) [5.12].

Przyjmując rozwiązanie Ocvirk'a równanie Reynolds'a dla $L < R$ (gdzie L jest długością łożyska, a R jego promieniem) przyjmuje postać [5.12]:

$$\frac{\partial}{\partial z} \left(h^3 \frac{\partial p}{\partial z} \right) = 6\nu\Omega \left(\frac{\partial h}{\partial x} \right) \quad (5.1)$$

gdzie, p jest rozkładem ciśnienia na powierzchni panwi, natomiast h jest funkcją grubości filmu olejowego. Przyjmując, że pochodna cząstkowa $\frac{\partial h}{\partial x}$ i h są niezależne od zmian rozkładu ciśnienia w kierunku poosiowym (np. łożyska bardzo krótkie w których przekoszenia panwi nie powodują znacznych zmian rozkładu ciśnienia wzdłuż panwi) można, po dwukrotnym całkowaniu, zapisać:

$$p = \frac{3\nu\Omega}{h^3} \left(\frac{\partial h}{\partial x} \right) z^2 + \frac{C1}{h^3} z + C2 \quad (5.2)$$

przy uwzględnieniu warunków brzegowych $p = 0$ dla $z = \pm L/2$ wzór na ciśnienie hydrodynamiczne przyjmuje postać:

$$p = \frac{3\nu\Omega}{h^3} \left(\frac{l^2}{4} - z^2 \right) \left(\frac{\partial h}{\partial x} \right) \quad (5.3)$$

ponieważ warstwa oleju w dowolnym przekroju ma grubość,

$$h = c(1 + \varepsilon \cos(\theta)) \quad (5.4)$$

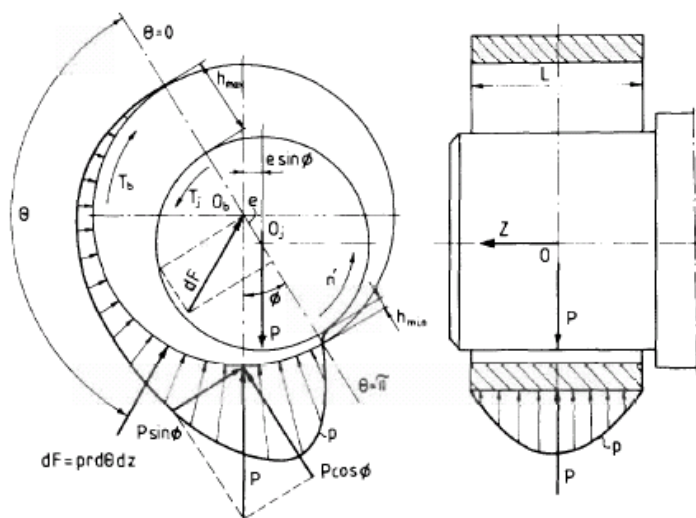
a jej pochodna dla dowolnego przekroju warstwy oleju jest równa

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial r\theta} = \frac{1}{r} \frac{\partial h}{\partial \theta} \quad (5.5)$$

można zapisać, że [5.12]

$$p = \frac{3\nu\Omega}{h^3} \left(\frac{l^2}{4} - z^2 \right) \frac{3\varepsilon \sin(\theta)}{(1 + \varepsilon \cos(\theta))^3} \quad (5.6)$$

Równanie to pokazuje rozkład ciśnienia zgodny z rysunkiem 5.2.



Rys. 5.2: Rozkład ciśnień i odpowiadający układ sił występujących w łożysku hydrodynamicznym [5.12]

Na rysunku 5.2 przedstawione zostały zaznaczone siły wynikające z ciśnienia hydrodynamicznego. Ciśnienia panujące w łożysku mają kierunek normalny do filmu olejowego i rozkładają się wzdłuż długości łożyska w ten sposób, że wszystkie elementarne siły $dF = prd\theta dz$ można zredukować do wypadkowej siły promieniowej ciśnienia P przechodzącej przez środek łożyska. Podobnie można zredukować wszystkie siły obciążenia, które wywierają ciśnienie, do wypadkowej siły przechodzącej przez środek czopa. Siły te muszą się równoważyć, mieć ten sam kierunek, przeciwne zwrotny i być do siebie równoległe. Czop wywiera moment ścinający T_j oddziaływujący na film olejowy zgodnie z kierunkiem obrotu czopa, tym samym łożysko wywiera moment oporu T_b przeciwdziałając temu momentowi. Momenty te nie są jednakowe i spełniają następujące równanie momentów względem środka czopa.

$$T_j = T_b + P \sin(\theta); \quad (5.7)$$

Siły hydrodynamiczne w rzeczywistości są siłami zależnymi od położenia czopa w łożysku hydrodynamicznym jak i zależnymi od szybkości tych zmian. We współrzędnych związanych z przekrojem dla minimalnej grubości warstwy filmu olejowego, zależność można zdefiniować w następujący sposób [5.4]:

$$P \sin(\theta) = 2 \int_0^l \int_0^{2\pi} (prd(\theta) dz \sin(\theta)) = \frac{\mu \Omega R L^3}{2h^2} (1 - 2\dot{\gamma}) \frac{2\varepsilon^2}{(1 - \varepsilon^2)^2} + \pi \dot{\varepsilon} \frac{1 + 2\varepsilon^2}{(1 - \varepsilon^2)^{\frac{5}{2}}} \quad (5.8)$$

$$P \cos(\theta) = 2 \int_0^l \int_0^{2\pi} (prd(\theta) dz \cos(\theta)) = \frac{\mu \Omega R L^3}{2h^2} - \frac{\pi i}{2} (1 - 2\dot{\gamma}) \frac{\varepsilon}{(1 - \varepsilon^2)^{\frac{3}{2}}} - \varepsilon \frac{4\varepsilon}{(1 - \varepsilon^2)^2} \quad (5.9)$$

gdzie pomiędzy składowymi w układzie Oxy , a układem związanym z przekrojem dla minimalnej grubości warstwy olejowej zachodzą następujące związki [5.4]:

$$F_x = P \cos(\theta) - P \sin(\theta) \quad (5.10)$$

$$F_y = P \sin(\theta) - P \cos(\theta) \quad (5.11)$$

Kontynuując rozważania z równania 5.6 i pomijając tym samym wpływ zmian prędkości na dynamikę sił hydrodynamicznych równania (5.8) i (5.9) przyjmują postać:

$$P \cos(\theta) = 2 \int_0^l \int_0^{2\pi} prd(\theta) dz \cos(\theta) = \frac{\mu \Omega L^3}{c^2} \frac{\varepsilon^2}{(1 - \varepsilon^2)^2} \quad (5.12)$$

$$P \sin(\theta) = 2 \int_0^l \int_0^{2\pi} prd(\theta) dz \sin(\theta) = \frac{\mu \Omega L^3}{c^2} \frac{\pi \varepsilon}{4(1 - \varepsilon^2)^{\frac{3}{2}}} \quad (5.13)$$

Otrzymanie powyższych równań pozwala na obliczenie wartości współczynników sztywności i tłumienia stosowanych w klasycznym ujęciu modelowania własności reologicznych filmu olejowego. W tym celu należy rozwiązać poszczególne pochodne cząstkowe zgodnie z poniższymi wzorami [5.4]:

$$k_{\alpha,\beta} = \left(\frac{\partial P_\alpha}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial w_\beta} \right) + \left(\frac{\partial P_\alpha}{\partial \gamma} \frac{\partial \gamma}{\partial w_\beta} \right) \quad (5.14)$$

$$d_{\alpha,\beta} = \left(\frac{\partial P_\alpha}{\partial \dot{\varepsilon}} \frac{\partial \dot{\varepsilon}}{\partial w_\beta} \right) + \left(\frac{\partial P_\alpha}{\partial \dot{\gamma}} \frac{\partial \dot{\gamma}}{\partial w_\beta} \right) \quad (5.15)$$

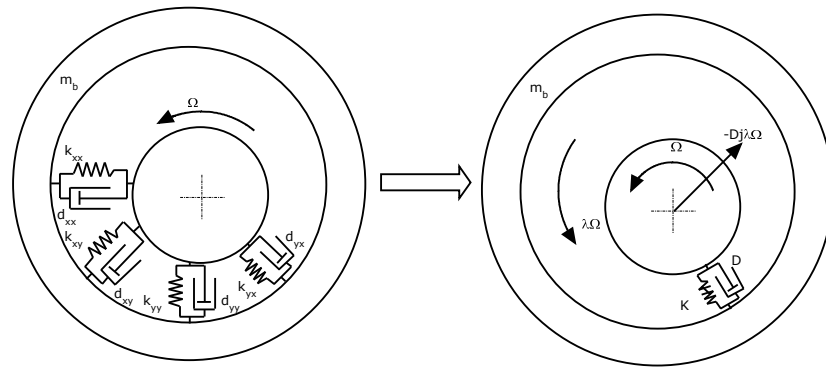
dla $\alpha, \beta = x, y$

Ujęcie to posiada jednak kilka wad i stąd też podjęto próbę identyfikacji współczynników modelu filmu olejowego w oparciu o mniej powszechnie stosowany model; nazywany modelem *BM* [5.10] od nazwisk badaczy. Na rysunku 5.3 przedstawiono porównanie sposobu modelowania filmu olejowego z wykorzystaniem modelu *BM* [5.10] w porównaniu do tradycyjnego ujęcia.

W modelu tym, w odróżnieniu od klasycznego ujęcia, gdzie własności filmu olejowego są reprezentowane za pomocą czterech stacjonarnych układów Kelvina-Voigta ułożonych w określony sposób, układ Kelvina-Voigta jest układem dynamicznym związanym z układem biegunowym w którym oś kierunkowa jest związana z obracającą się płaszczyzną przecinającą łożysko w punkcie o najmniejszej grubości warstwy olejowej.

W modelu *BM* [5.10] podstawową różnicą jest to, że zgodnie z interpretacją fizyczną siła tłumienia filmu olejowego nie tylko oddziałuje w kierunku promieniowym, ale również w kierunku stycznym, gdzie siła ta zależy od prędkości przepływu czynnika roboczego. Dodatkowymi zaletami stosowania modelu *BM* są [5.3]:

- współczynniki opisujące własności reologii cieczy w układzie *BM* są niezależne od układu współrzędnych,



Rys. 5.3: Prosty model wirnika

- w łatwy sposób można określić granicę stabilności dla rozpatrywanego modelu BM,
- w empiryczny sposób łatwo jest zidentyfikować wartości współczynników sztywności i tłumienia dla modelu BM.

Wadą przedstawionego podejścia jest konieczność prowadzenia złożonych obliczeń w przypadku braku danych empirycznych pozwalających na bezpośrednią identyfikację współczynników modelu łożyska hydrodynamicznego.

Model siły hydrodynamicznej stosowanej w modelu *BM* (w przypadku izotropowych własności) ma następującą postać [5.10] [5.11] [5.9][5.2] [5.2] [5.8]:

$$F_h = Kr + D(\dot{r} - j\lambda\Omega r) + M_f(\ddot{r} - 2j\lambda\Omega\dot{r} - \lambda^2\Omega^2\ddot{r}) \quad (5.16)$$

gdzie $r = x + jy$, K jest sztywnością dynamiczną, D jest tłumieniem dynamicznym, M_f określa inercję filmu olejowego, Ω jest prędkością kątową wirnika, λ określa średnią prędkość przepływu filmu olejowego, która jest zależna od położenia czopa:

$$\lambda = \lambda(\varepsilon) \quad (5.17)$$

gdzie $\varepsilon = \sqrt{x^2 + y^2}/c$ jest względnym współczynnikiem mimośrodowości położenia czopa względem środka łożyska. Wartość tego współczynnika jest w przybliżeniu odwrotnie proporcjonalna do liczby Sommerfelda:

$$S = \nu LR \frac{R^2}{c^2} \frac{\Omega}{P} \approx \frac{1}{\varepsilon} \quad (5.18)$$

gdzie ν jest lepkością dynamiczną filmu olejowego, L długością łożyska, R promieniem łożyska, c luzem promieniowym, P siłą promieniową oddziaływującą na czop wirnika. W celu uproszczenia obliczeń przyjęto, że wartość średniej prędkości filmu olejowego będzie stała i równa $\lambda\Omega$ oraz, że wpływ masy dynamicznej filmu olejowego jest na tyle niewielki w porównaniu do sił sztywności i tłumienia, że zostanie on w dalszych rozważaniach pominięty. Należy mieć na uwadze fakt, że przyjęty model jest odpowiedni w przypadku małej oraz średniej mimośrodowości względnej czopa w panwi. Rozważania, w których konieczne jest uwzględnienie większych wychyleń promieniowych czopa w panwi wymagają rozpatrzenia wpływu tych zmian na średnią obwodową prędkość filmu olejowego λ , jak również na uwzględnienie tarcia w przypadku gdy $\lambda(c) = 0$.

