

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

Krzysztof Ciupke

**Metoda selekcji
i redukcji informacji
w diagnostyce maszyn**

Gliwice 2001

Recenzenci

Dr hab. inż. Jan Maciej Kościelny, prof. nzw. w Politechnice Warszawskiej
Prof. dr hab. inż. Wojciech Cholewa, Politechnika Śląska

Redaktor Zeszytów

Wojciech Cholewa

Redaktor językowy

Ewa Opoka

Redaktor techniczny

Marek Wyleźoł

Projekt okładki

Wojciech Cholewa, Marek Wyleźoł

ISBN 83–906533–6–2

Wydawca

Katedra Podstaw Konstrukcji Maszyn
Wydział Mechaniczny Technologiczny
Politechnika Śląska
ul. Konarskiego 18a, 44-100 Gliwice
tel. (32) 237-14-67, fax (32) 237-13-60
<http://kpkm.mt.polsl.gliwice.pl>

Od autora

Zeszyt został opracowany na podstawie mojej pracy doktorskiej, wykonanej pod kierunkiem dr. hab. Wojciecha Moczulskiego, prof. nzw. w Politechnice Śląskiej. Publiczna obrona rozprawy odbyła się 29 maja 2001 roku przed Komisją powołaną przez Radę Wydziału Mechanicznego Technologicznego. W opracowaniu zostały uwzględnione uwagi recenzentów rozprawy doktorskiej: dr. hab. inż. Jana Macieja Kościelnego i prof. dr. hab. inż. Wojciecha Cholewy.

Składam serdeczne podziękowania Rodzinie, a także Koleżankom i Kolegom z Katedry Podstaw Konstrukcji Maszyn Politechniki Śląskiej za okazaną mi pomoc i życzliwość w czasie wykonywania pracy. Szczególnie zaś dziękuję promotorowi dr. hab. Wojciechowi Moczulskiemu oraz prof. Wojciechowi Cholewie za możliwość podjęcia i realizacji samodzielnych badań.

Gliwice, lipiec 2001

Krzysztof Ciupke

Spis treści

Wykaz ważniejszych oznaczeń	9
1 Wstęp	13
2 Problem badawczy	21
2.1 Pozyskiwanie wiedzy diagnostycznej	21
2.2 Pozyskiwanie wiedzy metodami „uczenia maszynowego”	22
2.3 Niedoskonałość danych	24
2.4 Cechy ilościowe i jakościowe	25
2.5 Reprezentacja danych i wiedzy	26
2.5.1 Reprezentacja danych	26
2.5.2 Reprezentacja wiedzy	28
2.6 Konwersja wartości atrybutów	29
2.7 Ograniczenie liczności zbioru atrybutów	32
2.8 Klasyfikator stanu	34
2.9 Weryfikacja klasyfikatora	34
2.10 Podsumowanie dotyczące problemu badawczego	35
3 Cel, teza i zakres rozprawy	37
3.1 Cel rozprawy	37
3.2 Teza rozprawy	38
3.3 Zakres rozprawy	38
4 Wybrane metody redukcji informacji diagnostycznej	41
4.1 Celowość redukcji informacji diagnostycznej	41
4.2 Metody jednoatrybutowe redukcji ilości informacji	43

4.2.1	Metoda równych szerokości przedziałów	43
4.2.2	Metoda równych częstości w przedziałach	43
4.2.3	Metody oparte na testach statystycznych i miarach prawdopodobieństwa	44
4.2.4	Metoda minimum entropii	50
4.2.5	Metoda grupowania	55
4.3	Metody wieloatrybutowe redukcji ilości informacji	55
4.3.1	Metoda grupowania	57
4.3.2	Metoda heurystyczna	57
4.4	Liczba wartości granicznych	61
5	Wybrane metody selekcji atrybutów diagnostycznych	65
5.1	Metody selekcji atrybutów bazujące na miarach prawdopodobieństwa	66
5.2	Metoda selekcji atrybutów oparta na teorii zbiorów przybliżonych	67
5.3	Metoda selekcji atrybutów bazująca na teorii algorytmów genetycznych	69
5.4	Metoda selekcji atrybutów z wykorzystaniem kryterium minimum entropii	72
5.5	Metoda selekcji atrybutów z zastosowaniem sieci neuronalnych	75
5.6	Metoda selekcji atrybutów z zastosowaniem „uczenia maszynowego”	79
6	Kryteria i sposoby oceny rozpatrywanych metod redukcji i selekcji informacji	85
6.1	Kryteria oceny rozpatrywanych metod redukcji i selekcji informacji	85
6.2	Sposoby oceny rozpatrywanych metod redukcji i selekcji informacji	88
6.2.1	Określenie stopnia spełnienia kryteriów oceny	88
6.2.2	Sposoby określania sprawności i błędów klasyfikacji	90
6.2.3	Proces realizacji badań weryfikacyjnych	91
7	Badania rozpatrywanych metod redukcji i selekcji informacji diagnostycznej	93
7.1	Wybór metody pozyskiwania wiedzy	93
7.2	Badania dotyczące wyników czynnego eksperymentu diagnostycznego	96
7.2.1	Zadanie diagnostyczne	96
7.2.2	Stanowisko do przeprowadzania czynnych eksperymentów diagnostycznych	96
7.2.3	Zbiór przykładów uczących	97

7.2.4	Wyznaczanie wartości granicznych	98
7.2.5	Selekcja atrybutów	101
7.2.6	Wyniki badań dotyczące czynnego eksperymentu diagnostycznego i ich analiza	102
7.3	Badania dotyczące wyników eksperymentu numerycznego	114
7.3.1	Zadanie diagnostyczne	114
7.3.2	Model obiektu zastosowanego do badań numerycznych	114
7.3.3	Zbiór przykładów uczących	115
7.3.4	Zastosowane metody redukcji i selekcji informacji	117
7.3.5	Wyniki dotyczące eksperymentu numerycznego i ich analiza . . .	122
7.4	Podsumowanie przeprowadzonych badań	128
7.5	Proponowana metoda redukcji i selekcji informacji diagnostycznej . . .	130
8	Podsumowanie i wnioski	133
	Literatura	137
	Dodatek Elementy teorii zbiorów przybliżonych	151
	Spis rysunków	154
	Spis tabel	157
	Spis algorytmów	159
	Streszczenie	160
	Summary	161

Wykaz ważniejszych oznaczeń

A	zbiór atrybutów
a, a_j	atrybut obiektu
a^q, a^c	atrybut jakościowy
A^c	zbiór atrybutów jakościowych
$a()$	wartość atrybutu a
$\arg \min f()$	wartość (dowolna) minimalizująca $f()$
b	wartość graniczna atrybutu określająca stan niezdatności maszyny
B	zbiór atrybutów
$\underline{B}X$	B-dolna aproksymacja zbioru X
$\overline{B}X$	B-górna aproksymacja zbioru X
$Bn_B(X)$	B-brzeg zbioru X
$\text{bel}()$	stopień przekonania
c_i^a	wartość graniczna atrybutu a
c_j	cecha obiektu
$\text{card}()$	moc zbioru
C	zbiór cech, zbiór atrybutów
\mathbb{C}_a	zbiór wartości granicznych atrybutu a
\mathcal{C}_k	klasyfikator
$\text{class}()$	funkcja zwracająca identyfikator klasy obiektu
$\text{CORE}()$	rdzeń zbioru
d	atrybut decyzyjny
D	zbiór atrybutów decyzyjnych
$\text{Dom}()$	dziedzina atrybutu

E, E^u, E^t	zbiór przykładów, podzbiory przykładów uczących i testowych
E^B	klasa równoważności
E_i	podzbiór zbioru przykładów
$H()$	entropia
\mathbb{H}	historia eksploatacji obiektu
I	liczba całkowita
$\text{IND}()$	relacja nierozróżnialności
k_a	liczba wartości jakościowych atrybutu a
l	lewy koniec przedziału wartości cechy
L	zbiór klas równoważności
$\text{nam}()$	nazwa (funkcja zwracająca nazwę swojego argumentu)
n	liczba przykładów
N	liczba całkowita
o	obiekt
O, O_k	zbiór obiektów, podzbiór obiektów
p	stwierdzenie
p^+	prawdopodobieństwo wystąpienia przykładu pozytywnego
p^-	prawdopodobieństwo wystąpienia przykładu negatywnego
P_g	współczynnik gotowości maszyny
r	prawy koniec przedziału wartości cechy
\mathfrak{R}	zbiór liczb rzeczywistych
R	redukt
$\text{RED}()$	rodzina reduktów systemu informacyjnego
S	klasa przykładów, klasa obiektów
\mathbb{S}	zbiór stanów diagnozowanego obiektu
\mathbb{S}_K	zbiór stanów związanych z konstrukcją maszyny
\mathbb{S}_P	zbiór stanów wywołanych procesem wytwarzania i/lub użytkowania maszyny
\mathcal{S}	system informacyjny

t	czas
u, u_i	przykłady
U, U_i	zbiór przykładów, podzbiór przykładów
$\text{val}()$	wartość (funkcja zwracająca wartość swojego argumentu)
v_i	wartość atrybutu
$V, V()$	zbiór wartości atrybutu
w	waga
W^X	liczba rozróżnialnych par elementów zbioru X
x, x_j	przykład
X	zbiór przykładów
\mathbb{X}	zbiór wejść obiektu
\mathbb{X}_U	zbiór wejść diagnozowanego obiektu związanych z jego działaniem użytecznym
\mathbb{X}_D	zbiór wejść obiektu określających warunki jego działania
\mathbb{X}_S	zbiór wejść interpretowanych jako szum (zakłócenia)
y	przykład
\mathbb{Y}	zbiór wyjść obiektu
\mathbb{Y}_U	zbiór wyjść związanych z użytecznym działaniem obiektu
\mathbb{Y}_R	zbiór wyjść związanych z występowaniem procesów resztkowych
\mathbb{Y}_N	zbiór wyjść nieznanych lub nieobserwowanych, będących skutkiem działania obiektu
β	poziom niesprzeczności tablicy decyzyjnej
δ	długość przedziału pomiędzy wartościami granicznymi
ϵ	błąd klasyfikacji
η	sprawność klasyfikatora
λ	liczba wartości atrybutu w danym przedziale konwersji
π_a	zbiór par wartości (przedziałów), określający przekształcenie cechy ilościowej a w cechę jakościową a^q
τ	czas bezwymiarowy
\leftarrow	symbol przypisania

Rozdział 1

Wstęp

Celem diagnostyki technicznej jest identyfikacja szeroko rozumianego stanu technicznego maszyn i urządzeń za pomocą odpowiednich metod i środków w celu osiągnięcia ich optymalnej trwałości, niezawodności i sprawności działania [13]. Stan maszyny lub urządzenia może być oceniany bezpośrednio na podstawie badań elementów danego obiektu i/lub badań współdziałania tych elementów lub pośrednio na podstawie obserwacji sygnałów związanych z działaniem badanego obiektu. Pośrednia ocena stanu maszyny może być realizowana między innymi przez tzw. *układy diagnozujące*.

Jeśli przyjąć, że stan obiektu oceniany będzie pośrednio, to można sformułować podstawowe zadanie diagnostyki maszyn jako *ocenę stanu technicznego* danej maszyny na podstawie zgromadzonych *symptomów diagnostycznych* oraz na podstawie *cech wejścia maszyny* (przede wszystkim jej warunków działania).

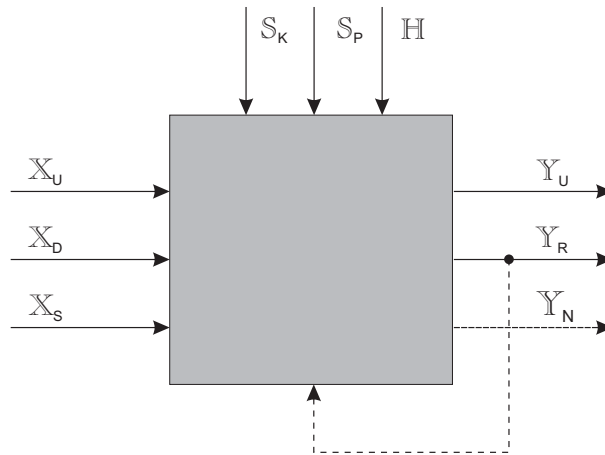
Aby przedyskutować sformułowane zadanie diagnostyki, rozważmy ogólny model diagnozowanego obiektu przedstawiony na rys. 1.1 [85]. Dla modelu tego wyróżniono zbiory wejść \mathbb{X} , stanów \mathbb{S} , uporządkowanego ciągu zdarzeń stanowiących historię eksploatacji obiektu \mathbb{H} i wyjść \mathbb{Y} :

$$\mathbb{X} = \mathbb{X}_U \cup \mathbb{X}_D \cup \mathbb{X}_S,$$

gdzie: \mathbb{X}_U — zbiór wejść związanych z użytecznym działaniem maszyny, \mathbb{X}_D — zbiór wejść określających warunki działania maszyny, \mathbb{X}_S — zbiór wejść interpretowanych jako szum,

$$\mathbb{S} = \mathbb{S}_K \cup \mathbb{S}_P,$$

gdzie: \mathbb{S}_K — zbiór stanów związanych z konstrukcją maszyny, \mathbb{S}_P — zbiór stanów



Rysunek 1.1: Model obiektu diagnozowania [85]

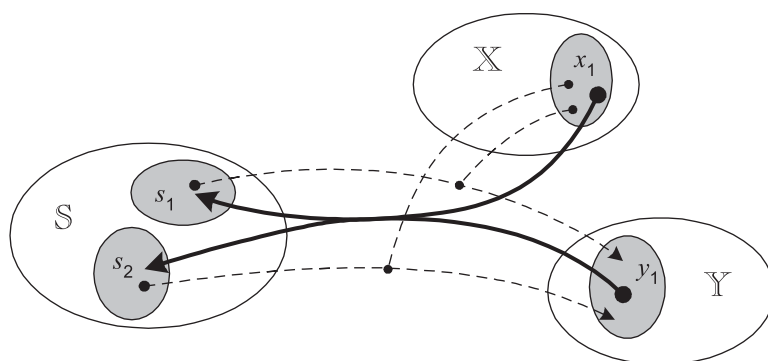
wywołanych procesem wytwarzania i/lub użytkowania maszyny,

$$\mathbb{Y} = \mathbb{Y}_U \cup \mathbb{Y}_R \cup \mathbb{Y}_N,$$

gdzie: \mathbb{Y}_U — zbiór wyjść związanych z użytecznym działaniem maszyny, \mathbb{Y}_R — zbiór wyjść związanych z występowaniem procesów resztkowych [13], \mathbb{Y}_N — zbiór wyjść nie znanych lub nie obserwowanych.

W praktycznych zastosowaniach wykorzystywany jest model uproszczony, w którym pomijane są niektóre z wymienionych podzbiorów. Dlatego w diagnostyce maszyn podstawowe znaczenie ma relacja przyczynowo-skutkowa *stan i warunki działania* \rightarrow *symptom* oraz relacja diagnostyczna *symptom i warunki działania* \rightarrow *stan* (por. rys. 1.2) [85]. Relacje te mogą być identyfikowane w ramach eksperymentów diagnostycznych. Z punktu widzenia celu i zadania diagnostyki maszyn istotna jest identyfikacja relacji odwrotnej, pozwalającej na ocenę stanu na podstawie analizy symptomów diagnostycznych (na rys. 1.2 przedstawiono ją linią grubą). Relacja ta jest relacją przybliżoną. Na podstawie wybranych warunków działania (np. zadana prędkość obrotowa wału maszyny) oraz na podstawie wybranych symptomów diagnostycznych (np. podwyższona wartość amplitudy składowej 1X drgań korpusu łożyska w kierunku promieniowym) możemy wnioskować o stanie maszyny jedynie w sposób przybliżony i zwykle niejednoznaczny (np. wirnik może być niewyrównoważony lub wał może być wygięty).

Wnioskowanie o relacjach między *stanem technicznym maszyny* a *symptomami diagnostycznymi* może być przeprowadzone, jeżeli [85]:



Rysunek 1.2: Relacje w diagnostyce maszyn [87]

- zostały zgromadzone odpowiednie *dane* (stanowiące *przesłanki*),
- znane są możliwe stany techniczne diagnozowanego obiektu,
- *układ diagnozujący* dysponuje odpowiednią *wiedzą*.

Obecnie dane często gromadzone są w sposób automatyczny. Możliwe stany techniczne badanego obiektu są dla tego obiektu swoiste i ich określenie zwykle nie przysparza problemów. Zgromadzona zaś wiedza w zasadniczy sposób determinuje jakość działania układu diagnozującego, zwłaszcza poprawność osiągniętych przez niego wniosków [85]. Wiedza ta może być gromadzona między innymi w bazach wiedzy diagnostycznych *systemów doradczych* [26] i zapisywana jest na przykład w postaci reguł określonych na podstawie zidentyfikowanych relacji diagnostycznych. Reguły takie wskazują możliwe przyczyny (np. uszkodzona panewka łożyska) obserwowanych skutków (np. wysoka amplituda drgań korpusu łożyska) [24]. Niestety, brak jest szczegółowej wiedzy diagnostycznej dotyczącej relacji diagnostycznych, dlatego też problem *pozyskiwania wiedzy* stał się jednym z głównych problemów diagnostyki [28].

Źródłami wiedzy z zakresu diagnostyki maszyn mogą być [85]:

- *specjaliści z zakresu danej dziedziny*:
 - *bezpośrednio* — poprzez osobisty udział w procesie pozyskiwania wiedzy,
 - *pośrednio* — specjaliści są autorami pozycji literatury fachowej, publikacji itp., wiedza ta musi jednak zostać odpowiednio zinterpretowana, np. przez tzw. inżyniera wiedzy,

- *bazy danych*:
 - zawierające *wyniki obserwacji diagnostycznych* badanego obiektu lub zbioru obiektów (zgromadzone w ramach eksperymentu biernego lub czynnego),
 - zawierające *wyniki obliczeń symulacyjnych* (zgromadzone w ramach tzw. eksperymentu numerycznego),
 - zawierające *przykłady uczące przygotowane przez specjalistów*.

Pozyskiwanie wiedzy od specjalistów jest mało efektywne. Często wiedza jest przekazywana w postaci opisów procedur postępowania (np. [136]). Próby ich bezpośredniego zastosowania w praktyce napotykają na poważne ograniczenia. Skutecznym sposobem pozyskiwania wiedzy od specjalistów może być odpowiednia organizacja procesu komunikacji ze specjalistami i umiejętne zapisywanie uzyskanych od nich informacji [18]. Działania takie podejmowane są m.in. w Katedrze Podstaw Konstrukcji Maszyn Politechniki Śląskiej (np. [85], [106], [135]). Istniejące ograniczenia powodują, iż poszukiwane są nowe źródła pozyskiwania wiedzy. Takim źródłem stały się *bazy danych diagnostycznych*.

Wyróżnić można dwie grupy metod pozyskiwania wiedzy z baz danych:

- *indukcyjne metody* tzw. *uczenia maszynowego* (ang. *machine learning*),
- *metody odkrywania zależności w bazach danych* (ang. *knowledge discovery*).

„Uczenie maszynowe” odbywa się na podstawie danych (wyników obserwacji) określanych mianem *przykładów*, które zostały uprzednio sklasyfikowane przez nauczyciela. Stąd metody te zwane są również metodami uczenia z *nauczycielem*.

Pozyskiwanie wiedzy metodami odkrywania zależności (nazywane też odkrywaniem wiedzy), w przeciwieństwie do metod indukcyjnych, nazywane jest uczeniem *bez nauczyciela*, gdyż przykłady nie są sklasyfikowane, lub też informacje o klasyfikacji nie są wykorzystywane. Zastosowanie tych metod ma na celu *identyfikację regularności* istniejących w analizowanym zbiorze danych. Regularność ta jest określana przez pewien *obraz* (ang. *pattern*) oraz *zakres*, w którym ten obraz występuje [140]. Przykładami obrazów są tablice kontyngencji, równania i równoważności logiczne. Badania związane z zastosowaniem tej grupy metod do pozyskiwania wiedzy diagnostycznej zostały zainicjowane przez W. Moczulskiego i J. Żytkowa [90].

Warto również zwrócić uwagę na zagadnienie *odwracania modeli diagnostycznych* [24] pozwalających na identyfikację relacji *symptom i warunki działania* \rightarrow *stan*. Podstawową trudność w stosowaniu modeli odwrotnych stanowi konieczność odwrócenia znanego modelu teoretycznego maszyny. Jednak, jak stwierdzono w [24, 57], technika pozyskiwania relacji diagnostycznych w postaci odwrotnych modeli obiektów ma szansę być zdecydowanie bardziej użyteczna, a tym samym perspektywiczna w diagnostyce maszyn.

W pracy autor skoncentrował się na problemach dotyczących pozyskiwania wiedzy metodami indukcyjnego „uczenia maszynowego”. Aby dane (wartości cech sygnałów diagnostycznych), gromadzone podczas eksperymentu diagnostycznego, mogły być zastosowane w procesie pozyskiwania wiedzy diagnostycznej na przykład metodami „uczenia maszynowego”, zwykle wskazane jest ich wstępne przetworzenie. Przetwarzanie to obejmuje najczęściej dwa etapy:

- *konwersję wartości cech sygnałów diagnostycznych,*
- *selekcję cech sygnałów diagnostycznych.*

Wartości cech sygnałów diagnostycznych są najczęściej wartościami ilościowymi. Konwersja pozwala zaś na przekształcenie ich dziedziny z ilościowej na jakościową, co powoduje redukcję (często znaczną) ilości informacji niesionej przez te wartości. Wspomniany wyżej etap przetwarzania wstępnego — *konwersja ilościowych wartości cech diagnostycznych na wartości jakościowe* lub inaczej *redukcja ilości informacji* — pozwala na:

- uogólnienie i uniezależnienie pozyskanej wiedzy od indywidualnych cech poszczególnych obiektów [85],
- poprawę skuteczności wnioskowania [109].

W pracy jako miary *skuteczności wnioskowania (klasyfikacji)* stosowane będą oceny błędów klasyfikacji. Jako skuteczne uznawane będzie takie wnioskowanie (z zastosowaniem pozyskanej wiedzy), które związane będzie z niskimi wartościami tych błędów.

Współczesne układy monitorujące pracę maszyny pozwalają na obserwację wielu sygnałów diagnostycznych jednocześnie. Dla każdego sygnału można z kolei wyznaczyć

wiele estymat (ocen) tego sygnału. Liczba cech opisujących badany obiekt, w szczególności złożony obiekt (np. turbozespół), może być bardzo duża. Nie wszystkie jednak cechy, z punktu widzenia wnioskowania diagnostycznego, są istotne. Ponadto, jak pokazano m.in. w [49], dla skończonego zbioru obiektów stosowanych do wyznaczenia danych źródłowych jakość pozyskanej wiedzy (w szczególności konstruowanego klasyfikatora) może pogarszać się po dołączeniu nieodpowiednich cech. Oznacza to, że wybór uwzględnianych cech jest zadaniem wymagającym szczególnej uwagi [22]. Zadanie to określane jest mianem *wyboru cech istotnych (relewantnych)* lub inaczej *selekcji informacji* [119].

Selekcja cech sygnałów diagnostycznych jest zadaniem, które realizowane jest co najmniej dwukrotnie. Pierwotny wybór cech sygnałów diagnostycznych następuje na początku badań. Obserwując i rejestrując sygnały diagnostyczne, specjalista na podstawie posiadanej wiedzy, określa zbiór wyznaczanych cech tych sygnałów. Wyznaczone wartości gromadzone są następnie w bazie danych. Przystępując do badań nad pozyskiwaniem relacji diagnostycznych, można mówić o wyborze cech dokonywanym spośród cech zgromadzonych w bazie danych. W tym przypadku zbiór wybieranych cech może być znany (np. na podstawie publikacji) lub wybór może być dokonywany heurystycznie. Wtórna selekcja cech dokonywana jest na początku procesu pozyskiwania wiedzy. Spośród pierwotnie określonego zbioru cech wybierane są te cechy, które pozwalają na identyfikację najbardziej przydatnych relacji diagnostycznych.

Analizując zagadnienia związane z pozyskiwaniem wiedzy, w szczególności z pozyskiwaniem wiedzy diagnostycznej, można zauważyć, że jednym z nierozwiązanych do końca zadań jest zadanie wstępnego przetwarzania danych. W celu praktycznego zastosowania metod pozyskiwania wiedzy z baz danych w dziedzinie diagnostyki maszyn pożądane byłoby zatem opracowanie metody pozwalającej na redukcję i selekcję informacji. Opracowanie takiej metody pozwoliłoby na bardziej skuteczne zastosowanie metod pozyskiwania wiedzy, a tym samym na bardziej skuteczne wnioskowanie o stanie technicznym maszyny na podstawie znajomości symptomów diagnostycznych, warunków działania oraz relacji diagnostycznych. Praca opisuje badania obejmujące analizę i porównanie wybranych metod wyznaczania wartości granicznych atrybutów ilościowych oraz wybranych metod selekcji atrybutów. Zaprezentowano zarówno metody zaczerpnięte z literatury jak również metody autora. Ponadto przedstawiono własną metodę redukcji i selekcji przydatną w diagnostyce maszyn.

Rozprawa została skomponowana w opisany w dalszym ciągu sposób.

Rozdział 1. stanowi krótkie wprowadzenie do problematyki pracy. Najważniejsze zdaniem autora, zagadnienia związane z tematem pracy, a wymagające dokładniejszego omówienia, opisano szerzej w rozdziale 2.

W rozdziale 3., na podstawie wcześniejszej analizy rozpatrywanego problemu sformułowano cel, tezę i zakres pracy.

Główna treść pracy została zawarta w czterech rozdziałach (rozdziały 4, 5, 6 i 7). W rozdziałach tych skupiono się na opisie wybranych metod redukcji i selekcji informacji, sposobie oceny tych metod, a także opisie badań weryfikacyjnych i uzyskanych wyników.

W rozdziale 4. przedstawiono opis wybranych metod wyznaczania wartości granicznych, a w rozdziale 5. — opis wybranych metod selekcji atrybutów. W obydwu rozdziałach oprócz podstaw teoretycznych przedstawiono szczegółowe algorytmy postępowania. W rozdziałach tych znalazły się między innymi opisy metod zaproponowanych przez autora.

W rozdziale 6. pokazano kryteria oraz sposób oceny metod redukcji i selekcji informacji. W szczególności stwierdzono, że ocena tych metod dokonywana będzie pośrednio na podstawie oceny pozyskanej wiedzy.

Opis badań weryfikacyjnych zamieszczono w rozdziale 7. Badania weryfikacyjne przeprowadzono, biorąc pod uwagę wyniki dwóch eksperymentów diagnostycznych: eksperymentu czynnego oraz eksperymentu numerycznego. Rozdział ten, oprócz analizy uzyskanych wyników, opisuje również proponowaną przez autora metodę redukcji i selekcji informacji w diagnostyce maszyn.

Rozdział 8. zawiera podsumowanie przeprowadzonych badań. W rozdziale tym zamieszczono wnioski i uwagi autora.

Rozdział 2

Problem badawczy

Przeprowadzone badania dotyczą zagadnień związanych z pozyskiwaniem wiedzy indukcyjnymi metodami „uczenia maszynowego”. Niektóre zagadnienia wymagają nieco dokładniejszego omówienia. Poniżej przedstawiono niektóre — zdaniem autora — ważniejsze problemy związane z tematem rozprawy.

Omówiono między innymi ogólne zagadnienia pozyskiwania wiedzy, reprezentacji danych i wiedzy, redukcji i selekcji informacji oraz, krótko, zagadnienia klasyfikacji i weryfikacji klasyfikatora.

2.1 Pozyskiwanie wiedzy diagnostycznej

Proces pozyskiwania wiedzy uznaje się za najtrudniejszy etap tworzenia nietrywialnych systemów „inteligentnych” [130]. W procesie tym można wyróżnić cztery główne etapy [85]:

- *identyfikacja źródeł wiedzy,*
- *wybór środków reprezentacji wiedzy,*
- *właściwe pozyskiwanie wiedzy,*
- *walidacja wiedzy.*

Możliwe do wykorzystania źródła wiedzy diagnostycznej zostały omówione wcześniej. W pracy autor skupił się nad pozyskiwaniem wiedzy z baz danych. W diagnostyce

maszyn gromadzi się wiele danych dotyczących parametrów wejść i stanu oraz warunków działania badanych obiektów. Sposób obserwacji sygnałów diagnostycznych powoduje, że w wartościach cech tych sygnałów oraz cech parametrów „procesowych” można zidentyfikować wpływ szumu [85].

Zadanie pozyskiwania wiedzy diagnostycznej może polegać na zidentyfikowaniu tej wiedzy na podstawie zgromadzonych danych. Jedną z grup metod pozyskiwania wiedzy na podstawie danych, wspomnianą wcześniej, jest grupa indukcyjnych metod „uczenia maszynowego”.

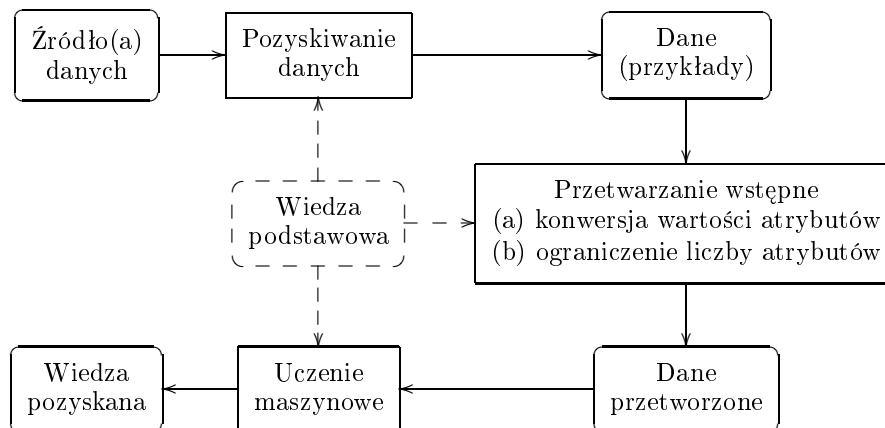
Do pozyskiwania wiedzy diagnostycznej indukcyjnymi metodami „uczenia maszynowego” stosowany może być tzw. *model atrybutowy* [85], zaś spośród metod „uczenia maszynowego” możliwe do zastosowania w badaniach (z uwagi na dostępność oprogramowania) są:

- indukcyjna metoda pozyskiwania wiedzy *za pomocą pokryć* według algorytmu A^Q [78],
- metoda generowania *drzew decyzyjnych* w sposób indukcyjny [107],
- metoda oparta na teorii *zbiorów przybliżonych* [103].

2.2 Pozyskiwanie wiedzy metodami „uczenia maszynowego”

Proces pozyskiwania wiedzy metodą „uczenia maszynowego” przedstawiany jest przez różnych autorów w różnym stopniu szczegółowości. Zdaniem autora proces ten należy przedstawiać jak na rys. 2.1.

Dane pozyskane mogą być z jednego obiektu lub kilku obiektów tej samej klasy. Aby jednak mogły być zastosowane w procesie „uczenia maszynowego”, powinny zostać przetworzone wstępnie. W dziedzinie diagnostyki technicznej gromadzone są najczęściej cechy diagnostyczne o wartościach ilościowych. Dla uogólnienia pozyskiwanej wiedzy i uniezależnienia jej od indywidualnych cech badanych obiektów konieczna jest konwersja wartości ilościowych na jakościowe (patrz: punkt 2.6). Rejestrowanych może być również bardzo wiele sygnałów, dla których wyznaczyć można wiele cech tych sygnałów. Nie wszystkie rejestrowane sygnały oraz nie wszystkie ich cechy niosą informację diagnostyczną. Wymagany jest zatem kolejny etap przetwarzania wstępnego — selekcja,



Rysunek 2.1: Pozyskiwanie wiedzy metodą „uczenia maszynowego”

czyli wybór istotnych, ze względu na zadanie diagnostyczne, cech sygnałów diagnostycznych. Dopiero tak przetworzone dane stanowią podstawę procesu pozyskiwania wiedzy metodami „uczenia maszynowego”.