W przypadku takiego modelu współczynniki K oraz D , które są zależne od położenia oraz prędkości czopa są najczęściej określane drogą empiryczną [5.9] [5.8]. W niniejszym przykładzie konieczne jest jednak uzyskanie tych współczynników na drodze analitycznej. W tym celu możliwe są dwa podejścia:

- poszukiwanie rozwiązania dla współczynników filmu olejowego na drodze klasycznej, a następnie ich transformacja w celu określenia współczynników dla modelu *BM* [5.3],
- poszukiwanie bezpośredniego rozwiązania analitycznego,

Dla celów dalszych rozważań drugą metodę wybrano jako najbardziej odpowiednią. Metoda ta polega na znalezieniu funkcji opisującej siły hydrodynamiczne w zależności od mimośrodowości względnej oraz prędkości jej zmian. W ten sposób rozwiązanie równania 5.1 dla łożysk krótkich można uzyskać poprzez określenie ciśnienia za pomocą nieskończonego szeregu w postaci [5.10]:

$$p = p_a \frac{L - z}{L} + \frac{p_b z}{L} + \frac{\nu \Omega}{c^2} \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{z^k}{R} \frac{L^{n-k}}{R} \quad (5.19)$$

która dla łożysk krótkich przyjmuje postać:

$$p = p_a \frac{L - z}{L} + \frac{p_b z}{L} + \frac{\nu \Omega}{c^2} a_0 \quad (5.20)$$

gdzie współczynniki a_{kn} określa się za pomocą poniższego wzoru:

$$a_{kn} = \left[\left(\frac{1}{2} - \frac{\dot{\varepsilon}}{\Omega} \right) \sin(\theta) - 6 \frac{d\varepsilon}{d\Omega t} \cos(\theta) \right] \frac{c^3}{h^3} \quad (5.21)$$

gdzie pierwsze dwa człony równania określają składowe ciśnienia w kierunku promieniowym oraz stycznym zgodnie z rysunkiem 5.2. Wartość ciśnienia hydrostatycznego określają z kolei pierwsze dwa człony równania (5.19).

Siły hydrodynamiczne można określić na podstawie składowych naprężeń normalnych i stycznych do powierzchni czopa wg. następujących wzorów:

$$F_r = \int_0^L \int_0^{2\pi} \sigma \cos(\theta) - \tau \sin(\theta) R \theta_{czopa} \quad (5.22)$$

$$F_t = \int_0^L \int_0^{2\pi} \tau \cos(\theta) + \sigma \sin(\theta) R \theta_{czopa} \quad (5.23)$$

gdzie r, z, θ_{czopa} , stanowią układ cylindryczny związany ze środkiem czopa. W celu określenia sił w oparciu o rozkład ciśnienia na powierzchni łożyska, konieczna jest transformacja układu początku układu cylindrycznego ze środka czopa do środka łożyska za pomocą następującego równania [5.10]:

$$R dz d\theta_{czopa} = R dz \left[1 + \varepsilon \frac{c}{R} \cos(\theta) + O\left(\frac{c^2}{R^2}\right) \right] d\theta \quad (5.24)$$

Korzystając z poniższych zależności dotyczących naprężeń normalnych i stycznych

$$\sigma = \sigma_r + 2\varepsilon \tau_{r\theta} \sin(\theta) + O\left(\frac{c^2}{R^2}\right) \quad (5.25)$$

$$\tau = \tau_{r\theta} + \varepsilon (\sigma_r + \sigma_r) \frac{c}{R} \sin(\tau) + O\left(\frac{c^2}{R^2}\right) \quad (5.26)$$

gdzie $\tau_{r\theta}, \sigma_r, \sigma_\theta$ są odpowiednimi składnikami tensora naprężeń we współrzędnych cylindrycznych, gdzie:

$$\sigma_r = -p + O\left(\frac{c^2}{R^2}\right) \quad (5.27)$$

$$\sigma_\theta = -p + O\left(\frac{c^2}{R^2}\right) \quad (5.28)$$

$$\tau_{r\theta} = \nu \left(\frac{1}{r} \frac{\partial V_r}{\partial \theta} + \frac{\partial V_\theta}{\partial r} - \frac{V_\theta}{r} \right) \left(\frac{c^2}{R^2} \right) \quad (5.29)$$

Ponieważ składowe prędkości filmu w kierunku r i θ olejowego są równe:

$$V_r = (R + c - r)^2 \left[-\frac{\varepsilon c \Omega \sin(\theta)}{2h^2} + \left(\frac{1}{R^2} \frac{\partial^2 p}{\partial \theta^2} + \frac{\partial^2 p}{\partial z^2} \right) \frac{2R - 2r - 3h}{12\nu} - \frac{\varepsilon c \sin(\theta)}{4\nu R^2} \theta \frac{\partial p}{\partial \theta} \right] \quad (5.30)$$

$$V_\theta = (R + c - r) \left[\frac{\Omega R}{h} + \left(\frac{\varepsilon c \cos(\theta) - r + R}{2\nu R} \right) \frac{\partial p}{\partial \theta} \right] \quad (5.31)$$

Stąd też podstawiając równania (5.30) i (5.31) do równania (5.29), oraz równania (5.28) i (5.27) do równań (5.26) i (5.25) do równań (5.22) i (5.23) otrzymujemy następujące równania całkowe.

$$F_t = \int_0^L dz \int_0^{2\pi} R \left(1 + \frac{c\varepsilon}{R} \cos(\theta) \right) \left\{ \left[-\frac{\nu \Omega R}{c} (1 + \varepsilon \cos(\theta)) - \frac{c}{2R} \frac{\partial P_{sin}}{\partial \theta} (1 - 2\varepsilon \cos(\theta)) \right] \cos(\theta) \right\} d\theta \quad (5.32)$$

$$F_r = \int_0^L dz \int_0^{2\pi} R \left(1 + \frac{c\varepsilon}{R} \cos(\theta) \right) \left\{ \frac{2\varepsilon c}{R} \sin(\theta) \nu \left(\frac{1}{r} \frac{\partial V_r}{\partial \theta} - \frac{V_\theta}{r} \right) \cos(\theta) - \left(\frac{1}{r} \frac{\partial V_r}{\partial \theta} - \frac{V_\theta}{r} \right) \sin(\theta) \right\} \theta \quad (5.33)$$

W równaniach tych pominięto wpływ członów niestacjonarnych, ponieważ w celu wyznaczenia współczynników D i K wystarczy rozwiązać fragment równania dla siły dynamicznej który odpowiada składowym stacjonarnym:

$$(K - jD\lambda\Omega)\varepsilon = \frac{F_t + F_r}{c} \quad (5.34)$$

stosując podstawienie Sommerfelda lub metodę Okazakiego [5.5] można uzyskać rozwiązanie całek (5.32) i (5.33). W całkach tych pominięto mniej znaczące wyrazy w których występują współczynniki c/R . Są one stosunkowo niewielkie w porównaniu do współczynników L/D . Rozwiązanie to jest zgodne z rozwiązaniem przedstawionym w pracy [5.1]:

$$D = \frac{\nu d L^3}{c^3} \frac{D_1 \varepsilon}{(1 - \varepsilon^2)^{\frac{5}{2}}} \quad (5.35)$$

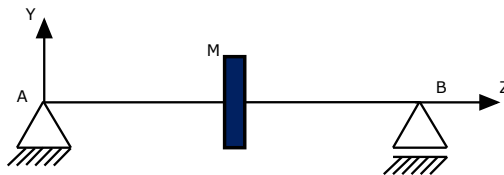
$$K = \frac{\nu \Omega d L^3}{c^3} \frac{K_1 \varepsilon}{(1 - \varepsilon^2)^{\frac{3}{2}}} \quad (5.36)$$

Powyższe równania określają część hydrodynamiczną, natomiast część hydrostatyczna jako mniej istotna dla dalszych rozważań została pominięta. W równaniach tych parametry K_1 oraz parametry D_1 stanowią stałe sztywności dynamicznej i tłumienia dynamicznego i są wykorzystywane w celu dalszego strojenia modelu. Powodem jest fakt, że funkcja $\lambda(\varepsilon)$ została przyjęta w dalszych rozważaniach jako stała.

Opracowane współczynniki sztywności i tłumienia dynamicznego zostały wykorzystane do budowy prostego modelu podpory łożyskowej.

5.1.2. Model

Model wirnika z łożyskiem hydrodynamicznym został opracowany na podstawie uproszczonego modelu wirnika Jefcott'a często nazywanego modelem de Laval'a od nazwisk badaczy. Postać przyjętego modelu przedstawia rysunek 5.4.

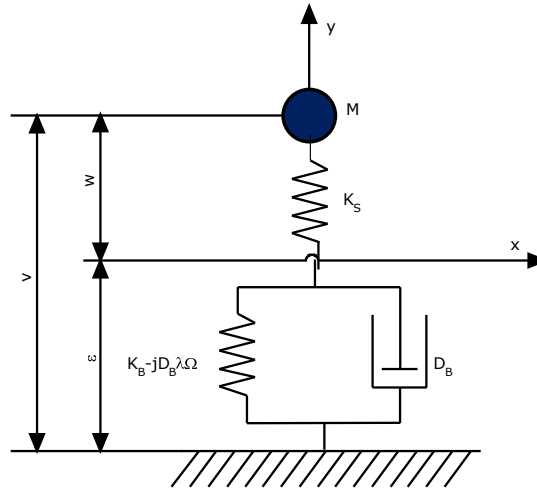


Rys. 5.4: Prosty model wirnika [5.1]

W modelu tym przyjęto, że wał jest bezmasowym elementem o sprężystości giętnej K_s , gdzie w połowie jego rozpiętości znajduje się tarcza o masie M . Ze względu na przeznaczenie modelu układ

podpory z łożyskiem hydrodynamicznym został zastosowany w punkcie B . Punkt A jest podporą bardzo sztywną umożliwiającą jedynie obrót wału wokół osi Z i może być interpretowany jako podpora z łożyskiem tocznym.

Na rysunku 5.5 przedstawiono model podpory B wirnika przedstawionego na rysunku 5.4 w którym model łożyska hydrodynamicznego został określony za pomocą modelu BM.



Rys. 5.5: Rozkład sił w łożysku hydrodynamicznym [5.12]

Przedstawiony model jest modelem cząstkowym bardziej złożonego wirnika wielopodporowego. Celem jego wyodrębnienia jest weryfikacja poprawności działania opracowanego modelu w oparciu o wiedzę dotyczącą typowych niesprawności związanych z łożyskami hydrodynamicznymi, które w przypadku rozpatrywania wirnika wielopodporowego, ze względu na liczne sprzężenia i związki nieliniowe pomiędzy podporami mogłyby być trudne do zidentyfikowania i uniemożliwiać weryfikację poprawności zbudowanego symulatora.

W proponowanym modelu można wyróżnić następujące podukłady:

- układ związany z własnościami mechanicznymi łożyska hydrodynamicznego,
- oraz układ związany z własnościami mechanicznymi wału oraz obciążenia.

Dla rozpatrywanego modelu BM [5.1] można zapisać następujące równania różniczkowe:

$$M\ddot{v} + K_s(\varepsilon, \Omega)(v - u) = Mr\Omega^2 e^{j(\Omega t + \alpha)} \quad (5.37)$$

$$K_s \varepsilon(w) = D_b(\varepsilon) \dot{u} + (K(\varepsilon, \Omega) - j\lambda(\varepsilon)\Omega D(\varepsilon))u \quad (5.38)$$

gdzie: zmienne v, w, u są zmiennymi zespolonymi, $v = w + u$, K_s jest sztywnością giętną wału, M jest masą tarczy osadzonej na bezmasowym wale, D_b jest tłumieniem wiskotycznym, K_b jest współczynnikiem sztywności dynamicznej filmu olejowego, r określa bezwzględną mimośrodowość osadzonej tarczy, Ω jest prędkością kątową niewyrównoważonej tarczy, a α jest kątem wyrażonym w radianach określającym kierunek wektora i zwrot wektora położenia środka tarczy względem linii łożysk.

5.1.3. Weryfikacja poprawności modelu

Opracowany prosty model wirnika z łożyskiem hydrodynamicznym został poddany weryfikacji. Jej podstawą była ocena zgodności z modelem analitycznym. W tym celu zweryfikowano charakterystykę czasowo-częstotliwościową modelu, sprawdzono jej zgodność z modelem pod względem częstotliwości rezonansowej oraz granicy utraty stabilności. Zbadano również, czy model ten pozwala na obserwację sił hydrodynamicznych w zakresie utraty stabilności i przede wszystkim czy możliwa jest obserwacja

modelu w zakresie małych drgań olejowych czyli wiru olejowego. Ze względu na brak uwzględnienia warunków przycierania, oraz zmienności średniej prędkości obwodowej filmu olejowego w przypadku dużych mimośrodowości $\varepsilon \cong 1$ obserwacja sił hydrodynamicznych w zakresie dużych drgań olejowych czyli tzw. bicia olejowego nie są możliwe.

Dla potrzeb weryfikacji modelu przyjęto następujące parametry wejściowe:

- masa tarczy $M = 11[kg]$,
- obciążenie statyczne $X = 0[N]$, $Y = 0[N]$,
- mimośrodek bezwzględna tarczy $a = 26,67[\mu m]$,
- średnica czopa $19,05[cm]$,
- sztywność właściwa wału $K_s = 400000[N/m]$,
- luz promieniowy $c = 38,1[\mu m]$,
- współczynnik lepkości dynamicznej $0.655[Ns/m^2]$,
- długość łożyska $= 3,81[cm]$,
- współczynnik średniej obwodowej prędkości filmu olejowego $\lambda = 0.48(1 - \frac{\varepsilon}{4})$,
- współczynnik korekcyjny $K_1 = 0.0001$,
- współczynnik korekcyjny $D_1 = 0.01$.