W procesie „uczenia maszynowego” wyróżnia się także tzw. wiedzę podstawową. Według [69] obejmuje ona m.in. informacje o języku stosowanym do opisu danych i wiedzy. Do wiedzy podstawowej zaliczyć również należy informacje o klasyfikacji przykładów uczących. Zdaniem autora, wynika z tego również, że wiedza ta powinna obejmować także informacje o obiekcie badań, jego możliwych stanach oraz cechach stosowanych do opisu tych stanów. Również wiedzę dotyczącą pozyskiwania danych, zagadnień wstępnego ich przetwarzania oraz wiedzę na temat metod „uczenia maszynowego” autor zalicza do wiedzy podstawowej.

Przy omawianiu procesu pozyskiwania wiedzy należałoby zwrócić uwagę na fakt, iż pozyskana wiedza może stanowić podstawę do podjęcia działań zmierzających do modyfikacji źródeł danych. Modyfikacja może polegać np. na zmianie (zmniejszeniu lub zwiększeniu liczby, zmianie lokalizacji) punktów pomiarowych, ale może także polegać na zmianie konstrukcji badanego obiektu.

Pozyskiwanie wiedzy metodami „uczenia maszynowego” może odbywać się według różnych algorytmów. Ogólnie algorytmy te dzieli się na dwie grupy [69]:

- metody działające na zasadzie „czarnej skrzynki” (np. sieci neuronalne, metody statystyczne),
- metody „zorientowane na wiedzę”.

Metody pozyskiwania wiedzy działające na zasadzie „czarnej skrzynki”, choć mające ogromne znaczenie (również w diagnostyce maszyn), cechuje zasadnicza wada — pozyskana tymi metodami wiedza jest trudna do zinterpretowania.

W pracy autor skupił się na metodach, które pozwalają użytkownikowi na zrozumienie pozyskanej wiedzy, a w efekcie m.in. również jej częściowe aktualizowanie (metody spełniające tzw. silne kryterium¹ według [82]) bez konieczności aktualizacji całej bazy wiedzy, jak to ma miejsce w wypadku metod działających na zasadzie „czarnej skrzynki”.

W ramach grupy metod „zorientowanych na wiedzę” stosowane są między innymi metody wymienione w punkcie 2.1.

Przykłady zastosowań omawianych metod w diagnostyce technicznej można znaleźć w wielu pracach, m.in. [101, 31, 85, 33, 75, 37, 61]. Wyniki (uwzględniające skuteczność klasyfikacji) uzyskane przy stosowaniu tych metod do celów pozyskiwania wiedzy diagnostycznej są zbliżone (np. [37]). Tak więc można przypuszczać, że każdą z tych metod można z powodzeniem stosować do tworzenia np. baz wiedzy diagnostycznych systemów doradczych.

Ważne jest to, by algorytmy „uczenia maszynowego” potrafiły działać na niedoskonałych danych [69]. Zakłada się, że nie tylko dane uczące mogą być niedoskonałe, lecz również wiedza podstawowa nie musi koniecznie być doskonała.

2.3 Niedoskonałość danych

Zwykle przyjmuje się [16], że dane w diagnostyce maszyn są wartościami cech sygnałów, za pomocą których obserwowane są własności i właściwości obiektu badań. Liczba obserwowanych sygnałów uwarunkowana jest celem postępowania diagnostycznego i zależy m.in. od złożoności badanego obiektu. Jednak efektywność wnioskowania diagnostycznego zależy przede wszystkim od *wrażliwości diagnostycznej*² mierzonych sygnałów, a nie od ich liczby [91]. Dlatego istotna jest *wrażliwość* dostępnych sygnałów diagnostycznych na niesprawności badanej maszyny objętej nadzorem.

Dane gromadzone podczas obserwacji diagnostycznej maszyn zwykle są niedoskon-

¹Metody pozwalające na stosowanie nowych przykładów do poprawy jakości bazy wiedzy (słabe kryterium) oraz pozwalające na dokonywanie modyfikacji wiedzy w określonej, symbolicznej formie.

²Sygnał uznawany jest za *wrażliwy* diagnostycznie, jeżeli jego cecha (cechy) wykazuje znaczną zależność wartości od występowania określonego stanu diagnozowanego obiektu [64].

ale. Dzieje się tak dlatego, ponieważ mogą one być [47]:

- *niekompletne*,
- *niepewne*,
- *niedokładne*,
- *błędne*.

Niekompletność danych może być spowodowana m.in. nieuwzględnieniem cech istotnych ze względu na ocenę stanu maszyny, brakiem wartości niektórych cech lub brakiem przykładów, opisujących w dostateczny sposób określony stan lub stany maszyny [47].

Niepewność danych związana jest z wpływem szumów, pojawiających się na wejściach i wyjściach maszyny, na wyniki obserwacji [47].

Niedokładność danych zależy od własności układów pomiarowych oraz technik określania wartości cech [47].

Błędne dane pojawiają się wówczas, kiedy albo jedna wartość danej cechy jest niepoprawna, albo dane mogą być *sprzeczne* (np. z powodu rozbieżnych opinii specjalistów) [47].

2.4 Cechy ilościowe i jakościowe

Cechy są elementami zbioru, który stanowi orzeczenie o przedmiocie poznania lub tworzenia [39]. Są to elementy odróżniające lub charakteryzujące pod jakimś względem istoty lub przedmioty, ich czynności i stany w otaczającej rzeczywistości [118]. Innymi słowy, dowolny abstrakt lub konkret może być opisany za pomocą zbioru cech $C = \{c_j\}$ gdzie [24]:

$$c_j = \langle \text{nam}(c_j), \text{val}(c_j) \rangle, \quad (2.1)$$

przy czym $\text{nam}()$ określa nazwę cechy, a $\text{val}()$ jej wartość. Zbiór wszystkich możliwych cech opisujących dany obiekt (lub klasę obiektów) nazywany będzie *przestrzenią cech*.

Zadanie redukcji i selekcji informacji wymaga modyfikacji pojęcia cechy. W pracy przyjęto³, że cecha reprezentowana będzie w postaci:

$$c_j = \langle a_j, \text{val}(a_j) \rangle. \quad (2.2)$$

³Sugestia W. Moczulskiego.

Przyjęto, że atrybut $a_j \in A$ jest funkcją, której wartości $a_j(o) \in V$ opisują obiekt o należący do pewnej klasy (zbioru) obiektów, natomiast $\text{val}(a_j)$ oznacza wartość tego atrybutu. Zbiór wszystkich możliwych wartości danego atrybutu $a_j(o)$ nazywany jest przestrzenią wartości atrybutu lub jego dziedziną $\text{Dom}(a_j)$:

$$\text{Dom}(a_j) = \{\text{val}(a_j)\}. \quad (2.3)$$

Ze względu na dziedzinę atrybutu, rozróżnia się:

- atrybuty ilościowe,
- atrybuty jakościowe,

a tym samym *cechy ilościowe* i *cechy jakościowe*.

Wartości atrybutów ilościowych stanowią zwykle bezpośredni wynik obliczenia lub pomiaru wielkości fizycznych.

Wartości atrybutów jakościowych mogą być określane w wyniku *kodowania symbolicznego* lub uzyskane w wyniku *konwersji*⁴ atrybutów ilościowych na jakościowe [119, 15, 40]. Najczęściej stosuje się kodowanie w postaci liczb całkowitych nieujemnych lub w postaci termów. Ponadto w praktyce dąży się do tego, by moc $\text{card}(V(a))$ zbioru wartości atrybutów była niewielka. Wówczas wartości atrybutów mogą być zastąpione wartościami lingwistycznymi (np. mały, średni, wysoki), co powoduje, że wiedza pozyskana na podstawie takich danych jest bardziej zrozumiała.

2.5 Reprezentacja danych i wiedzy

Termin *reprezentacja* jest przyjmowany dość dowolnie przez różnych autorów. W pracy terminem tym określane będzie ogólny formalizm zapisywania i gromadzenia dowolnego zasobu danych (reprezentacja danych) oraz wiedzy (reprezentacja wiedzy) [26]. Według [94] przyjęcie właściwej reprezentacji stanowi klucz do rozwiązania trudnych problemów.

2.5.1 Reprezentacja danych

Wyniki obserwacji diagnostycznych gromadzone są w bazach danych. Do podstawowych struktur baz danych zalicza się:

⁴W literaturze stosowane jest często określenie *dyskretyzacja*.

- strukturę relacyjną,
- strukturę sieciową, której szczególnym przypadkiem jest struktura hierarchiczna,
- strukturę obiektową.

Ciekawym przykładem definicji relacyjnej bazy danych jest baza danych projektu *MIMOSA* (Machinery Information Management Open System Alliance) [83]. Celem projektu jest opracowanie otwartego protokołu wymiany danych związanych z monitorowaniem stanu obiektów technicznych [20].

Dla rozpatrywanej klasy zagadnień bardzo ważne jest pojęcie *stwierdzenia* [26]. Stwierdzenie jest zdaniem, które stanowi orzeczenie o wartości $\text{val}(c)$ cechy o nazwie $\text{nam}(c)$ danego obiektu o i zapisywane jest za pomocą uporządkowanej trójki [19]:

$$p = \langle \text{nam}(o), \text{nam}(c), \text{val}(c) \rangle, \quad (2.4)$$

lub przy uwzględnieniu (2.2):

$$p = \langle \text{nam}(o), a, \text{val}(a) \rangle. \quad (2.5)$$

Taki sposób zapisu danych zastosowano w niniejszej pracy.

Stwierdzeniu może zostać przypisany *stopień prawdziwości* lub *stopień przekonania* $\text{bel}()$:

$$p = \langle \text{nam}(o), a, \text{val}(a), \text{bel}(p) \rangle. \quad (2.6)$$

Stwierdzenie takie określa się mianem *stwierdzenia przybliżonego* [26].

Stwierdzenia umożliwiają między innymi zapis wartości różnych cech obiektu i nie wymagają modyfikacji ze względu na zwiększenie stopnia szczegółowości opisu (np. poprzez wprowadzenie nowych cech) [85].

Jak wspomniano wcześniej, w zagadnieniach praktycznych dotyczących pozyskiwania wiedzy z zakresu diagnostyki maszyn, do reprezentacji danych stosowany jest zwykle *model atrybutowy*. W modelu tym zbiór danych (najczęściej ilościowych) reprezentowany jest w postaci macierzy [88]:

$$E = \begin{pmatrix} \text{val}(a_{1,1}) & \dots & \text{val}(a_{1,m}) & \text{val}(d_{11}) & \dots & \text{val}(d_{1n}) \\ \text{val}(a_{2,1}) & \dots & \text{val}(a_{2,m}) & \text{val}(d_{21}) & \dots & \text{val}(d_{2n}) \\ \dots & & & & & \\ \text{val}(a_{N,1}) & \dots & \text{val}(a_{Nm}) & \text{val}(d_{N1}) & \dots & \text{val}(d_{Nn}) \end{pmatrix}, \quad (2.7)$$

gdzie wartości $\text{val}(a_{ij}), i = 1, \dots, N \wedge j = 1, \dots, m$ oznaczają wartości m atrybutów nazywanych *atrybutami warunkowymi*, natomiast wartości $\text{val}(d_{ij}), i = 1, \dots, N \wedge j = 1, \dots, n$ są wartościami n *atrybutów decyzyjnych*. Dla $n = 1$ macierz (2.7) jest określana mianem *tablicy decyzyjnej*⁵ [104]. Jak wynika z powyższego, dane (przykłady) zapisane w macierzy (2.7) są sklasyfikowane. Klasyfikacji tej może dokonać np. specjalista z danej dziedziny, a wartości atrybutów decyzyjnych d_{ij} są w diagnostyce technicznej najczęściej identyfikatorami stanu badanego obiektu.

Jeżeli macierz (2.7) zawiera przykłady opisujące obiekt o , wówczas zapisane w niej stwierdzenia dotyczące np. atrybutów warunkowych można odczytywać w sposób następujący:

$$p_i = \langle \text{nam}(o), a_{ij}, \text{val}(a_{ij}) \rangle \quad i = 1, \dots, N \wedge j = 1, \dots, m.$$

Z tablicą decyzyjną związane jest pojęcie *poziomu niesprzeczności* (ang. *consistency level*) tablicy, określonego jako [104]:

$$\beta = \frac{n_{d\pm}}{N}, \quad (2.8)$$

gdzie $n_{d\pm}$ oznacza liczbę przykładów o wartościach atrybutów warunkowych różniących się między sobą co najwyżej o zadaną dokładność ε , a jednocześnie mających różne wartości atrybutu decyzyjnego, N jest liczbą wszystkich przykładów tablicy decyzyjnej.

2.5.2 Reprezentacja wiedzy

Spośród środków reprezentacji wiedzy deklaratywnej (do której pozyskiwania ograniczono się w pracy) należy wymienić [18]:

- tablice decyzyjne reguł,
- stwierdzenia,
- reguły,
- drzewa decyzyjne,
- sieci semantyczne i sieci stwierdzeń,

⁵Patrz przypis 1 na stronie 151.

- obiekty abstrakcyjne (np. ramy),
- hipertekst.
- sieci neuronalne.

W większości zastosowań wymienione środki reprezentacji wiedzy nie występują w izolacji od pozostałych; stosowane są techniki mieszane [18]. W pracy, pozyskiwana wiedza reprezentowana będzie za pomocą:

- drzew decyzyjnych,
- reguł.

Stwierdzenia, które omówiono w punkcie 2.5.1, mogą być środkiem reprezentacji zarówno danych jak i wiedzy.

Zbiory stwierdzeń nie wystarczają do zapisania wiedzy o określonej dziedzinie. Wiedza diagnostyczna jest zapisywana często w postaci reguł [21]:

Jeżeli przesłanka to konkluzja,

gdzie przesłanka jest wyrażeniem złożonym z prostych zdań logicznych połączonych funktorami **i**, **lub**. Stosowanie funktora **lub** nie jest zalecane ze względu na wymagania stawiane przez układy wnioskujące [21].

Drzewo decyzyjne to struktura, która jest [108]:

- albo *liściem* wskazującym klasę,
- albo *węzłem decyzyjnym* określającym pewien test do przeprowadzenia na wartości atrybutu przypisanego do tego węzła; z węzła wychodzi tyle krawędzi, ile jest możliwych wyników wspomnianego testu, a każda z krawędzi jest zakończona poddrzewem [85].

2.6 Konwersja wartości atrybutów

Niech a będzie atrybutem ilościowym, którego dziedzina $\text{Dom}(a) = [l, r] \subset \mathfrak{R}$. Zbiór par:

$$\pi_a = \{[v_0, v_1), [v_1, v_2), \dots, [v_{k-1}, v_k]\}, \quad (2.9)$$

gdzie:

$$v_0 = l, \quad v_{i-1} < v_i, \quad v_k = r, \quad i = 1, 2, \dots, k, \quad (2.10)$$

określa liniowe przekształcenie (konwersję) dziedziny atrybutu ilościowego a w dziedzinę atrybutu jakościowego a^q :

$$f : \text{Dom}(a) \rightarrow \text{Dom}(a^q). \quad (2.11)$$

Przekształcenie to jest zdefiniowane jako [15]:

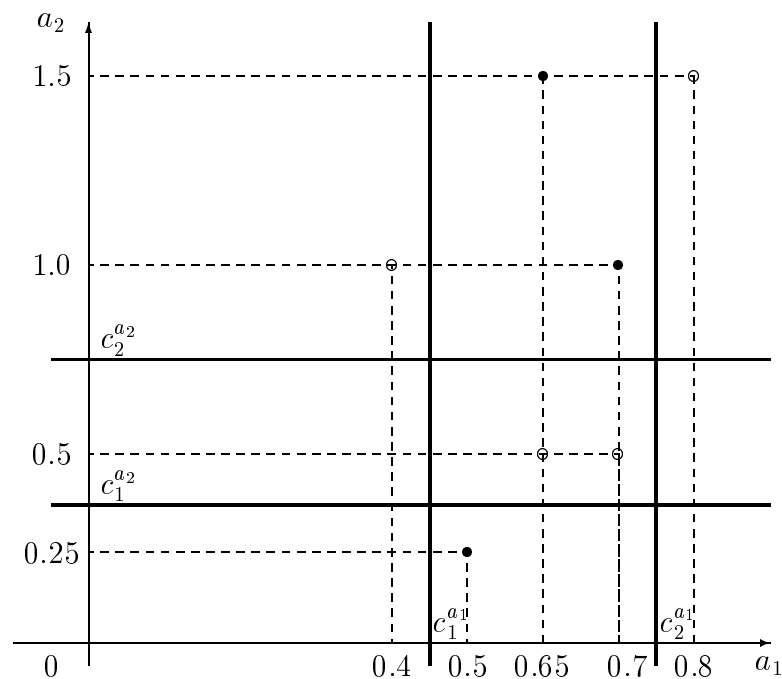
$$f(\nu) = \begin{cases} j, & \text{gdy } v_j \leq \nu < v_{j+1}, \quad \text{dla } j = 0, 1, \dots, k-2, \\ k-1, & \text{gdy } v_{k-1} \leq \nu \leq v_k \end{cases} \quad (2.12)$$

a wartości v_j , ($j = 1, \dots, k$) nazywane są *wartościami granicznymi* atrybutu a i oznaczane c^a . Zbiór wszystkich wartości granicznych atrybutu a oznaczany będzie jako \mathbb{C}_a . Interpretację geometryczną wartości granicznych pokazano na rys. 2.2. Przedstawiono zbiór przykładów należących do dwóch klas. Przykłady opisane są przez wartości dwóch atrybutów a_1 i a_2 . Dla danego zbioru przykładów wyznaczono wartości graniczne $c_1^{a_1}, c_2^{a_1}$ atrybutu a_1 , oraz $c_1^{a_2}, c_2^{a_2}$ atrybutu a_2 . Wartości te pozwalają na przekształcenie dziedziny atrybutów w taki sposób, że zachowany zostaje poziom niesprzeczności tablicy decyzyjnej zawierającej omawiane przykłady.

W zależności od kryterium podziału, rozróżnia się metody wyznaczania wartości granicznych atrybutów ilościowych:

- *jednoatrybutowe* i *wieloatrybutowe* [15],
- *bez nauczyciela* i *z nauczycielem* [40],
- *z zadawaną liczbą wartości granicznych* i *z doбором liczby wartości granicznych na podstawie danych* [58].

Metoda *jednoatrybutowa* polega na wyznaczaniu wartości granicznych dla każdego atrybutu oddzielnie, podczas gdy metoda *wieloatrybutowa* pozwala na wyznaczenie wartości granicznych jednocześnie dla wszystkich atrybutów. W literaturze stosowane są również pojęcia metod lokalnych i globalnych (między innymi [15, 40, 109]), choć



Rysunek 2.2: Geometryczna interpretacja wartości granicznych [97]

zdaniem autora nie do końca są one adekwatne i powinny być rozumiane nieco inaczej (np. [30, 95]).

Metoda wyznaczania wartości granicznych *bez nauczyciela* (czasami określana jako *bez nadzoru*) polega na przekształceniu dziedziny atrybutu (atrybutów) bez wykorzystania informacji o klasyfikacji zgromadzonych przykładów (danych), w przeciwieństwie do metody *z nauczycielem* (*z nadzorem*), gdzie informację o klasyfikacji przykładów bierze się pod uwagę podczas wyznaczania wartości granicznych [40].

Metody *z zadawaną liczbą wartości granicznych*⁶ wymagają określenia liczby poszukiwanych wartości granicznych lub przynajmniej maksymalnej jej wartości. Niekiedy wartość ta musi być zadawana dla każdego atrybutu oddzielnie [40]. Metody *z doбором liczby wartości granicznych na podstawie danych*⁷ pozwalają zaś na poszukiwanie tej liczby równocześnie dla wszystkich atrybutów, przy zastosowaniu np. kryterium zachowania minimalnego poziomu niesprzeczności rozpatrywanej tablicy decyzyjnej [58].

Teoretycznym aspektem konwersji wartości atrybutów poświęcono wiele prac, m.in.

⁶W literaturze stosowane jest zwykle określenie *statyczne*.

⁷W literaturze stosowane jest określenie *dynamiczne*.

[119, 15, 40, 96, 109, 95, 48, 29, 97].

Praktyczne zastosowanie konwersji wartości atrybutów w zagadnieniach diagnostyki maszyn było przedmiotem stosunkowo niewielu badań (m.in. [14, 22, 100, 75, 63, 86]). Było ono natomiast przedmiotem badań autora ([31, 32, 33, 37, 34]). Stosowane najczęściej do tej pory metody wyznaczania wartości granicznych bardzo często prowadziły do niezadowolających wyników. Wykorzystanie zaleceń zawartych w normach (np. ISO 2372, API 617), choć w pewnych przypadkach jest możliwe (np. [125]), często, z powodu zbytniego uogólnienia, prowadzi do uzyskania niezadowolających rezultatów w diagnozowaniu stanów maszyn [12].

2.7 Ograniczenie liczności zbioru atrybutów

Wyznaczany podczas badań zbiór cech sygnałów (a tym samym zbiór atrybutów) diagnostycznych może być zbiorem bardzo licznym. Istnieje trudne do rozwiązania zadanie ograniczenia liczności tego zbioru. Podejmując to zadanie, należy brać pod uwagę fakt, że uwzględniane są wyłącznie związki pomiędzy rozpatrywanymi (będącymi wynikiem badań) wartościami atrybutów [16].

Ograniczanie liczności zbioru uwzględnianych atrybutów może być dokonywane ze względu na kryteria obejmujące między innymi następujące czynniki [22]:

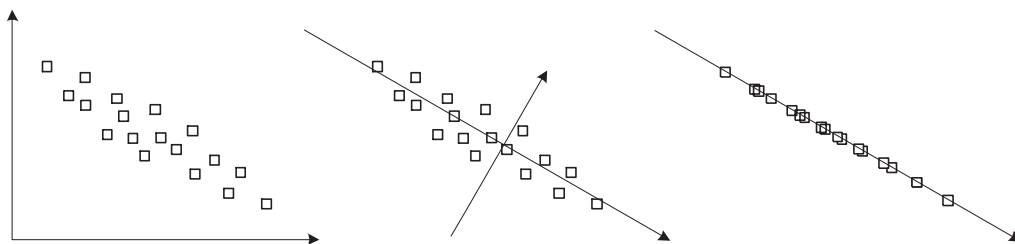
- informację stanowiącą podstawę wyboru:
 - wybór atrybutów dokonywany jest bez uwzględniania informacji o klasyfikacji przykładów (wybór bez nauczyciela),
 - wybór atrybutów dokonywany jest na podstawie analizy klasyfikacji przykładów (wybór z nauczycielem),
- rodzaj wybieranych atrybutów:
 - wybierane są wyłącznie atrybuty będące elementami rozpatrywanego zbioru atrybutów,
 - poszukiwany jest zbiór nowych atrybutów, wyznaczanych na podstawie transformacji atrybutów wyjściowych (por. np. [81, 134]),
- zależności między atrybutami:

- atrybuty wyjściowe są nieskorelowane i wybierane są te atrybuty, które charakteryzują się największą wariancją,
- atrybuty wyjściowe są skorelowane i wybór atrybutów poprzedzany jest identyfikacją zbioru atrybutów niezależnych (poszukiwanie ortogonalnej bazy przestrzeni wartości atrybutów), w którym odrzuca się następnie atrybuty charakteryzujące się najmniejszą wariancją.

Teoretycznym zagadnieniom ograniczania liczności zbioru atrybutów poświęcono wiele prac (m.in. [119, 116, 52, 5, 139, 70, 80, 68, 62, 126, 51, 76, 127]).

Natomiast w dziedzinie diagnostyki maszyn prace w całości poświęcone zagadnieniom wyboru atrybutów to przede wszystkim: [10, 16, 1, 27, 122, 64]. Zagadnienia te rozpatrywane były również w pracach: [54, 100, 101, 23, 56, 85, 45, 63, 65] oraz w pracach autora: [31, 32, 33, 37].

Ciekawym przykładem rozwiązania zadania ograniczenia liczności zbioru rozpatrywanych atrybutów diagnostycznych jest metoda, zaproponowana przez W. Cholewę [22]. Polega ona na wyznaczeniu osi głównych przestrzeni wartości atrybutów, a następnie określeniu ograniczonej przestrzeni wartości atrybutów głównych (por. rys. 2.3).



Rysunek 2.3: Przykład ograniczenia liczności zbioru uwzględnianych atrybutów [22]

W niniejszej pracy ograniczono się do zagadnienia wyboru atrybutów diagnostycznych spośród atrybutów istniejących. Nie rozpatrywano zagadnień związanych ze wspomnianym wyżej przykładem, jak również nie rozpatrywano zagadnień związanych z tak zwaną *konstruktywną indukcją* [81], pozwalającą na wyznaczanie nowych atrybutów na podstawie transformacji atrybutów wyjściowych.

2.8 Klasyfikator stanu

Klasyfikacją nazywa się przyporządkowanie obiektu do jednej klasy spośród rozpatrywanego zbioru klas [22]. Przyporządkowanie to odbywa się na podstawie cech przysługujących klasyfikowanemu obiektowi.

Zakłada się [22], że istnieje wzorcowy podział populacji maszyn O na K klas $O_k \subset O$, oraz że:

$$O = \bigcup_{k=1, \dots, K} O_k. \quad (2.13)$$

Podział populacji O jest określony za pomocą rodziny nieznanych *klasyfikatorów* [22]:

$$\mathcal{C}_k : O \rightarrow [0, 1]; k = 1, \dots, K, \quad (2.14)$$

gdzie:

$$\begin{aligned} \mathcal{C}_k(o) = 1 &\Leftrightarrow o \in O_k, \\ 0 < \mathcal{C}_k(o) < 1 &\Leftrightarrow o \tilde{\in} O_k, \\ \mathcal{C}_k(o) = 0 &\Leftrightarrow o \notin O_k, \end{aligned} \quad (2.15)$$

a symbol $\tilde{\in}$ oznacza, że obiekt częściowo należy do danego zbioru.

Należy podkreślić [85], że w ogólnym przypadku nie można zakładać, że klasy $O_k, k = 1, \dots, K$ są rozłączne. Ponadto należy zwrócić uwagę na możliwość wystąpienia zjawiska *synergii*, które w wypadku uszkodzeń oznacza, że łączny skutek występowania dwóch (skorelowanych) przyczyn może się istotnie różnić od sumy skutków tych przyczyn.

Klasyfikator (2.14) stosowany jest następnie do klasyfikacji nowych danych — przykładów, opisujących własności i właściwości badanych obiektów.

2.9 Weryfikacja klasyfikatora

Klasyfikator \mathcal{C}_k wyznaczany jest na podstawie skończonego zbioru danych. Jedną z metod weryfikacji tego klasyfikatora jest metoda polegająca na zastosowaniu zbioru przykładów zarówno w procesie tworzenia klasyfikatora, jak i w procesie jego weryfikacji.

Uporządkowany zbiór przykładów (2.7) dzielony jest na dwa podzbiory:

$$E = E^u \cup E^t, \quad (2.16)$$

gdzie: podzbiór E^u nazywany jest *zbiorem przykładów uczących (zbiorem uczącym)*, natomiast podzbiór E^t — *zbiorem przykładów testowych (zbiorem testowym)*.

Ocena merytoryczna wiedzy (w tym wypadku klasyfikatora), pozyskanej na podstawie zbioru przykładów uczących E^u , dokonywana jest poprzez określenie wartości estymatora *sprawności klasyfikatora*, w odniesieniu do odpowiedniego zbioru przykładów testowych E^t [85]. Zakłada się przy tym, że rozkład przykładów jest losowy i reprezentatywny dla oceny klasyfikatora.

Miarę *sprawności klasyfikatora* określa się jako:

$$\eta = 1 - \frac{n_{err}}{\text{card}(E^t)}, \quad (2.17)$$

gdzie n_{err} oznacza liczbę przykładów błędnie sklasyfikowanych.

W wypadku, gdy proces weryfikacji jest procesem iteracyjnym (liczba kroków $N > 1$), ocena sprawności uzyskiwana jest jako wartość średnia:

$$\bar{\eta} = \frac{1}{N} \sum_{i=1}^N \eta_i, \quad (2.18)$$

gdzie wartość η_i obliczana jest dla każdego kroku według 2.17.

2.10 Podsumowanie dotyczące problemu badawczego

Przeprowadzona analiza problemu na podstawie badań literaturowych pokazuje, iż obecnie rozwój metod pozyskiwania wiedzy diagnostycznej następuje silnie w kierunku wykorzystania do tego celu baz danych. Dane pochodzące z obserwacji obiektów diagnozowanych gromadzone są w coraz większej liczbie. Obserwacji podlega nie tylko coraz więcej obiektów, lecz również sygnałów, co umożliwił między innymi rozwój układów pomiarowych. W efekcie, diagnostyczne bazy danych zawierają bardzo liczne zbiory danych, w których zarówno liczba rekordów (przykładów) jak również liczba pól (atrybutów) jest znaczna. Możliwość i celowość stosowania metod pozyskiwania wiedzy z diagnostycznych baz danych nie budzi już wątpliwości, aczkolwiek liczba przeprowadzonych badań z tego zakresu w dziedzinie diagnostyki maszyn nie jest zbyt wielka.

Z badań literaturowych, jak również badań własnych autora wynika, że szczególne trudności w zastosowaniu metod uczenia maszynowego w dziedzinie diagnostyki maszyn związane są z dwoma etapami procesu pozyskiwania wiedzy:

- *etapem wyboru atrybutów relewantnych*, które zostaną zastosowane w procesie uczenia maszynowego, a dalej — będą służyły do zapisu pozyskanej wiedzy,
- *etapem wyznaczania wartości granicznych atrybutów*, pozwalającym na zastosowanie atrybutów jakościowych, co ma wpływ między innymi na uogólnienie pozyskanej wiedzy oraz na uzyskiwaną jej jakość (sprawność klasyfikatora).

Na podstawie badań literaturowych można zatem sformułować następujące wnioski:

- zastosowanie metod uczenia maszynowego jest, przy obecnej złożoności badanych obiektów, celowe dla pozyskania wiedzy diagnostycznej dotyczącej tych obiektów,
- istotne jest przyjęcie odpowiedniego sposobu reprezentacji zarówno danych jak i wiedzy,
- ogromna liczba cech sygnałów diagnostycznych możliwych do analizy wymaga ograniczenia, z uwagi na złożoność procesu pozyskiwania wiedzy i późniejszego wnioskowania,
- niedoskonałość gromadzonych danych wymaga uwzględnienia w procesie pozyskiwania wiedzy,
- brak jest szczegółowych badań w zakresie wyboru istotnych cech diagnostycznych z zastosowaniem technik sztucznej inteligencji,
- brak jest badań dotyczących metod wyznaczania granicznych wartości atrybutów, które to metody byłyby przydatne w dziedzinie diagnostyki maszyn i bazowałyby na technikach sztucznej inteligencji.

Rozdział 3

Cel, teza i zakres rozprawy

Podstawowe zadanie diagnostyki maszyn polega na określeniu lub rozpoznaniu stanu technicznego pewnego układu technicznego. Istotą tego działania jest wnioskowanie diagnostyczne, będące działaniem na informacji [85]. Warunkiem koniecznym skutecznego wnioskowania jest posiadanie odpowiedniej wiedzy, od której jakości zależy skuteczność wnioskowania diagnostycznego. W celu efektywnego pozyskiwania wiedzy stosuje się m.in. tzw. uczenie maszynowe, w szczególności zaś indukcyjne uczenie na podstawie przykładów. Przykłady stosowane w procesie indukcyjnego uczenia maszynowego opisywane są poprzez pewien zbiór cech, będących najczęściej estymatami pewnych sygnałów diagnostycznych. Zbiór ten jest pewnym podzbiorem nieskończenie licznego zbioru własności i właściwości opisujących stan badanego układu [39]. Ze względu na cel badań oraz ograniczone możliwości poznania dokonuje się wyboru uwzględnianego podczas badań podzbioru cech [122].

3.1 Cel rozprawy

Celem rozprawy jest zgromadzenie, analiza i porównanie metod selekcji oraz metod redukcji informacji diagnostycznej, pozwalających na ograniczenie liczności zbioru rozpatrywanych cech diagnostycznych oraz określenie sposobów wyznaczania wartości progowych konwersji wartości tych cech. W wyniku porównania zostanie opracowana metoda redukcji i selekcji informacji przydatna do zastosowania w dziedzinie diagnostyki maszyn.

3.2 Teza rozprawy

Możliwe jest przyjęcie takiej metody redukcji i selekcji informacji diagnostycznej, której zastosowanie w procesie pozyskiwania wiedzy dotyczącej relacji diagnostycznych powoduje wzrost efektywności tego procesu oraz zwiększa skuteczność wnioskowania.