Granica stabilności dla łożysk hydrodynamicznych została określana na podstawie zależności [5.10]:

$$\lambda\Omega = \sqrt{\frac{K_z}{M}} \quad (5.39)$$

gdzie K_z jest zastępczą sztywnością dynamiczną, która dla układu przedstawionego na rysunku 5.5 wynosi:

$$K_z = \frac{1}{\frac{1}{K_s} + \frac{1}{K_b}} \quad (5.40)$$

stąd granica stabilności jest równa

$$\lambda\Omega = \sqrt{\frac{1}{M} \frac{1}{\frac{1}{K_s} + \frac{1}{K_b}}} \quad (5.41)$$

Stąd przybliżone wartości dla rozpatrywanego układu wynoszą odpowiednio:

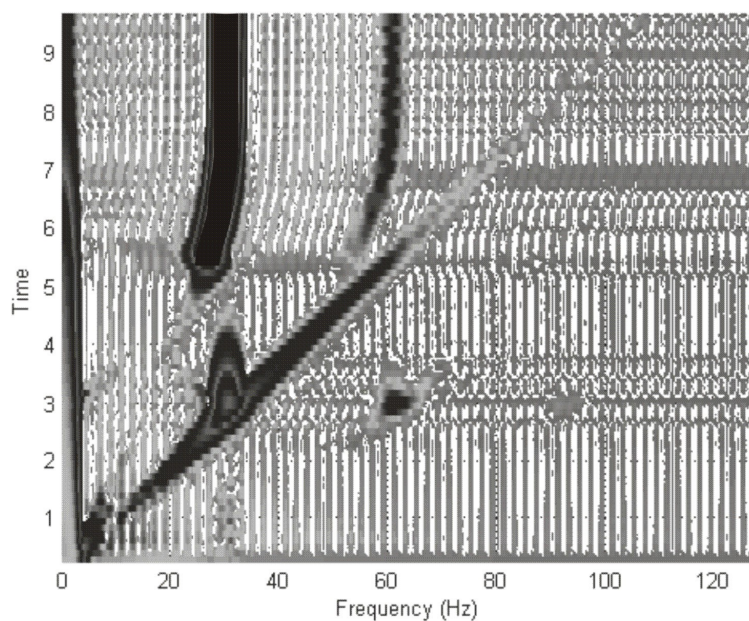
$$\omega_{rez} = 30,36[Hz] \quad (5.42)$$

$$\Omega_{stab} = 63,25[Hz] \quad (5.43)$$

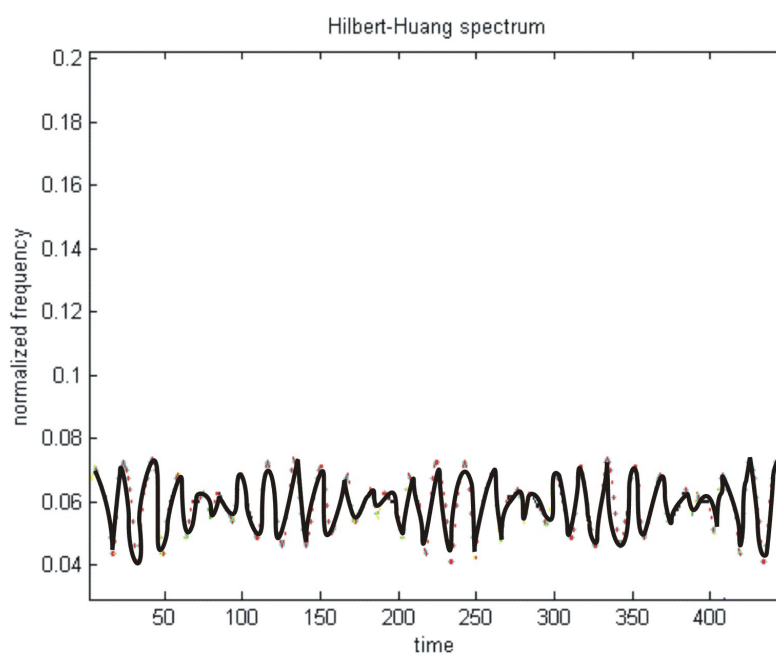
Powyższe wyniki są zgodne z wynikami przedstawionymi w formie wykresu 5.6 czasowo-częstotliwościowego STFT drgań czopa w łożysku hydrodynamicznym dla warunków rozruchu. Na rysunku 5.6 można zauważyć składową 1X oraz składową podharmoniczną, która pojawia się w po przekroczeniu granicy stabilności, a następnie krótki odcinek wiru olejowego oraz asymptotycznie zniekształconą składową podharmoniczną w stosunku do częstotliwości rezonansowej co odpowiada powstaniu bicia olejowego. Utrata stabilności oraz małe drgania olejowe powodują zmianę częstotliwości chwilowej co zostało przedstawione na widmie Hilberta na rysunku 5.7.

5.2. Model wirnika wielopodporowego

Opracowywany model wirnika wielopodporowego składa się z dwóch giętkich wałów połączonych ze sobą za pomocą sprzęgła w połowie długości całego zespołu wirnika. Układ ten jest łożyskowany w łożyskach hydrodynamicznych. Niniejszy podrozdział przedstawia budowę takiego modelu ze szczególnym opisem sposobu modelowania linii wału oraz budowy podpór. Ze względu na swoją złożoność, w modelu tym przyjęto szereg założeń upraszczających, które przedstawiono poniżej.



Rys. 5.6: Widmo STFT dla drgań czopa w łożysku hydrodynamicznym



Rys. 5.7: Widmo Hilberta w zakresie małych drgań olejowych

5.2.1. Założenia

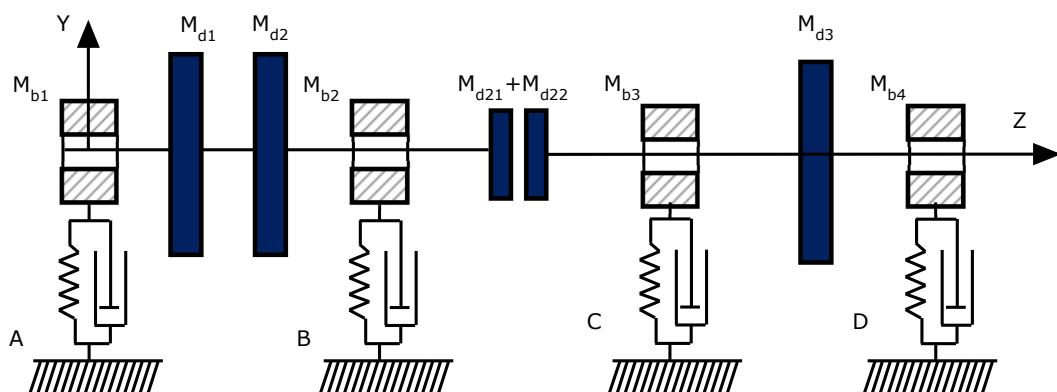
Dla uproszczenia modelu przyjęto następujące założenia:

- badania są prowadzone dla określonych warunków działania związanego ze stałą prędkością obrotową wirnika,
- wpływ zjawisk termicznych na właściwości cieczy został pominięty (przyjęto izotermiczne warunki pracy),

- wektor prędkości jest zawsze równoległy do geodezyjnej linii wirnika (oś panwi jest równoległa do osi czopa), co oznacza, że pomijane są momenty skrętne, a film olejowy nie powoduje powstawania sił poosiowych,
- oddziaływania filmu olejowego silnie zależą od położenia środka czopa oraz od jego chwilowej prędkości liniowej,
- nie występuje zjawisko kawitacji,
- przekroje poprzeczne panwi i czopa są kołowe, czop jest objęty panwiami walcowymi na całym obwodzie,
- łożysko jest stosunkowo krótkie $L < D$,
- grubość filmu olejowego h jest mała w porównaniu z pozostałymi wymiarami łożyska,
- ciśnienie hydrodynamiczne wzdłuż grubości filmu olejowego jest stałe,
- cząsteczka cieczy bezpośrednio przylegająca do powierzchni ograniczającej ma taką samą prędkość jak ta powierzchnia,
- ciecz smarna (olej) jest cieczą newtonowską i nieściśliwą,
- przepływ przez szczelinę jest laminarny,
- pomija się przepływ oleju w kierunku osiowym,
- lepkość dynamiczna oleju jest stała,
- pominięto uwzględnienie efektu giroskopowego tarcz, ponieważ jak wynika z badań przedstawionych w [5.4], nie posiada odpowiedniego uzasadnienia dla uzyskanych wyników, które nieznacznie różnią się od wyników z uwzględnieniem tego zjawiska,
- ciężar sprzęgła jest niewielki w porównaniu do tarcz osadzonych na wałach, stąd też bezwładność sprzęgła została w dalszych rozważaniach pominięta,
- układ jest modelem o parametrach skupionych, gdzie masa całego układu (tarcz, opraw i wału) została skupiona do trzech środków ciężkości tarcz, czterech środków ciężkości opraw łożyskowych oraz czterech środków ciężkości czopów zlokalizowanych w węzłach łożyskowych.

Przedmiotem rozważań jest zatem układ wirnika o 22 stopniach swobody.

Na rysunku 5.8 przedstawiono rysunek schematyczny modelu układu wirnika w jednej płaszczyźnie. Pamiętając jednak z budowy układu przedstawionego w poprzednim podrozdziale, wiadomo, że układ ten jest układem przestrzennym.



Rys. 5.8: Schematyczny układ wirnika przedstawiony w jednej płaszczyźnie

Literami ABCD oznaczono kolejne podpory, znakami M_{d1} , M_{d2} , M_{d3} oznaczono masy tarcz osadzonych na kolejnych segmentach wału. Współczynnikami M_{b1} , M_{b2} , M_{b3} , M_{b4} oznaczono masę całkowitą kolejnych podpór wraz z łożyskami usytuowanymi wewnątrz tych podpór. Zgodnie z wcze-

śniejszymi założeniami bezwładność mas połówek sprzęgła M_{d21}, M_{d22} zostały pominięte. Ponadto uwzględniono bewładność czopów łożyskowych którym przyporządkowano następująca notację $M_{jA}, M_{jB}, M_{jC}, M_{jD}$, gdzie indeksy A,B,C,D oznaczają położenie czopa w układzie wirnika.

5.2.2. Obciążenia

Niewyrównoważone tarcze stanowią podstawowe źródło wymuszenia dynamicznego w układzie rozpatrywanego wirnika. Jeżeli m_n jest masą tarczy niewyrównoważonej, a r_n jest promieniem określającym mimośrodowość masy względem środka geometrycznego to siła niewyrównoważenia jest równa:

$$m_n r_n \Omega^2 e^{j(\Omega t + \alpha_n)} \quad (5.44)$$

Pozostałymi siłami oddziaływującymi bezpośrednio na tarcze są:

- siły ciężkości $-j m_n g$
- siły promieniowe asynchroniczne stałe co kierunku i wartości R_n ,
- reakcje ugiętego wału s_n ,

Siłami wyróżnionymi w układzie podpór łożyskowych są:

- składowa hydrodynamiczna związana ze "sztywnością" warstwy olejowej,
- składowa hydrodynamiczna cyrkulacyjna związana z przepływem środka medium roboczego w łożysku,
- obciążenia statyczne związane z siłami ciężkości $-j m_b g$,
- reakcje ugiętego wału,

Różniczkowe równania ruchu dla tarcz mają postać:

$$m_n \ddot{v}_n + s_n = m_n r_n \Omega^2 e^{i(\Omega t + \alpha_n)} + R_n - j m_n g \quad (5.45)$$

gdzie v_n jest przemieszczeniem uogólnionym.

Różniczkowe równania ruchu dla podpór zostały omówione w punkcie 5.1.2.

Ponadto na układ wirnika, w przypadku rozoosiowania promieniowego działają, na pierwszy wał, następujące siły synchroniczne (w opisie kartezjańskiego układu współrzędnych): dla przemieszczeń poziomych

$$-\frac{K_4 K_5}{K_4 + K_5} \delta \cos(\phi) + \frac{K_4 K_5}{K_4 + K_5} \delta \quad (5.46)$$

dla przemieszczeń pionowych

$$-\frac{K_4 K_5}{K_4 + K_5} \delta \sin(\phi) \quad (5.47)$$

gdzie δ oznacza wielkość promienia korby. Analogicznie równania (5.46, 5.47) wymuszeń obowiązują dla wału drugiego lecz o przeciwnym znaku.