3.3 Zakres rozprawy

Praca obejmuje opis metod redukcji i selekcji informacji oraz metodologię oceny tych metod. Pokazano przykłady zastosowania metod redukcji i selekcji informacji diagnostycznej i omówiono wyniki. Przedstawiono zaproponowaną metodę redukcji i selekcji informacji diagnostycznej.

W zakresie przedmiotu badań ograniczono się do oceny stanu maszyn wirnikowych. Rozpatrywano wybrane stany maszyn wirnikowych. Do badań zastosowano zbiór cech ogólnie przyjętych w diagnostyce maszyn, nie wprowadzając nowych cech, w szczególności nie omówiono zagadnień związanych z tzw. *konstruktywną indukcją* [81] pozwalającą na tworzenie nowych atrybutów warunkowych. Ponadto przyjęto, iż ze względu na możliwości uogólniania wiedzy, celowe jest zastosowanie głównie cech jakościowych.

Omawiając metody redukcji i selekcji informacji diagnostycznej pominięto zagadnienia związane z teorią zbiorów rozmytych [138]. Badania w tym zakresie były i są prowadzone w Katedrze Podstaw Konstrukcji Maszyn Politechniki Śląskiej (np. [16]). Pominięto również metody redukcji bazujące na pojęciu hiperpłaszczyzny [43]. Metody te obecnie są coraz szerzej rozwijane (np. [98]), lecz w praktyce uzyskane wyniki są trudne do zinterpretowania.

Badania weryfikacyjne zaproponowanej metody przeprowadzono dla dwóch zbiorów danych, będących wynikiem: eksperymentu czynnego na modelu maszyny wirnikowej *Rotor Kit* (opis w [9]) oraz eksperymentu numerycznego dla modelu stanowiska do badań drgań wirników (opis w [112] i [55]).

W pracy ograniczono się do zagadnień związanych z pozyskiwaniem *wiedzy deklaratywnej*, pomijając zagadnienia związane z pozyskiwaniem *wiedzy proceduralnej*. W zakresie metod pozyskiwania wiedzy ograniczono się do metod indukcyjnych. Uzasadniono wybór przyjętej metody pozyskiwania wiedzy.

W pracy pominięto podstawowe zagadnienia związane z pomiarem, rejestracją i analizą sygnałów (zob. np. [25]). Pominięto również zagadnienia przetwarzania cech sygnałów i podstawy wnioskowania diagnostycznego (zob. np. [22]).

Rozdział 4

Wybrane metody redukcji informacji diagnostycznej

Jak opisano w punkcie 2.6, metody wyznaczania wartości granicznych dzieli się między innymi na:

- *jednoatrybutowe* — w danej chwili wyznaczane są wartości graniczne tylko jednego atrybutu,
- *wielogatrybutowe* — wartości graniczne wszystkich atrybutów wyznaczane są równocześnie.

W pracy przyjęto właśnie taki podział metod wyznaczania wartości granicznych atrybutów. W pierwszej kolejności jednak należałoby krótko uzasadnić potrzebę stosowania redukcji informacji.

4.1 Celowość redukcji informacji diagnostycznej

Gdy rozpatruje się zagadnienia związane z pozyskiwaniem wiedzy na podstawie danych, pojawia się pytanie, czy i dlaczego należy dokonywać redukcji informacji diagnostycznej. Wprowadzając jakiegokolwiek zmiany w zbiorze atrybutów i ich wartości, nie można bowiem w żaden sposób poprawić opisu zgromadzonych przykładów. Nie da się zwiększyć ilości informacji niesionej przez zgromadzone dane, a zmiany w zbiorze danych mogą jedynie zmniejszyć ilość tej informacji.

Jak wynika z danych literaturowych i badań własnych autora, rozpatrywać należy dwa aspekty zagadnienia redukcji ilości informacji: efektywność obliczeniową oraz (a raczej

przede wszystkim) *jakość pozyskanej wiedzy*. Przez jakość wiedzy autor rozumie możliwość podejmowania trafnych decyzji przy zastosowaniu tej wiedzy. Jakość ta może być określana poprzez sprawność klasyfikatora. Łączny nakład obliczeniowy na pozyskanie porcji wiedzy jest w każdym przypadku znacznie mniejszy po konwersji wartości atrybutów. Ponadto pozyskane reguły (lub drzewo decyzyjne) są prostsze i bardziej zrozumiałe. Oprócz tego, reguły te (lub np. drzewo decyzyjne) są bardziej ogólne, co zapobiega nadmiernemu dopasowaniu pozyskanych reguł do danych uczących [30]. Na korzyść stosowania redukcji informacji przemawia również fakt, iż bardzo często gromadzone dane są niedoskonałe (patrz s. 25). Przekształcenie wartości może w tych przypadkach poprawić jakość pozyskanej wiedzy.

Z praktycznego punktu widzenia istotny jest również fakt, że niektórych algorytmów „uczenia maszynowego” nie da się zastosować dla danych ilościowych lub ich zastosowanie powoduje uzyskanie wyników trudnych do interpretacji, np. przez operatora maszyny.

W diagnostyce maszyn wirnikowych opracowano wiele zaleceń dotyczących wartości granicznych niektórych cech sygnałów diagnostycznych. Znane są m.in. normy ISO 2372, API 617 czy też zalecenia producentów. Jednak bezpośrednie zastosowanie tych zaleceń do określania wartości granicznych w rozumieniu (2.12) nie daje oczekiwanych rezultatów, o czym wspomniano w rozdziale 2.6. Dlatego też podejmowane są próby określania wartości granicznych w inny sposób. Najczęściej wykorzystywane metody wyznaczania wartości granicznych do celów nadzoru maszyn bazują na miarach prawdopodobieństwa, a wyznaczone wartości graniczne stosowane są do wykrywania stanów niezdatności maszyn. Przykładem może być tutaj [14], gdzie pokazano metody wyznaczania wartości granicznych dla wykrywania stanów zdatności i niezdatności maszyn. Jak się jednak okazuje, metoda sprawdzona dla jednych danych diagnostycznych [101], może dla innych danych dawać niezadowalające rezultaty [124]. Dlatego też zastosowanie opisywanych w literaturze metod określania wartości granicznych cech sygnałów diagnostycznych w „uczeniu maszynowym” wymaga indywidualnego podejścia w zależności od danych. Podjęte przez autora badania mają na celu m.in. znalezienie metody umożliwiającej wyznaczanie wartości granicznych różnych cech sygnałów diagnostycznych przy jednoczesnym założeniu, że liczba wartości granicznych powinna być stosunkowo niewielka.

4.2 Metody jednoatrybutowe redukcji ilości informacji

W niniejszym podrozdziale przedstawiono wybrane jednoatrybutowe metody wyznaczania wartości granicznych.

4.2.1 Metoda równych szerokości przedziałów

Aby zastosować metodę równych szerokości przedziałów (ang. *equal width method*) wymagana jest znajomość dziedziny $\text{Dom}(a) = [l_a, r_a]$ atrybutu $a \in A$. Istotą metody jest podział dziedziny atrybutu a na k_a równych podprzedziałów.

Przedział $[l_a, r_a]$ dzieli się na k_a równych podprzedziałów, gdzie wartość $k_a \in N$ jest zadawana przez użytkownika. Długość wszystkich podprzedziałów jest równa i wynosi:

$$\delta = \frac{l_a - r_a}{k_a}. \quad (4.1)$$

Wówczas wartości graniczne można obliczyć jako:

$$c_i^a = l_a + i \cdot \delta, \quad (4.2)$$

gdzie $i = 1, \dots, k_a - 1$.

Sposób wyznaczania wartości granicznych metodą równych szerokości przedziałów przedstawiono za pomocą algorytmu 4.1 [15].

4.2.2 Metoda równych częstości w przedziałach

W metodzie równych częstości (ang. *equal frequency method*), wartości atrybutu ilościowego $a \in A$ są sortowane w ten sposób, aby $l = v_1^a \leq v_2^a \leq \dots \leq v_{n_a}^a = r$, i dzielone na k_a podprzedziałów (wartość k_a zadawana jest przez użytkownika). Każdy podprzedział zawiera:

$$\lambda = \left\lceil \frac{n_a}{k_a} \right\rceil \quad (4.3)$$

wartości (gdzie symbol $\lceil \cdot \rceil$ oznacza zaokrąglenie do najbliższej wartości całkowitej), a kolejne wartości graniczne można obliczyć jako:

$$c_i^a = \frac{v_{i \cdot \lambda}^a + v_{i \cdot \lambda + 1}^a}{2}, \quad (4.4)$$

Algorytm 4.1 Sposób wyznaczania wartości granicznych metodą równych szerokości przedziałów

Funkcja Równe szerokości

argumenty: $V(a)$ - zbiór wartości atrybutu a

k - liczba przedziałów

zwracane: \mathbb{C}_a - macierz wartości granicznych atrybutu a

Początek

$l_a \leftarrow \min(V(a))$

$r_a \leftarrow \max(V(a))$

$\delta \leftarrow \frac{r_a - l_a}{k}$

Dla $i = 1$ **do** $k - 1$

$\mathbb{C}_a[i] \leftarrow l_a + i * \delta$

Koniec

Koniec

dla $i = 1, \dots, k_a - 1$.

Sposób wyznaczania wartości granicznych metodą równych częstości można zapisać za pomocą algorytmu 4.2 [15].

4.2.3 Metody oparte na testach statystycznych i miarach prawdopodobieństwa

Liczna grupa metod wyznaczania wartości granicznych jest oparta na testach statystycznych i miarach prawdopodobieństwa.

Metoda χ^2

Dowolna wartość graniczna $c^a \in \mathbb{C}_a$ dzieli zbiór wartości $[l_a, r_a]$ atrybutu a na dwa podprzedziały: $[l_a, c^a]$ oraz $[c^a, r_a]$. Testy statystyczne pozwalają określić niezależność statystyczną pomiędzy podziałem przykładowym dokonany przez atrybut decyzyjny, a podziałem dokonany poprzez określoną wartość graniczną c . Oceny stopnia niezależności dokonać można testem χ^2 :

Algorytm 4.2 Sposób wyznaczania wartości granicznych metodą równych częstości**Funkcja** Równe częstości

argumenty: $V(a)$ - zbiór wartości atrybutu a
 U - zbiór przykładów
 k - liczba wartości granicznych

zwracane: \mathbb{C}_a - uporządkowany zbiór wartości granicznych atrybutu

Początek

$n \leftarrow \text{card}(U)$

Sortuj $V(a)$

$\lambda \leftarrow \left\lceil \frac{n}{k} \right\rceil$

$\lambda_r \leftarrow n \text{ modulo } k$

$\mathbb{C}_a \leftarrow \{\min(V(a)), \max(V(a))\}$

$l \leftarrow 1$

Dopóki $l \leq n$

$i \leftarrow l$

Dopóki $i \leq n \wedge (i - l) \leq \lambda$

$i \leftarrow i + 1$

Koniec**Jeżeli** $\lambda_r > 0$

$i \leftarrow i + 1$

$\lambda_r \leftarrow \lambda_r - 1$

Koniec

$\mathbb{C}_a \leftarrow \mathbb{C}_a \cup \left\{ \frac{V_i(a) + V_{i+1}(a)}{2} \right\}$

$l \leftarrow i$

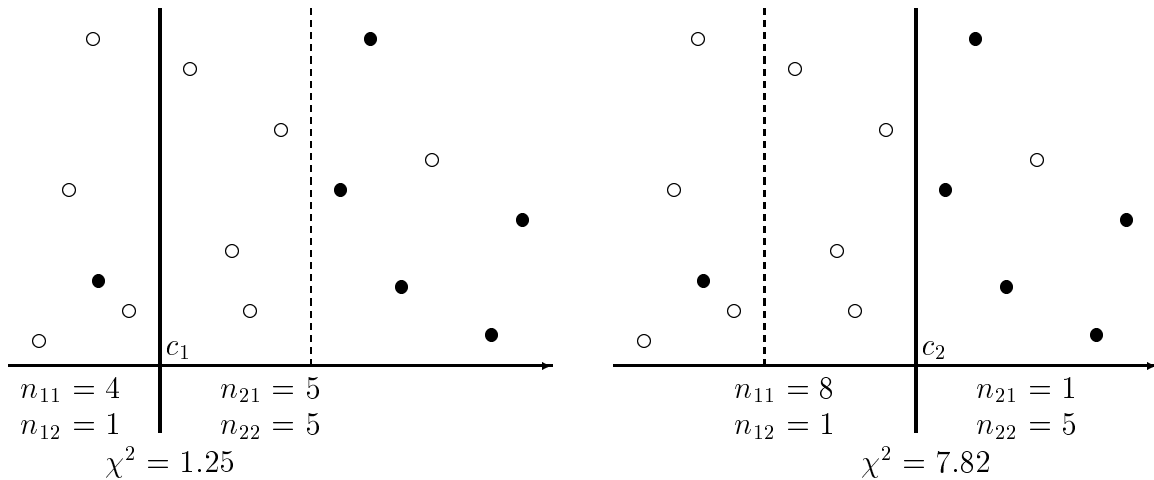
Koniec**Koniec**

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^k \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (4.5)$$

gdzie k — liczba klas, N — całkowita liczba przykładów, n_{ij} — liczba przykładów j -tej klasy w i -tym przedziale, E_{ij} — przewidywana częstość przykładów j -tej klasy w i -tym przedziale:

$$E_{ij} = \frac{n_i \cdot n_j}{N}.$$

W trakcie wyznaczania wartości granicznych metodą χ^2 wybierane są tylko te wartości graniczne, dla których wartość funkcji testowej jest wyższa od pewnej zadanej wartości α_χ . Jednocześnie liczba wartości granicznych nie może przekroczyć α_c . Rys. 4.1 ilustruje to zadanie. Dla przedstawionego zbioru przykładów należących do dwóch klas obliczono wartość testu χ^2 dla dwóch różnych wartości granicznych. Im podział przykładów na klasy przez rozpatrywaną wartość graniczną jest lepszy, tym wartość testu χ^2 jest większa.



Rysunek 4.1: Geometryczna interpretacja metody χ^2 [97]

Istota metody polega na tym, że uporządkowany zbiór wartości rozpatrywanego atrybutu a dzielony na $\text{card}(V(a))$ podprzedziałów, a następnie, podprzedziały te łączą się¹ do momentu, gdy spełnione jest kryterium zatrzymania obliczeń. Kryterium to można przedstawić następująco [30]:

¹Dlatego też metoda ta często jest określana mianem *ChiMerge*.

$$\text{card}(\mathbb{C}_a) \leq \alpha_c \wedge \min_{\substack{\pi_1, \pi_2 \in \pi_a \\ l_2 = r_1}} \chi^2(V(a), \pi_a) \geq \alpha_\chi \quad (4.6)$$

gdzie α_c i α_χ są zadawanymi przez użytkownika parametrami, π_1, π_2 są przedziałami do połączenia natomiast l_2 i r_1 określają lewy i prawy koniec przedziałów odpowiednio π_2 oraz π_1 .

Wybór przyległych przedziałów, które należy połączyć, dokonywany jest z kolei dla tych przedziałów, dla których wartość statystyki χ^2 jest najmniejsza [30]:

$$\arg \min_{\substack{\pi_1, \pi_2 \in \pi_a \\ l_2 = r_1}} \chi^2(V(a), \pi_a) \quad (4.7)$$

Opisany sposób postępowania określa algorytm 4.3.

Algorytm 4.3 Sposób wyznaczania wartości granicznych metodą χ^2

Funkcja ChiMerge

argumenty: $V(a)$ - wartości atrybutu a

α_c, α_χ - parametry kryterium zatrzymania obliczeń

zwracane: π_a - uporządkowany zbiór przedziałów konwersji

Początek

Zainicjuj π_a jako zbiór przedziałów zawierających dokładnie jedną wartość atrybutu a występującą w zbiorze $V(a)$

Powtarzaj

$\pi_1, \pi_2 \leftarrow \text{wybierz-przedziały-do-połączenia}(V(a), \pi_a)$

$\pi_a \leftarrow \pi_a \setminus \{\pi_1, \pi_2\} \cup \{\pi_1 \cup \pi_2\}$

Dopóki nie spełnione kryterium-zatrzymania($V(a), \pi_a, \alpha_c, \alpha_\chi$)

Koniec

Metody oparte na innych testach statystycznych

C. Cempel w [14] zaproponował metody wyznaczania wartości krytycznych (związanych ze zmianą stanu maszyny) oparte na nierówności Czebyszewa (metoda C) oraz rozkładach Weibulla (metoda W) i Pareto (metoda P). W [100] zaproponowano przyjęcie wartości krytycznych jako wartości granicznych w procesie redukcji informacji. Postępowanie

takie, zastosowane dla danych pochodzących z eksperymentu diagnostycznego, dało zadowalające rezultaty [100].

W pierwszym kroku wyznaczana jest wartość graniczna b atrybutu a określająca stan niezdatności maszyny. Wartość tę oblicza się w sposób następujący [14]:

- metoda C :

$$b = \bar{v} + \sigma \sqrt{\frac{P_g}{2A}} \quad (4.8)$$

gdzie \bar{v} jest wartością średnią wartości rozpatrywanego atrybutu obliczaną jako:

$$\bar{v} = \frac{\sum_{i=1}^n a(u_i)}{n}, \quad (4.9)$$

n jest liczbą przykładów, σ — odchyleniem standardowym wartości atrybutu:

$$\sigma = \frac{\sqrt{\sum_{i=1}^n (a(u_i) - \bar{v})^2}}{n}, \quad (4.10)$$

P_g jest współczynnikiem gotowości maszyny określanym jako iloraz czasu pracy t_r maszyny do sumy czasu pracy i czasu postoju (remontów) t_n :

$$P_g = \frac{t_r}{t_r + t_n}, \quad (4.11)$$

A jest prawdopodobieństwem zbędnego remontu,

- metoda P :

$$b = (1 - \gamma^{-1}) \bar{v} \gamma \sqrt{\frac{P_g}{A}}, \quad (4.12)$$

gdzie γ jest współczynnikiem rozkładu Pareto:

$$\gamma = 1 + \sqrt{1 + \left(\frac{\bar{v}}{\sigma}\right)^2}, \quad (4.13)$$

- metoda W :

$$b = v_{min} + (\bar{v} - v_{min}) \Gamma^{-1}(1 + k^{-1}) \sqrt[k]{\ln\left(\frac{P_g}{A}\right)}, \quad (4.14)$$

gdzie v_{min} jest wartością minimalną rozpatrywanego atrybutu a , k jest współczynnikiem rozkładu Weibulla:

$$k = \frac{\bar{v} - v_{min}}{\sigma}, \quad (4.15)$$

Γ jest funkcją gamma Eulera,

Wówczas wartości graniczne atrybutu a definiuje się jako [14]:

$$c_1^a = \bar{v} + 0.25 (b - \bar{v}), \quad (4.16)$$

$$c_2^a = \bar{v} + 0.50 (b - \bar{v}), \quad (4.17)$$

$$c_3^a = \bar{v} + 0.75 (b - \bar{v}). \quad (4.18)$$

Przyjęcie tak zdefiniowanych wartości granicznych pozwala zatem na uzyskanie czterech wartości jakościowych atrybutów.

Przykładowy sposób obliczenia wartości granicznych metodą C pokazano jako algorytm 4.4.

Metoda rozkładu częstości wartości atrybutu

Metoda wyznaczania wartości granicznych na podstawie rozkładu częstości wartości atrybutu (histogramu) została zaproponowana przez autora [34]. Polega ona na określeniu wartości granicznych według rozkładu częstości wartości poszczególnych cech sygnałów diagnostycznych. Szczegółowa analiza histogramów utworzonych dla wielu różnych wartości cech sygnałów diagnostycznych pozwoliła autorowi stwierdzić, że cechy diagnostyczne nie przyjmują wartości równomiernie z całej swej dziedziny. Można zaobserwować, że pewne wartości występują częściej, inne zaś nie występują wcale. Autor zaproponował, podjęcie próby zastosowania tych własności histogramów dla celów konwersji wartości atrybutów diagnostycznych.

Rozpatrując histogram wartości pewnego atrybutu, można na nim wskazać pewne lokalne minima — wartości, dla których liczba przykładów jest istotnie mniejsza od liczby przykładów dla wartości sąsiednich (por. rys. 4.2). Wartości, w których występują minima lokalne, można potraktować jako wartości graniczne i z ich zastosowaniem dokonać konwersji wartości atrybutu.

Problemy, jakie można tutaj napotkać, związane są między innymi z określeniem liczby przedziałów, dla których wyznaczany jest histogram, jak również z określeniem, czy dane minimum lokalne jest na tyle istotne, by przyjąć określającą go wartość za wartość graniczną.

Przykładowy sposób wyznaczania wartości granicznych na podstawie histogramu przedstawiono jako algorytm 4.5.

Algorytm 4.4 Sposób wyznaczania wartości granicznych metodą C**Funkcja** Czebyszew

argumenty: $V(a)$ - zbiór wartości atrybutu $a \in A$

t_r - czas pracy maszyny

t_n - czas remontów maszyny

P_r - prawdopodobieństwo zbędnego remontu

zwracane: \mathbb{C}_a - uporządkowany zbiór wartości granicznych atrybutu a

Początek

$$P_g \leftarrow \frac{t_r}{t_r + t_n}$$

$$n \leftarrow \text{card}(V(a))$$

$$\bar{v} \leftarrow \frac{\sum_{i=1}^n V_i(a)}{n}$$

$$\sigma \leftarrow \frac{\sqrt{\sum_{i=1}^n (a(u_i) - \bar{v})^2}}{n}$$

$$b \leftarrow \bar{v} + \sigma \sqrt{\frac{n}{2P_r}}$$

$$c_1 \leftarrow \bar{v} + 0,25(b - \bar{v})$$

$$c_2 \leftarrow \bar{v} + 0,50(b - \bar{v})$$

$$c_3 \leftarrow \bar{v} + 0,75(b - \bar{v})$$

$$\mathbb{C}_a \leftarrow \{c_1\} \cup \{c_2\} \cup \{c_3\}$$

Koniec**4.2.4 Metoda minimum entropii**

Jedną z najbardziej licznych grup metod wyznaczania wartości granicznych atrybutów ilościowych jest grupa metod bazujących na minimalizacji entropii zbioru przykładów (ang. *minimal class entropy method*). Entropia $H_{a,c^a}(U)$ zbioru przykładów U uwzględniająca podział dziedziny atrybutu a wartością graniczną c^a jest zdefiniowana jako [40]:

$$H_{a,c^a}(U) = \frac{|U_1|}{n} H(U_1) + \frac{|U_2|}{n} H(U_2), \quad (4.19)$$

Algorytm 4.5 Sposób wyznaczania wartości granicznych według histogramu**Funkcja** Histogram

argumenty: $V(a)$ - zbiór wartości atrybutu $a \in A$
 α_e - wymagany poziom istotności
 α_k - maksymalna liczba wartości granicznych
 α_c - liczba przedziałów dla histogramu
zwracane: \mathbb{C}_a - uporządkowany zbiór wartości granicznych

PoczątekSortuj rosnąco $V(a)$ Utwórz histogram $\text{Hst}(V(a), \alpha_c)$ Znajdź wszystkie minima lokalne m histogramu Hst , $m \in M$ **Dla** $i = 1$ **do** $\text{card}(M)$ $e_i \leftarrow \text{PoziomIstotności}(m_i)$ $E \leftarrow E \cup \{e_i\}$ **Koniec**Sortuj rosnąco E **Dla** $i = 1$ **do** $\text{card}(E)$ **Jeżeli** $e_i \geq \alpha_e$ **i** $\text{card}(\mathbb{C}_a) < \alpha_k$ $\mathbb{C}_a \leftarrow \mathbb{C}_a \cup \{e_i\}$ **Koniec****Koniec****Koniec**

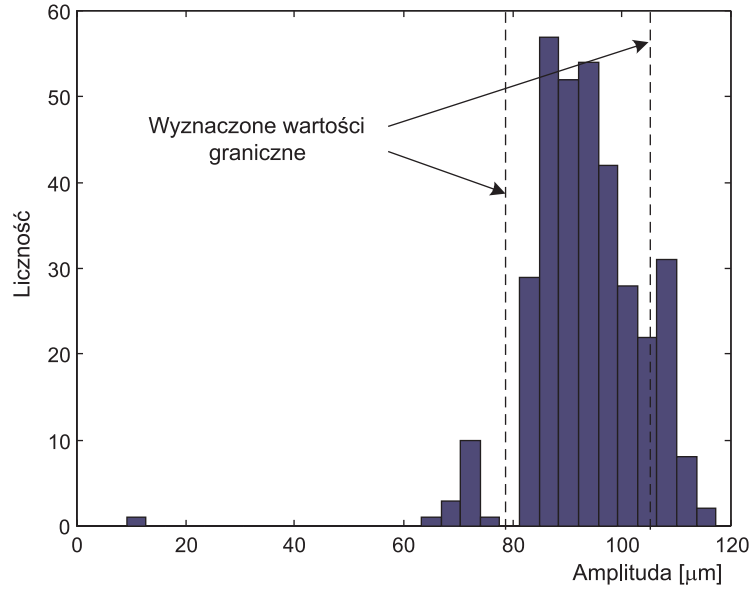
gdzie n oznacza liczbę przykładów w U a U_1, U_2 są podzbiorami zbioru U wyznaczonymi przez wartość graniczną c^a :

$$U_1 = \{u \in U : a(u) \leq c^a\}, \quad (4.20)$$

$$U_2 = \{u \in U : a(u) > c^a\}. \quad (4.21)$$

Wartości entropii $H(U_i)$ wyznacza się zwykle na podstawie wzoru [69]:

$$H(U_i) = -p_i^+ \log_2 p_i^+ - p_i^- \log_2 p_i^-. \quad (4.22)$$



Rysunek 4.2: Przykład wyznaczania wartości granicznych cechy diagnostycznej na podstawie histogramu

Prawdopodobieństwa p_i^+ oraz p_i^- obliczyć można następująco [69]:

$$p_i^+ = \frac{n_i^+}{n_i^+ + n_i^-}, \quad (4.23)$$

$$p_i^- = \frac{n_i^-}{n_i^+ + n_i^-}, \quad (4.24)$$

gdzie n_i^+ oznacza liczbę przykładów pozytywnych w U_i , a n_i^- oznacza liczbę przykładów negatywnych w U_i .

Przykładowy sposób wyznaczania wartości granicznych metodą minimalizacji entropii przedstawiono jako algorytm 4.6. Jako algorytm 4.7 pokazano metodę poszukiwania najlepszej wartości granicznej [15]. W algorytmie tym P oznacza uporządkowany zbiór początkowych wartości granicznych p .

Wśród licznej grupy metod opartych na minimalizacji entropii warto wymienić metodę rekurencyjnej minimalizacji długości kodu [44], w której zastosowano, jako kryterium jakości podziału, kryterium minimalnej długości kodu *MDL* (*Minimum Description Length*) [110]. Przyjęcie takiego kryterium oznacza, że wybierany będzie taki przypadek, który w efekcie dostarczy krótszych reguł. W praktyce poszukuje się rozwiązań, które są optymalne ze względu na kryteria długości reguł oraz ich jakości [30].

Algorytm 4.6 Sposób wyznaczania wartości granicznych metodą minimum entropii**Funkcja** Minimum entropii

argumenty: $V(a)$ - zbiór wartości atrybutu a
 U - uporządkowany zbiór przykładów
 k - liczba wartości granicznych
zwracane: \mathbb{C}_a - uporządkowany zbiór wartości granicznych atrybutu a

PoczątekSortuj $V(a)$ $C \leftarrow \emptyset$ Dla $i = 1$ do $\text{card}(U) - 1$ **Jeżeli** $V_i(a) \neq V_{i+1}(a)$ $C \leftarrow C \cup \left\{ \frac{V_i(a) + V_{i+1}(a)}{2} \right\}$ **Koniec****Koniec** $\mathbb{C}_a \leftarrow \{\min(V(a)), \max(V(a))\}$ **Dopóki** $(\text{card}(\mathbb{C}_a) - 1) < k$ $Hc \leftarrow 0$ $c \leftarrow 0$ **Dla każdego** $l, r \in \mathbb{C}_a$ **takiego że** $l < r \wedge (r - l)$ osiąga minimum **Wywołaj** *Znajdź najlepszą wartość graniczną* argumenty: $C, l, r, V(a)$ zwracane: p, Hp **Jeżeli** $Hp < Hc$ $c \leftarrow p$ $Hc \leftarrow Hp$ **Koniec** **Koniec** $\mathbb{C}_a \leftarrow \mathbb{C}_a \cup \{c\}$ $C \leftarrow C \setminus \{c\}$ **Koniec****Koniec**

Algorytm 4.7 Znajdowanie najlepszej wartości granicznej dla metody minimum entropii

Funkcja Znajdź najlepszą wartość graniczną

argumenty: C - uporządkowany zbiór początkowych wartości granicznych
 l, r - wartości brzegowe przedziału konwersji
 U - uporządkowany zbiór przykładów
 V - uporządkowany zbiór wartości atrybutu a

zwracane: c - najlepsza wartość graniczna atrybutu a
 Hc - entropia dla wartości c

Początek

$Hc \leftarrow \infty$

$P \leftarrow \emptyset$

Dla $i = 2$ **do** $\text{card}(V)$

$P \leftarrow P \cup \left\{ \frac{V_i + V_{i-1}}{2} \right\}$

Koniec

Dla każdego $p \in P$ **takiego że** $l < p < r$

Oblicz $U_1 \subseteq U$ **tak by**

$x \in U \wedge l < V(x) < p \Rightarrow x \in S_1$

Oblicz $U_2 \subseteq U$ **tak by**

$x \in U \wedge p < V(x) < r \Rightarrow x \in S_2$

$H_p \leftarrow \frac{|S_1|}{|S_1 \cup S_2|} H(S_1) + \frac{|S_2|}{|S_1 \cup S_2|} H(S_2)$

Jeżeli $H_p < Hc$

$c \leftarrow p$

$Hc \leftarrow H_p$

Koniec

Koniec

Koniec

4.2.5 Metoda grupowania

Do wyznaczania wartości granicznych stosowane są również metody grupowania (ang. *cluster analysis method*). Stosuje się wiele różnych metod grupowania wartości atrybutów. Jednymi z najbardziej popularnych są [129]:

- k -means,
- mini-max,
- k -najbliższych sąsiadów.

Metody te można stosować zarówno jako jednoatrybutowe, jak i wieloatrybutowe. Stosowanie tych metod wymaga jednak narzucenia liczby poszukiwanych wartości granicznych. Jednym ze sposobów określenia tej liczby — zaproponowanym przez autora — jest zastosowanie histogramów wartości poszczególnych atrybutów [34].

Zastosowanie metod grupowania wartości atrybutów do wyznaczania wartości granicznych jest zadaniem stosunkowo prostym w przypadku metod jednoatrybutowych. Rozpatrywana jest wówczas przestrzeń jednowymiarowa. Sposób postępowania można przedstawić w postaci algorytmu 4.8.

4.3 Metody wieloatrybutowe redukcji ilości informacji

Cechą jednoatrybutowych metod wyznaczania wartości granicznych jest to, że wymagają narzucenia liczby wartości granicznych. W praktyce jedynie specjalista z danej dziedziny potrafi określić tę liczbę, a i tak nie zawsze. Konieczne zatem jest przy stosowaniu metod jednoatrybutowych poszukiwanie nie tylko najbardziej skutecznej metody oraz optymalnej liczby wartości przedziałów, lecz również powtarzanie tych kroków k -krotnie dla każdego atrybutu oddzielnie.