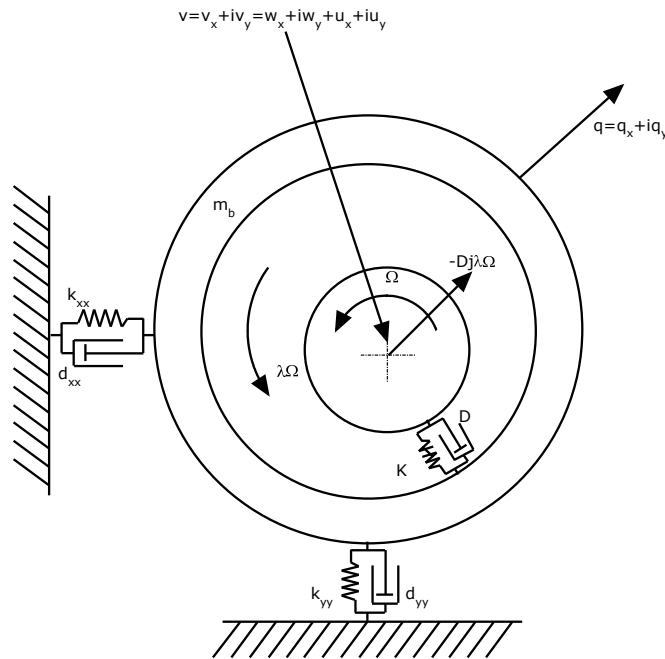
5.2.3. Model podpory

Model podpory został przedstawiony na rysunku 5.9. Układ samej podpory jest realizowany poprzez dwa ortogonalnie rozmieszczone układy Kelvina-Voigta, co umożliwia modelowanie anizotropii tych podpór. W związku z tym, że większa sztywność podpór jest związana z kierunkiem pionowym, zastosowano odpowiednie wartości współczynników określające model podpory. W przypadku ruchu podpór przyjmuje się, że masa podpory i łożyska stanowi jedną masę m_b . Przyjmuje się, że położenie tej masy jest zredukowane do środka łożyska. Natomiast siły działające w łożysku są zredukowane do punktu związanego ze środkiem czopa, są to m.in. siły hydrodynamiczne związane ze sztywnością filmu olejowego oraz z cyrkulacją tegoż filmu.

Równanie ruchu dla przedstawionego układu ma następującą postać:

$$m^b \ddot{q}^b + (c^b \dot{q}^b + k^b q^b) + r^b = 0 \quad (5.48)$$

gdzie r^b oznacza reakcję hydrodynamiczną oddziaływującą na panwie łożyska i podporę.



Rys. 5.9: Schematyczny układ wirnika przedstawiony w jednej płaszczyźnie

5.2.4. Równania ruchu

Zarówno równania ruchu podpór jak i równania ruchu tarcz są sprzężone ze sobą poprzez reakcje oddziaływujące poprzez wał wirnika, który jest liniowo-sprężystą belką o sztywności na zginanie EI . Ugięcie wału jest w tym przypadku wypadkową ugięć wywołanych przez siły hydrodynamiczne oddziaływujące w węzłach łożyskowych oraz siły poprzeczne wynikające z obciążeń w węzłach tarczowych. W oznaczeniach przyjęto następującą notację dotyczącą przemieszczeń uogólnionych:

- v_{n1}, v_{n2}, v_{n3} , dla przemieszczeń bezwzględnych tarcz w węzłach tarczowych,
- $v_{b1}, v_{b2}, v_{b3}, v_{b4}$, dla przemieszczeń bezwzględnych czopów łożyskowych w węzłach łożyskowych,
- q_{b1}, q_{b2}, q_{b3} , dla przemieszczeń bezwzględnych panwi i podpór w węzłach łożyskowych.

Pamiętając o tym, że przemieszczenia środka łożyska q_b i v_b są wzajemnie niezależne można określić przemieszczenie czopa względem ruchomej panwi, dla każdego węzła łożyskowego b jako:

$$\varepsilon_b = q_b - v_b \quad (5.49)$$

stąd dla kolejnych węzłów łożyskowych przyjęto oznaczenia $\varepsilon_{j1}, \varepsilon_{j2}, \varepsilon_{j3}$, dla przemieszczeń względnych czopów łożyskowych. Dla wszystkich wymienionych wcześniej zmiennych dla uproszczenia opisu przyjęto opis w dziedzinie zespolonej

$$v = v_x + j v_y \quad (5.50)$$

Ostatecznie otrzymujemy następujący układ równań różniczkowych dla rozpatrywanego wirnika. Równania ruchu tarcz (przy założeniu izotropowości) mają postać:

$$M_{n1} \ddot{v}_{n1} + S_{n1} = M_{n1} r_1 \Omega^2 e^{j(\Omega t + \alpha_1)} - j M_{n1} g + R_{n1} \quad (5.51)$$

$$M_{n2} \ddot{v}_{n2} + S_{n2} = M_{n2} r_2 \Omega^2 e^{j(\Omega t + \alpha_2)} - j M_{n2} g + R_{n2} \quad (5.52)$$

$$M_{n3} \ddot{v}_{n3} + S_{n3} = M_{n3} r_3 \Omega^2 e^{j(\Omega t + \alpha_3)} - j M_{n3} g + R_{n3} \quad (5.53)$$

gdzie zmienne S_{n1}, S_{n2}, S_{n3} , są zdefiniowane w następujący sposób:

$$S_{n1} = D_{s1} \dot{v}_{n1} + (K_1 + K_2) v_{n1} - K_1 v_{b1} - K_2 v_{n2} \quad (5.54)$$

$$S_{n2} = D_{s2}\dot{v}_{n2} + (K_2 + K_3)v_{n2} - K_2v_{n1} - K_3v_{b2} \quad (5.55)$$

$$S_{n3} = D_{s3}\dot{v}_{n3} + (K_6 + K_7)v_{n3} - K_6v_{n1} - K_7v_{b4} \quad (5.56)$$

gdzie sztywności $K_1 \dots K_7$ odpowiadają sztywnością kolejnych segmentów wału (od lewej) pomiędzy kolejnymi węzłami łożyskowymi i tarczowymi, z uwzględnieniem węzła w którym występuje sprzęgło.

Natomiast równania ruchu czopów łożyskowych mają postać:

$$M_{j1}\ddot{v}_{b1} + D_{b1}(\dot{\varepsilon}_{b1}) + (K_{b1}(\varepsilon_{b1}, \Omega) - j\lambda(\varepsilon_{b1})\Omega D_{b1})v_{b1} - K_1v_{n1} = 0 \quad (5.57)$$

$$M_{j2}\ddot{v}_{b2} + D_{b2}(\dot{\varepsilon}_{b2}) + (K_{b2}(\varepsilon_{b2}, \Omega) - j\lambda(\varepsilon_{b2})\Omega D_{b2})v_{b2} - K_3v_{n2} - \frac{K_4K_5}{K_4 + K_5}v_{b3} = KO_1 \quad (5.58)$$

$$M_{j3}\ddot{v}_{b3} + D_{b3}(\dot{\varepsilon}_{b3}) + (K_{b3}(\varepsilon_{b3}, \Omega) - j\lambda(\varepsilon_{b3})\Omega D_{b3})v_{b3} - K_5v_{n3} - \frac{K_4K_5}{K_4 + K_5}v_{b2} = KO_2 \quad (5.59)$$

$$M_{j4}\ddot{v}_{b4} + D_{b4}(\dot{\varepsilon}_{b4}) + (K_{b4}(\varepsilon_{b4}, \Omega) - j\lambda(\varepsilon_{b4})\Omega D_{b4})v_{b4} - K_7v_{n3} = 0 \quad (5.60)$$

gdzie wymuszenia KO_1, KO_2 odpowiadają wymuszeniom rozosiowania przedstawionym w równaniu (5.46), (5.47).

Natomiast równania ruchu podpór:

$$M_{b1}\ddot{q}_{b1} + (k_{b1}\dot{q}_{b1} + k_{b1}q_{b1}) = -r_{b1} \quad (5.61)$$

$$M_{b2}\ddot{q}_{b2} + (k_{b2}\dot{q}_{b2} + k_{b2}q_{b2}) = -r_{b2} \quad (5.62)$$

$$M_{b3}\ddot{q}_{b3} + (k_{b3}\dot{q}_{b3} + k_{b3}q_{b3}) = -r_{b3} \quad (5.63)$$

$$M_{b4}\ddot{q}_{b4} + (k_{b4}\dot{q}_{b4} + k_{b4}q_{b4}) = -r_{b4} \quad (5.64)$$

gdzie $k_b = k_{xx} + jk_{yy}$ (zgodnie z opisem na rysunku 5.9). Oznaczenie r_b oznacza reakcję hydrodynamiczną łożyska. Reakcje odpowiadają równaniu (5.65):

$$(K(\varepsilon, \Omega)\varepsilon + D(\varepsilon, \Omega)(\dot{\varepsilon} - j\lambda\Omega\varepsilon)) = r \quad (5.65)$$

gdzie współczynniki K i D :

$$D = \frac{\nu d L^3}{c^3} \frac{D_1 \varepsilon}{(1 - \varepsilon^2)^{\frac{5}{2}}} \quad (5.66)$$

$$K = \frac{\nu \Omega d L^3}{c^3} \frac{K_1 \varepsilon}{(1 - \varepsilon^2)^{\frac{3}{2}}} \quad (5.67)$$

gdzie:

$$\varepsilon = \sqrt{\varepsilon_x^2 + \varepsilon_y^2}/c \quad (5.68)$$

gdzie c jest wielkością koła luzów.

5.2.5. Symulator MLH i MSR

Numeryczne odpowiedniki modeli przedstawionych w punktach 5.1 i 5.2 zostały wykonane z wykorzystaniem środowiska obliczeniowego MATLAB (SIMULINK), w wersji 6.6(R2007a). Symulatorowi układu wirnika Foeppla z łożyskiem hydrodynamicznym przyporządkowano oznaczenie MLH (*mlh.mdl*). W przypadku modelu rozbudowanego przyjęto nazwę MSR *msr.mdl*. Opracowane symulatory umożliwiają w sposób przyjazny wprowadzanie zmian dotyczących stanu, cech geometrycznych modelowanych obiektów, jak również warunków działania.

Bibliografia

- [5.1] Bently D.E., Goldman P., Eldridge T. The death of whirl – what the sfcb can do for the stability of rotating machinery. *Orbit, Bently Nevada Co.*, strony 10–13, 2001 (1).
- [5.2] Bently D.E., Goldman P., Eldridge T. The death of whirl and whip -use of externally pressurized bearings and seals for control of whirl and whip instability. *Orbit, Bently Nevada Co.*, strony 42–47, 2004 (1).
- [5.3] Bently D.E., Goldman P., Yu J. The advantages of dynamic stiffness parameters over classical bearing coefficients. *Orbit, Bently Nevada Co.*, strony 23–28, 2001 (3).
- [5.4] Gryboś R. *Dynamika maszyn wirnikowych*. PWN, Warszawa, 1995.
- [5.5] Hori Y. *Hydrodynamic lubrication*. Springer Verlag, Tokyo, 2005.
- [5.6] Kiciński J. *Dynamika wirników i łożysk ślizgowych*. Wydawnictwo IMP PAN, Gdańsk, 2005.
- [5.7] Kiciński J., redaktor. *Modelowanie i diagnostyka oddziaływań mechanicznych, aerodynamicznych i magnetycznych w turbozespołach energetycznych*. Wydawnictwo IMP PAN, Gdańsk, 2005.
- [5.8] Minhui He., Byrne M. Fundamentals of fluid film journal bearing operation and modelling. *Proceedings of thirty-fourth turbomachinery simposium*, strony 155–176, 1997.
- [5.9] Muszynska A. Modal testing of rotors with fluid interaction. *International Journal of Rotating Machinery*, strony 83–116, 1995, Vol. I, No. 2,.
- [5.10] Muszyńska A. *Rotordynamics*. CRC Press, New York, 2005.
- [5.11] Sabin S. Understanding and using dynamic stiffness – a tutorial. *Orbit, Bently Nevada Co.*, strony 44–54, 2000 (2).
- [5.12] Stolarski T.A. *Tribology in machine design*. Butterworth and Heidemann, Oxford, 1990.

Rozdział 6

Przykład zastosowania proponowanej metody

Tomasz ROGALA

W rozdziale tym przedstawiono przykład zastosowania koncepcji modelu BNBM. Omówiono szczegółowo kolejne etapy procesu identyfikacji modelu, podając istotne wskazówki dotyczące jakości uzyskanego klasyfikatora. Przedstawiono również metodę uzgadniania wyjść zastosowaną w drugim stopniu modelu oraz proponowaną procedurę strojenia modelu. Opisano następujące działania związane z budowaniem omawianego modelu:

- pozyskanie danych uczących oraz zbioru dodatkowych zmiennych,
- dekompozycja modelu globalnego i parametryzacja ziarnistości zmiennych dodatkowych,
- wybór cech użytecznych,
- identyfikacja klasyfikatorów jednoklasowych,
- identyfikacja sieci przekonań,
- optymalizacja ewolucyjna modelu BNBM,
- zastosowanie II stopnia modelu BNBM.

6.1. Dane uczące

Identyfikacja diagnostycznego modelu BNBM wymaga przeprowadzenia eksperymentu symulacyjnego pozwalającego na pozyskanie przykładów uczących. W tym celu w badaniach symulacyjnych posłużono się modelem maszyny wirnikowej wielopodporowej, który został przedstawiony w rozdziale 5. Dla zadanych wartości cech stanu, wartości warunków działania oraz wartości cech konstrukcyjnych obiektu rejestrowano sygnały szybkozmiennne drgań w podporach tego obiektu. Wszystkie zebrane w wyniku eksperymentu symulacyjnego sygnały szybkozmiennne zostały przetworzone do zbioru wartości opisujących je cech. W tym celu dla każdego sygnału określono, na podstawie opracowania [6.1], zbiór cech punktowych własnych i wzajemnych w dziedzinie czasu i częstotliwości dla przemieszczeń względnych czopów łożyskowych w kierunku X i Y oraz cech punktowych dla trajektorii centralnych.

Ostatecznie przygotowano zbiór danych uczących DU oraz DB, gdzie liczba przykładów wyniosła:

$$|DU| = |DA| + |DV| = 314 + 120, \quad (6.1)$$

$$|DB| = 314, \quad (6.2)$$

gdzie zbiór DA jest zbiorem przykładów trenujących, zbiór DV jest zbiorem przykładów testujących, a zbiór DB jest zbiorem przykładów dla przestrzeni wartości zmiennych dodatkowych \underline{z} .

Jeżeli zbiór zmiennych \underline{x} jest wektorem zmiennych wejściowych dla modelu BNBM, a zbiór zmiennych \underline{y} jest wektorem zmiennych wyjściowych, to zbiór danych DA reprezentowany jest jako:

$$DA = \{(\underline{x}_1, \underline{y}_1), \dots, (\underline{x}_N, \underline{y}_K)\}, \quad (6.3)$$

natomiast DB

$$DB = \{(\underline{x}_1, \underline{z}_1), \dots, (\underline{x}_N, \underline{z}_J)\}. \quad (6.4)$$

gdzie N określa liczbę przykładów uczących, K określa liczbę rozpatrywanych klas niesprawności, J jest liczbą klas wartości zmiennych dodatkowych.

Rozpatrywany zbiór danych trenujących DA oraz testujących DV zostały odpowiednio sklasyfikowane. Przyjętymi klasami jest klasa określająca stan bazowy oraz klasy określające stany niesprawności. Te ostatnie to przemieszczenia podpór łożyskowych względem wymaganej postaci kinostatycznej linii wału oraz rozosiowanie promieniowe wałów w sprzęgle. Rozpatrywane stany zostały sklasyfikowane do następujących 5 klas stanów:

- stan bazowy -N
- przemieszczenie podpory pierwszej -P1,
- przemieszczenie podpory drugiej -P2,
- przemieszczenie podpory trzeciej -P3,
- przemieszczenie podpory czwartej -P4,
- rozosiowanie promieniowe spowodowane niewłaściwym połączeniem połówek sprzęgła -COU.

Poszczególne stany związane z przemieszczeniem podpór zostały uszczegółowione do czterech kierunków przemieszczeń oraz przemieszczenia nominalnego związanego z przemieszczeniem montażowym w następujący sposób:

- przemieszczenie podpory X w prawo klasa -PXP,
- przemieszczenie podpory X w górę klasa -PXG,
- przemieszczenie podpory X w lewo klasa -PXL,
- przemieszczenie podpory X w dół klasa -PXD,
- przemieszczenie nominalne podpory X, klasa -PXN.

Dla wszystkich wymienionych klas niesprawności przyjęto charakterystyczne dla tej grupy obiektów uszkodzenia multiplikatywne o różnym stopniu zaawansowania. Są one reprezentowane w postaci dyskretnych punktów stanu w przestrzeni stanu obiektu. Dane te zostały zebrane dla tych samych warunków działania oraz tych samych cech konstrukcyjnych obiektu.

Ostatecznie jako dane trenujące zebrano przebiegi czasowe w postaci sygnałów zespolonych (dla dwóch promieniowych kierunków do siebie prostopadłych) przemieszczeń względnych czopów łożyskowych dla każdej podpory. W tym:

- 20 przykładów dla rozosiowania spowodowanego niewłaściwym połączeniem połówek sprzęgła,
- po 20 przykładów dla każdego rozpatrywanego kierunku przemieszczenia podpory 1,
- po 20 przykładów dla każdego rozpatrywanego kierunku przemieszczenia podpory 2,
- po 20 przykładów dla każdego rozpatrywanego kierunku przemieszczenia podpory 3,
- po 20 przykładów dla każdego rozpatrywanego kierunku przemieszczenia podpory 4,
- 1 przykład bazowy, reprezentujący stan odniesienia.

Przedstawiony powyżej zbiór danych opisujący klasy stanów w postaci przemieszczeń względnych czopów łożyskowych wraz z informacjami o warunkach działania obiektu i jego cechach konstrukcyjnych stanowi podstawowy zbiór wejść i wyjść dla identyfikowanego modelu.

6.2. Proces dekompozycji i parametryzacja ziarnistości zmiennych dodatkowych

Rozpatrywane stany niesprawności są powodem powstawania dużych sił oddziaływujących na linię wału na jej całej długości. Przeciwwagą pozwalającą na zachowanie quasistateczności takiego układu wielopodporowego są reakcje oddziaływujące w węzłach łożyskowych. Reakcje stanowią zatem zbiór danych, w rzeczywistości najczęściej niedostępnych pomiarowo, które określają bezpośrednie przyczyny powstawania, pomiarowo dostępnych drgań tychże podpór. Znajomość reakcji dla rozpatrywanego układu może być zatem bardzo pomocnym narzędziem w identyfikacji omawianych stanów niesprawności. Z powyższego względu jako zmienne dodatkowe wybrano reakcje łożyskowe. Przyjęcie założenia upraszczającego, że reakcje są od siebie statystycznie niezależne pozwoliło na dokonanie dekompozycji przestrzennej modelu globalnego do modeli lokalnych sprzęgniętych z podporami łożyskowymi.

Zbiór wartości zmiennych dodatkowych \underline{z} został pozyskany z numerycznego modelu obiektu badań. W tym celu zarejestrowano sygnały szybkozmiennych reakcji łożyskowych na podstawie których wyznaczono wartości reakcji łożyskowych w poszczególnych podporach. Eksperyment symulacyjny prowadzony był w sposób sterowany umożliwiając dla zadanych wartości zmiennych \underline{x} pozyskanie równomiernie rozmieszczonych przykładów uczących DB zmiennych \underline{z} . Przyjęto, że dla każdej przestrzeni zmiennych dodatkowych $Z_1 \dots Z_J$ przykłady będą występowały z prawdopodobieństwem

$$p((\underline{x}, z) \in Z_i) = \begin{cases} \frac{1}{b-a}, & \text{dla } a \leq z \leq b \\ 0, & \text{dla } z > b \text{ lub } z < a \end{cases} \quad (6.5)$$

co oznacza, że każda klasa wartości zmiennych dodatkowych Z_i oparta jest na zbiorze równomiernie rozmieszczonych przykładów w dziedzinie wartości $< a, b >$ stanowiących jej granicę.

Każda podprzestrzeń zmiennych pomocniczych związanych z daną podporą łożyskową została podzielona na następujące klasy:

- duża reakcja podporowa w podporze X w prawo -klasa RXP,
- duża reakcja podporowa w podporze X w górę -klasa RXG,
- duża reakcja podporowa w podporze X w lewo -klasa RXL,
- duża reakcja podporowa w podporze X w dół -klasa RXD,
- mała reakcja w podporze X -klasa RXN.

przy czym klasy te zostały zdefiniowane w oparciu o różnice wartości reakcji względem reakcji bazowych odpowiadających stanowi nominalnemu.

Przykładowy podział w podprzestrzeni zmiennych dodatkowych odpowiadający wybranej podporze został przedstawiony na rysunku 6.1. Granicę małych wartości reakcji podporowych określa elipsa, natomiast pozostałe klasy wartości reakcji są od siebie oddzielone za pomocą prostych.

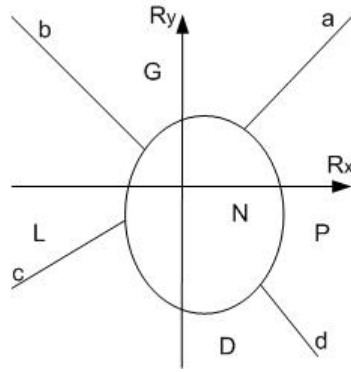
Dla rozpatrywanego przykładu parametrami ziarnistości klas wartości zmiennych dodatkowych mogą być:

- kąty pochylenia linii a,b,c,d względem linii Rx $\alpha_1, \alpha_2, \alpha_3, \alpha_4$,
- długość małej osi elipsy d_1 ,
- długość dużej osi elipsy d_2 ,
- położenie środka elipsy $p(R_x, R_y)$.

Parametry te mogą być różne w zależności od rozpatrywanej podprzestrzeni zmiennych dodatkowych związanych z założoną dekompozycją przestrzenną modelu BNBM.

Ostatecznie ze względu na charakterystyczny rozkład przykładów uczących w obrębie danej podprzestrzeni zmiennych dodatkowych zdecydowano, że parametry ziarnistości dotyczące kątów pochylenia linii będą związane następującymi zależnościami.

$$\alpha_3 = \alpha_1 + 180^\circ \quad (6.6)$$



Rys. 6.1: Podział podprzestrzeni lokalnej zmiennych dodatkowych na klasy reakcji podporowych

$$\alpha_4 = \alpha_2 + 180^\circ \quad (6.7)$$

gdzie $\alpha_1 = 45^\circ$, natomiast $\alpha_2 = 135^\circ$

Uzyskanie definicji klas wartości zmiennych dodatkowych może być również wynikiem zastosowania specjalnych metod przetwarzania tych samych przykładów uczących (choć nie koniecznie), np. zastosowania technik ślepej separacji celem dostarczenia istotnej wiedzy do zdefiniowania zmiennych dodatkowych, a następnie klas ich wartości. Ponadto klasy takie mogą być wynikiem wprowadzenia wiedzy eksperckiej przez eksperta z danej dziedziny. Wykorzystując to ostatnie podejście zdefiniowano omawiane klasy wartości zmiennych dodatkowych. Wstępne wartości parametrów określających ziarnistość tych klas zostały określone na podstawie oceny rozrzutu wartości reakcji podporowych należących do różnych klas w przestrzeni zmiennych dodatkowych.

6.3. Cechy użyteczne

Dla dużych zbiorów cech sygnałów wyznaczonych dla wielu przykładów uczących koniecznością staje się wybór tych cech, które niosą najwięcej informacji o danej klasie stanu obiektu. Uzyskanie takiego zbioru cech wymaga stosowania technik wstępnego przetwarzania sygnałów. Można wymienić m.in. takie metody jak analiza składowych głównych, analiza składowych niezależnych, analiza wrażliwości, skalowanie wielowymiarowe itp..

Dla celów niniejszych badań wybrano jedną z najczęściej stosowanych metod, metodę analizy składowych głównych [6.2]. Jest to metoda pozwalająca na transformację n-wymiarowej przestrzeni danych tj. cech sygnałów w przestrzeń o mniejszym wymiarze określoną na podstawie nieskorelowanych składowych głównych.

Wybór cech użytecznych został przeprowadzony z uwzględnieniem podziału danych DB na klasy w przestrzeni zmiennych dodatkowych tak, że dla każdej przestrzeni tych zmiennych $Z_{i...J}$:

$$\underline{x} \longrightarrow \underline{x}_{rel} \quad (6.8)$$

gdzie $|\underline{x}| > |\underline{x}_{rel}|$, natomiast \underline{x}_{rel} jest zbiorem zmiennych użytecznych. Uwzględnienie podziału przestrzeni zmiennych dodatkowych przy wyborze cech użytecznych pozwala na uzyskanie różnych przestrzeni cech dla każdej klasy wartości zmiennych dodatkowych. Zaletą takiego postępowania jest możliwość uzyskania większej rozróżnialności rozpatrywanych stanów.

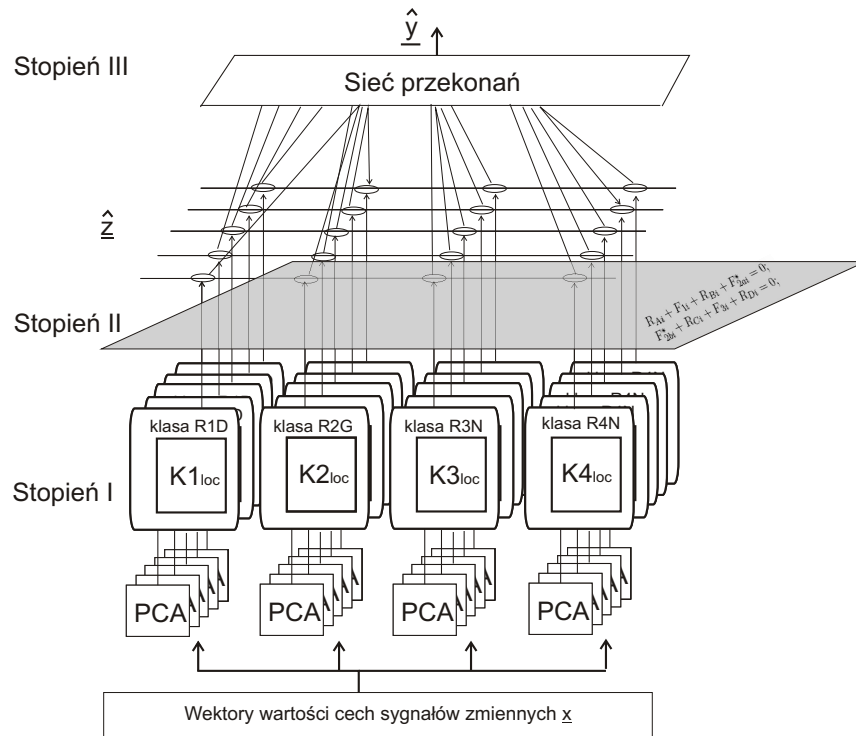
Po dokonaniu transformacji do przestrzeni składowych głównych uzyskano zbiór przykładów uczących

$$DB_{rel} = \{(\underline{x}_{rel_1}, \underline{z}_1), \dots, (\underline{x}_{rel_N}, \underline{z}_J), \}. \quad (6.9)$$

przeznaczonych do trenowania klasyfikatorów jednoklasowych.