Analizując te problemy, można stwierdzić, że skuteczna metoda wyznaczenia wartości granicznych to taka, która [15]:

- jest kompletna, tzn. za jej pomocą można określić wartości graniczne wszystkich atrybutów,

Algorytm 4.8 Sposób wyznaczania wartości granicznych jednoatrybutową metodą grupowania

Funkcja Grupowanie jednoatrybutowe

argumenty: $V(a)$ - zbiór wartości atrybutu a

zwracane: \mathbb{C}_a - uporządkowany zbiór wartości granicznych atrybutu a

Początek

Normalizuj wartości zbioru $V(a)$, nowe wartości zapisz jako $V^G(a)$

Zastosuj algorytm grupowania dla wartości $V^G(a)$

Przypisz etykiety wyznaczonych grup elementom zbioru $V(a)$

Sortuj rosnąco $V(a)$

Dla $i = 1$ **do** $\text{card}(V(a))$

Jeżeli nastąpiła zmiana etykiety grupy

$$\mathbb{C}_a \leftarrow \mathbb{C}_a \cup \left\{ \frac{V_{i-1}(a) + V_i(a)}{2} \right\}$$

Koniec

Koniec

Koniec

- daje w wyniku małą liczbę wartości granicznych: im moc dziedziny atrybutów jest mniejsza, tym pozyskana wiedza jest prostsza, reguły są krótsze, a co za tym idzie — bardziej ogólne.
- zachowuje wyjściowy poziom niesprzeczności danych: z reguły tablica decyzyjna, która zawiera przykłady pochodzące z rzeczywistych badań (z atrybutami o wartościach ilościowych) jest niesprzeczna, wskazane jest zachowanie wyjściowego poziomu niesprzeczności (2.8).

Na podstawie powyższych stwierdzeń podjęto, w pierwszej kolejności, próby rozszerzenia metod jednoatrybutowych (np. [15]), a w kolejnym kroku — opracowano metody wieloatrybutowe określania wartości granicznych atrybutów ilościowych.

4.3.1 Metoda grupowania

Metodę grupowania dla przypadku jednoatrybutowego omówiono wcześniej. Poniżej przedstawiona zostanie w skrócie idea metody wieloatrybutowej.

Rozpatrywana jest przestrzeń k -wymiarowa ($k = \text{card}(A)$). Grupowanych jest n punktów k -wymiarowej przestrzeni. Wartości graniczne wyznaczone są w sposób przedstawiony na rys. 4.3a. Wartości: maksymalna i minimalna każdego atrybutu w danej grupie przyjmowane są jako wartości graniczne. Sposób ten prowadzi jednak do generowania nadmiernej liczby wartości granicznych.

W [15] zaproponowano modyfikację tej metody. Modyfikacja polega na połączeniu niektórych przedziałów konwersji, co pozwala między innymi na znaczną redukcję liczby wartości granicznych. Łączenia wyznaczonych metodą grupowania przedziałów dokonuje się tylko wówczas, jeżeli poziom niesprzeczności rozpatrywanej tablicy decyzyjnej nie zostanie pogorszony. Przykładowy zbiór wartości po dokonaniu opisanych zmian przedstawiono na rys. 4.3b.

Sposób postępowania w przypadku zastosowania metody grupowania wieloatrybutowego można przedstawić w postaci algorytmu 4.9 [15].

4.3.2 Metoda heurystyczna

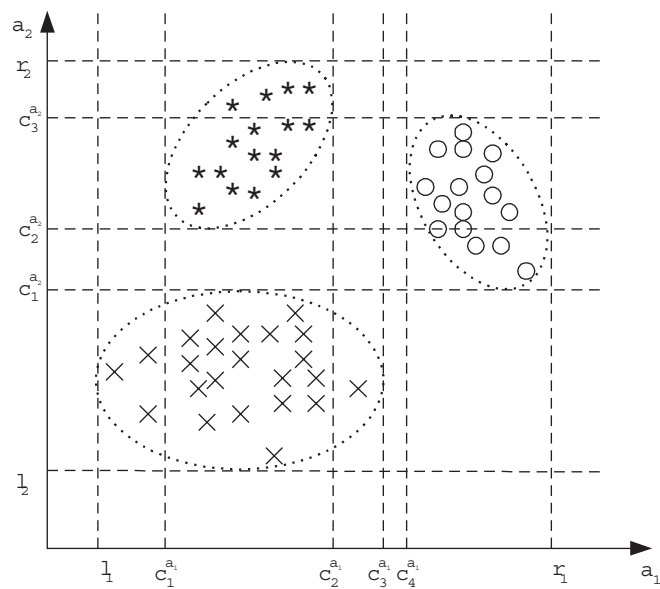
Metoda, nazywana w literaturze metodą heurystyczną [96], oparta jest częściowo na teorii *zbiorów przybliżonych* [104]. Elementy tej teorii przedstawiono w dodatku do rozprawy.

Niech dana będzie niesprzeczna (w sensie (2.8)) tablica decyzyjna $\mathcal{S} = (U, A \cup \{d\})$ gdzie $U = \{x_1, x_2, \dots, x_n\}$, $A = \{a_1, a_2, \dots, a_k\}$ oraz $d : U \rightarrow \{1, \dots, r\}$. Załóżmy, że dla każdego $a \in A$ zachodzi $V_a = [v_1^a, \dots, v_{n_a}^a] \subset \mathfrak{R}$.

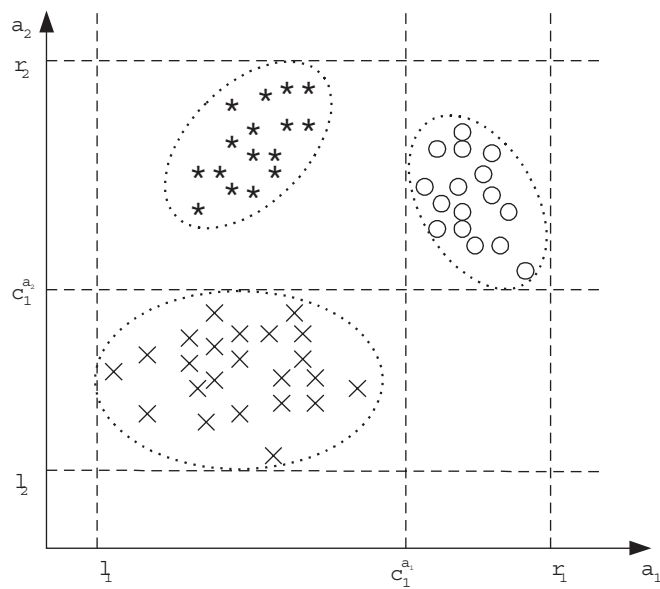
Ponadto niech tablica decyzyjna $\mathcal{S}_1 = (U_1, A_1 \cup \{d_1\})$ będzie zdefiniowana następująco [98]:

$$\begin{aligned}
 U_1 &= \{(x, y) \in U \times U : d(x) \neq d(y)\} \cup \{\mathbf{z}\}, \quad \text{gdzie: } \mathbf{z} \notin U \times U, \\
 d_1 : U_1 &\rightarrow \{0, 1\} \quad \text{przy czym: } d_1(u) = \begin{cases} 0 & \text{jeśli } u = \mathbf{z} \\ 1 & \text{w przeciwnym wypadku} \end{cases}, \quad (4.25) \\
 A_1 &= \bigcup_{a \in A} \left\{ c_i^a = \frac{v_i^a + v_{i+1}^a}{2} \right\}, \quad \text{gdzie: } 1 \leq i \leq n_a - 1.
 \end{aligned}$$

a)



b)



Rysunek 4.3: Przykład wyznaczania wartości granicznych cech diagnostycznych metodą grupowania [15]: a) początkowe wartości graniczne, b) wartości graniczne po redukcji ich liczby

Algorytm 4.9 Sposób wyznaczania wartości granicznych wieloatrybutową metodą grupowania

Funkcja Grupowanie wieloatrybutowe

argumenty: $V(A)$ - zbiór wartości atrybutów $a \in A$

U - zbiór przykładów $u \in U$

β_w - wymagany poziom niesprzeczności

zwracane: π_A - uporządkowany zbiór przedziałów konwersji atrybutów $a \in A$

Początek

Normalizuj wartości zbioru $V(A)$, nowe wartości zapisz jako $V^G(A)$

Zastosuj algorytm grupowania dla wartości $V^G(A)$

Przypisz etykiety wyznaczonych grup elementom zbioru $V(A)$

Dla $i = 1$ **do** $\text{card}(U)$

Sortuj rosnąco $V(a_i)$

Dla $j = 1$ **do** $\text{card}(V(a_i))$

Jeżeli nastąpiła zmiana etykiety grupy

$$\mathbb{C}_{a_i} \leftarrow \mathbb{C}_{a_i} \cup \left\{ \frac{V_{j-1}(a_i) + V_j(a_i)}{2} \right\}$$

Koniec

Koniec

Przekształć \mathbb{C}_{a_i} w π_{a_i}

Koniec

$M \leftarrow \emptyset$

Dla $i = 1$ **do** $\text{card}(U)$

Dla $j = 1$ **do** $\text{card}(\pi_{a_i})$

$$G \leftarrow \{u \in U : c_{j-1}^{a_i} \leq a_i(u) \leq c_{j+1}^{a_i}\}$$

Jeżeli $H(G) = 0$

$$\pi_{a_i} \leftarrow \pi_{a_i} \setminus \{[c_{j-1}^{a_i}, c_j^{a_i}), [c_j^{a_i}, c_{j+1}^{a_i})\} \cup \{[c_{j-1}^{a_i}, c_{j+1}^{a_i})\}$$

W przeciwnym razie

$$M \leftarrow M \cup \{i, j, H(G)\}$$

Koniec

Koniec

Algorytm 4.9 (c.d.)

/* $m \in M$ oraz $M = \{m_1, m_2, m_3\}$ */

Dopóki $M \neq \emptyset$

Znajdź $m \in M$ dla którego m_3 jest minimalne

$$\pi_{a_{m_1}}^s \leftarrow \pi_{a_{m_1}}$$

$$\pi_{a_{m_1}} \leftarrow (\pi_{a_{m_1}} \setminus \{[c_{m_2-1}, c_{m_2}], [c_{m_2}, c_{m_2+1}]\}) \in \{[c_{m_2-1}, c_{m_2+1}]\}$$

Oblicz β dla bieżących wartości granicznych

Jeżeli $\beta_w \leq \beta$

Znajdź $w \in M$ takie, że $w_1 = m_1$ i $m_2 = w_2 + 1$

$$M \leftarrow M \setminus \{w\}$$

Znajdź $p \in M$ takie, że $p_1 = m_1$ i $p_2 = m_2 - 1$

Znajdź $q \in M$ takie, że $q_1 = m_1$ i $q_2 = m_2 + 2$

Aktualizuj $p, m \in M$, gdzie $p_1 = q_1 = m_1$ i $p_2 = m_2 - 1$ i $q_2 = m_2 + 2$

W przeciwnym razie

$$\pi_{a_{m_1}} \leftarrow \pi_{a_{m_1}}^s$$

Koniec

Koniec

Koniec

Element z zbioru U_1 pozwala na utworzenie prostego kryterium zatrzymania obliczeń. Zbiór A_1 jest początkowym zbiorem wartości granicznych c^a atrybutu a .

Dla danych wartości granicznych $c_m^a \in A_1$ atrybutu a , gdzie $a \in A$ i $1 \leq m \leq n_a$, oraz dla danego podzbioru $X \subseteq U$ wprowadza się następujące wielkości dla $j = 1, \dots, r$ [98]:

$$\begin{aligned} l_j^X(c_m^a) &= \text{card}(\{x \in X : (a(x) < c_m^a) \wedge (d(x) = j)\}), \\ r_j^X(c_m^a) &= \text{card}(\{x \in X : (a(x) > c_m^a) \wedge (d(x) = j)\}), \\ l^X(c_m^a) &= \sum_{j=1}^r l_j^X(c_m^a) = \text{card}(\{x \in X : a(x) < c_m^a\}), \\ r^X(c_m^a) &= \sum_{j=1}^r r_j^X(c_m^a) = \text{card}(\{x \in X : a(x) > c_m^a\}), \end{aligned} \tag{4.26}$$

Wówczas dla dowolnej wartości $c_m^a \in A_1$ i $X \subseteq U$, liczba $W^X(c_m^a)$ par przykładów zbioru X rozróżnialnych przez c_m^a określana jest jako :

$$W^X(c_m^a) = l^X(c_m^a) \cdot r^X(c_m^a) - \sum_{i=1}^r l_i^X(c_m^a) \cdot r_i^X(c_m^a). \tag{4.27}$$

Dowolny zbiór wartości granicznych $\mathbb{C} = \bigcup_{a \in A} \mathbb{C}_a$ tablicy decyzyjnej \mathcal{S} pozwala określić nową tablicę decyzyjną:

$$\mathcal{S}^{\mathbb{C}} = (U, A^{\mathbb{C}} \cup \{d\}), \quad (4.28)$$

w której $A^{\mathbb{C}} = \{a^{\mathbb{C}} : a^{\mathbb{C}} \in A\}$ oraz $a^{\mathbb{C}}(x) = m \Leftrightarrow a(x) \in [c_m^a, c_{m+1}^a)$ dla dowolnego $x \in U$.

Określając relację równoważności $\text{IND}(A^{\mathbb{C}})$ dla tablicy $\mathcal{S}^{\mathbb{C}}$ wyznaczyć można klasy równoważności $E_1^{\mathbb{C}}, E_2^{\mathbb{C}}, \dots, E_p^{\mathbb{C}}$. Elementy (przykłady) zbioru $E_i^{\mathbb{C}}$, gdzie $i = 1, \dots, p$, są rozróżnialne przez dowolną wartość ze zbioru \mathbb{C} . Stąd liczba par przykładów rozróżnialnych przez wartość $c \notin \mathbb{C}$, lecz nierozróżnialnych przez wartości ze zbioru \mathbb{C} jest równa:

$$W_{\mathbb{C}}(c) = W^{E_1^{\mathbb{C}}}(c) + W^{E_2^{\mathbb{C}}}(c) + \dots + W^{E_p^{\mathbb{C}}}(c). \quad (4.29)$$

Jeżeli istnieje tylko jeden element $x \in U$ taki, że $a(x) \in [c_m^a, c_{m+1}^a)$ i jeżeli $d(x) = t$ i ponadto $x \in E_i^{\mathbb{C}}$, wówczas (co udowodniono w [95]):

$$\begin{aligned} W_{\mathbb{C}}(c_{m+1}^a) - W_{\mathbb{C}}(c_m^a) &= W^{E_i^{\mathbb{C}}}(c_{m+1}^a) - W^{E_i^{\mathbb{C}}}(c_m^a) = \\ &= \left(r^{E_i^{\mathbb{C}}}(c_m^a) - l^{E_i^{\mathbb{C}}}(c_m^a) \right) - \left(r_t^{E_i^{\mathbb{C}}}(c_m^a) - l_t^{E_i^{\mathbb{C}}}(c_m^a) \right). \end{aligned} \quad (4.30)$$

Dla tak wprowadzonych wielkości, algorytm 4.10 przedstawia metodę wyznaczania wartości granicznych atrybutów danej tablicy decyzyjnej \mathcal{S} . Algorytm oparty jest na tzw. strategii Johnsona [53]. Poszukiwana jest taka wartość graniczna $c \in A_1$, która rozróżnia największą liczbę par elementów zbioru U_1 . Wówczas wartość c jest dodawana do zbioru wartości granicznych \mathbb{C} , a ze zbioru U_1 usuwane są wszystkie pary elementów rozróżnialne przez wartość c . Algorytm powtarzany jest tak długo, aż $U_1 = \{\mathbf{z}\}$.

4.4 Liczba wartości granicznych

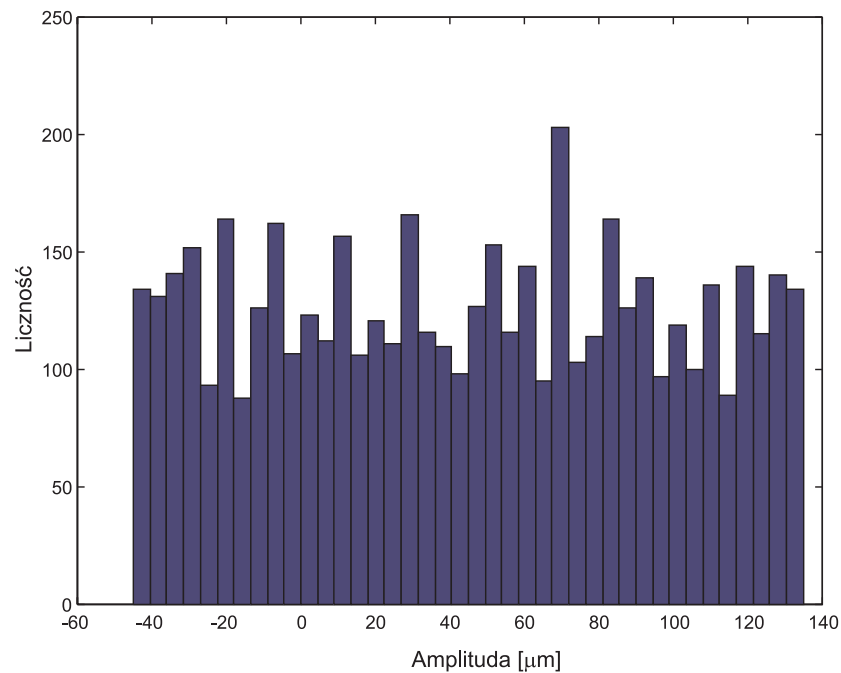
Mankamentem wielu metod wyznaczania wartości granicznych jest to, że określenie liczby tych wartości zależne jest od osoby prowadzącej badania. Tak jest np. w przypadku metody równych częstości lub metody grupowania. Niektóre metody wyznaczania wartości granicznych z pozoru pozbawione są tej niedogodności i pozwalają na

Algorytm 4.10 Sposób wyznaczania wartości granicznych metodą heurystyczną**Funkcja** Wybór heurystycznyargumenty: \mathcal{S} - niesprzeczna tablica decyzyjnazwracane: \mathbb{C} - zbiór wartości granicznych**Początek** $\mathbb{C} \leftarrow \emptyset$ $L \leftarrow \{U\}$ /* L jest zbiorem klas równoważności $[\text{IND}(A^{\mathbb{C}})]$ */ $A_1 \leftarrow$ początkowy zbiór wartości granicznych tablicy \mathcal{S} **Powtarzaj**Dla każdego $c \in A_1$ Oblicz $W_{\mathbb{C}}(c)$ **Koniec**Wybierz wartość c_{max} dla której $W_{\mathbb{C}}(c_{max})$ osiąga wartość maksymalną $\mathbb{C} \leftarrow \mathbb{C} \cup \{c_{max}\}$ $A_1 \leftarrow A_1 \setminus \{c_{max}\}$ Dla każdego $E^{\mathbb{C}} \in L$ Jeżeli c_{max} dzieli zbiór $E^{\mathbb{C}}$ na $E_1^{\mathbb{C}}, E_2^{\mathbb{C}}$ Usuń $E^{\mathbb{C}}$ z L Dodaj do L dwa zbiory $E_1^{\mathbb{C}}, E_2^{\mathbb{C}}$ **Koniec****Koniec****Dopóki** wszystkie zbiory w L nie zawierają elementów wyłącznie jednej klasy**Koniec**

wyznaczenie liczby wartości granicznych na podstawie danych. Wymagają one jednak określenia parametrów wpływających na kryterium zatrzymania działania algorytmu. Przy zbyt małym zróżnicowaniu rozkładu wartości atrybutów algorytm ten mogą powodować wyznaczanie dużej liczby wartości granicznych.

Próbą rozwiązania tego zagadnienia jest zaproponowane przez autora zastosowanie rozkładów częstości wartości atrybutów (histogramów) do wyznaczania liczby wartości granicznych (por. rys. 4.2) [34]. W pewnych przypadkach określenie wartości granicznych

wymaga jednak podjęcia dodatkowych badań. Przypadek taki ilustruje histogram pokazany na rys. 4.4. Trudno jest na tym histogramie wskazać wyraźne minima lokalne, a tym samym trudno jest określić wartości graniczne.



Rysunek 4.4: Przykład rozkładu częstości wartości atrybutu diagnostycznego AdifTD (por. tab. 7.17)

Do określenia liczby wartości granicznych autor zaproponował zastosowanie jednoa-trybutowej metody grupowania oraz ocenę jakości dokonanego podziału. Oceny jakości podziału można dokonać między innymi na podstawie [128]:

- ilorazu średniej odległości elementów w grupie i średniej odległości grup,
- ilorazu sumy „momentów bezwładności” grup i „momentu bezwładności” wszystkich elementów.

W obydwu przypadkach mniejsza wartość oceny oznacza lepszą jakość podziału.

Rozdział 5

Wybrane metody selekcji atrybutów diagnostycznych

Podczas obserwacji maszyny rejestruje się wiele sygnałów. Dla każdego sygnału diagnostycznego można wyznaczyć liczny zbiór jego cech (estymat). Nie wszystkie cechy są jednak wykorzystywane w procesie diagnozowania - nie każda bowiem niesie odpowiednią ilość informacji diagnostycznej. Być może dla określenia stanu badanego obiektu lub dla określenia rodzaju powstałej niesprawności nie ma również potrzeby gromadzenia wszystkich rejestrowanych dotychczas sygnałów diagnostycznych. Stąd też w procesie pozyskiwania wiedzy dąży się do tego, by wybrać te atrybuty, które są najbardziej wrażliwe diagnostycznie. Proces wyboru atrybutów określany jest mianem *selekcji (wyboru) atrybutów* lub też *ekstrakcji atrybutów* (określenia nowych cech).

Autor skupił się nad zagadnieniem wyboru atrybutów ze zbioru atrybutów tradycyjnie stosowanych w diagnostyce maszyn. Główne metody wyboru atrybutów można podzielić na:

- bazujące na miarach prawdopodobieństwa np. [35],
- oparte na teorii zbiorów przybliżonych np. [92],
- bazujące na teorii algorytmów genetycznych np. [132],
- oparte na minimalizacji entropii zbioru przykładów np. [79],
- z zastosowaniem sieci neuronalnych np. [121],

- z zastosowaniem algorytmu indukcyjnego np. [59],
- z zastosowaniem analizy wrażliwości np. [16].

Założmy, że rozpatrywać będziemy zbiór przykładów, w którym atrybut decyzyjny d jest atrybutem binarnym, takim, że $d = \{\oplus, \ominus\}$. Ponadto założmy, że wyboru atrybutów będziemy dokonywać na podstawie analizy wartości pewnej funkcji. Funkcja ta powinna spełniać następujące wymagania [69]:

1. osiąga swoje maksimum wtedy, kiedy wszystkie uzyskane podzbiory przykładów są homogeniczne¹,
2. osiąga minimum wtedy, gdy w każdym podzbiorze znajduje się 50% przykładów klasy \oplus i 50% przykładów klasy \ominus ,
3. jest silnie monotoniczna w pobliżu ekstremum (100% przykładów klasy \oplus i 0% przykładów klasy \ominus lub odwrotnie) oraz płaska w okolicy przypadku 50%—50% przykładów klasy \oplus i \ominus .

5.1 Metody selekcji atrybutów bazujące na miarach prawdopodobieństwa

Jedną z metod ograniczania liczby uwzględnianych atrybutów jest metoda polegająca na analizie macierzy kowariancji wartości atrybutów poszczególnych przykładów [16]. Wyznaczyć można wówczas wektory własne i wartości własne tej macierzy i na podstawie analizy wartości własnych ograniczać liczbę uwzględnianych atrybutów.

Inna metoda polega na analizie korelacji wartości atrybutów dla poszczególnych przykładów [36]. Dla ograniczenia liczby uwzględnianych atrybutów grupuje się te atrybuty, których wartości są skorelowane ze sobą. Następnie, do dalszej analizy, z każdej grupy wybierany jest jeden atrybut (reprezentant). Wyboru dokonuje się na podstawie zaleceń diagnostycznych (np. [123]). Zalecenia te pozwalają ustalić, jaka cecha diagnostyczna jest charakterystyczna dla rozpatrywanego zbioru stanów maszyny. W ten sposób zostaje wskazany atrybut, który w myśl tych zaleceń powinien być najbardziej wrażliwy diagnostycznie.

¹W podzbiorze znajdują się tylko przykłady z jednej klasy

Przykładowy sposób wyznaczania atrybutów relewantnych z zastosowaniem opisanej wyżej metody korelacji wartości atrybutów przedstawiono jako algorytm 5.1.

5.2 Metoda selekcji atrybutów oparta na teorii zbiorów przybliżonych

Jak już wspomniano, elementy teorii zbiorów przybliżonych opisano w dodatku do rozprawy.

Dany jest system informacyjny $\mathcal{S} = (U, C)$ według (A.1) oraz zbiór atrybutów $A \subseteq C$. Atrybut $a \in A$ jest *redundantny* (*nadmiarowy*) w zbiorze atrybutów A systemu \mathcal{S} , jeżeli [117]:

$$\text{IND}(A) = \text{IND}(A \setminus \{a\}), \quad (5.1)$$

w przeciwnym razie atrybut a jest *relewantny* (*istotny*) w A . Jeżeli wszystkie atrybuty $a \in A$ są istotne, wówczas zbiór A jest nazywany *ortogonalnym*.

Zbiór wszystkich atrybutów istotnych zbioru A nazywa się *rdzeniem* systemu \mathcal{S} , i oznacza się jako $\text{CORE}(\mathcal{S})$.

Jeżeli $B \subseteq A$, i każdy atrybut $a \in B$ jest istotny w B oraz $\text{IND}(B) = \text{IND}(A)$, wówczas podzbiór B nazywa się *reduktem* w systemie \mathcal{S} i oznacza R . Innymi słowy, redukt jest minimalnym podzbiorem zbioru atrybutów A , które rozróżniają wszystkie przykłady $u_j \in U$ w systemie informacyjnym \mathcal{S} . Rodzinę wszystkich reduktów systemu \mathcal{S} oznacza się $\text{RED}(\mathcal{S})$. Zależność pomiędzy rdzeniem a reduktami systemu informacyjnego jest następująca:

$$\text{CORE}(\mathcal{S}) = \bigcap_{R \in \text{RED}(\mathcal{S})} R. \quad (5.2)$$

Zadanie wyboru atrybutów istotnych można przedstawić w sposób następujący [96]. Dany jest system informacyjny $\mathcal{S} = (U, A)$ według (A.1). Niech $[\text{IND}^X(B)]$ oznacza rodzinę relacji równoważności $\text{IND}^X(B)$ określonych dla $B \subseteq A$ i $X \subseteq U$.

Rozważać będziemy atrybut $a \in A$, zbiór elementów $X \subseteq U$ i zbiór $a(X) = \{a(u) : u \in X\}$ wartości atrybutu a na X . Niech $[\text{IND}^X(a)] = \{E_1^B, \dots, E_m^B\}$ (por. opis w dodatku na stronie 152). Ponadto niech e, e_1, \dots, e_m oznaczają liczności zbiorów odpowiednio X, E_1^B, \dots, E_m^B . Jeżeli $e = e_1 + \dots + e_m$, liczbę $W^X(a)$ par elementów zbioru X rozróżnialnych przez atrybut a można obliczyć jako [96]:

Algorytm 5.1 Sposób selekcji atrybutów metodą opartą na analizie korelacji**Funkcja** Korelacja

argumenty: U - zbiór przykładów
 A - zbiór atrybutów
 α_c - maksymalna korelacja atrybutów relewantnych
 α_k - maksymalna liczba atrybutów relewantnych
zwracane: $R \subseteq A$ - zbiór atrybutów relewantnych

Początek

Dla $i = 1$ **do** $\text{card}(A)$

Dla $j = i + 1$ **do** $\text{card}(A)$

$k_{ij} \leftarrow \text{ObliczKorelację}(a_i(), a_j())$

$K \leftarrow K \cup \{k_{ij}\}$

Koniec

Koniec

Sortuj rosnąco K

Dla $i = 1$ **do** $\text{card}(K)$

Dla $j = 1$ **do** $\text{card}(K)$

Jeżeli $k_{ij} \geq \alpha_c$ **i** $\text{card}(P) < \alpha_k$

$P \leftarrow P \cup \{a_i, a_j\}$ /* $p \in P$ oraz $P = \{a_1, a_2\}$ */

Koniec

Koniec

Koniec

Dla $i = 1$ **do** $\text{card}(P)$

$a_i \leftarrow \text{WskażAtrybut}(a_1, a_2)$, gdzie $\{a_1, a_2\} \in p_i$

Jeżeli $a_i \notin R$

$R \leftarrow R \cup \{a_i\}$

Koniec

Koniec

Koniec

$$W^X(a) = \frac{\sum_{\substack{i=1 \\ i \neq j}}^m e_i e_j}{2} = \frac{e^2 - \sum_{i=1}^m e_i^2}{2}. \quad (5.3)$$

Można wykazać [96], że jeżeli $[\text{IND}^X(B)] = \{E_1^B, \dots, E_m^B\}$ i $a \in A \setminus B$, wówczas:

$$[\text{IND}^X(B \cup \{a\})] = [\text{IND}^{E_1^B}(a)] \cup \dots \cup [\text{IND}^{E_m^B}(a)], \quad (5.4)$$

$$W_B^X(a) = W^{E_1^B}(a) + \dots + W^{E_m^B}(a). \quad (5.5)$$

Sposób obliczania reduktu minimalnego przedstawia algorytm 5.2 [96], zaś sposób obliczania reduktu losowego przedstawia algorytm 5.3 [96].

5.3 Metoda selekcji atrybutów bazująca na teorii algorytmów genetycznych

Idea algorytmów genetycznych (ang. *genetic algorithms*) oparta jest na darwinowskiej teorii doboru naturalnego [50]. Załóżmy, że dane są elementy pewnego zbioru Z i dana jest funkcja celu $f(z)$. Rozpatrywane jest zdanie poszukiwania maksimum (lub minimum) funkcji f . Algorytm genetyczny symuluje przebieg naturalnej ewolucji [77]:

1. elementy $z \in Z$ reprezentowane są w postaci tzw. „chromosomów”,
2. początkowa populacja, jako zbiór „chromosomów”, generowana jest w sposób losowy,
3. rolę środowiska odgrywa funkcja $f(z)$, która traktowana jest jako funkcja dopasowania (przystosowania) „chromosomu” do warunków tego środowiska,
4. nowa populacja generowana jest losowo poprzez mutacje, krzyżowanie i inne operacje, przy czym chromosomy o niższej wartości funkcji dopasowania zastępowane są przez „chromosomy”, dla których funkcja ta przyjmuje wyższe wartości,
5. algorytm kończy działanie po zadanej liczbie „pokoleń” lub po osiągnięciu zadowalającego rezultatu; wynikiem działania algorytmu jest „chromosom” o największej znalezionej wartości funkcji celu.

Algorytm 5.2 Sposób obliczania reduktu minimalnego**Funkcja** Redukt minimalny

argumenty: U - uporządkowany zbiór przykładów
 A - uporządkowany zbiór atrybutów
 zwracane: $R \subseteq A$ - redukt

Początek $R \leftarrow \emptyset$ $L \leftarrow \{U\}$ **Powtarzaj**Dla każdego $a \in A$ Dla każdego $E_i^B \in L$ Szukaj $[\text{IND}^{E_i^B}(a)]$ Oblicz $W^{E_i^B}(a)$ **Koniec** $W_R^U(a) \leftarrow W^{E_1^B}(a) + \dots + W^{E_m^B}(a)$ **Koniec**Wybierz atrybut a dla którego $W_R^U(a)$ osiąga maksimum $A \leftarrow A \setminus \{a\}$ $R \leftarrow R \cup \{a\}$ $L \leftarrow [\text{IND}^{E_1^B}(a)] \cup \dots \cup [\text{IND}^{E_m^B}(a)]$ **Dopóki** $W_R^U(a) = 0 \vee A = \emptyset$ **Koniec**

W zadaniu selekcji atrybutów ($Z = A$) „chromosom” może być reprezentowany przez ciąg binarny — jak w klasycznym algorytmie genetycznym. Wówczas każdy „chromosom” odpowiada podzbiorowi atrybutów w ten sposób, że:

$$1001001100 \leftrightarrow \{a_1, a_4, a_7, a_8\}.$$

Dla zbioru „chromosomów” binarnych można zastosować klasyczne operacje mutacji i krzyżowania, przy czym określone są prawdopodobieństwa wystąpienia tych operacji. Zwykle przyjmuje się [77, 93], że prawdopodobieństwo p_c operacji krzyżowania

Algorytm 5.3 Obliczanie reduktu losowego

Funkcja Redukt losowy

argumenty: U - uporządkowany zbiór przykładów
 A - uporządkowany zbiór atrybutów
zwracane: $R \subseteq A$ - redukt

Początek $R \leftarrow \emptyset$ $L \leftarrow \{U\}$ **Powtarzaj****Dla każdego** $a \in A$ **Dla każdego** $E_i^B \subset L$ Szukaj $[\text{IND}^{E_i^B}(a)]$ Oblicz $W^{E_i^B}(a)$ **Koniec** $W_R^U(a) \leftarrow W^{E_1^B}(a) + \dots + W^{E_m^B}(a)$ **Koniec**Wybierz losowo atrybut a , którego prawdopodobieństwo wynosi

$$P(a) \leftarrow \frac{W_R^U(a)}{\sum W_R^U(a_j)} \text{ dla } a \in A$$

 $A \leftarrow A \setminus \{a\}$ $R \leftarrow R \cup \{a\}$ $L \leftarrow [\text{IND}^{E_1^B}(a)] \cup \dots \cup [\text{IND}^{E_m^B}(a)]$ **Dopóki** $W_R^U(a) = 0 \vee A = \emptyset$ **Koniec**

jest znacznie większe od prawdopodobieństwa p_m operacji mutacji. Szerzej zagadnienia związane z określaniem tych prawdopodobieństw omówione są w literaturze (m.in. [77, 38, 84, 99]).