Rozmiar przestrzeni składowych głównych został przyjęty na podstawie określonej wartości proęgowej współczynnika jakości odwzorowania [6.2], który określa utratę informacji spowodowaną przeprowadzoną transformacją. Liczba uzyskanych składowych głównych wahała się w przedziale od 4

do 17 w zależności od transformowanego zbioru uczącego przeznaczonego do klasyfikacji. Na takie podejście pozwala stosowanie klasyfikatorów jednoklasowych, gdzie każda klasa może być reprezentowana w dowolnej wielowymiarowej przestrzeni cech, podczas gdy zastosowanie klasyfikatora wieloklasowego wymaga wyboru odpowiedniej liczby wymiarów przestrzeni cech wspólnej dla wszystkich rozpatrywanych klas. Ilustrację struktury modelu BNBM dla rozpatrywanego przykładu wirnika wielopodporowego przedstawia rysunek 6.2.



Rys. 6.2: Struktura modelu BNBM dla rozpatrywanego wirnika wielopodporowego

Głównym celem prowadzonych badań nie było uzyskanie najbardziej efektywnego zbioru cech użytecznych. Nie stosowano w tym celu żadnych dodatkowych metod badania wrażliwości cech. Również dla zastosowanej metody redukcji liczby cech z wykorzystaniem transformacji PCA nie przeprowadzono badań, które miałyby na celu odpowiedź, jakim cechom ze zbioru cech sygnałów odpowiadają wyznaczone, trudne do zinterpretowania fizycznie, wyznaczone składowe główne.

Przygotowany zbiór wektorów wartości cech relewantnych został następnie wykorzystany w procesie uczenia klasyfikatorów jednoklasowych.

6.4. Klasyfikatory jednoklasowe (BNBM - stopień I)

Ze względu na zastosowany podział przestrzenny obiektu, wynikający z punktów pomiarowych zlokalizowanych w podporach łożyskowych, każdej z podpór przyporządkowano przybliżony klasyfikator jednoklasowy związany z daną klasą zmiennych dodatkowych. Klasyfikatory te realizują przekształcenie:

$$OCC_i : \underline{x} \longrightarrow \underline{z}_i \quad (6.10)$$

Jako wynik działania warstwy klasyfikatorów, uzyskiwany jest wektor stopni przynależności \underline{z} reprezentujący współrzędne przestrzenne obrazu klas wartości zmiennych dodatkowych.

Dla przykładów uczących pogrupowanych wg. zdefiniowanych klas wartości zmiennych dodatkowych przeprowadzono proces uczenia klasyfikatorów.

Proces uczenia może być realizowany na dwa sposoby:

- proces uczenia obejmuje przykłady należące do rozpatrywanej klasy dodatkowej oraz wszystkie pozostałe przykłady nie należące do tej klasy,
- proces obejmuje jedynie przykłady należące do rozpatrywanej klasy.

W badaniach zastosowano trzy typy klasyfikatorów jednoklasowych przybliżonych:

- klasyfikator *NN* działający na podstawie algorytmu najbliższych sąsiadów,
- klasyfikator *SVM1* działający na podstawie tzw. wektorów wspierających z wykorzystaniem radialnej gaussowskiej funkcji jądra oraz z uwzględnieniem w procesie uczenia przykładów nie należących do rozpatrywanej klasy,
- klasyfikator *SVM2* z wykorzystaniem radialnej gaussowskiej funkcji jądra bez uwzględnienia w procesie uczenia przykładów nie należących do rozpatrywanej klasy,

W przypadku klasyfikatorów *SVM* jako parametry uczenia określano frakcję dla przykładów należących do rozpatrywanej klasy, które mogą być niepoprawnie sklasyfikowane. Na przykład w celu uwzględnienia przykładów, w których mogły wystąpić błędy numeryczne. Drugi parametr uczenia określa kształt spłaszczenia radialnej funkcji gaussowskiej. Parametry te zostały dobrane w sposób iteracyjny ażeby uzyskać wymaganą jakość klasyfikatora.

W przypadku klasyfikatora *NN* zadanie doboru parametru wymagało jedynie określenia wspomnianej frakcji dla przykładów należących do rozpatrywanej klasy.

Dla opracowanych klasyfikatorów przeprowadzono weryfikację określając podstawowe parametry oceny jakościowej wyników klasyfikacji w oparciu o parametry przedstawione w [6.6]. Uzyskane oceny jakości klasyfikacji zostały opracowane na podstawie 4-punktowej walidacji krzyżowej. W tabeli 6.1 przedstawiono wyniki uśrednione wszystkich stosowanych klasyfikatorów dzięki którym uzyskiwany jest obraz przynależności do klas dodatkowych (patrz 6.2).

Na podstawie uzyskanych wyników przeprowadzono porównanie wybranych typów klasyfikatorów. Porównania dokonano na podstawie oceny parametrów określających jakość klasyfikacji takich jak:

- współczynnik *FNF* określający stopień sklasyfikowanych przykładów niesprawności jako stanów normalnych (nominalnych), gdzie $FNF = \frac{FN}{TP+FN}$
- współczynnik *FPF* określający stopień sklasyfikowanych przykładów stanów normalnych jako niesprawności, gdzie $FPF = \frac{FP}{TN+FP}$
- współczynnik trafności *T* oznaczający frakcję poprawnie sklasyfikowanych przykładów nienależących do rozpatrywanej klasy gdzie $T = \frac{TN}{TN+FP}$,
- parametr *F1* określony zależnością przedstawioną w (6.11).

$$F1 = \frac{2ST}{S+T} \quad (6.11)$$

gdzie *S* jest czułością określoną wzorem $S = \frac{TP}{TP+FN}$, natomiast

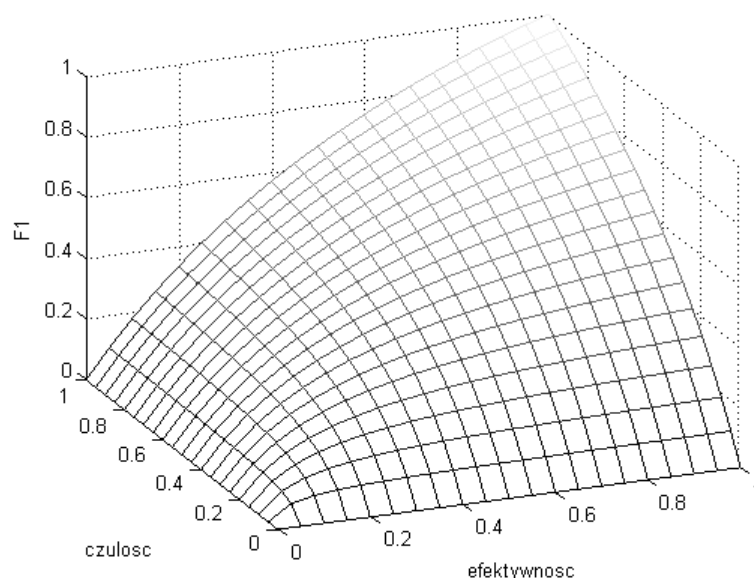
- *TP* - jest liczbą poprawnie sklasyfikowanych przykładów stanu niesprawności,
- *TN* - jest liczbą poprawnie sklasyfikowanych przykładów stanu nominalnego,
- *FN* - jest liczbą błędnie rozpoznanych przykładów stanów niesprawności jako stan nominalny,
- *FP* - jest liczbą błędnie rozpoznanych przykładów dla stanu nominalnego jako stanu niesprawności.

Parametr *F1* określa zrównoważoną ocenę jakości klasyfikatora, gdzie zarówno brana jest pod uwagę trafność *T* jak i czułość *S* (efektywność) klasyfikatora. Interpretacja graficzna tego parametru została przedstawiona na rysunku 6.3.

Uzyskane wyniki dla obydwu typów klasyfikatorów wskazują na przewagę klasyfikatora najbliższego sąsiada, dla którego uzyskano lepsze wyniki pod względem parametru *FNF* - parametru bardziej

Tab. 6.1: Wyniki wstępnej weryfikacji klasyfikatorów SVM1 i NN na podstawie 4-punktowej walidacji krzyżowej (wartości uśrednione dla zbioru wszystkich klasyfikatorów lokalnych)

	FNF	FPF
SVD	0.487	0.016
NN	0.075	0.568



Rys. 6.3: Interpretacja graficzna zrównoważonej oceny F1 jakości klasyfikatora

istotnego z punktu diagnostycznego przeznaczenia budowanego klasyfikatora. Wartość parametru T dla klasyfikatora NN wyniosła 0.268 natomiast parametr $F1$ wyniósł 0.342. Przy czym wartość parametru FPF jest nie mniej istotna, ponieważ system diagnostyczny będzie przekazywał informację o uszkodzeniach i niesprawnościach, które nie miały miejsca.

Obserwacja wizualna uzyskanych obszarów dla niektórych przykładów, które zostały określone w przestrzeni dwuwymiarowych składowych głównych pokazały, że o wiele częściej w przypadku klasyfikatorów NN, obszary te są mniej spójne i bardziej rozległe. Wynika to prawdopodobnie z konstrukcji tej metody, która dla niewielkiej liczby danych tworzy bardziej rozległe obszary klas w przestrzeni cech. Przykładowe obszary wyznaczone przez klasyfikatory SVM1 i NN dla tych samych danych zostały przedstawione na rysunku 6.4.

Na rysunku 6.4 widoczny jest rozległy obszar określony przez klasyfikator najbliższego sąsiada, co może bezpośrednio wpłynąć na jakość uzyskanych stopni przynależności w przypadku danych zaszumionych.

W zależności od wyników jakości klasyfikacji, może być stosowany bardziej korzystny typ klasyfikatora. W przypadku stosowania w tym samym systemie diagnozującym dwóch typów klasyfikatorów konieczne jest opracowanie jednoznacznego sposobu obliczania stopnia przynależności będącego wyjściem z klasyfikatora w obydwu przypadkach np. na podstawie takich samych typów funkcji przynależności. Na przykład stopień przynależności do klas wartości zmiennych dodatkowych może być określony na podstawie wartości funkcji Gaussa:

$$\nu(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \quad (6.12)$$

gdzie x określa dane, dla których obliczany jest stopień przynależności, natomiast parametr σ określa

szerokość spłaszczenia funkcji Gaussa. Przy czym, w realizowanym zadaniu stopień przynależności został wyskalowany w taki sposób, że jego wartość jest z przedziału od zera do jeden przyjmując wartość zero poza obszarem wyznaczonym przez wartość progową.

W wyniku działania klasyfikatorów uzyskiwany jest zbiór zmiennych \hat{z} wykorzystywany przez kolejne stopnie modelu BNBM.

Opisane badania zostały przeprowadzone przy wykorzystaniu środowiska obliczeniowego Matlab z dodatkowym pakietem narzędziowym PRTTools [6.5].

6.5. Sieć przekonań (BNBM - stopień III)

Wnioskowanie diagnostyczne realizowane jest przez trzeci stopień modelu diagnostycznego BNBM reprezentowany przez sieć przekonań, która przyjmując dla węzłów informacyjnych informacje o stopniach przynależności wyznaczonych przez pierwszy stopień modelu \hat{z} jako wynik przedstawia zmienne \hat{y} będące stopniami przekonania o możliwym wystąpieniu niesprawności.

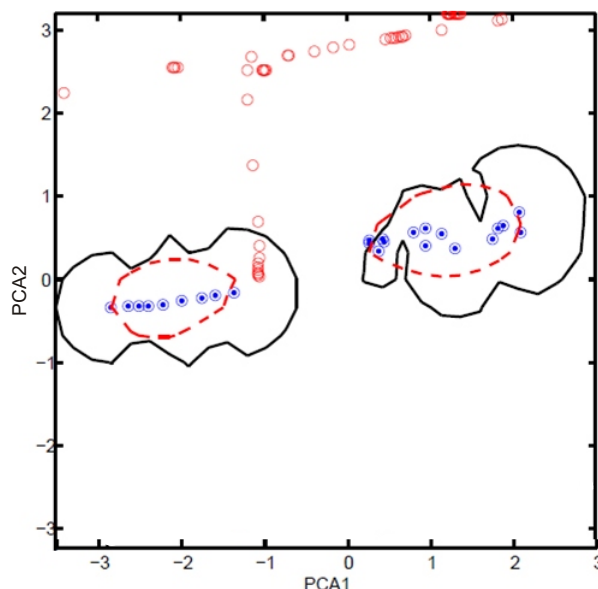
Struktura sieci została określona na podstawie podstawowej zależności przyczynowych zachodzących pomiędzy zmiennymi, w których można wyróżnić:

- zmienne określające symptomy, które zawarte są w informacyjnych węzłach sieci,
- zmienne określające stan (niesprawności), które są reprezentowane w postaci węzłów obserwowanych,
- zmienne pomocnicze, które zawarte są w węzłach ukrytych, a w których zawarto dodatkową wiedzę ekspercką.