Funkcję dopasowania podzbioru atrybutów $B \subseteq A$ można definiować w różny sposób. Dla przykładu, w [133] przyjęto, że:

$$f(B) = \frac{k - L_B}{k} + \frac{2C_B}{n^2 - n}, \quad (5.6)$$

gdzie k jest liczbą atrybutów, L_B określa liczbę jedynek w podzbiorze B , C_B — liczbę par elementów (o różnych wartościach atrybutu decyzyjnego) rozróżnianych przez atrybutu podzbioru B , natomiast n jest liczbą wszystkich elementów.

W [137] funkcję tę zdefiniowano następująco:

$$f(B) = \eta(B) - \frac{\text{cost}(B)}{\eta(B) + 1} + \text{cost}_{max}, \quad (5.7)$$

gdzie $\eta(B)$ jest sprawnością klasyfikatora uzyskaną dla podzbioru B atrybutów, $\text{cost}(B)$ oznacza sumaryczny koszt pomiarów wartości atrybutów z podzbioru B , cost_{max} jest natomiast określany jako maksymalny koszt jakim może być „obciążony” podzbiór B (w najprostszym przypadku jest to suma kosztów dla wszystkich atrybutów ze zbioru A).

Sposób postępowania w przypadku selekcji atrybutów metodą bazującą na teorii algorytmów genetycznych przedstawia algorytm 5.4. Sposób ten przedstawiono również graficznie na rys. 5.1.

5.4 Metoda selekcji atrybutów z wykorzystaniem kryterium minimum entropii

Entropia określa stopień nieuporządkowania danych. Matematycznie można wykazać, że ilość informacji wzrasta, gdy entropia maleje [80].

Rozpatrzmy podzbiór zbioru przykładów $U_i \subset U$. Niech p_i^+ oznacza prawdopodobieństwo wystąpienia w zbiorze przykładu pozytywnego, p_i^- oznacza prawdopodobieństwo wystąpienia w zbiorze przykładu negatywnego. Entropię podzbioru U_i oblicza się następująco [79]:

$$H(U_i) = -p_i^+ \log_2 p_i^+ - p_i^- \log_2 p_i^-. \quad (5.8)$$

Prawdopodobieństwa p_i^+ oraz p_i^- obliczyć można następująco:

$$p_i^+ = \frac{n_i^+}{n_i^+ + n_i^-}, \quad (5.9)$$

Algorytm 5.4 Selekcja atrybutów oparta na teorii algorytmów genetycznych

Funkcja Genet

argumenty: A - zbiór atrybutów
 α_n - rozmiar populacji
 α_p - maksymalna liczba pokoleń
 p_c - prawdopodobieństwo krzyżowania
 p_m - prawdopodobieństwo mutacji
zwracane: $R \subseteq A$ - zbiór atrybutów relewantnych

Początek

Dla $i = 1$ **do** α_n

Wygeneruj losowo chromosom (podzbiór atrybutów) $B_i \in A$

$B \leftarrow B \cup \{B_i\}$

Koniec

$R' \leftarrow \{\emptyset, 0\}$ /* $R' = \{r_1, r_2\}$ */

Dopóki $i \leq \alpha_p$

Dla $j = 1$ **do** α_n

Oblicz $f(B_i)$

$B' \leftarrow B' \cup \{B_i, f(B_i)\}$ /* $b \in B'$ oraz $b = \{b_1, b_2\}$ */

Koniec

OperacjeGenetyczne(B', p_c, p_m)

Znajdź $b' \in B'$ takie, że b_2 jest maksymalne

Jeżeli $f(b') \geq r_2$

$R' \leftarrow \{b', f(b')\}$

Koniec

$B \leftarrow \{b_1 \in B'\}$

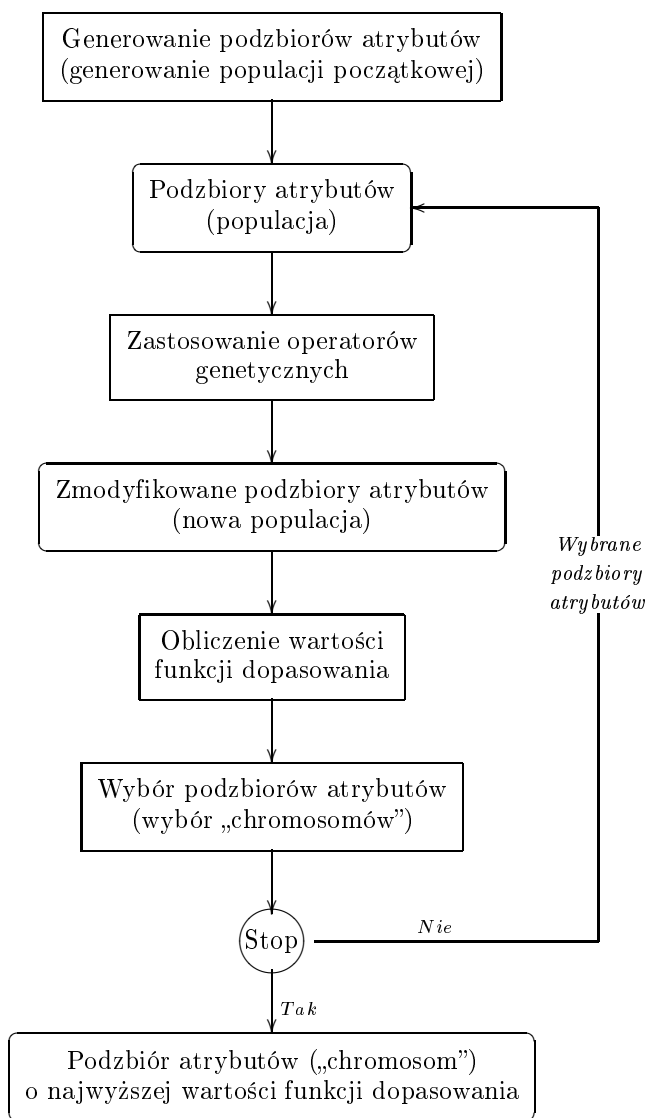
$B' \leftarrow \{\emptyset, 0\}$

$i \leftarrow i + 1$

Koniec

$R \leftarrow r_1$

Koniec



Rysunek 5.1: Sposób selekcji atrybutów metodą bazującą na teorii algorytmów genetycznych

$$p_i^- = \frac{n_i^-}{n_i^+ + n_i^-}, \quad (5.10)$$

gdzie n_i^+ oznacza liczbę przykładów pozytywnych w U_i , a n_i^- oznacza liczbę przykładów negatywnych w U_i .

Niech wartości atrybutu a dzielą zbiór U na podzbiory $U_i, i = 1, \dots, k$, gdzie k jest liczbą wartości atrybutu a . Entropię zbioru U podzielonego na podzbiory U_i można obliczyć jako:

$$H(U, a) = \sum_{i=1}^k P(U_i) \cdot H(U_i), \quad (5.11)$$

gdzie $P(U_i)$ jest prawdopodobieństwem wystąpienia przykładu z podzbioru U_i i może być obliczone jako:

$$P(U_i) = \frac{\text{card}(U_i)}{\text{card}(U)} \quad (5.12)$$

Zatem zmiana entropii spowodowana podziałem zbioru U wartościami atrybutu a wynosi:

$$I(U, a) = H(U) - H(U, a). \quad (5.13)$$

Atrybuty, dla których $I(U, a)$ osiąga najwyższe wartości, traktowane są jako atrybuty relewantne.

Sposób wyznaczenia atrybutów relewantnych omawianą metodą przedstawiono za pomocą algorytmu 5.5.

5.5 Metoda selekcji atrybutów z zastosowaniem sieci neuronalnych

Sieci neuronalne są przydatne przy realizacji bardzo wielu zadań praktycznych. W dziedzinie diagnostyki technicznej najwięcej prac z zastosowaniem sieci neuronalnych dotyczy detekcji uszkodzeń [67]. Sieć neuronalna jest modelem typu „czarnej skrzynki”. Łącząc model neuronalny z modelem rozmytym, można wyeliminować tę wadę. Jednak wówczas liczba uzyskiwanych reguł gwałtownie rośnie wraz ze wzrostem liczby wejść [66]. Sieci neuronalne mogą być również wykorzystane do ograniczenia liczby wejść poprzez ograniczanie liczności rozpatrywanego zbioru atrybutów (np. [114, 8, 120]).

Algorytm 5.5 Sposób selekcji atrybutów metodą minimum entropii**Funkcja** Minimum entropii

argumenty: A - zbiór atrybutów $a \in A$
 $V(A)$ - zbiór wartości atrybutów $a \in A$
 U - zbiór przykładów $u \in U$
 α_k - maksymalna wymagana liczba atrybutów
 α_I - minimalna wymagana zamiana entropii
zwracane: $R \subseteq A$ - zbiór atrybutów relewantnych

Początek

Dla $i = 1$ **do** $\text{card}(A)$

Sortuj rosnąco U według $V(a_i)$

$k \leftarrow 0$

Dla $j = 1$ **do** $\text{card}(U)$

Jeżeli nastąpiła zmiana wartości $a_i(u_j)$ **lub** $k = 0$

$k \leftarrow k + 1$

$U_k \leftarrow \emptyset$

Koniec

$U_k \leftarrow U_k \cup \{u_j\}$

Koniec

$$H(U, a_i) \leftarrow \sum_{i=1}^k P(U_i) \cdot H(U_i)$$

$$I(U, a_i) \leftarrow H(U) - H(U, a_i)$$

$$J \leftarrow J \cup \{I(U, a_i), i\}$$

Koniec

Sortuj J według $I(U, a_i)$

$R \leftarrow \emptyset$

Dla $i = 1$ **do** $\text{card}(J)$

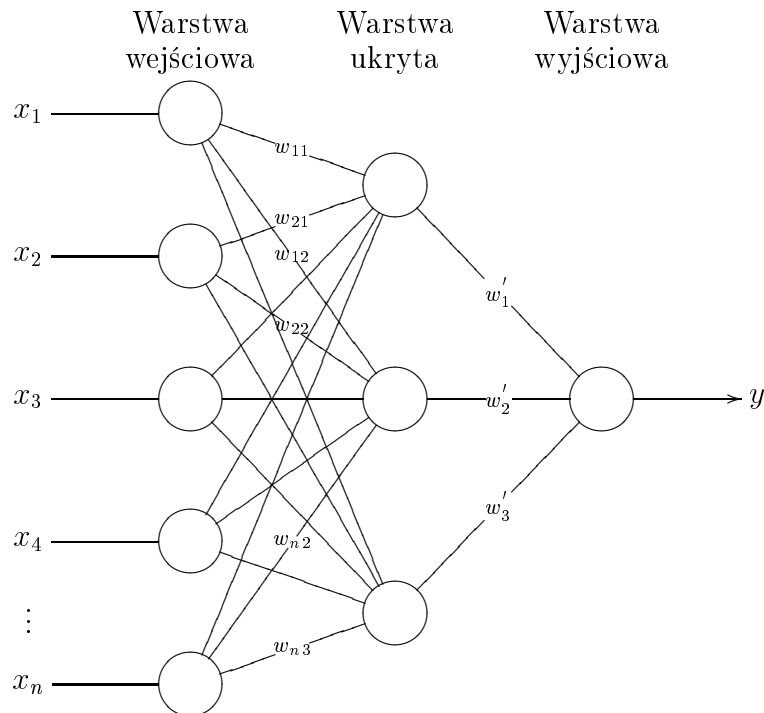
Jeżeli $J_i \geq \alpha_I$ **i** $\text{card}(R) \leq \alpha_k$

$R \leftarrow R \cup \{a_i\}$

Koniec

Koniec

Koniec



Rysunek 5.2: Sieć neuronalna zawierająca warstwę wejściową, jedną warstwę ukrytą i jeden węzeł w warstwie wyjściowej

Sieci neuronalne

Sieć neuronalna (ang. *neural network*), nazywana często sztuczną siecią neuronową, jest komputerowym modelem ludzkiej sieci neuronowej i podobnie jak ona zbudowana jest z połączonych w odpowiedni sposób węzłów (neuronów). Każdą sieć tworzą warstwy węzłów — wejściowa, jedna lub więcej warstw ukrytych i wyjściowa — jak pokazano na rys. 5.2. Połączeniom węzłów przypisane są wagi w_{ij} .

Ogólnie, warstwa wejściowa zawiera wyłącznie węzły umowne [17], które przesyłają wartości x_i ich wejść bezpośrednio na wyjścia. Każdy węzeł warstwy ukrytej lub wyjściowej l przekształca wyjścia $x_{l-1,i}$ warstwy poprzedzającej o numerze $(l - 1)$ w swoje wyjście $x_{l,j}$:

$$x_{l,j} = f \left(\sum_{i=1}^n w_{l,j,i} \cdot x_{l-1,i} \right), \quad (5.14)$$

Znanych jest wiele modeli sieci neuronalnych. Wśród nich należałoby wymienić [42]:

- sieci z wsteczną propagacją błędów FFBP (*Feed Forward Back Propagation*) [111],
- sieci samoorganizujące się (np. sieci Kohonena) [60],

- sieci GMDH (*Group Method of Data Handling*) [105],
- sieci modularne [4].

Najbardziej popularnym modelem jest obecnie sieć z wsteczną propagacją błędu [42]. Taki też model stosowany jest do rozwiązywania zadań związanych z selekcją atrybutów [2, 121].

Selekcja atrybutów

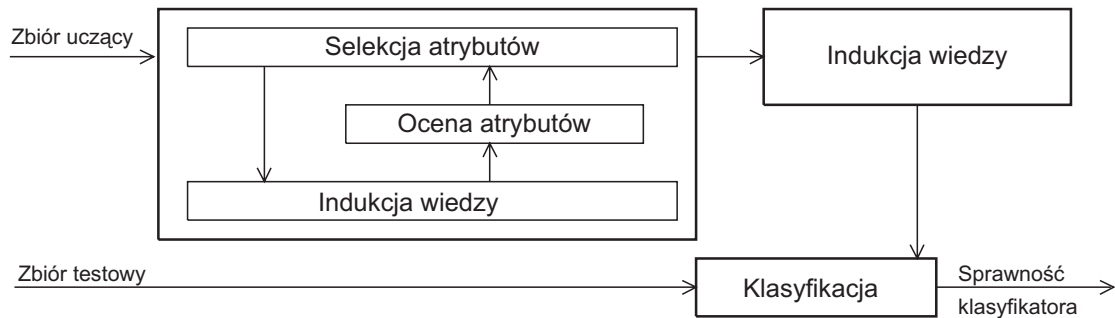
W przypadku zastosowania metod selekcji atrybutów z zastosowaniem sieci neuronalnych liczba wejść równa jest liczbie atrybutów warunkowych. Wejściami sieci neuronalnych są wartości poszczególnych atrybutów. Wyjściem sieci jest natomiast wartość określająca np. numer (kod) klasy.

Zastosowanie sieci neuronalnych do selekcji atrybutów zaproponowano w [2]. W przypadku sieci neuronalnej ze wsteczną propagacją błędu selekcja atrybutów realizowana może być z zastosowaniem [121]:

- metody „przycinania” (pruning) wag,
- metody sumy wag,
- zmodyfikowanej metody Karina,
- analizy wrażliwości.

Zastosowanie metody „przycinania” wag sprowadza się do prób odrzucania wag łączących komórki wejściowe z warstwą ukrytą. Przyjmuje się, że możliwość odrzucenia stosunkowo dużej liczby wag wychodzących z danej komórki wejściowej świadczy o mniejszej istotności analizowanego wejścia. W metodzie sumy wag istotność danego wejścia oceniana jest na podstawie sumy wartości bezwzględnych wag wychodzących z tego wejścia. Kolejna metoda bazuje na algorytmie Karina, który przeznaczony jest w zasadzie do oceny istotności poszczególnych wag w sieci [120]. Natomiast istotność wejść do sieci wyznaczana jest jako maksimum istotności wag łączących dane wejście z warstwą ukrytą. Ostatnia z metod bazuje na klasycznej analizie wrażliwości.

Innym możliwym podejściem jest jednoczesne zastosowanie wyżej wymienionych metod. Każda z metod może prowadzić do uzyskania innych wyników. Podjęto więc próbę [120] wyznaczenia „wypadkowej” istotności wejść (atrybutów) przez obliczenie średniej arytmetycznej i średniej geometrycznej istotności wejść wyznaczonych z zastosowaniem różnych metod.



Rysunek 5.3: Selekcja atrybutów z zastosowaniem algorytmu indukcyjnego „uczenia maszynowego” [59]

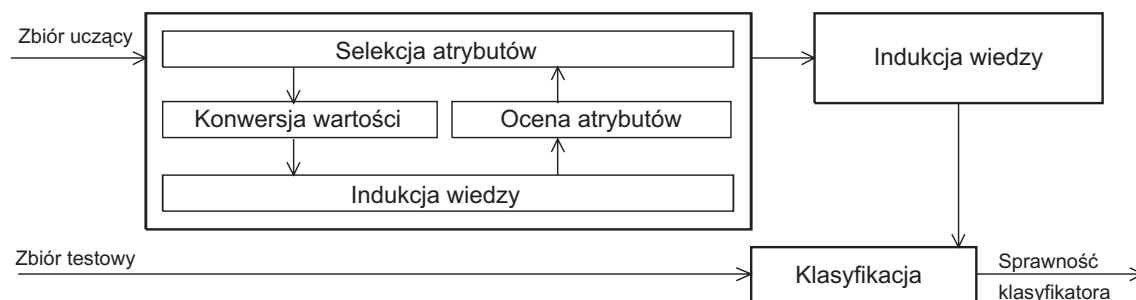
5.6 Metoda selekcji atrybutów z zastosowaniem algorytmu indukcyjnego „uczenia maszynowego”

Idea selekcji atrybutów z zastosowaniem algorytmu indukcyjnego (ang. *wrapper approach*) została przedstawiona na rys. 5.3. Algorytm indukcji wiedzy, stosowany do pozyskiwania wiedzy, został tutaj wykorzystany również podczas selekcji atrybutów jako sposób oceny podzbioru atrybutów. Odpowiedni algorytm selekcji atrybutów umożliwia zaś właściwy dobór atrybutów na podstawie tak otrzymanej oceny [52].

Proponowana w literaturze metoda selekcji ma — zdaniem autora — pewną wadę. Konwersji wartości atrybutów należy dokonać przed zastosowaniem tej metody, a selekcji atrybutów z zastosowaniem tej metody dokonuje się na podstawie atrybutów jakościowych. Pewna ilość informacji niesiona przez atrybuty ilościowe jest więc tracona. Ponadto wartości graniczne wyznaczone metodami wieloatrybutowymi zależą od liczby atrybutów. Zatem dla wybranego podzbioru atrybutów zbiór wartości granicznych może nie być zbiorem optymalnym. Autor zaproponował więc modyfikację tej metody polegającą na tym, że etap konwersji wartości atrybutów ilościowych na jakościowe jest włączony do procesu selekcji atrybutów (por. rys. 5.4) .

Sposób selekcji atrybutów zgodnie z omawianą metodą dla wersji z zastosowaniem losowego wyboru atrybutów [71] można przedstawić tak jak w algorytmie 5.6.

W algorytmie tym A jest wyjściowym zbiorem atrybutów, B i C są podzbiórmi atrybutów, natomiast err, err', err'' są wartościami błędów klasyfikacji. Algorytm uczący, który na podstawie zadanego podzbioru atrybutów B lub C i dla zadanych zbiorów przykładów uczących E^u , umożliwia pozyskanie wiedzy, może być dowolnym



Rysunek 5.4: Metoda selekcji atrybutów z zastosowaniem algorytmu indukcyjnego „uczenia maszynowego” i konwersją wartości atrybutów

algorytmem indukcyjnym pozyskiwania wiedzy.

Zmodyfikowaną wersję poprzedniego algorytmu, w której zamiast oceny podzbioru według błędu klasyfikacji, zastosowano sprawdzanie niesprzeczności, przedstawiono jako algorytm 5.7 [72]. Powyższy algorytm wymaga, by w każdym kroku był losowo generowany podzbiór atrybutów B . Jeżeli liczba atrybutów K' w zbiorze B jest mniejsza od liczby atrybutów w najmniejszym z dotychczas znalezionych podzbiorów, wówczas sprawdzany jest poziom niesprzeczności zbioru przykładów U opisanych atrybutami B . Jeżeli poziom niesprzeczności jest niższy od zadanego poziomu γ , wtedy zbiór B jest najlepszym z dotychczas znalezionych rozwiązań ($C = B$). Iteracja jest powtarzana I -krotnie. Funkcja `LosujPodzbiór()` zwraca losowo wybrany podzbiór zbioru atrybutów A . Funkcja `PoziomNiespN()` zwraca wartość poziomu niesprzeczności przykładów U opisanych atrybutami B .

Kolejna modyfikacja polega na zastosowaniu selekcji przyrostowej (por. algorytm 5.8). W algorytmie 5.8 funkcja `PoziomNiespI()` ma działanie podobne do funkcji `PoziomNiespN()` występującej w algorytmie 5.7. Dodatkowo sprzeczne przykłady ze zbioru E_1 zapisywane są w *SpData*. Najczęściej zadawana wartość γ jest równa 0 lub jest wartością poziomu niesprzeczności przy zastosowaniu wszystkich atrybutów $\gamma = \text{PoziomNiespN}(A, U)$. W algorytmie tym dane dzieli się na dwa podzbiory: E_0 — zawierający p procent losowo wybranych przykładów ze zbioru U , oraz E_1 — zawierający pozostałe przykłady ze zbioru U . Podzbiór atrybutów C uzyskany na podstawie algorytmu selekcji atrybutów według kryterium niesprzeczności dla E_0 stanowi podstawę do określenia poziomu niesprzeczności dla podzbioru zbioru przykładów E_1 . Wiadomo, że $\beta \leq \gamma$. Stąd, jeżeli poziom niesprzeczności dla E_1 nie przekracza $\gamma - \beta$, obliczenia

Algorytm 5.6 Losowa selekcja atrybutów

Funkcja SWAP

argumenty: U - zbiór przykładów
 A - uporządkowany zbiór atrybutów
 I - liczba iteracji
 K - początkowa minimalna liczba atrybutów

zwracane: $C \subseteq A$ - minimalny podzbiór atrybutów
 err - błąd klasyfikacji

Początek $err' \leftarrow 0$ $i \leftarrow 0$ **Dla** $i = 1$ **do** K $B \leftarrow \text{LosujPodzbiór}(A)$ $K' \leftarrow \text{card}(B)$ $err'' \leftarrow \text{Błąd}(\text{AlgorytmUczący}(B, E^u, \text{Null}))$ **Jeżeli** $err'' < err' \vee (err'' = err' \wedge K' < K)$ $k \leftarrow 0$ $err' \leftarrow err''$ $K \leftarrow K'$ $C \leftarrow B$ **Koniec** $i \leftarrow i + 1$ **Koniec** $err \leftarrow \text{Błąd}(\text{AlgorytmUczący}(C, E^u, E^t))$ **Koniec**

zostają zakończone. W przeciwnym wypadku, przykłady $SpData$ dołączane są do E_0 i usuwane z E_1 . Selekcja jest powtarzana do momentu znalezienia zbioru minimalnego. Jeżeli taki zbiór nie może być znaleziony, rozwiązaniem jest cały zbiór atrybutów A .

Algorytm 5.7 Selekcja atrybutów wg kryterium niesprzeczności

Funkcja SWAN

argumenty: U - uporządkowany zbiór przykładów
 A - uporządkowany zbiór atrybutów
 I - liczba iteracji
 γ - dopuszczalny poziom niesprzeczności

zwracane: $C \subseteq A$ - minimalny podzbiór atrybutów
 β - poziom niesprzeczności

Początek $K \leftarrow \text{card}(A)$ **Dla** $i = 1$ **do** I $B \leftarrow \text{LosujPodzbiór}(A)$ $K' \leftarrow \text{card}(B)$ $\beta \leftarrow \text{PoziomNiespN}(B, U)$ **Jeżeli** $K' < K \wedge \beta < \gamma$ $C \leftarrow B$ $K \leftarrow K'$ **W przeciwnym razie jeżeli** $K' = K \wedge \beta < \gamma$ $C \leftarrow B$ **Koniec****Koniec****Koniec**

Algorytm 5.8 Przyrostowa selekcja atrybutów

Funkcja SWAI

argumenty: U - uporządkowany zbiór przykładów
 A - uporządkowany zbiór atrybutów
 p - procent przykładów stosowanych do selekcji atrybutów
 γ - dopuszczalny poziom niesprzeczności
 I - liczba iteracji
zwracane: $C \subseteq A$ - minimalny podzbiór atrybutów

Początek

$E_0 \leftarrow p$ procent losowo wybranych przykładów z U

$E_1 \leftarrow U \setminus E_0$ /* pozostałe przykłady */

Powtarzaj

$\beta \leftarrow \text{SWAN}(E_0, A, I, \gamma, C)$

Jeżeli $\text{PoziomNiespl}(C, E_1, SpData) \leq \gamma - \beta$

Zwróć C

W przeciwnym razie

Dopisz($SpData, E_0$)

Usuń($SpData, E_1$)

Koniec**Koniec****Koniec**

Rozdział 6

Kryteria i sposoby oceny rozpatrywanych metod redukcji i selekcji informacji

Jednym z istotnych zadań w przeprowadzonych badaniach stała się ocena badanych metod redukcji i selekcji informacji. Poniżej przedstawiono zbiór kryteriów oraz sposoby oceny tych metod.

6.1 Kryteria oceny rozpatrywanych metod redukcji i selekcji informacji

Metody wyznaczania wartości granicznych pozwalające na konwersję wartości ilościowych, a tym samym na redukcję informacji, a także metody selekcji atrybutów relewantnych, trudno jest porównywać i oceniać bezpośrednio. Możliwa jest jednak pośrednia ocena tych metod, polegająca na ocenie jakości wiedzy pozyskanej na podstawie przykładów, dla których dokonano redukcji i selekcji informacji.

Do oceny metod konieczne jest określenie kryteriów, według których ocena ta będzie dokonywana. Autor przyjął następujący zbiór kryteriów oceny wyżej wymienionych metod:

- poprawność pozyskanej wiedzy określana na podstawie poprawności działania klasyfikatora, w tym przede wszystkim przez:

- łączny błąd klasyfikacji (sprawność klasyfikatora),
- błędy klasyfikacji (błąd pominięcia i błąd niesłusznego zaliczenia),
- mała liczba wartości granicznych,
- mała liczba atrybutów relewantnych,
- mała liczba pozyskanych reguł.

Sprawność klasyfikatora (lub związany z nią błąd klasyfikacji) jest najbardziej znanym i najprostszym kryterium merytorycznej oceny pozyskiwanej wiedzy. Uzyskanie wysokiej sprawności klasyfikacji jest często głównym kryterium uznania metod pozyskiwania wiedzy za przydatne w danej dziedzinie. Sprawność klasyfikatora oblicza się według (2.17).

Oprócz sprawności klasyfikatora ważną rolę odgrywają błędy: błąd pominięcia i błąd niesłusznego zaliczenia [3]. Błędy te często różnią się od błędu łącznego, a pozwalają dokonać merytorycznej oceny wiedzy pod względem braku klasyfikacji lub złej klasyfikacji przykładów. Dla zbioru przykładów, dla którego liczba klas $K > 1$, określić można następujące oceny błędu klasyfikatora [3]:

1. *Łączny błąd klasyfikacji:*

$$\epsilon = \frac{n_{err}}{\text{card}(E^t)} = 1 - \eta, \quad (6.1)$$

gdzie n_{err} oznacza liczbę przykładów błędnie sklasyfikowanych.

2. *Błąd pominięcia:*

$$\epsilon_o = \frac{1}{K} \sum_{k=1}^K \frac{\text{card}(E_k^+)}{\text{card}(E_k^t)}, \quad (6.2)$$

gdzie E_k^+ oznacza podzbiór przykładów które nie zostały sklasyfikowane poprawnie, E_k^t jest podzbiorem przykładów klasy o indeksie k .

3. *Błąd niesłusznego zaliczenia:*

$$\epsilon_c = \frac{1}{K} \sum_{k=1}^K \frac{1}{\text{card}(E^t \setminus E_k^t)} \cdot \sum_{\substack{j=1 \\ j \neq k}}^K \text{card}(E_{kj}^-), \quad (6.3)$$

gdzie E_{kj}^- jest zbiorem przykładów klasy j niesłuszenie zaklasyfikowanych do klasy k .

W. Moczulski w [85] wskazał na istotną wadę powyższych estymatorów, a mianowicie fakt nieuwzględniania empirycznego prawdopodobieństwa wystąpienia przykładu należącego do danej klasy. Aby usunąć tę wadę, w [85] zaproponowano wprowadzenie wagi:

$$w_k = \frac{\text{card}(E^t)}{K \text{card}(E_k^t)}, \quad k = 1, \dots, K, \quad (6.4)$$

oraz:

$$w = \sum_{k=1}^K w_k. \quad (6.5)$$

Wówczas błędy (6.1) do (6.3) można zastąpić następującymi błędami ważonymi [85]:

1. *Łączny ważony błąd klasyfikacji:*

$$\epsilon^w = \frac{\sum_{k=1}^K w_k n_{err}^k}{\text{card}(E^t)}, \quad (6.6)$$

gdzie n_{err}^k oznacza liczbę błędnie sklasyfikowanych przykładów klasy k .

2. *Ważony błąd pominięcia:*

$$\epsilon_o^w = \frac{1}{w} \sum_{k=1}^K w_k \frac{\text{card}(E_k^+)}{\text{card}(E_k^t)}. \quad (6.7)$$

3. *Ważony błąd niestusznego zaliczenia:*

$$\epsilon_c^w = \frac{1}{w} \sum_{k=1}^K \frac{w_k}{\text{card}(E^t \setminus E_k^t)} \cdot \sum_{\substack{j=1 \\ j \neq k}}^K \text{card}(E_{kj}^-). \quad (6.8)$$

Z praktycznego punktu widzenia, istotnym elementem oceny metod wyznaczania wartości granicznych jest uzyskanie małej liczby wartości granicznych, a co za tym idzie, po dokonaniu konwersji wartości, małej liczby wartości jakościowych atrybutów. W efekcie pozwala to na przedstawienie pozyskanej wiedzy w czytelnej dla specjalistów formie. Ułatwia to weryfikację pozyskanej wiedzy przez specjalistów z danej dziedziny zastosowania.

Uzyskanie małej liczby atrybutów relewantnych może być korzystne z dwóch powodów. Po pierwsze, mniejsza liczba atrybutów oznacza najczęściej uzyskanie prostszych, tzn. krótszych reguł. Po drugie, redukcja liczby atrybutów pociąga często za sobą skutki

ekonomiczne, gdyż może się wiązać z redukcją liczby punktów pomiarowych na badanej maszynie lub z redukcją liczby estymowanych cech sygnałów diagnostycznych.

Wydaje się, że im opis relacji zachodzących w maszynie jest prostszy, tym lepiej. Prostotę opisu relacji zachodzących w maszynie można oceniać poprzez wielkość pozyskanej porcji wiedzy. Wiedza najczęściej jest pozyskiwana w formie reguł, stąd ocena ilości pozyskanej wiedzy może polegać na ocenie liczby pozyskanych reguł, a także ich długości, określonej jako liczba przesłanek elementarnych, tworzących złożoną przesłankę reguły.