Wszystkie węzły mają postać węzłów deterministycznych dwustanowych lub pięciostanowych.

Sposób połączenia węzłów za pomocą gałęzi oraz pełna struktura została przedstawiona na rysunku 6.5, gdzie węzłami informacyjnymi są węzły: $R1, R2, R3, R4$ natomiast węzłami obserwowanymi są węzły $P1, P2, P3, P4, Cou, S$. Gdzie akronimy :

$R1...R4$ oznaczają kolejno węzły w których poszczególne stany odpowiadają klasom wartości reakcji dla kolejnych podpór 1...4,



Rys. 6.4: Porównanie klasyfikatorów SVM1 i NN. Przykładowe obszary dla klasy *duże reakcje podporowe w podporze czwartej* w dwu-wymiarowej przestrzeni składowych głównych wyznaczone przez klasyfikator SVM1 (linia przerywana) oraz klasyfikator NN (linia ciągła) dla podpory trzeciej

$P1...P4$ oznaczają kolejno węzły w których poszczególne stany odpowiadają klasom przemieszczeń dla kolejnych podpór 1...4,

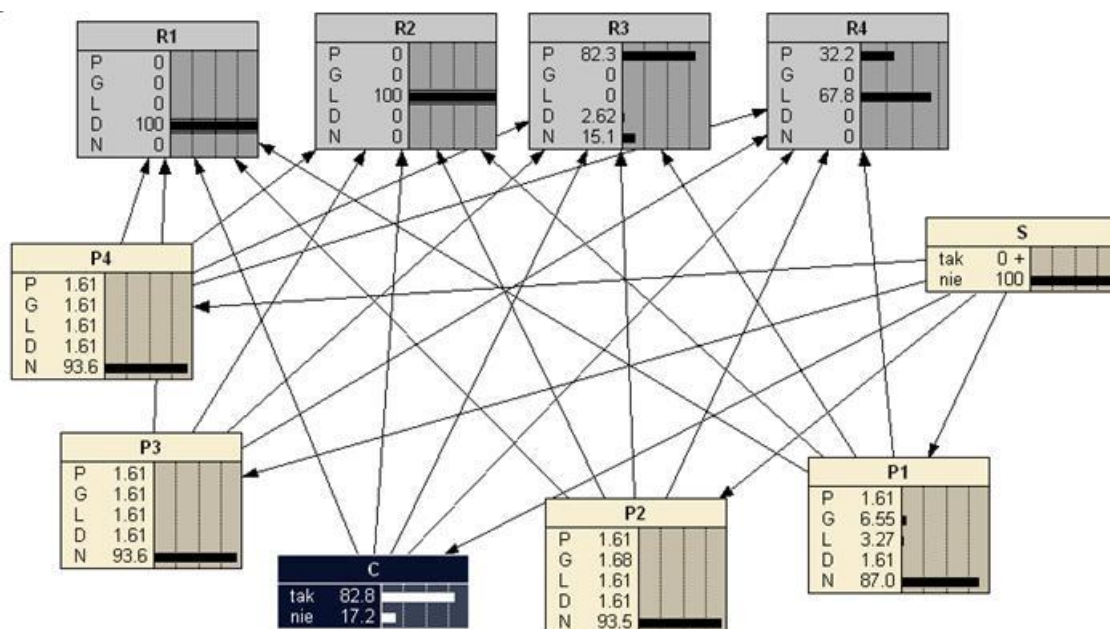
Cou oznacza węzeł opisujący stan zajścia lub braku zajścia stanu związanego z rozoosiowaniem spowodowanym niewłaściwym połączeniem połówek sprzęgieł,

S węzeł ze stanami oznaczającymi stan zdalny i niezdalny.

Węzeł S może być rozpatrywany jako pomocniczy węzeł ukryty, który pozwala na wnioskowanie o stanie zdalności przede wszystkim na podstawie wartości stanów nominalnych reakcji w węzłach $R1, R2, R3, R4$.

Dla zbioru wektorów stopni przynależności uzyskanych w wyniku przetworzenia przez I stopień modelu uzyskano zbiór wektorów stopni przynależności stanowiących w konwencji sieci bayesowskich stopnie przekonania o wystąpieniu danego stanu.

Ponieważ stopnie przekonania o wystąpieniu danego stanu w obrębie danego węzła sumują się do jedynki, stąd stopnie przynależności dotyczące tego węzła wymagają przeprowadzenia odpowiedniego przeskalowania.



Rys. 6.5: Sieć przekonań dla przykładu wystąpienia rozoosiowania promieniowego w sprzęgle. Kolorem szarym zaznaczono węzły informacyjne ze stopniami przekonania o wystąpieniu danej klasy wartości zmiennych dodatkowych (reakcji). Kolorem czarnym zaznaczono jeden z węzłów obserwowanych w którym stopień przekonania wskazuje na wystąpienie testowanej niesprawności. (reprezentacja graficzna z wykorzystaniem interfejsu programu NETICA firmy Norsys)

Wiedza zaimplementowana w sieci przekonań oprócz gałęzi reprezentowana jest w postaci tablic prawdopodobieństw warunkowych oraz prawdopodobieństw apriori. Możliwe jest zdefiniowanie tablic prawdopodobieństw przez eksperta lub grupę ekspertów. Ponadto możliwe jest uczenie węzłów informacyjnych oraz obserwowanych na podstawie tzw. przykładów uczących.

W celu konfiguracji sieci, zdefiniowano wstępnie, na podstawie dostępnej wiedzy literaturowej, opisującej relacje diagnostyczne stan-symptom, tablice prawdopodobieństw warunkowych. W kolejnym etapie przystąpiono do uczenia węzłów obserwowanych oraz informacyjnych na podstawie przykładów uczących przetworzonych do postaci stopni przynależności przez pierwszą warstwę klasyfikatorów jednoklasowych.

W wyniku wstępnego testowania modelu BNBM na podstawie przykładów testujących uzyskano wyniki dotyczące jakości klasyfikatora BNBM. Wyniki zostały przedstawione w tabeli 6.2. Chociaż

Tab. 6.2: Skuteczność klasyfikatora BNBM dla wstępnie założonego podziału w przestrzeni zmiennych dodatkowych. Testowanie przeprowadzono z wykorzystaniem zbioru danych testujących.

	Sprawność klasyfikacji	Czułość	Trafność
Przemieszczenia w podporze I	79%	19%	99%
Przemieszczenia w podporze II	83%	60%	93%
Przemieszczenia w podporze III	84%	40%	94%
Przemieszczenia w podporze IV	81%	30%	99%
Rozosiowanie (sprzęgło)	95%	53%	100%

Tab. 6.3: Analiza szczegółowych wyników skuteczności klasyfikacji dla stanów związanych z przemieszczeniem podpory trzeciej -wyniki wstępne dla klasyfikatora BNBM. Testowanie przeprowadzono z wykorzystaniem zbioru danych testujących.

Stan rozpoznany					Stan Bieżący
P3P	P3G	P3L	P3D	P3N	
6	0	0	0	14	P3P
0	8	1	0	11	P3G
0	0	10	0	10	P3L
0	0	0	14	6	P3D
0	0	3	1	257	P3N

wyniki sprawności klasyfikatorów są zadowalające, to szczegółowa analiza parametrów czułości klasyfikatora oraz liczby przypadków błędnie sklasyfikowanych jest niezadowalająca. Przykładowa analiza wyników dla przemieszczeń w podporze trzeciej została przedstawiona w tabeli 6.3

Należy zwrócić uwagę, że na poprawność uzyskanego rozwiązania ma wpływ nie tylko proces uczenia sieci i konfiguracji węzłów ukrytych, ale również odpowiednia jakość wyników klasyfikacji uzyskana w pierwszej warstwie modelu BNBM. Wynika stąd wniosek, że możliwe jest prowadzenie procesu strojenia modelu BNBM poprzez dodatkowe kryterium optymalizacyjne w celu strojenia nie tylko modelu pierwszej warstwy i uczeniu sieci przekonań, ale również zmian polegających na zmianie definicji klas wartości zmiennych dodatkowych. Badania przeprowadzone w tej sekcji zostały przeprowadzone przy wykorzystaniu środowiska obliczeniowego Matlab z dodatkowym pakietem narzędziowym BNTtoolbox, Netica, oraz Netica API [6.4].

6.6. Optymalizacja ewolucyjna modelu BNBM

W poprzednim podrozdziale klasy wartości zmiennych dodatkowych zostały ustalone na podstawie wstępnej oceny rozrzutu przykładów w przestrzeni zmiennych dodatkowych. Jednak ustalenie odpowiedniej wartości progowej dla danej klasy wartości zmiennych dodatkowych jest zagadnieniem kluczowym, warunkującym uzyskanie optymalnego podziału w tej przestrzeni. Ma to istotny wpływ na uzyskaną jakość klasyfikatora BNBM. Odpowiednia definicja klas wartości zmiennych dodatkowych na podstawie wartości progowej może być prowadzona z wykorzystaniem odpowiednich metod optymalizacyjnych. W tym celu posłużono się zastosowaniem algorytmów genetycznych, które poprzez ewolucyjny dobór wartości parametrów ziarnistości poszukują najlepszego rozwiązania. Parametrami sterowanymi w procesie optymalizacji są parametry ziarnistości omówione w poprzednich podrozdziałach. Funkcja przystosowania, dla której poszukiwana była maksymalna wartość, zdefiniowana została w następujący sposób:

$$FF = \frac{1}{\sum_k Q(k)} \quad (6.13)$$

gdzie

$$Q(k) = \frac{TP(k) + TN(k)}{TP + TN + FP + FN} \quad (6.14)$$

gdzie $Q(k)$ jest sprawnością klasyfikacji k -tego obserwowanego węzła, natomiast $TP(k) + TN(k)$ są poprawnie sklasyfikowanymi przykładami dla stanów odpowiadających węzłowi k . Wyznaczenie funkcji przystosowania dla każdego osobnika w każdym pokoleniu związane było z:

- wyznaczeniem zbioru cech relewantnych dla nowego podziału klas zmiennych dodatkowych $\{Z_1 \dots Z_j\}$
- identyfikacją klasyfikatorów jednoklasowych $\{OCC_1 \dots OCC_m\}$
- identyfikacją sieci przekonań na podstawie nowego zbioru wektorów stopni przynależności \hat{z} do klas pomocniczych,
- ocenie uzyskanej sprawności klasyfikatora BNBM na podstawie oceny wartości (6.14)

Przed przystąpieniem do wyżej opisanej procedury dokonano następujących działań:

- wyznaczono w sposób iteracyjny optymalną wartości współczynnika jakości odwzorowania dla analizy składowych głównych przy niezmiennym podziale przestrzeni zmiennych dodatkowych. Funkcją oceniającą jest w tym przypadku średnia wartość jakości klasyfikatorów lokalnych uzyskana przy zastosowaniu walidacji krzyżowej czteropunktowej,
- wyznaczono w sposób iteracyjny najlepsze parametry klasyfikatorów dla stałego podziału przestrzeni zmiennych dodatkowych. Funkcją oceniającą jest sprawność tych klasyfikatorów,

W wyniku przeprowadzonej optymalizacji uzyskano poprawę czułości klasyfikatora BNBM. Wyniki dotyczące oceny skuteczności klasyfikacji przy zastosowaniu metody LOO (ang. leave one out) zostały zestawione w tabeli 6.4 i 6.5. Ilustracją prowadzonej optymalizacji podziału zmiennych dodatkowych jest rysunek 6.6.