Spełnienie powyższych kryteriów warunkuje uznanie badanej metody wyznaczania wartości granicznych lub metody selekcji atrybutów za optymalną do zastosowania w dziedzinie diagnostyki maszyn

6.2 Sposoby oceny rozpatrywanych metod redukcji i selekcji informacji

Rozpatrując zagadnienie oceny metod redukcji i selekcji informacji dla przyjętych kryteriów, należy określić sposób, w jaki będzie oceniany stopień spełnienia poszczególnych kryteriów oraz sposób, w jaki będą wyznaczane błędy klasyfikacji przyjęte jako jedno z kryteriów oceny badanych metod.

6.2.1 Określenie stopnia spełnienia kryteriów oceny

Przyjęte kryteria oceny metod redukcji i selekcji informacji nie są oczywiście równoważne. Nieco inne kryteria odnoszą się do metod redukcji, a nieco inne do metod selekcji informacji. Konieczne zatem jest określenie ważności przyjętych kryteriów, przy czym, zdaniem autora, należy tego dokonać oddzielnie dla metod redukcji oraz metod selekcji informacji.

W tab. 6.1 porównano kryteria oceny metod redukcji informacji pomiędzy sobą. W ten sposób określono ważność każdego z nich. Przyjęto, że kryterium otrzymuje 2 punkty wówczas, gdy jego spełnienie jest ważniejsze od spełnienia innego kryterium, 0 — w wypadku przeciwnym. Gdy kryteria uznano za równoważne, przyznawano im po 1 punkcie. Dopuszczono możliwość stosowania ocen pośrednich. Suma tak przyznanych punktów pozwoliła na wyznaczenie wag poszczególnych kryteriów.

Tablica 6.1: Określenie wagi przyjętych kryteriów oceny metod redukcji informacji

	k_1^R	k_2^R	k_3^R	k_4^R	k_5^R
Niski błąd klasyfikacji (k_1^R)	-	0	1	1	0
Niski ważony błąd pominięcia (k_2^R)	2	-	2	1	1
Niski ważony błąd niesłusznego zaliczenia (k_3^R)	1	0	-	0	0
Mała liczba wartości granicznych (k_4^R)	1	1	2	-	0
Mała liczba reguł (k_5^R)	2	1	2	2	-
Waga kryterium	6	2	7	4	1

Przyjęto, że naczelnymi kryteriami są kryteria minimalnego ważonego błędu niesłusznego zaliczenia oraz minimalnego łącznego błędu klasyfikacji. W razie wykrycia niesprawności czy uszkodzenia badanego obiektu konsekwencje postawienia złej diagnozy przez system wnioskujący mogą być znacznie poważniejsze niż wtedy, kiedy system wnioskujący nie będzie w stanie podjąć decyzji. Stosunkowo istotna wydaje się również mała liczba wartości jakościowych atrybutów. Jak wspomniano wcześniej, pozwala to na pozyskiwanie wiedzy, która jest łatwa do zinterpretowania (np. przez operatora).

Po przeprowadzonej ocenie, uszeregowane według ważności wraz z wagami kryteria oceny metod redukcji informacji przedstawiają się następująco:

- niski ważony błąd niesłusznego zaliczenia: $w_{k_3^R} = 7$,
- niski łączny błąd klasyfikacji: $w_{k_1^R} = 6$,
- mała liczba wartości granicznych: $w_{k_4^R} = 4$,
- niski ważony błąd pominięcia: $w_{k_2^R} = 2$,
- mała liczba reguł: $w_{k_5^R} = 1$.

Z teoretycznego punktu widzenia, usunięcie atrybutów redundantnych (nadmiarowych) nie powinno powodować zwiększenia błędów klasyfikacji, a więc otrzymane błędy klasyfikacji powinny być takie same jak dla wyjściowego zbioru atrybutów. W związku z powyższym autor postanowił dla oceny metod selekcji informacji zastosować jedynie trzy kryteria, których porównanie przedstawiono w tab. 6.2.

Tablica 6.2: Określenie wagi przyjętych kryteriów oceny metod selekcji informacji

	k_1^S	k_2^S	k_3^S
Niski łączny błąd klasyfikacji (k_1^S)	-	0	0
Mała liczba atrybutów relewantnych (k_2^S)	2	-	0,5
Mała liczba reguł (k_3^S)	2	1,5	-
Waga kryterium	4	1,5	0,5

Przyjęto, podobnie jak poprzednio, że postawienie właściwej diagnozy jest kryterium naczelnym. Za najmniej istotne kryterium uznano kryterium małej liczby reguł. Koszty związane z przechowywaniem i analizą większej liczby reguł są, zdaniem autora, znacznie mniejsze od kosztów związanych np. z instalacją dodatkowego punktu pomiarowego.

Uszeregowane według ważności kryteria oceny metod selekcji informacji przedstawiają się następująco:

- niski łączny błąd klasyfikacji: $w_{k_1^S} = 4$,
- mała liczba atrybutów relewantnych: $w_{k_2^S} = 1,5$,
- mała liczba reguł: $w_{k_3^S} = 0,5$.

Ocena poszczególnych metod według przyjętych kryteriów będzie zatem ważoną sumą ocen spełnienia każdego z kryterium. Można przyjąć, że spełnienie danego kryterium oceniane będzie w skali 0–5 punktów, przy czym 5 punktów oznacza, że metoda całkowicie spełnia przyjęte kryterium, 0 punktów oznacza, że metoda w ogóle nie spełnia przyjętego kryterium.

6.2.2 Sposoby określania sprawności i błędów klasyfikacji

Do określenia błędów klasyfikacji stosuje się cztery podstawowe techniki (pozostawiono oryginalne angielskie nazwy tych technik ze względu na powszechne ich stosowanie i trudności z ich tłumaczeniem) [131]:

1. *Hold-out* polega na jednorazowym podziale zbioru przykładów E według (2.16), utworzeniu klasyfikatora na podstawie zbioru E^u i ocenie sprawności przy wykorzystaniu zbioru testowego E^t . Stosowanie tej techniki wymaga dużej liczby przykładów $\text{card}(E)$.
2. *Random subsampling* polega na N -krotnym wykonaniu czynności określonych przez *Hold-out*, przy czym podziały (2.16) dokonywane są niezależnie w kolejnych iteracjach. Oceny sprawności dokonuje się, stosując zależność (2.18). Uśredniona ocena ma mniejsze odchylenie standardowe niż ocena uzyskana techniką *hold-out*.
3. *Leaving-one-out* jest techniką, w której liczba iteracji jest równa liczbie przykładów ($N = \text{card}(E)$). W kolejnych iteracjach o numerach $i, i = 1, \dots, N$ ze zbioru przykładów E usuwa się jeden przykład u_i , tworząc zbiór uczący $E_i^u = E \setminus \{u_i\}$ oraz jednoelementowy zbiór testowy $E_i^t = u_i$. Technika ta jest obliczeniowo najbardziej kosztowna.
4. *k-fold cross-validation* różni się od techniki *leaving-one-out* tym, że w kolejnych iteracjach ze zbioru E usuwa się nie jeden przykład, a wieloelementowy zbiór przykładów E_i . Wówczas $E_i^u = E \setminus E_i$ oraz $E_i^t = E_i$, natomiast $E = E_1 \cup \dots \cup E_k$. Podobnie jak poprzednio, do oceny sprawności klasyfikatora stosuje się zależność (2.18).

Wybór odpowiedniej techniki uzależniony jest od wielkości zbioru przykładów. Dla stosunkowo dużych zbiorów przykładów najczęściej przeprowadza się testowanie techniką *10-fold cross-validation* [131].

6.2.3 Proces realizacji badań weryfikacyjnych

Proces realizacji badań prowadzących do wyboru metod/metody wyznaczania wartości granicznych oraz metod/metody selekcji atrybutów podzielić można na kilka etapów. Realizacja procesu weryfikacji stanowi fragment procesu pozyskiwania wiedzy przedstawionego na rys. 2.1 i może być przedstawiona w sposób pokazany na rys. 6.1. Schemat ten nie obejmuje metod selekcji atrybutów pozwalających na selekcję atrybutów ilościowych (np. metody bazujące na miarach prawdopodobieństwa), dla których etapy związane z selekcją i redukcją mogą być realizowane w odwrotnej kolejności.



Rysunek 6.1: Proces oceny i wyboru metod redukcji i selekcji informacji

Zdaniem autora, dokonując wyboru metody wyznaczania wartości granicznych (w trakcie procesu weryfikacji metod redukcji i selekcji informacji przedstawionego na rys. 6.1), nie należy wyboru ograniczyć do jednej metody. Nie zawsze zastosowanie „najlepszej” metody wyznaczania wartości granicznych musi spowodować w dalszym etapie wybór „najlepszego” podzbioru atrybutów relewantnych. W pracy przyjęto, że weryfikacja metod selekcji atrybutów będzie prowadzona dla zbiorów atrybutów, dla których dokonano konwersji wartości dwoma „najlepszymi” metodami.

Rozdział 7

Badania rozpatrywanych metod redukcji i selekcji informacji diagnostycznej

Aby potwierdzić przydatność badanych metod redukcji i selekcji informacji w diagnostyce maszyn, należało przeprowadzić liczne badania weryfikacyjne, w których ograniczono się do oceny stanu lub oceny rodzaju niesprawności maszyn wirnikowych. Przeprowadzone badania opierają się na wynikach dwóch eksperymentów diagnostycznych:

- czynnego eksperymentu diagnostycznego (w dalszym ciągu oznaczanego jako RK),
- eksperymentu numerycznego (w dalszym ciągu oznaczanego jako GD).

Do przeprowadzenia badań niezbędne było określenie metody „uczenia maszynowego”, co stanowiło pierwszy etap badań weryfikacyjnych.

7.1 Wybór metody pozyskiwania wiedzy

Spośród metod „uczenia maszynowego” trudno wskazać metodę lepszą lub gorszą od innych. Można natomiast próbować oceniać poszczególne oprogramowanie komputerowe działające na podstawie określonej metody. Badania w tym zakresie były prowadzone między innymi przez autora [37]. Do znanych, powszechnie stosowanych i dostępnych, programów można zaliczyć:

- w zakresie indukcji drzew decyzyjnych: *C4.5* [107] oraz *C5/See5* [109, 113],
- w zakresie pozyskiwanie reguł z zastosowaniem teorii zbiorów przybliżonych: *LEERS* [46], *ROSETTA* [102], *RSES* [6],
- w zakresie pozyskiwania reguł metodą generowania pokryć: oprogramowanie serii A^Q [78].

Analizując możliwości każdego z tych programów, można wskazać na ich zróżnicowanie pod względem, między innymi:

- oferowanych możliwości wstępnego przetwarzania danych,
- liczności możliwego do zastosowania zbioru uczącego,
- oferowanych możliwości testowania pozyskanej wiedzy,
- wymaganej od użytkownika wiedzy z zakresu danej metody pozyskiwania wiedzy,
- liczności zbioru pozyskiwanych reguł.

Spośród wymienionych programów najbardziej zaawansowane możliwości wstępnego przetwarzania danych oferuje oprogramowanie *ROSETTA*, które oparte jest na bibliotekach oprogramowania *RSES* i rozszerza ich możliwości. Dostępne są zarówno metody konwersji wartości atrybutów, jak i metody selekcji atrybutów. *See5* i *LEERS* (w wersji na komputery klasy IBM PC) oferują w tym zakresie jedynie własną metodę wyznaczania wartości granicznych stosowaną dla uogólnienia pozyskanych reguł. Oprogramowanie serii A^Q w zależności od wersji pozwala na wyznaczanie wartości granicznych oraz selekcję atrybutów.

Ze względu na licznosc możliwego do zastosowania zbioru uczącego, wymienione oprogramowanie pozwala na zastosowanie zbiorów uczących o różnej licznosci. Stosowanie oprogramowania *ROSETTA*, w jego pełnej wersji, wymaga zakupu licencji. Z kolei oprogramowanie *LEERS*, w dostępnej do badań naukowych wersji, jest oprogramowaniem działającym w systemie operacyjnym MS DOS. Jego zastosowanie ograniczone jest zatem zasobami pamięci operacyjnej w tym systemie. Oprogramowanie *RSES* jest dostępne w pełnej wersji, bez jakichkolwiek ograniczeń. A^Q może być stosowany również bez ograniczeń dla celów naukowych i fakt, że działa także w środowisku UNIX pozwala

na stosowanie zbiorów danych, których licznosc jest uzalezniona jedynie rozmiarem pamieci operacyjnej. Natomiast na oprogramowanie *See5* dzialajace w srodowisku *Windows NT* dostepna jest licencja, stad tez mozliwe jest zastosowanie tego oprogramowania dla zbiorow danych dowolnej wielkosci (zaleznej jedynie od dostepnej pamieci komputera).

W zakresie testowania pozyskanej wiedzy spore mozliwosci oferuje oprogramowanie *See5*, przy czym testowanie odbywa sie w prosty sposob, poprzez wybor jednej z opcji programu. Programy *ROSETTA*, *RSES*, *LEERS* i A^Q pozwalaja na wykonywanie testow, wymaga to jednak odpowiedniego przygotowania danych i czesto opracowania dodatkowego programu sterujacego przebiegiem badan.

Kazdy z wymienionych programow w rozny stopniu wymaga znajomosci odpowiedniej metody pozyskiwania wiedzy. Zarowno A^Q jak i *ROSETTA* oferuja duze mozliwosci w zakresie przetwarzania danych i pozyskiwania wiedzy. Wymagana jest jednak przy tym stosunkowo glęboka wiedza z zakresu metod „uczenia maszynowego”. W przypadku osob, ktore nie sa bezposrednio związane z tą dziedzina nauki, zastosowanie tych programow w praktyce moze okazać sie dosc kłopotliwe, a wyniki budzace zastrzezenia. Wymagane bowiem jest okreczenie wielu parametrów, ktore warunkuja optymalne dzialanie programu dla analizowanych danych. Najmniejsze watpliwosci budzi zastosowanie oprogramowania *LEERS* i *See5*, w korych liczba parametrów jest niewielka, a ich znaczenie ogolnie znane.

Zastosowanie kazdego z wymienionych programow pozwala na pozyskanie wiedzy w postaci regul. W zalezności od zastosowanego oprogramowania licznosc zbioru regul jest rozna. O ile bowiem *LEERS*, A^Q i *See5* pozwalaja na pozyskanie zbioru regul, ktorego licznosc jest zbliżona dla kazdego z tych programow i minimalna dla danego zbioru uczajacego, o tyle zastosowanie oprogramowania *ROSETTA* lub *RSES* powoduje pozyskanie zbioru regul, ktorego licznosc czesto przekracza licznosc zbioru przykladów uczajacych i jest co najmniej o rzad wyzsza od licznosci zbiorow regul pozyskanych przy zastosowaniu poprzednio wymienionych programow.

Na podstawie powyższych rozważań dotyczących cech dostępnego oprogramowania, autor zalozył, ze do pozyskiwania wiedzy i jej testowania zastosowane zostanie oprogramowanie *See5*, natomiast oprogramowanie *ROSETTA* i *RSES* w miare mozliwosci zostanie zastosowane do zadan związanych z przetwarzaniem wstepnym.

7.2 Badania dotyczące wyników czynnego eksperymentu diagnostycznego

Prowadzenie czynnych eksperymentów diagnostycznych umożliwia pozyskiwanie empirycznych relacji diagnostycznych, wiążących znany stan badanego obiektu z symptomami, które są cechami obserwowanych sygnałów diagnostycznych.

W przeprowadzonych badaniach celowo zastosowano wyniki czynnego eksperymentu diagnostycznego, gdyż obserwowane sygnały oprócz składowych uwarunkowanych stanem technicznym i warunkami działania zawierają także składowe szumowe. W trakcie badań rejestrowano symptomy diagnostyczne opisywane w literaturze. Istniała zatem realna możliwość selekcji atrybutów ze względu na postawioną hipotezę roboczą, że zbiór tych atrybutów jest nadmiarowy.

7.2.1 Zadanie diagnostyczne

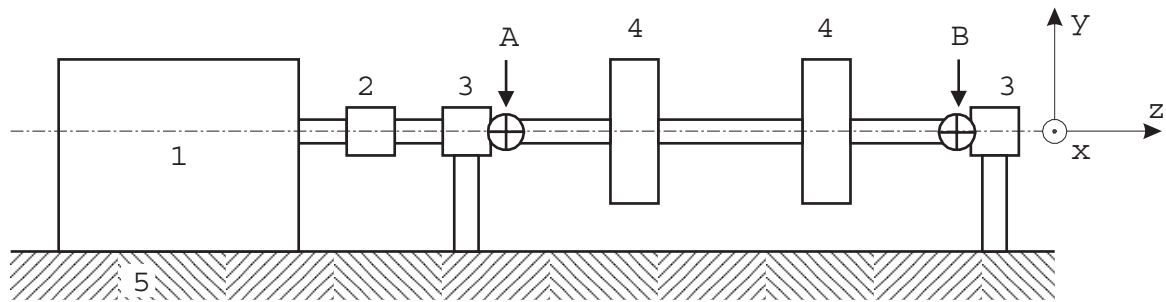
Postawione zadanie diagnostyczne polegało na rozpoznawaniu rodzaju niesprawności wirnika badanego obiektu.

Określenie niesprawności winno nastąpić na podstawie znajomości symptomów diagnostycznych, którymi są cechy sygnałów przemieszczeń względnych wału.

Aby identyfikacja niesprawności wirnika mogła być wspomagana komputerowo (np. za pomocą systemu doradczego), konieczne jest pozyskanie odpowiedniej wiedzy diagnostycznej.

7.2.2 Stanowisko do przeprowadzania czynnych eksperymentów diagnostycznych

Obiektem badań był model maszyny wirnikowej *Rotor Kit* [9], umożliwiający obserwację układu wirnik–łożyska–podpory. Na wale zamocowane były dwie tarcze, umożliwiające zadawanie w tym układzie niewyrównoważenia. Ponadto stanowisko wyposażone było w dodatkowy osprzęt, umożliwiający zadawanie innych niesprawności jak przycieranie i przeciążenie. W przypadku zamocowania specjalnego łożyska hydrodynamicznego istniała możliwość wywoływania wiru olejowego. Układ pomiarowy do obserwacji drgań modelu składał się z czterech czujników drgań względnych, czujnika położenia



Rysunek 7.1: Model maszyny wirnikowej *Rotor Kit*. 1 – silnik, 2 – sprzęgło, 3 – łożyska, 4 – tarcze, 5 – fundament, A, B – punkty pomiaru drgań względnych.

znacznika kąтового wału oraz przetwornika analogowo-cyfrowego *SigLab* [115] wraz z odpowiednim oprogramowaniem.

7.2.3 Zbiór przykładów uczących

Zbiór przykładów uczących określono na podstawie przeprowadzonych pomiarów sygnałów diagnostycznych [124]. Pomiarów dokonywano dla pięciu wybranych rozłącznych klas niesprawności wirnika:

1. niewyrównoważenie jednopłaszczyznowe (N1P),
2. niewyrównoważenie dwupłaszczyznowe (N2P),
3. przeciążenie (PRZE),
4. przycieranie (PRZY),
5. wir olejowy (WIR).

Rejestrowano cztery sygnały drgań względnych wału w dwóch płaszczyznach prostopadłych do geometrycznej osi wału, przy czym w każdej z tych płaszczyzn pomiar następował w dwóch wzajemnie prostopadłych kierunkach x i y (por. rys. 7.1). W sumie wykonano 341 pomiarów [124]. Liczbę pomiarów wykonanych dla każdej klasy niesprawności przedstawiono w tabeli 7.1.

Dla rozpatrywanego zbioru niesprawności, na podstawie informacji zawartych w literaturze (np. [13, 73, 91, 123]), określono zbiór cech sygnałów diagnostycznych, które powinny umożliwić identyfikację tych niesprawności. Dla rejestrowanych 4 sygnałów

Tablica 7.1: Liczba pomiarów przeprowadzonych dla danego rodzaju (klasy) niesprawności (eksperyment RK)

Klasa (IDN)	N1P	N2P	PRZE	PRZY	WIR	Suma
Liczba pomiarów	40	98	95	60	48	341

diagnostycznych wyznaczono 34 cechy tych sygnałów [124]. Rozpatrywano zatem 34 atrybuty, których opis przedstawiono w tabeli 7.2.

Rozpatrywane są między innymi atrybuty określające różnicę faz dla składowych 1X obserwowanych w punktach A i B wału (por. rys. 7.1 i tab. 7.2). W przypadku wybranej metody „uczenia maszynowego” nie ma możliwości określania tzw. atrybutów cyklicznych, pozwalających między innymi na zadawanie miar kątów. Podanie kątów np. 1° oraz 359° powoduje, iż algorytm nie potraktuje ich jako bardzo bliskich i różniących się od siebie o 2° , lecz jako wartości różniące się o 358° . Zastosowano więc rozwiązanie polegające na zastąpieniu jednej miary kąta dwoma wartościami sin i cos danego kąta¹ określającymi jednoznacznie tę wartość, a likwidującymi opisaną niedogodność. W badaniach zastosowano zatem 36 atrybutów. W dalszej części pracy, w celu zwiększenia czytelności opisu atrybutów, pozostawiono jednak oznaczenia różnicy faz podane w tab. 7.2. Traktować je jednak należy jako łączne oznaczenie dwóch atrybutów.

7.2.4 Wyznaczanie wartości granicznych

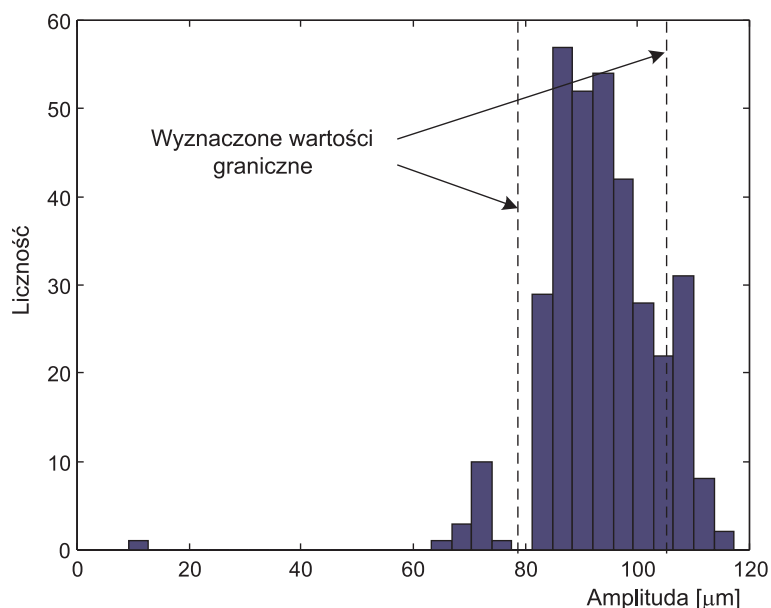
W przeprowadzonych badaniach zastosowano kilka metod wyznaczania wartości granicznych. Zastosowano metody:

1. minimum entropii (ME) — metoda jednoatrybutowa z nauczycielem,
2. heurystyczną (HR) — metoda wieloatrybutowa z nauczycielem,
3. równych częstości (EF) — metoda jednoatrybutowa bez nauczyciela,
4. statystyczną (ST) — metoda jednoatrybutowa bez nauczyciela,
5. według histogramów (HST) — metoda jednoatrybutowa bez nauczyciela,

¹Koncepcja zaproponowana przez W. Cholewę.

Tablica 7.2: Wykaz atrybutów stosowanych w badaniach dotyczących wyników czynnego eksperymentu diagnostycznego RK

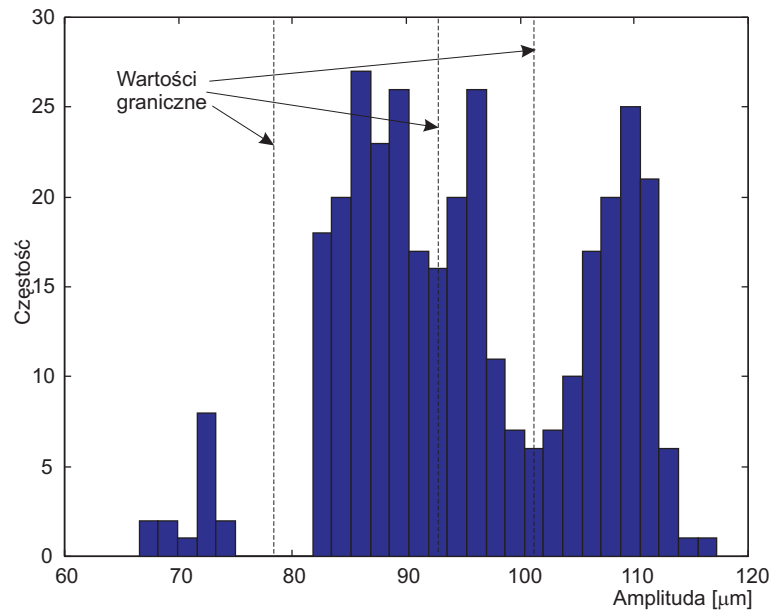
Oznaczenie atrybutu	Opis atrybutu
Ax_Max, Ay_Max, Bx_Max, By_Max	Maksymalna amplituda przemieszczeń wału w punkcie A lub B w kierunkach x i y dla składowych z zakresu 0,2 – 0,5 częstotliwości obrotowej
1X_Ax, 1X_Ay, 1X_Bx, 1X_By, 1.5X_Ax, 1.5X_Ay, 1.5X_Bx, 1.5X_By, 2X_Ax, 2X_Ay, 2X_Bx, 2X_By, 2.5X_Ax, 2.5X_Ay, 2.5X_Bx, 2.5X_By, 3X_Ax, 3X_Ay, 3X_Bx, 3X_By, 3.5X_Ax, 3.5X_Ay, 3.5X_Bx, 3.5X_By, 4X_Ax, 4X_Ay, 4X_Bx, 4X_By	Amplituda przemieszczeń wału dla składowych odpowiednio 1X, 1,5X, ..., 4X w punktach A i B mierzona odpowiednio w kierunkach x lub y, gdzie X jest częstością wirowania wału
R_faz_X, R_faz_Y	Różnica faz dla składowych 1X obserwowanych w punktach A i B w kierunku poziomym x lub pionowym y



Rysunek 7.2: Przykład wyznaczania wartości granicznych na podstawie histogramu dla atrybutu $1X_{Ax}$ (eksperyment RK)

6. grupowanie jednoatrybutowe (GRL) — metoda jednoatrybutowa bez nauczyciela,
7. grupowanie wieloatrybutowe (GRG) — metoda wieloatrybutowa bez nauczyciela.

Jednoatrybutową metodę minimum entropii (opisaną w punkcie 4.2.4) zastosowano przy wykorzystaniu oprogramowania *ROSETTA* [102]. Metodę heurystyczną (por. punkt 4.3.2) zastosowano w systemie *RSES* [6]. Natomiast metoda oparta na miarach prawdopodobieństwa, nazwana tu metodą statystyczną, opisana na str. 47, została opracowana na podstawie [14, 100, 101] i zastosowana w pracy [124] (wykonanej pod opieką autora). Opracowano własne oprogramowanie i zastosowano również znaną metodę równych częstości wartości atrybutu pomiędzy wartościami granicznymi na podstawie [95, 15] (por. punkt 4.2.2). Metoda wyznaczania wartości granicznych na podstawie rozkładu częstości (histogramów) wartości rozpatrywanej cechy została zaproponowana i opracowana przez autora [33] (por. rys 7.2 i 7.3). Również dwie kolejne metody zostały zaproponowane i opracowane (wraz z oprogramowaniem) przez autora: grupowanie jednoatrybutowe z zastosowaniem zbiorów rozmytych (por. punkt 4.2.5) [34, 35] oraz grupowanie wieloatrybutowe (opis w punkcie 4.3.1). Wieloatrybutowa metoda grupowania zaproponowana przez autora została opracowana na podstawie [15].



Rysunek 7.3: Przykład wyznaczania wartości granicznych na podstawie histogramu dla atrybutu Ax_Max (eksperyment RK)

Dla wartości granicznych, wyznaczonych sposobami zgodnymi z opisanymi metodami, dokonywano konwersji wartości atrybutów według (2.12).

7.2.5 Selekcja atrybutów

Zbiór atrybutów poddany został analizie, mającej na celu ograniczenie jego liczności. Na podstawie badań literaturowych dotyczących między innymi wyników uzyskiwanych przy zastosowaniu poszczególnych metod selekcji atrybutów, a także biorąc pod uwagę dostępność gotowego oprogramowania, spośród wielu metod selekcji atrybutów zastosowano cztery:

1. metodę bazującą na teorii algorytmów genetycznych **GA**,
2. metodę opartą na teorii zbiorów przybliżonych **RS**,
3. metodę bazującą na miarach prawdopodobieństwa — analizę korelacji **K**,
4. metodę z zastosowaniem algorytmu indukcyjnego **WA**.

Metoda bazująca na teorii algorytmów genetycznych (por. punkt 5.3) [132, 137], została zastosowana przy wykorzystaniu oprogramowania *RSES* [6]. Do selekcji atrybutów metodą opartą na teorii zbiorów przybliżonych (opisaną w punkcie 5.2) zastosowano kilka programów: *ROSETTA* [102], *RSES* [6] oraz program *REDUKTY* opracowany przez autora.

Jak wynika z literatury (np. [62]), analizowane dane warto na wstępie poddać badaniom statystycznym. Zgodnie z tym zaleceniem, dla zgromadzonych danych, przeprowadzono analizę korelacji i kowariancji wartości wszystkich atrybutów [36]. O ile metody opisane wcześniej wymagają konwersji wartości atrybutów, o tyle metody statystyczne selekcji atrybutów można stosować dla wartości ilościowych.

W przeprowadzonych badaniach zastosowano analizę korelacji wartości atrybutów (por. punkt 5.1). Badano różne dopuszczalne poziomy skorelowania wartości poszczególnych atrybutów, odrzucając atrybuty, dla których poziom skorelowania ich wartości był wyższy od zadanego. W efekcie do badań zastosowano dwa dopuszczalne poziomy skorelowania: 0,8 oraz 0,7.

7.2.6 Wyniki badań dotyczące czynnego eksperymentu diagnostycznego i ich analiza

W pierwszym etapie badań wiedzę w postaci klasyfikatorów pozyskiwano na podstawie pełnego zbioru atrybutów stosowanego do oceny metod wyznaczania wartości granicznych. W drugim etapie wiedzę pozyskiwano na podstawie ograniczonego zbioru atrybutów. W każdym z tych etapów klasyfikatory testowano techniką *10-fold cross-validation* (por. str. 91).

Wyniki i analiza wyników redukcji informacji

Wyniki badań pierwszego etapu — redukcji informacji — przedstawiono poniżej. Uzyskaną sprawność klasyfikatora w zależności od metody wyznaczania wartości granicznych przedstawiono w tab. 7.3. W przypadku metody równych częstości wyznaczono dwa zbiory wartości granicznych: zbiór zawierający cztery wartości graniczne oraz zbiór zawierający pięć wartości granicznych. Oznaczono to w tabelach odpowiednio jako (4) i (5).

Tablica 7.3: Sprawność klasyfikatora w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment RK)

Metoda określania wartości granicznych	Sprawność klasyfikatora [%]
Metoda minimum entropii	85,2
Metoda równych częstości (4)	77,1
Metoda równych częstości (5)	81,3
Metoda histogramów	77,4
Metoda statystyczna	62,6
Metoda heurystyczna	80,7
Metoda grupowania jednoatrybutowego	81,7
Metoda grupowania wieloatrybutowego	74,9

Metoda statystyczna, jako metoda dająca najgorsze wyniki klasyfikacji (co pokazano w [124]), została w dalszych rozważaniach pominięta.

W tab. 7.4 podano wartości łącznego błędu klasyfikacji ϵ , błędu pominięcia ϵ_o oraz błędu niesłusznego zaliczenia ϵ_c . W tab. 7.5 podano zaś błędy względne: łączny ważony błąd klasyfikacji ϵ^w , ważony błąd pominięcia ϵ_o^w oraz ważony błąd niesłusznego zaliczenia ϵ_c^w .