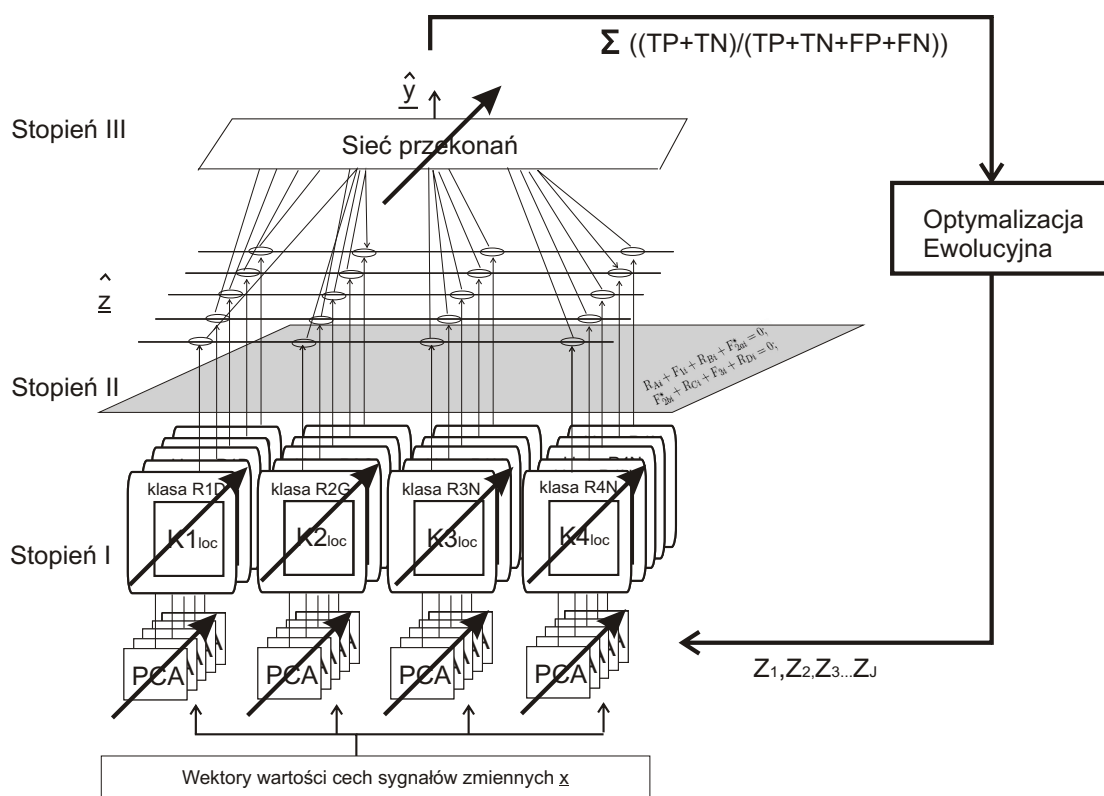
Tab. 6.4: Skuteczność klasyfikatora BNBM po optymalizacji podziału w przestrzeni zmiennych dodatkowych. Wyniki uzyskano stosując metodę testowania LOO

	Sprawność klasyfikacji	Czułość	Trafność
Przemieszczenia w podporze I	88%	66%	99%
Przemieszczenia w podporze II	87%	52%	100%
Przemieszczenia w podporze III	89%	77%	93%
Przemieszczenia w podporze IV	88%	72%	99%
Rozosiowanie (sprzęgło)	98%	85%	100%

6.7. Uzgadnianie wartości zmiennych dodatkowych (BNBM - stopień II)

Drugi stopień modelu BNBM umożliwiający uzgadnianie wartości zmiennych dodatkowych nie jest elementem obligatoryjnym. Zastosowanie II stopnia zapewnia jednak możliwość jawnego uwzględnienia wiedzy o rozpatrywanym obiekcie.

Zastosowana dekompozycja przestrzenna obiektu badań dla zmiennych dodatkowych umożliwia budowanie klasyfikatorów lokalnych OCC. Podejście takie, chociaż słuszne z punktu modelowania diagnostycznego, powoduje, że uzyskane stopnie przynależności do określonych klas wartości zmiennych dodatkowych związanych z daną podporą są niezależne od stopni przynależności w pozostałych



Rys. 6.6: Schemat procesu optymalizacji ewolucyjnej dla rozpatrywanego modelu BNBM

podporach. Sytuacja taka jest skutkiem przyjętego wcześniej założenia o możliwości przestrzennej dekompozycji rozpatrywanego modelu maszyny wirnikowej na niezależnie modelowane podpory łóżyiskowe. Może to prowadzić do uzyskania wyników niezgodnych z podstawowymi prawami fizyki. Niedogodność tą można zniwelować stosując II stopień modelu BNBM, poprzez wprowadzanie wiedzy na temat sposobu uzgadniania wartości zmiennych dodatkowych.

Jeżeli rozpatrywany model wirnika wielopodporowego jest modelem quasistatycznym to konieczne jest zachowanie następujących równań równowagi statycznej obiektu.

$$\sum_{i=1}^k F_{xi} + \sum_{i=1}^k R_{xi} = 0 \quad (6.15)$$

$$\sum_{i=1}^k F_{yi} + \sum_{i=1}^k R_{yi} = 0 \quad (6.16)$$

Tab. 6.5: Szczegółowa analiza błędów dla klasyfikacji przemieszczeń w podporze III -po optymalizacji. Wyniki uzyskano stosując metodę testowania LOO

	Stan rozpoznany					Stan Bieżący
	P3P	P3G	P3L	P3D	P3N	
14	0	0	0	0	6	P3P
0	16	0	0	0	4	P3G
0	0	16	0	0	4	P3L
0	0	0	16	0	4	P3D
3	7	3	3	3	245	P3N

$$\sum_{i=1}^k F_{xi} l_i + \sum_{i=1}^{k-1} R_{xi} l_i = 0 \quad (6.17)$$

$$\sum_{i=1}^k F_{yi} l_i + \sum_{i=1}^{k-1} R_{yi} l_i = 0 \quad (6.18)$$

gdzie $F_{xi}, F_{yi}, R_{xi}, R_{yi}$ są składowymi poziomymi i pionowymi sił i reakcji, natomiast l_i są poszczególnymi odległościami geometrycznymi w których przyłożone są wektory sił i reakcji.

W celu rozwiązania zadania polegającego na uzgodnieniu wartości stopni przynależności do rozpatrywanych klas stanu wartości zmiennych pomocniczych zastosowano metodę wyrównawczą pokazaną w [6.3]. Zmodyfikowaną postać tej metody opisano poniżej.

Niech każda klasa wartości zmiennych dodatkowych

$$Z_i = \{z_1, z_2, \dots, z_N\} \quad (6.19)$$

będzie określona elementem reprezentatywnym $\eta_{i...J}$ takim, że dla każdej klasy wartości zmiennych dodatkowych $Z_{i...J}$:

$$\eta_{i...J} = \max\{z_1, z_2, \dots, z_N\} \quad (6.20)$$

Zatem dla każdego przykładu uczącego i testującego dane są wartości (nie rozmyte) dla klas wartości zmiennych dodatkowych $\eta_{1...J}$ oraz odpowiadające im stopnie przynależności $\nu_{1...J}$.

Dane te dla każdego przykładu uczącego tworzą parę $\{\eta, \nu\}$. Przy czym jako wartość obserwowaną uznaje się wartość iloczynu pary tych liczb w postaci.

$$L = [o_1, o_2, \dots, o_J]; \quad (6.21)$$

gdzie:

$$o_i = \eta_i \nu_i; \quad (6.22)$$

przy czym ($\eta_i = \text{idem}$).

Znając powyższe dane oraz r równań warunkowych w postaci:

$$F_i(o_1 + v_1, o_2 + v_2, \dots, o_n + v_n) = 0 \quad (6.23)$$

dla $i = 1...r$ gdzie r jest liczbą równań warunkowych a ν poprawką pozwalającą na wyrównanie obserwacji poszukiwane są wartości poprawek pozwalające na spełnienie równań warunkowych. Ponieważ uzyskane rozwiązanie może posiadać nieskończenie wiele rozwiązań poszukiwane są wartości ν_i spełniające kryterium najmniejszych kwadratów dla wartości poprawek.

$$[\mathbf{vv}] = v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2 \Rightarrow \min \quad (6.24)$$

Mając równania warunkowe w postaci liniowej równania odchyłek przyjmują formę:

$$[\mathbf{a}_1 \mathbf{v}] = a_{1,1} v_1 + a_{1,2} v_2 + \dots + a_{1,n} v_n + \omega_1 = 0 \quad (6.25)$$

$$[\mathbf{a}_2 \mathbf{v}] = a_{2,1} v_1 + a_{2,2} v_2 + \dots + a_{2,n} v_n + \omega_2 = 0$$

$$\dots \quad (6.26)$$

$$[\mathbf{a}_r \mathbf{v}] = a_{r,1} v_1 + a_{r,2} v_2 + \dots + a_{r,n} v_n + \omega_r = 0$$

gdzie ω_i

$$\omega_i = F_i(o_1 + v_1, o_2 + v_2, \dots, o_n + v_n) - F_i(o_1, o_2, \dots, o_n) \quad (6.27)$$

Układ równań (6.25) w postaci macierzowej ma postać:

$$\mathbf{A}_{[rxn]} \mathbf{V}_{[nx1]} + \omega_{[rx1]} = 0 \quad (6.28)$$

Spełnienie warunku (6.24) przy uwzględnieniu stosownych równań warunkowych wymaga znalezienia ekstremum funkcji F_L :

$$F_L(\nu_{1...n}, \lambda_{1..r}) = \mathbf{v}\mathbf{v} - 2\lambda_1[\mathbf{a}_1\mathbf{v}] - 2\lambda_2[\mathbf{a}_2\mathbf{v}] - \dots - 2\lambda_r[\mathbf{a}_r\mathbf{v}] \quad (6.29)$$

przy czym spełniony jest warunek $v > r$.

Podstawowy warunek konieczny dla istnienia ekstremum jest równy

$$\frac{\partial F_L}{\partial v_i} = 2v_i - 2k_1a_{1,i} - 2k_2a_{2,i} - \dots - 2k_ra_{r,i} = 0 \quad (6.30)$$

stąd

$$v_i = k_1a_{1,i} + k_2a_{2,i} + \dots + k_ra_{r,i} = 0 \quad (6.31)$$

Podstawiając równania (6.32) do równań (6.25) uzyskiwany jest układ równań. Postać macierzowa tych równań dana jest zależnością:

$$\mathbf{A}\mathbf{A}_{[R*R]}^T\mathbf{K}_{[R*1]} + \omega_{[R*1]} = 0 \quad (6.32)$$

Wektor współczynników Lagrange'a jest równy:

$$\mathbf{K}_{[R*1]} = (\mathbf{A}\mathbf{A}_{[R*R]}^T)^{-1}\omega_{[R*1]} \quad (6.33)$$

natomiast poszukiwany wektor poprawek jest równy:

$$\mathbf{V}_{[N*1]} = (\mathbf{A}_{[N*R]}^T)\mathbf{K}_{[R*1]} \quad (6.34)$$

Dla omawianego w poprzedniej sekcjach zadania dotyczącego identyfikacji modelu BNBM, zastosowano powyższą metodę wyrównawczą. Jako wynik działania II stopnia modelu uzyskano zmodyfikowane wartości stopni przynależności uwzględniające statyczne równania równowagi rozpatrywanego układu maszyny wirnikowej. Metoda ta, ze względu na zbyt małą liczbę klas wartości w przestrzeni zmiennych dodatkowych nie przyniosła wymiernych korzyści w postaci poprawy skuteczności klasyfikatora BNBM.

Konieczne jest zatem prowadzenie dalszych badań z uwzględnieniem bardziej szczegółowego podziału w przestrzeni zmiennych dodatkowych. W tym celu klasy wartości powinny nie tylko uwzględniać kierunek reakcji, ale też ich wartość przybliżoną np. reakcja bardzo duża, reakcja duża, reakcja mała, reakcja dopuszczalna itp..

6.8. Wnioski

Na podstawie przeprowadzonych badań, stwierdzono przydatność proponowanej metody jako metody modelowania diagnostycznego. Podstawową zaletą tej metody jest możliwość implementowania wiedzy zarówno w postaci relacji diagnostycznych np. jak w powyższym przykładzie w stopniu trzecim, jak i wiedzy ogólnej, czyli np. praw fizyki w drugim stopniu modelu. Uzyskane wyniki dla przedstawionego modelu można uznać za zadowalające jak również potwierdzające własności uogólniające omawianego modelu. Istotną zaletą klasyfikatora BNBM, w porównaniu z prostymi klasyfikatorami, jest jego wysoka trafność połączona z dobrą dokładnością klasyfikacji. Na jakość klasyfikatora w znacznym stopniu wpływa optymalizacja podziału zmiennych dodatkowych.

Bibliografia

- [6.1] Cholewa W., Ryzdzik S. Wektor wartości cech sygnałów. Kiciński J., redaktor, *Modelowanie i diagnostyka oddziaływań mechanicznych, aerodynamicznych i magnetycznych w turbozespołach energetycznych*. Wydawnictwo IMP PAN, Gdańsk, 2005.

- [6.2] Cholewa W., Skupnik D., Urbanek G, Sławik D. Cechy relewantne. Kiciński J., redaktor, *Modelowanie i diagnostyka oddziaływań mechanicznych, aerodynamicznych i magnetycznych w turbozespołach energetycznych*. Wydawnictwo IMP PAN, Gdańsk, 2005.
- [6.3] Gałda M. *Obliczenia geodezyjne i rachunek wyrównawczy*. PWN, W-wa, 1972.
- [6.4] Murphy K.P. A brief introduction to graphical models and bayesian networks. <http://www.cs.berkeley.edu/murphyk/Bayes/bayes.html>.
- [6.5] Heijden van der F., Duin R.P.W., Ridder de D., Tax D.M.J. Prtools software. <http://www.prtools.org/>.
- [6.6] Heijden van der F., Duin R.P.W., Ridder de D., Tax D.M.J., redaktorzy. *Classification, Parameter Estimation and State Estimation*. John Wiley and Sons Ltd, England, Londyn, 2006.