Tablica 7.4: Błędy klasyfikatora, wyrażone w [%], dla wybranych metod wyznaczania wartości granicznych (eksperyment RK)

Metoda wyznaczania wartości granicznych	ϵ	ϵ_o	ϵ_c
Metoda minimum entropii	14,8	17,1	4,0
Metoda równych częstości (4)	22,9	27,2	6,2
Metoda równych częstości (5)	18,7	26,9	6,1
Metoda histogramów	22,6	27,8	6,7
Metoda heurystyczna	19,3	21,1	4,9
Metoda grupowania jednoatrybutowego	18,3	20,4	3,3
Metoda grupowania wieloatrybutowego	25,1	28,5	7,4

Tablica 7.5: Błędy ważone klasyfikatora, wyrażone w [%], dla wybranych metod wyznaczania wartości granicznych (eksperyment RK)

Metoda wyznaczania wartości granicznych	ϵ^w	ϵ_o^w	ϵ_c^w
Metoda minimum entropii	16,5	19,3	3,4
Metoda równych częstości (4)	27,0	31,3	4,9
Metoda równych częstości (5)	25,1	29,7	5,0
Metoda histogramów	24,8	31,8	4,0
Metoda heurystyczna	20,6	21,2	3,9
Metoda grupowania jednoatrybutowego	21,4	28,7	3,6
Metoda grupowania wieloatrybutowego	27,6	32,3	4,2

Istotnym elementem oceny metody określania wartości granicznych jest liczba wyznaczonych wartości granicznych z zastosowaniem wybranej metody. Wyniki tej analizy przedstawiono w tab. 7.6. Podano średnią liczbę wartości granicznych przypadających na jeden atrybut oraz minimalną i maksymalną liczbę wartości granicznych wyznaczonych za pomocą danej metody. W przypadku metody heurystycznej uzyskana minimalna liczba wartości granicznych wynosi 0. Oznacza to, że atrybut, dla którego liczba wartości granicznych wynosi 0, nie ma wpływu na klasyfikację przykładów — po konwersji przyjmuje jedną wartość dla wszystkich przykładów. Metoda heurystyczna pozwala zatem już na etapie konwersji wartości atrybutów dokonać selekcji atrybutów.

Stosując różne metody wyznaczania wartości granicznych, uzyskiwano klasyfikatory, którym odpowiadały zbiory reguł o różnej liczności. Liczność zbioru reguł stanowi kolejny element oceny metod wyznaczania wartości granicznych. Uzyskane liczby reguł przedstawiono w tab. 7.7.

Na podstawie uzyskanych wyników pośredniej oceny metod redukcji informacji diagnostycznej można stwierdzić, że:

- większość z zastosowanych metod pozwala na uzyskanie stosunkowo wysokiej sprawności klasyfikacji,
- liczba wyznaczonych wartości granicznych jest niska, a w przypadku metody heurystycznej osiąga minimum; jedynie zastosowanie metody minimum entropii powoduje wyznaczenie dużej liczby wartości granicznych.

Tablica 7.6: Liczba wartości granicznych w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment RK)

Metoda określania wartości granicznych	Liczba wartości granicznych		
	średnia	minimalna	maksymalna
Metoda minimum entropii	103,5	18	185
Metoda równych częstości (4)	4,0	4	4
Metoda równych częstości (5)	5,0	5	5
Metoda histogramów	4,2	2	5
Metoda heurystyczna	1,8	0	6
Metoda grupowania jednoatrybutowego	4,0	2	6
Metoda grupowania wieloatrybutowego	7,5	5	8

Poszukiwana metoda wyznaczania wartości granicznych winna spełniać kryteria opisane w tab. 6.1. W świetle przyjętych kryteriów, na podstawie uzyskanych wyników (por. rys. 7.4) dokonano oceny zastosowanych metod. Wyniki oceny przedstawiono w tab. 7.8, 7.9 oraz na rys. 7.5.

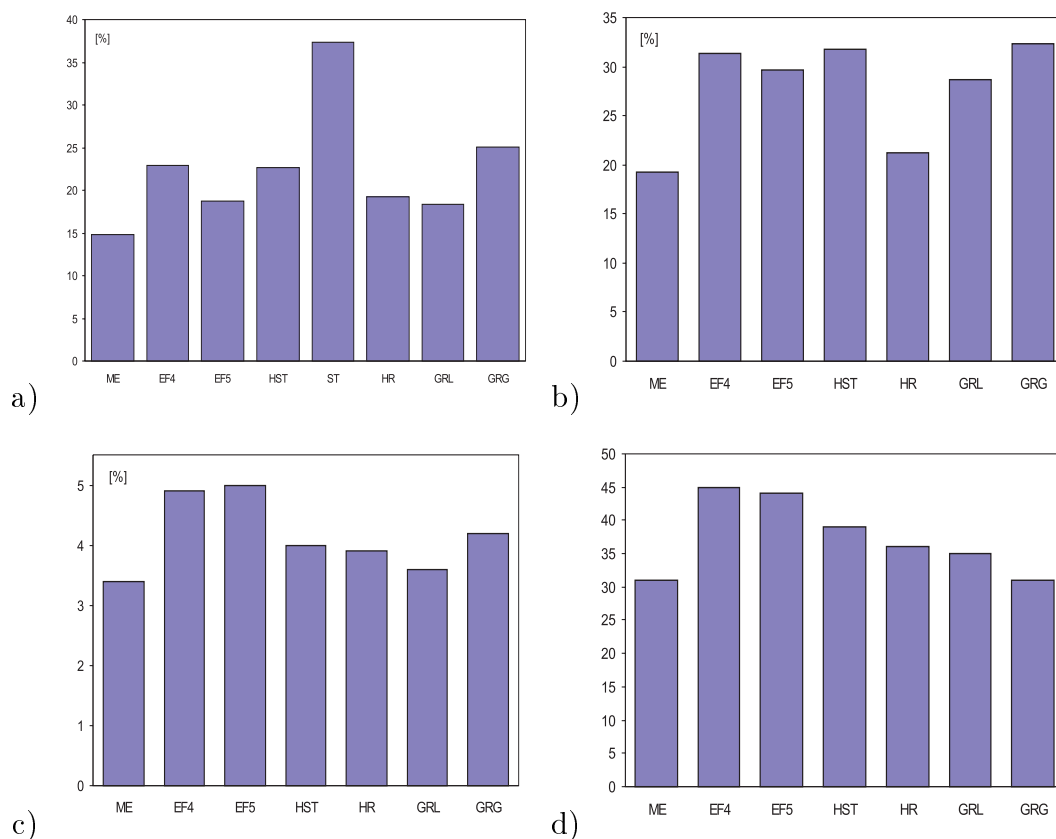
Metoda heurystyczna jest metodą optymalną ze względu na przyjęte kryteria. W dalszych badaniach, dla porównania wyników, zastosowano dane po konwersji dokonanej według dwóch metod:

- metody heurystycznej,
- metody grupowania jednoatrybutowego.

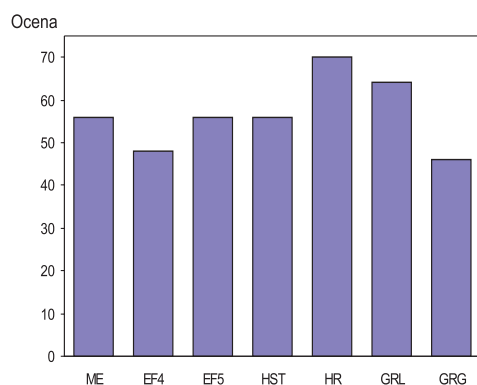
Wyniki selekcji atrybutów i ich analiza

W wyniku zastosowania algorytmów selekcji atrybutów licznosc zbioru atrybutów została znacznie ograniczona. Uzyskane liczby atrybutów po ich selekcji przedstawiono w tab. 7.10 — dla redukcji metodą heurystyczną oraz w tab. 7.11 — dla redukcji metodą grupowania jednoatrybutowego.

Do selekcji atrybutów zastosowano między innymi analizę korelacji ich wartości. Zastosowano dwa dopuszczalne maksymalne poziomy korelacji wartości atrybutów: 0,8 i 0,7 (por. opis str. 102). Metoda analizy korelacji wartości nie opiera się ani na za-



Rysunek 7.4: Graficzna prezentacja wyników redukcji informacji diagnostycznej dla przypadku eksperymentu czynnego RK: a) łączny błąd klasyfikacji, b) względny błąd pominięcia, c) względny błąd niesłusznego zaliczenia, d) uzyskana liczba reguł



Rysunek 7.5: Wyniki oceny metod redukcji informacji diagnostycznej (eksperyment RK)

Tablica 7.7: Liczby reguł uzyskane dla różnych metod wyznaczania wartości granicznych (eksperyment RK)

Metoda określania wartości granicznych	Liczba reguł
Metoda minimum entropii	31
Metoda równych częstości (4)	45
Metoda równych częstości (5)	44
Metoda histogramów	39
Metoda heurystyczna	36
Metoda grupowania jednoatrybutowego	35
Metoda grupowania wieloatrybutowego	31

chowaniu poziomu niesprzeczności w zbiorze uczącym (jak to ma miejsce w przypadku metody selekcji z zastosowaniem algorytmu genetycznego i metody opartej na teorii zbiorów przybliżonych), ani na zachowaniu sprawności klasyfikatora (metoda z zastosowaniem algorytmu indukcyjnego). Stąd też w tab. 7.10 i 7.11 podano dodatkowo, w celu oceny tej metody, uzyskane sprawności klasyfikatorów. Zauważyć należy, iż z uwagi na możliwość wykorzystania informacji niesionej przez wartości atrybutów ilościowych zdecydowano, że analiza korelacji wykonana zostanie właśnie dla wartości ilościowych atrybutów, które dopiero po tej analizie poddane zostały konwersji. Metoda z zastosowaniem algorytmu indukcyjnego pozwoliła na uzyskanie najlepszego rezultatu — zdecydowanie najmniej liczne zbioru atrybutów relewantnych.

Zastosowano również metodę z wykorzystaniem algorytmu indukcyjnego i z konwersją wartości atrybutów (por. opis na str. 79). Do wyznaczania wartości granicznych zastosowano metodę heurystyczną, natomiast jako algorytm indukcyjny „uczenia maszynowego” wykorzystywany w procesie selekcji atrybutów zastosowano *See5*. W wyniku selekcji atrybutów tą metodą uzyskano wynik identyczny jak w przypadku selekcji metodą z zastosowaniem algorytmu indukcyjnego bez konwersji wartości atrybutów. Oznaczać to może, że niemożliwe jest wyznaczenie innych wartości granicznych atrybutów metodą heurystyczną niż te, które zostały wyznaczone przy uwzględnieniu wyjściowego zbioru atrybutów.

Tablica 7.8: Ocena metod określania wartości granicznych na podstawie przyjętych kryteriów (eksperyment RK)

Metoda	k_1^R	k_2^R	k_3^R	k_4^R	k_5^R
Minimum entropii	4	4	3	0	3
Równych częstości (4)	2	2	2	4	2
Równych częstości (5)	3	3	2	4	2
Histogramów	2	2	3	4	3
Heurystyczna	3	4	3	5	3
Grupowania jednoatrybutowego	3	3	3	4	3
Grupowania wieloatrybutowego	1	2	3	3	3

Tablica 7.9: Ważona ocena metod określania wartości granicznych (eksperyment RK)

Metoda	k_1^R	k_2^R	k_3^R	k_4^R	k_5^R	Suma
Minimum entropii	24	8	21	0	3	56
Równych częstości (4)	12	4	14	16	2	48
Równych częstości (5)	18	6	14	16	2	56
Histogramów	12	4	21	16	3	56
Heurystyczna	18	8	21	20	3	70
Grupowania jednoatrybutowego	18	6	21	16	3	64
Grupowania wieloatrybutowego	6	4	21	12	3	46

Tablica 7.10: Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów dla redukcji metodą heurystyczną (eksperyment RK)

Metoda selekcji atrybutów	Liczba atrybutów relewantnych	Sprawność klasyfikatora [%]
Analiza korelacji (0,7)	10	71,7
Analiza korelacji (0,8)	13	75,2
Z zastosowaniem algorytmu genetycznego	13	80,7
Oparta na teorii zbiorów przybliżonych	13	80,7
Z zastosowaniem algorytmu indukcyjnego	6	80,7

Tablica 7.11: Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów dla redukcji metodą grupowania jednoatrybutowego (eksperyment RK)

Metoda selekcji atrybutów	Liczba atrybutów relewantnych	Sprawność klasyfikatora [%]
Analiza korelacji (0,7)	10	76,2
Analiza korelacji (0,8)	13	78,1
Z zastosowaniem algorytmu genetycznego	9	79,3
Oparta na teorii zbiorów przybliżonych	9	79,3
Z zastosowaniem algorytmu indukcyjnego	7	80,7

Zbiory atrybutów wyznaczone w wyniku zastosowania metod selekcji atrybutów relewantnych pokazano w tab. 7.12 (opis atrybutów — patrz tab. 7.2). Oznaczenie $K_{0,7}$ określa metodę statystyczną opartą na analizie korelacji o progu 0,7.

Przy stosowaniu metody analizy korelacji ważne jest, by zbiór relewantnych atrybutów diagnostycznych wyznaczony został na podstawie zaleceń zawartych w literaturze. Dlatego też atrybuty relewantne wyznaczone z zastosowaniem tej metody są najczęściej atrybutami charakterystycznymi dla rozważanych stanów niesprawności badanej maszyny wirnikowej.

Tabela 7.12. (c.d.)

Atrybut diagnostyczny	Metoda redukcji informacji							
	HR				GRL			
	Metoda selekcji				Metoda selekcji			
	K0,7	GA	RS	WA	K0,7	GA	RS	WA
3.5X_Ax								
3.5X_Ay								
3.5X_Bx						x		
3.5X_By								
4X_Ax				x				
4X_Ay						x		
4X_Bx				x				
4X_By		x	x					
R_faz_X	x	x	x		x			
R_faz_Y	x	x	x		x			

Zastosowanie pozostałych metod selekcji atrybutów powoduje, że uzyskane zbiory atrybutów relewantnych różnią się pomiędzy sobą. Wybór atrybutów uzależniony jest nie tylko od metody ich selekcji, ale również od zastosowanej wcześniej metody redukcji informacji. Zbiórów atrybutów relewantnych można wyznaczyć wiele. W przypadku metody opartej na algorytmach genetycznych uzyskiwany wynik nie jest powtarzalny — każdorazowo uzyskuje się nieco inne zbiory atrybutów relewantnych. Przy zastosowaniu metody opartej na teorii zbiorów przybliżonych wynik jest powtarzalny, pod warunkiem zachowania kolejności atrybutów w zbiorze uczącym, co potwierdzono wcześniej między innymi w [74]. Liczba zbiorów atrybutów relewantnych otrzymana przy zastosowaniu obydwu metod może być stosunkowo duża (kilkadziesiąt zbiorów atrybutów), zależna od przyjętych parametrów działania algorytmu.

Jak wynika z przeprowadzonych badań, uzyskana sprawność klasyfikatora zależy od zastosowanego zbioru atrybutów relewantnych (por. tab. 7.10 i 7.11).

Liczba reguł uzyskanych przy zastosowaniu poszczególnych metod selekcji atrybutów została przedstawiona w tab. 7.13. Obserwować można niewielki wzrost liczby reguł, szczególnie przy wykorzystaniu metod opartych na algorytmach genetycznych

Tablica 7.13: Liczba pozyskanych reguł w zależności od zastosowanej metody selekcji atrybutów (eksperyment RK)

Metoda selekcji atrybutów	Liczba reguł dla redukcji metodą	
	HR	GRL
Analiza korelacji (0,7)	36	46
Analiza korelacji (0,8)	35	44
Z zastosowaniem algorytmu genetycznego	31	51
Oparta na teorii zbiorów przybliżonych	31	50
Z zastosowaniem algorytmu indukcyjnego	35	38

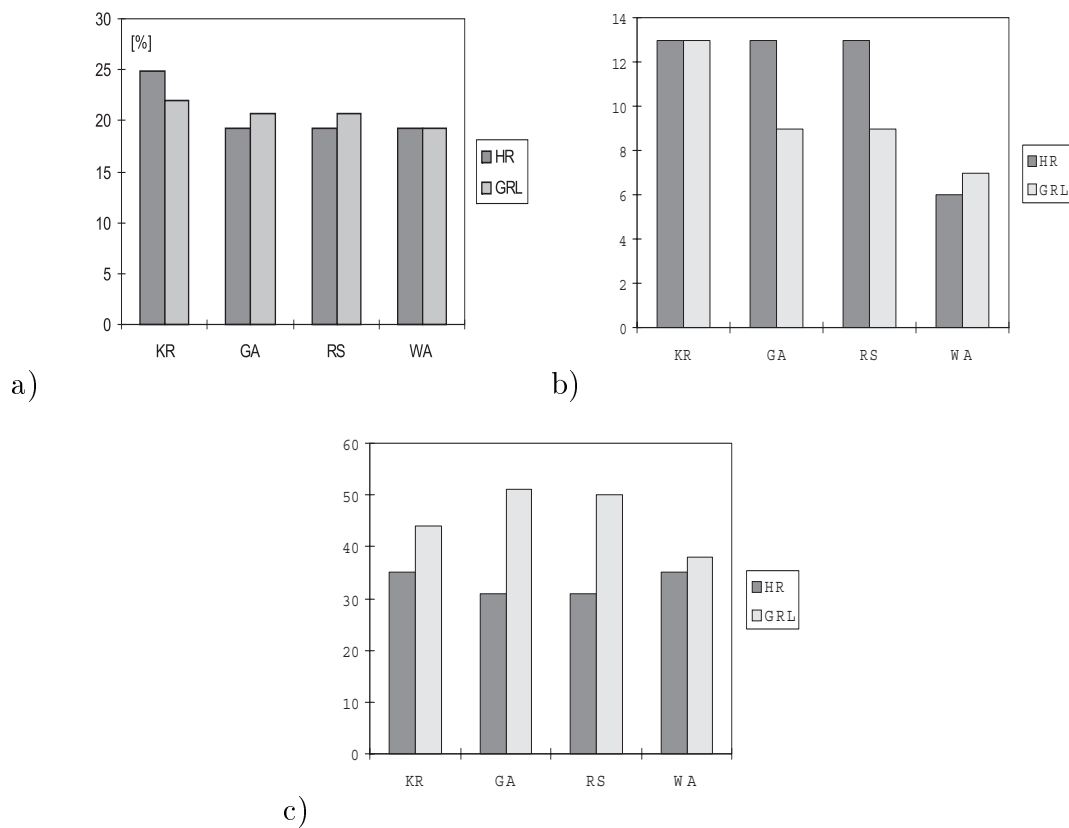
Tablica 7.14: Ocena metod selekcji atrybutów (eksperyment RK)

Metoda	k_1^S	k_2^S	k_3^S
Analiza korelacji	2	3	4
Z zastosowaniem algorytmu genetycznego	4	4	4
Oparta na teorii zbiorów przybliżonych	4	4	4
Z zastosowaniem algorytmu indukcyjnego	5	5	5

i teorii zbiorów przybliżonych. Ograniczenie liczby atrybutów powoduje zatem, że liczba reguł rośnie, choć uzyskane liczby reguł są — zdaniem autora — nadal stosunkowo niskie.

Zastosowane metody selekcji atrybutów poddano ocenie według kryteriów opisanych w punkcie 6.2.1. Wyniki selekcji przedstawiono graficznie na rys. 7.6, natomiast wyniki oceny metod selekcji przedstawiono w tab. 7.14 oraz w tab. 7.15.

W świetle przyjętych kryteriów metoda selekcji z zastosowaniem algorytmu indukcyjnego jest metodą optymalną. Nie tylko pozwala na ograniczenie liczebności zbioru atrybutów, lecz także na uzyskanie małej liczby reguł.



Rysunek 7.6: Graficzna prezentacja wyników selekcji atrybutów dla przypadku eksperymentu czynnego RK: a) łączny błąd klasyfikacji, b) liczba atrybutów relewantnych, c) liczba reguł

Tablica 7.15: Ważona ocena metod selekcji atrybutów (eksperyment RK)

Metoda	k_1^S	k_2^S	k_3^S	Suma
Analiza korelacji	8	4,5	2	14,5
Z zastosowaniem algorytmu genetycznego	16	6	2	24,0
Oparta na teorii zbiorów przybliżonych	16	6	2	24,0
Z zastosowaniem algorytmu indukcyjnego	20	7,5	2,5	30,0

7.3 Badania dotyczące wyników eksperymentu numerycznego

7.3.1 Zadanie diagnostyczne

Zadanie diagnostyczne polegało na identyfikacji stanu niewyrównoważenia wirnika podpartego w dwóch łożyskach hydrodynamicznych.

Określenie niesprawności winno nastąpić na podstawie znajomości symptomów diagnostycznych, którymi w tym wypadku są cechy trajektorii ruchu (względego i bezwzględnego) wybranych węzłów wału.

7.3.2 Model obiektu zastosowanego do badań numerycznych

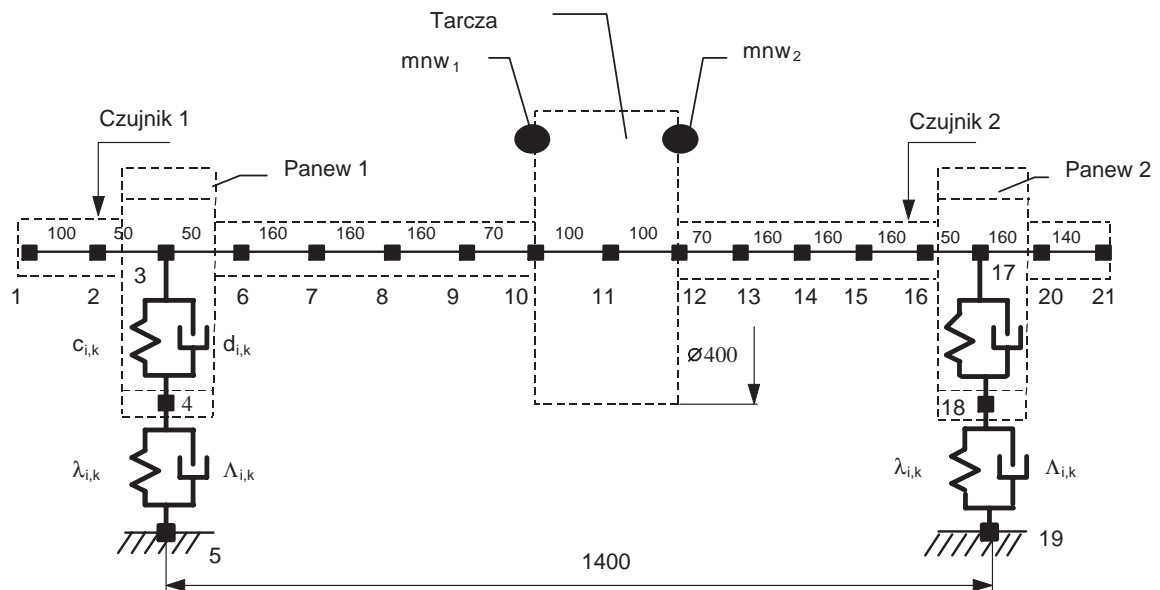
Przyjęty do badań obiekt stanowi część linii wirników wielkogabarytowego stanowiska do badania wirników, będącego własnością Instytutu Maszyn Przepływowych PAN w Gdańsku. Zasadniczą częścią stanowiska jest modelowy wirnik, podparty w dwóch łożyskach ślizgowych i napędzany silnikiem elektrycznym. Łożyska osadzone są w stojakach łożyskowych posadowionych w ramie. Rama oparta jest na ciężkim bloku betonowym pełniącym rolę fundamentu i układu odniesienia dla mierzonych przemieszczeń bezwzględnych elementów stanowiska [24].

Dla opisanego wyżej układu w Instytucie Maszyn Przepływowych PAN w Gdańsku opracowano model teoretyczny (rys. 7.7) [55]. Do analizy dynamiki badanego układu wykorzystano metodę elementów skończonych. Wał zamodelowano jako belkowe elementy skończone, natomiast podpory jako punktowe elementy skończone.

Na podstawie przyjętego modelu opracowano systemu programów *MESWIR* [24] pozwalający na symulację zachowania się badanego układu. Model opisuje właściwości układu konkretnego, stąd możliwe było bezpośrednie porównanie wyników obliczeń symulacyjnych z wynikami pomiarów przeprowadzonych na stanowisku.

W przeprowadzonych badaniach zastosowano dane uzyskane w wyniku symulacji niesprawności (niewyrównoważenia) opisanego układu.

Autorzy oprogramowania przyjęli, że przedmiotem analizy będą zjawiska zachodzące poniżej granicy stabilności hydrodynamicznej układu, a więc przy prędkościach obrotowych wirnika, dla których nie występuje np. niestabilność hydrodynamiczna. Za-



Rysunek 7.7: Teoretyczny model układu wirnik–łożyska–podpory łożyskowe: $\lambda_{i,k}$, $\Lambda_{i,k}$ — współczynniki sztywności i tłumienia utwierdzeń zewnętrznych panwi, $c_{i,k}$, $d_{i,k}$ — współczynniki sztywności i tłumienia filmu olejowego łożysk [55]

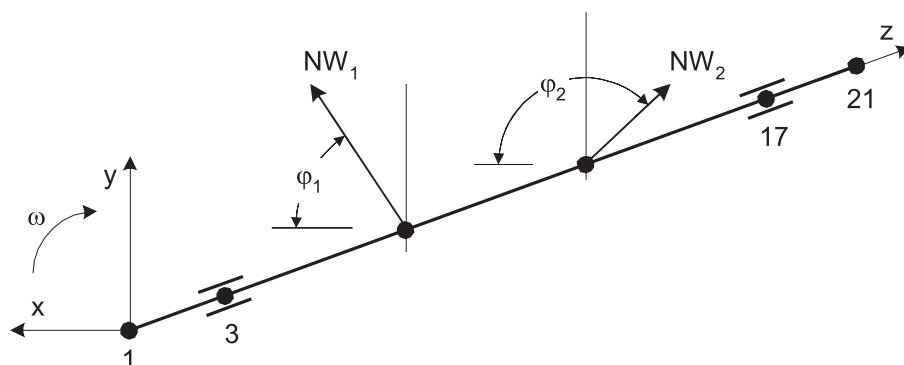
łożono ponadto, że przemieszczenia względne czopów i panwi w łożyskach będą dostatecznie małe w stosunku do koła luzów. Powyższe założenia pozwalają wówczas na zastosowanie modelu liniowego [24]. Przyjęcie modelu liniowego powoduje, że trajektorie ruchu środka czopa wału są elipsami.

Przyjęto [55], że zmieniającymi się cechami modelu teoretycznego będą (por. rys. 7.8):

- wartości niewywagi: NW_1 , NW_2 ,
- kąty położenia niewywagi: φ_1 , φ_2 ,
- położenie niewywag wzdłuż osi wirnika,
- prędkość obrotowa wirnika: ω .

7.3.3 Zbiór przykładów uczących

Opracowane w Instytucie Maszyn Przepływowych oprogramowanie umożliwiło wygenerowanie danych, które stanowiły zbiór uczący do badań związanych z „uczeniem maszynowym” [75].



Rysunek 7.8: Przykładowe rozmieszczenie niewyważ [24]

Rozpatrywano następujące stany niewyrównowazenia układu:

1. niewyrównowazenie statyczne (S),
2. niewyrównowazenie quasi-statyczne (Q),
3. niewyrównowazenie momentowe (M),
4. niewyrównowazenie dynamiczne (D),
5. zgrubne wyrównowazenie wirnika (W).

Dla każdego stanu niewyrównowazenia układu wygenerowano zestaw wyników. Liczby przykładów dla każdej z klas przedstawiono w tab. 7.16. Dane zostały wygenerowane dla trzech prędkości obrotowych wału: 2050, 2972, 3509 [min^{-1}] oraz dla mas niewyrównowazenia m_{nw} : 0,004, 0,005, 0,006, 0,0095, 0,0125, 0,0155, 0,0150, 0,020, 0,025 [kg] na promieniu r_{nw} 0,18 [m]. Stąd niewyważa $NW = m_{nw} \times r_{nw}$. Kąty pomiędzy wektorami niewyważ zmieniano skokowo w granicach od 0° do 180° .

Tablica 7.16: Liczba przykładów w poszczególnych klasach (eksperyment GD)

Klasa (NW)	S	Q	M	D	W	Suma
Liczba przykładów	972	972	972	1080	1080	5076

Uzyskane dane symulacyjne określają wartości czterech współczynników A, B, C, D w pełni opisujące trajektorie drgań wymuszonych wybranych węzłów układu [24] :

$$\begin{aligned}x &= A \sin \tau + B \cos \tau, \\y &= C \sin \tau + D \cos \tau,\end{aligned}\tag{7.1}$$

gdzie $\tau = \omega t$ oznacza bezwymiarowy czas, x jest składową poziomą elipsy drgań, y — składową pionową elipsy drgań.

Pomimo że zależności (7.1) w pełni opisują trajektorie ruchu środka wału względem panewki łożyska, opis ten dla potrzeb diagnostyki jest nieprzydatny [85]. Stąd też wartości współczynników A, B, C, D przekształcono w zbiór cech opisujących kształt trajektorii ruchu środka czopa wału (por. rys. 7.9) [89] takich jak:

- długość dłuższej półosi (D1),
- długość krótszej półosi (D2),
- kąt pochylenia dłuższej półosi (TD),
- kąt położenia znacznika (TZ).

Dodatkowo uwzględniono prędkość obrotową wału oznaczoną przez $RtSp$. Otrzymano zatem 17 atrybutów, których opis przedstawiono w tab. 7.17.

Analizując zastosowany sposób reprezentacji trajektorii będących wynikiem obliczeń z wykorzystaniem modelu liniowego, można stwierdzić, że zastosowany zbiór atrybutów jest minimalnym zbiorem opisującym kształt trajektorii ruchu środka czopa wału. Usunięcie jakiegokolwiek atrybutu spowoduje, że część informacji o kształcie trajektorii zostanie utracona. Pomimo to, aby sprawdzić działanie algorytmów selekcji atrybutów dla takich przypadków, autor postanowił zastosować wybrane metody selekcji do opisanego wyżej zbioru danych.

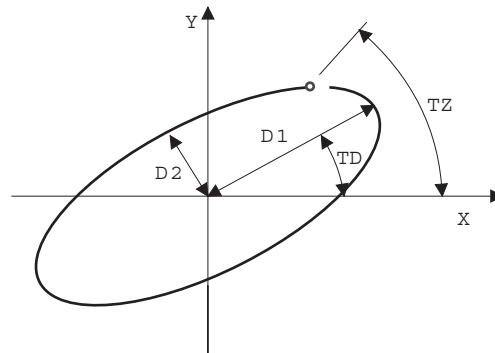
7.3.4 Zastosowane metody redukcji i selekcji informacji

W przeprowadzonych badaniach zastosowano następujące metody wyznaczania wartości granicznych:

1. heurystyczną (HR),
2. równych częstości (EF),
3. histogramów (HST),

Tablica 7.17: Wykaz atrybutów stosowanych w badaniach dotyczących wyników eksperymentu numerycznego GD

Nazwa atrybutu	Opis atrybutu
RtSp	Prędkość obrotowa
A04D1, A18D1	długość większej półosi elipsy, określona na podstawie drgań bezwzględnych obserwowanych w węźle nr 4 lub 18
A04D2, A18D2	długość mniejszej półosi elipsy, określona na podstawie drgań bezwzględnych obserwowanych w węźle nr 4 lub 18
A04TD	kąt pochylenia większej półosi elipsy, określona na podstawie drgań bezwzględnych obserwowanych w węźle nr 4
A04TZ	faza początkowa znacznika elipsy, określona na podstawie drgań bezwzględnych obserwowanych w węźle nr 4
AdifTD	= A04TD - A18TD
AdifTZ	= A04TZ - A18TZ
R02D1, R16D1	długość większej półosi elipsy, określona na podstawie drgań względnych obserwowanych w węźle nr 2 lub 16
R02D2, R16D2	długość mniejszej półosi elipsy, określona na podstawie drgań względnych obserwowanych w węźle nr 2 lub 16
R02TD	kąt pochylenia większej półosi elipsy, określona na podstawie drgań względnych obserwowanych w węźle nr 2
R02TZ	faza początkowa znacznika elipsy, określona na podstawie drgań względnych obserwowanych w węźle nr 2
RdifTD	= R02TD - R16TD
RdifTZ	= R02TZ - R16TZ

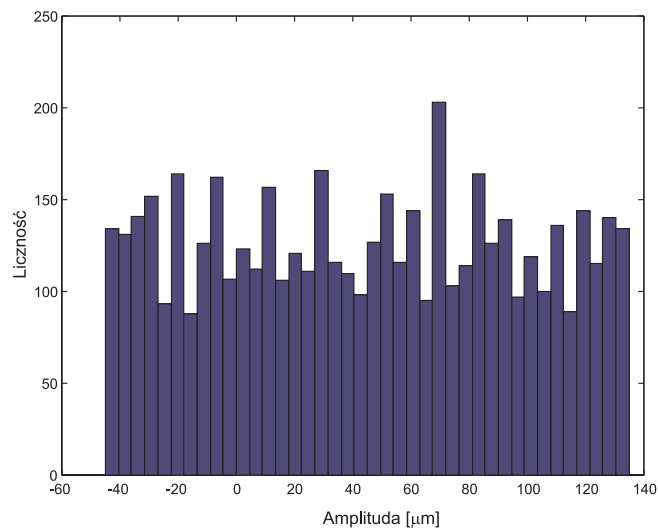


Rysunek 7.9: Cechy trajektorii zastosowane w badaniach GD [89]

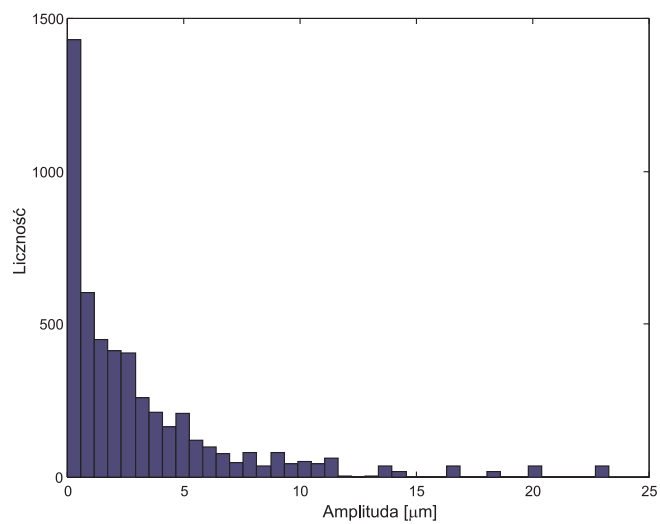
4. grupowania jednoatrybutowego (GRL),
5. grupowania wieloatrybutowego (GRG).

Wadą wielu z tych metod jest to, iż konieczne jest określenie oczekiwanej liczby wartości granicznych. Jak wspomniano wcześniej, autor zaproponował określanie tej liczby na podstawie rozkładu częstości wartości poszczególnych atrybutów. O ile w poprzednich badaniach, dotyczących wyników czynnego eksperymentu diagnostycznego, określenie liczby wartości granicznych było dość proste, o tyle w badaniach dotyczących wyników eksperymentu numerycznego okazało się to sprawą skomplikowaną. Rozkłady częstości wartości dla większości atrybutów przedstawiają się bowiem podobnie jak w wypadku atrybutów przedstawionych na rys. 7.10 i 7.11 i określenie na ich podstawie liczby wartości granicznych jest trudne. Oznacza to, że w przypadku metody określania wartości granicznych według histogramów przyjęte wartości graniczne i ich liczba najprawdopodobniej nie są optymalne ze względu na kryterium zachowania poziomu niesprzeczności danych.

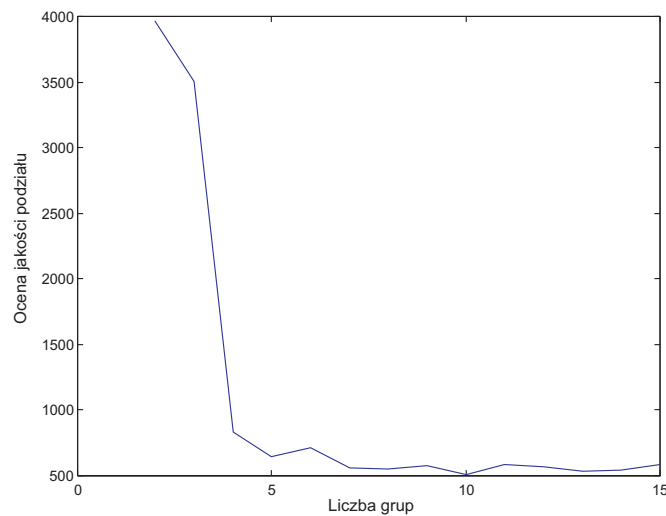
W przypadku metody grupowania, do określenia optymalnej liczby wartości granicznych zastosowano ocenę jakości podziału na podstawie ilorazu średniej odległości elementów w grupie i średniej odległości grup [128]. Przykładowe wyniki tej oceny dla atrybutów AdifTD i R02D2 przedstawiono na rys. 7.12 i 7.13. Dla atrybutu AdifTD najlepszy podział, z punktu widzenia przyjętej oceny jakości podziału, to podział na 10 grup. Przyjęto jednak, że zastosowany zostanie podział na 5 grup. Przyjęcie takiego podziału oznacza, że mamy jedynie 4 wartości graniczne (zamiast 9), a wzrost jakości podziału dla większej liczby grup jest stosunkowo niewielki. Podobnie dla atrybutu R02D2 zas-



Rysunek 7.10: Rozkład częstości wartości atrybutu AdifDT (eksperyment GD)



Rysunek 7.11: Rozkład częstości wartości atrybutu R02D2 (eksperyment GD)



Rysunek 7.12: Wyniki oceny jakości podziału wartości atrybutu `AdifTD` na grupy (eksperyment GD)

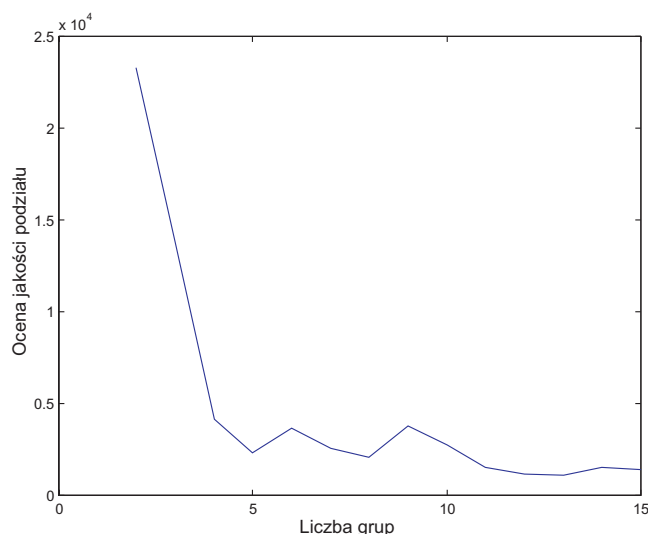
tosowano podział na 5 grup. Dla tego atrybutu dopiero podział na 11 grup daje poprawę oceny jakości, co jest liczbą — zdaniem autora — zbyt wielką.

Pozostałe metody zastosowano w sposób podobny jak wobec wyników czynnego eksperymentu diagnostycznego opisanego w punkcie 7.2.

W zakresie metod selekcji atrybutów zastosowano trzy metody opierające się na:

1. teorii algorytmów genetycznych (GA),
2. teorii zbiorów przybliżonych (RS),
3. zastosowaniu algorytmu indukcyjnego (WA).

Analiza korelacji wartości atrybutów wykazała, że wartości te nie są skorelowane ze sobą. Nie uwzględniono więc w dalszych badaniach metody analizy korelacji wartości atrybutów. Nie wyklucza to, oczywiście, związków pomiędzy wartościami poszczególnych atrybutów. Analiza korelacji pozwala bowiem na wykrycie jedynie związków liniowych pomiędzy wartościami atrybutów, a w rzeczywistości związki te mogą być wyższego rzędu.



Rysunek 7.13: Wyniki oceny jakości podziału wartości atrybutu R02D2 na grupy (eksperyment GD)

7.3.5 Wyniki dotyczące eksperymentu numerycznego i ich analiza

Badania dotyczące wyników eksperymentu numerycznego, podobnie jak badania dotyczące wyników eksperymentu czynnego, podzielono na dwa etapy: pierwszy — ocena metod wyznaczania wartości granicznych oraz drugi — ocena metod selekcji atrybutów. Podobnie jak poprzednio, uzyskane klasyfikatory testowano techniką *10-fold cross-validation*.

Wyniki redukcji informacji i ich analiza

Wyniki redukcji informacji obejmujące uzyskaną sprawność klasyfikatora oraz liczby wartości granicznych w zależności od metody wyznaczania wartości granicznych przedstawiono w tab. 7.18 do 7.21.

Podobnie jak w poprzednich badaniach liczba uzyskanych reguł zależała od zastosowanej metody wyznaczania wartości granicznych (por. tab. 7.22).

Graficznie wyniki redukcji informacji przedstawiono na rys. 7.14. Ocenę wyników redukcji informacji diagnostycznej dla danych zgromadzonych w ramach eksperymentu numerycznego przedstawiono w tab. 7.23 i 7.24 oraz na rys. 7.15.

Tablica 7.18: Sprawność klasyfikatora w zależności od zastosowanej metody wyznaczenia wartości granicznych (eksperyment GD)

Metoda określania wartości granicznych	Sprawność klasyfikatora [%]
Metoda równych częstości	93,8
Metoda histogramów	85,1
Metoda heurystyczna	92,2
Metoda grupowania jednoatrybutowego	85,9
Metoda grupowania wieloatrybutowego	70,2

Tablica 7.19: Błędy klasyfikatora, wyrażone w [%], dla wybranych metod wyznaczenia wartości granicznych (eksperyment GD)

Metoda wyznaczenia wartości granicznych	ϵ	ϵ_o	ϵ_c
Metoda równych częstości	6,2	6,1	1,5
Metoda histogramów	14,9	14,9	3,7
Metoda heurystyczna	7,8	7,0	2,0
Metoda grupowania jednoatrybutowego	14,1	14,2	3,5
Metoda grupowania wieloatrybutowego	29,8	30,2	4,9

Tablica 7.20: Błędy ważone klasyfikatora, wyrażone w [%], dla wybranych metod wyznaczenia wartości granicznych (eksperyment GD)

Metoda wyznaczenia wartości granicznych	ϵ^w	ϵ_o^w	ϵ_c^w
Metoda równych częstości	6,3	6,2	1,5
Metoda histogramów	14,8	14,9	3,7
Metoda heurystyczna	7,4	6,4	1,8
Metoda grupowania jednoatrybutowego	14,1	14,3	3,5
Metoda grupowania wieloatrybutowego	29,9	30,0	4,8

Tablica 7.21: Liczba wartości granicznych w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment GD)

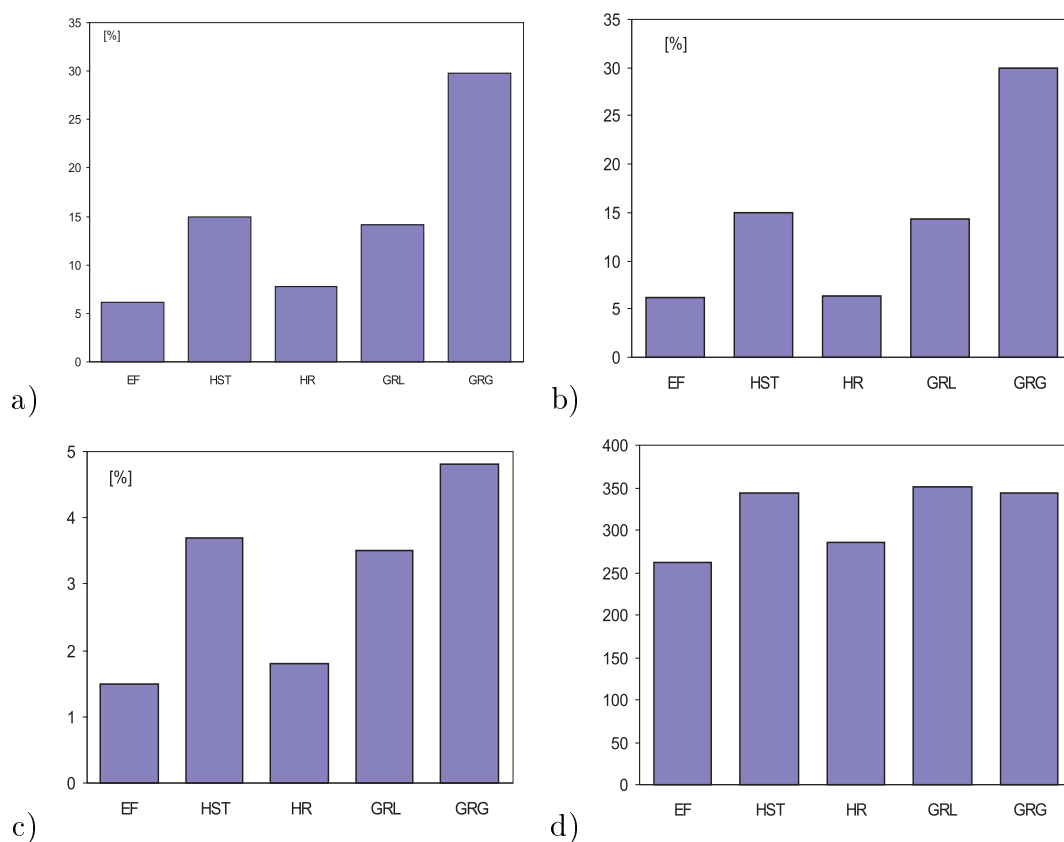
Metoda określania wartości granicznych	Liczba wartości granicznych		
	średnia	min	max
Metoda równych częstości	4,0	4	4
Metoda histogramów	3,8	2	6
Metoda heurystyczna	2,5	1	9
Metoda grupowania jednoatrybutowego	4,7	1	7
Metoda grupowania wieloatrybutowego	4,5	1	7

Tablica 7.22: Liczby reguł uzyskane dla różnych metod wyznaczania wartości granicznych (eksperyment GD)

Metoda określania wartości granicznych	Liczba reguł
Metoda równych częstości	261
Metoda histogramów	344
Metoda heurystyczna	286
Metoda grupowania jednoatrybutowego	351
Metoda grupowania wieloatrybutowego	344

Tablica 7.23: Ocena metod redukcji informacji diagnostycznej (eksperyment GD)

Metoda	k_1^R	k_2^R	k_3^R	k_4^R	k_5^R
Równych częstości	4	3	3	4	3
Histogramów	3	2	2	4	2
Heurystyczna	4	3	3	4	3
Grupowania jednoatrybutowego	3	2	2	3	2
Grupowania wieloatrybutowego	1	1	1	3	2



Rysunek 7.14: Graficzna prezentacja wyników redukcji informacji diagnostycznej dla przypadku eksperymentu numerycznego GD: a) łączny błąd klasyfikacji, b) ważony błąd pominięcia, c) ważony błąd niesłusznego zaliczenia, d) uzyskana liczba reguł

Tablica 7.24: Ważona ocena metod redukcji informacji diagnostycznej (eksperyment GD)

Metoda	k_1^R	k_2^R	k_3^R	k_4^R	k_5^R	Suma
Równych częstości	24	6	21	16	3	70
Histogramów	18	4	14	16	2	54
Heurystyczna	24	6	21	16	3	70
Grupowania jednoatrybutowego	18	4	14	12	2	50
Grupowania wieloatrybutowego	6	2	7	12	2	29



Rysunek 7.15: Wyniki oceny metod redukcji informacji diagnostycznej dla przypadku eksperymentu numerycznego GD

Na podstawie uzyskanych wyników można stwierdzić, że:

- dla danych pochodzących z eksperymentu numerycznego można uzyskać bardzo wysoką sprawność klasyfikatora,
- nie wszystkie metody wyznaczania wartości granicznych pozwalają na uzyskanie wysokiej sprawności klasyfikatora,
- liczba wyznaczonych wartości granicznych jest stosunkowo niska.

W opisywanych wynikach badań wyraźnie widać, że zastosowanie jednej z dwu metod, metody heurystycznej lub metody równych częstości, oprócz tego, że pozwala na uzyskanie bardzo wysokiej sprawności, pozwala również na uzyskanie bardzo niskiej średniej liczby wartości granicznych. Przyjęto zatem, że do dalszych badań zostaną zastosowane wartości graniczne wyznaczone tymi dwoma metodami.

Wyniki selekcji atrybutów i ich analiza

Zastosowanie różnych metod selekcji atrybutów, w przypadku badań dotyczących eksperymentu numerycznego, potwierdziło prawdziwość twierdzenia zamieszczonego na stronie 117. Zastosowanymi metodami selekcji nie można ograniczyć liczności zbioru atrybutów. Spośród 17 atrybutów służących do opisu przykładów, zastosowanie metod selekcji pozwoliło na odrzucenie jedynie jednego atrybutu (A18D2), zarówno z wyko-

rzystaniem danych po redukcji metodą heurystyczną jak i metodą równych częstości (por. tab. 7.25).

Tablica 7.25: Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów (eksperyment GD)

Metoda wyboru atrybutów	Liczba atrybutów relewantnych
Z zastosowaniem algorytmu genetycznego	16
Oparta na teorii zbiorów przybliżonych	16
Z zastosowaniem algorytmu indukcyjnego	17

Autor postanowił podjąć próbę zastosowania własnej, opisanej na stronie 79, metody selekcji atrybutów z zastosowaniem algorytmu indukcyjnego i z konwersją wartości atrybutów. Jako algorytm indukcyjny „uczenia maszynowego” wykorzystywany w procesie selekcji atrybutów zastosowano ten sam algorytm, który posłużył w kolejnym etapie do pozyskiwania wiedzy — *See5*. Zastosowanie opracowanej przez autora metody selekcji atrybutów powoduje, że liczba wartości granicznych dla każdego podzbioru atrybutów może być inna, co jest związane z koniecznością spełnienia kryterium zachowania niesprzeczności w zbiorze danych. Ponieważ podobnie zmieniać się może liczba reguł, zatem i sprawność klasyfikatora może być różna. W przeprowadzonych badaniach do wyznaczania wartości granicznych zastosowano metodę heurystyczną.

Wyniki obejmujące liczbę atrybutów relewantnych, liczbę wartości granicznych, liczbę reguł oraz sprawność klasyfikatora podano w tab. 7.26.

W tab. 7.27 przedstawiono atrybuty, które według metody selekcji z zastosowaniem algorytmu indukcyjnego i z konwersją wartości atrybutów są relewantne. Wspomnieć jednak należy, że zastosowana metoda selekcji była metodą sekwencyjną. W kolejnych iteracjach ze zbioru uwzględnianych atrybutów usuwano ten atrybut, którego odrzucenie nie powodowało obniżenia sprawności uzyskiwanego klasyfikatora. Kolejne podzbiory atrybutów były zatem podzbiórami poprzednich, różniącymi się jedynie o jeden atrybut, który został usunięty. Można wykazać (np. [41]), że postępowanie to nie zawsze jest najlepsze i należałoby podjąć próbę niezależnego wyboru atrybutów.

Zaproponowana metoda selekcji atrybutów, połączona z metodą wyznaczania wartości granicznych, zastosowana dla danych pochodzących z eksperymentu numerycznego, poz-

Tablica 7.26: Wyniki selekcji atrybutów dla metody z zastosowaniem algorytmu indukcyjnego i z konwersją wartości atrybutów (eksperyment GD)

Liczba atrybutów relewantnych	10	9	8
Sprawność klasyfikatora [%]	95,2	95,4	93,8
Średnia liczba wartości granicznych	4,8	5,4	6,0
Minimalna liczba wartości granicznych	1	1	2
Maksymalna liczba wartości granicznych	8	10	10
Liczba reguł	273	265	250

woliła na znaczną redukcję liczby atrybutów przy jednoczesnym niewielkim wzroście liczby wartości granicznych, a równocześnie spadku liczby reguł. Ponadto ograniczenie liczności zbioru atrybutów o połowę nie spowodowało spadku sprawności klasyfikatora (dla podzbioru dziewięciu atrybutów uzyskano nawet nieco wyższą sprawność). Zauważono również prawidłowość, którą można było przewidzieć — im mniejsza liczba atrybutów, tym większa liczba wartości granicznych.

7.4 Podsumowanie przeprowadzonych badań

Badania wybranych metod redukcji i selekcji informacji w diagnostyce maszyn przeprowadzono dla dwóch zbiorów danych zawierających zarówno wyniki czynnego eksperymentu diagnostycznego, jak również wyniki eksperymentu numerycznego. Oceniono zastosowane metody redukcji i selekcji informacji pod kątem przydatności w dziedzinie diagnostyki maszyn, dowodząc tej przydatności w sposób indukcyjny — dla wymienionych zbiorów danych. Ocena ma charakter pośredni — podlegała jej wiedza pozyskana na podstawie zgromadzonych danych, dla których, zgodnie z wybranymi metodami, przeprowadzono konwersję wartości atrybutów oraz dokonano selekcji tych atrybutów.

Zaproponowano kilka własnych metod redukcji i selekcji. Opracowano metody:

- wyznaczania wartości granicznych:
 - metodę histogramów,
 - metodę grupowania jednoatrybutowego,

Tablica 7.27: Atrybuty relewantne uzyskane metodą z zastosowaniem algorytmu indukcyjnego i konwersją wartości atrybutów (eksperyment GD)

Atrybut	Liczba atrybutów		
	10	9	8
RtSp	x	x	x
A04D1	x	x	x
A18D1	x	x	x
A04D2	x		
A18D2			
A04TD			
A04TZ	x	x	x
AdifTD	x	x	x
AdifTZ	x	x	x
R02D1			
R16D1			
R02D2	x	x	x
R16D2	x	x	
R02TD	x	x	x
R02TZ			
RdifTD			
RdifTZ			

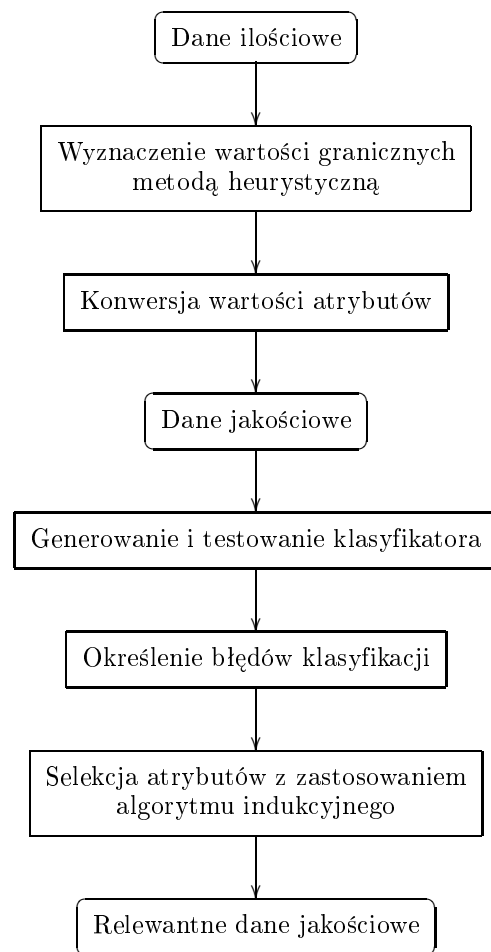
- metodę grupowania wieloatrybutowego,
- selekcji atrybutów:
 - metodę z zastosowaniem algorytmu indukcyjnego i z konwersją wartości atrybutów.

Uzyskane wyniki pokazują, iż niektóre z metod zaproponowanych przez autora należy poddać dalszej weryfikacji, aby dostarczyć dodatkowych dowodów ich przydatności w praktyce. Ponadto — zdaniem autora — stosowanie zarówno metod redukcji jak i metod selekcji wymaga ogromnej ostrożności, znajomości zagadnienia oraz wnikliwej analizy. Bezkrytyczne stosowanie metod redukcji informacji może prowadzić w praktyce, na przykład, do uzyskania zbyt dużej liczby wartości jakościowych, a w konsekwencji do pozyskania wiedzy trudnej do interpretacji. Również stosowanie metod selekcji atrybutów może spowodować nadmierne ograniczenie liczebności zbioru atrybutów, a w konsekwencji utratę pewnej ilości informacji. Ponadto stwierdzono, że kierowanie się w ocenie pozyskiwanej wiedzy wyłącznie sprawnością klasyfikatora (która to ocena jest najczęściej stosowana w literaturze) nie jest postępowaniem poprawnym. Jak wykazano w przeprowadzonych badaniach, wyższa sprawność klasyfikatora nie musi oznaczać mniejszych błędów ważonych: błędu pominięcia i niesłusznego zaliczenia.

7.5 Proponowana metoda redukcji i selekcji informacji diagnostycznej

Na podstawie analizy wyników przeprowadzonych badań autor zaproponował metodę redukcji i selekcji informacji diagnostycznej. Opiera się ona na metodzie heurystycznej wyznaczania wartości granicznych oraz na metodzie selekcji atrybutów z zastosowaniem algorytmu indukcyjnego. Metoda heurystyczna, zastosowana zarówno do wyników czynnego eksperymentu diagnostycznego, jak i do wyników eksperymentu numerycznego, pozwoliła na uzyskanie dobrych wyników. Jednocześnie metoda ta nie wymaga zadawania liczby wartości granicznych. Metoda selekcji z zastosowaniem algorytmu indukcyjnego pozwala zaś na uzyskanie bardzo małej liczby atrybutów relewantnych, przy jednoczesnym spełnieniu warunku zachowania niskich błędów klasyfikacji.

Zaproponowaną metodę można przedstawić następująco (por. rys. 7.16):



Rysunek 7.16: Proponowana metoda redukcji i selekcji informacji dla potrzeb diagnostyki maszyn

1. Wyznaczyć wartości graniczne metodą heurystyczną.
2. Dokonać konwersji ilościowych wartości atrybutów.
3. Dokonać selekcji atrybutów.
4. Pozyskać i przetestować pozyskany klasyfikator.
5. Określić błędy klasyfikacji.
6. Dokonać selekcji atrybutów z zastosowaniem metody wykorzystującej algorytm indukcyjny „uczenia maszynowego”.

Metoda ta powinna umożliwić uzyskanie wysokiej sprawności klasyfikatora, a równocześnie małej wartości błędów ważonych, małej liczby wartości jakościowych atrybutów oraz minimalnej liczby atrybutów relewantnych.

Rozdział 8

Podsumowanie i wnioski

Przedstawiona praca jest wynikiem zainteresowań autora związanych z metodami sztucznej inteligencji, a w szczególności z pozyskiwaniem wiedzy w diagnostyce maszyn. Autor, zajmując się problemami pozyskiwania wiedzy metodami „uczenia maszynowego”, dostrzegł istotną lukę w tej dziedzinie — brak prac związanych z zastosowaniem nowoczesnych metod redukcji i selekcji informacji oraz prac związanych z metodologią oceny tych metod. Wskazując na potrzebę redukcji i selekcji informacji diagnostycznej, autor podjął próbę syntezy obejmującej metody redukcji i selekcji informacji oraz sposoby oceny tych metod.

W pracy opisano wybrane metody selekcji atrybutów oraz wybrane metody wyznaczania wartości granicznych, przedstawiając dla każdej z nich zarówno podstawy teoretyczne, jak i szczegółowe algorytmy postępowania. Sformułowano kryteria, jakie powinny spełniać wymienione metody. Kryteria te oparto między innymi na pośredniej ocenie metod poprzez ocenę pozyskanej wiedzy.

Aby zrealizować cel pracy i zweryfikować postawioną tezę, autor przeprowadził badania weryfikacyjne z zastosowaniem wyników czynnego eksperymentu diagnostycznego oraz wyników eksperymentu numerycznego.

Sformułowany na wstępie cel pracy został osiągnięty:

- w zakresie *metod redukcji informacji diagnostycznej* — poprzez szczegółową identyfikację i opis wybranych metod wyznaczania wartości granicznych, a także ich analizę obejmującą aspekty teoretyczne i praktyczne zastosowanie tych metod,

- w zakresie *metod selekcji atrybutów diagnostycznych* — poprzez szczegółową identyfikację i opis wybranych metod selekcji atrybutów, ich analizę teoretyczną i praktyczną weryfikację możliwości ich zastosowania,
- w zakresie *kryteriów i sposobu oceny metod redukcji i selekcji informacji* — poprzez określenie kryteriów oceny metod oraz określenie sposobu oceny tych metod,
- w zakresie *oceny przydatności badanych metod redukcji i selekcji informacji w dziedzinie diagnostyki maszyn* — poprzez zastosowanie wybranych metod redukcji i selekcji informacji do redukcji wartości atrybutów i selekcji atrybutów diagnostycznych; metody te zweryfikowano z zastosowaniem wyników czynnego eksperymentu diagnostycznego oraz wyników eksperymentu numerycznego.

Analizując metody redukcji i selekcji informacji opisywane w literaturze oraz możliwości ich zastosowania w dziedzinie diagnostyki maszyn, autor zaproponował własne metody redukcji informacji oraz modyfikację metody selekcji atrybutów relewantnych.

Analizując wyniki zastosowania metod redukcji i selekcji informacji i biorąc pod uwagę specyfikę dziedziny diagnostyki maszyn, autor wskazał na niebezpieczeństwa związane z bezkrytycznym stosowaniem tych metod. Jak wykazano w badaniach weryfikacyjnych, liczby wartości granicznych, uzyskiwane przy zastosowaniu niektórych metod ich wyznaczania, mogą w praktyce być nie do przyjęcia. Zbyt duża liczba wartości granicznych, a tym samym zbyt duża liczba wartości jakościowych powoduje, że pozyskiwana wiedza staje się trudna do zinterpretowania. Również zastosowanie metod selekcji atrybutów może prowadzić w praktyce do uzyskania takich wyników, których przyjęcie przez specjalistów może okazać się niemożliwe. Przykładem mogą być wyniki uzyskane podczas badań dotyczących czynnego eksperymentu diagnostycznego. W uzyskanym zbiorze atrybutów relewantnych brak jest niektórych atrybutów uznawanych w literaturze za atrybuty niosące informację o rozpatrywanych stanach badanej maszyny. Pojawiły się natomiast atrybuty, których literatura nie podaje jako charakterystyczne dla tych stanów. Nie zawsze zatem pozytywna ocena sformułowana na podstawie badań weryfikacyjnych pozyskanej wiedzy może być wystarczająca do uznania pozyskanej wiedzy (a tym samym i metody redukcji i selekcji informacji) za poprawną.

Na podstawie przeprowadzonych badań i analizy uzyskanych wyników można sformułować następujące wnioski dotyczące postawionej tezy pracy:

- 1) możliwe jest określenie metody redukcji informacji diagnostycznej,
- 2) zastosowanie redukcji ilości informacji w procesie pozyskiwania wiedzy dotyczącej relacji diagnostycznych powoduje wzrost efektywności tego procesu oraz zwiększa skuteczność wnioskowania z użyciem pozyskanej wiedzy,
- 3) możliwe jest określenie metody selekcji atrybutów diagnostycznych,
- 4) zastosowanie wybranych metod selekcji atrybutów powoduje znaczne zmniejszenie liczności zbioru atrybutów i nie powoduje jednocześnie obniżenia sprawności klasyfikatora, uzyskanego w wyniku indukcji na podstawie danych o ograniczonej liczbie atrybutów,
- 5) możliwa jest pośrednia ocena metod redukcji i selekcji informacji poprzez ocenę jakości pozyskanej wiedzy, rozumianej jako zdolność do uzyskiwania wysokiej skuteczności klasyfikacji, czyli niskich wartości błędów klasyfikacji,
- 6) zaproponowany sposób postępowania umożliwia przyjęcie (dla danego zadania, polegającego na pozyskaniu wiedzy dotyczącej określonej dziedziny problemowej) optymalnej metody postępowania w zakresie redukcji informacji i selekcji atrybutów; metoda ta pozwala na skuteczne pozyskiwanie wiedzy deklaratywnej zapisanej w postaci reguł wnioskowania lub drzew decyzyjnych.

Oceniając wykonaną pracę, autor uważa, że najważniejszymi i oryginalnymi jej elementami są:

- 1) zebranie, usystematyzowanie i szczegółowy opis metod redukcji i selekcji informacji,
- 2) propozycje własnych metod wyznaczania wartości granicznych, sposobów określania liczby tych wartości, a także propozycja modyfikacji metody selekcji atrybutów z zastosowaniem algorytmu indukcyjnego,
- 3) określenie kryteriów i sposobu oceny metod redukcji i selekcji informacji,

- 4) wskazanie na niebezpieczeństwa w bezkrytycznym stosowaniu metod selekcji atrybutów diagnostycznych.

Przeprowadzone badania mogą stanowić podstawę do podjęcia dalszych działań zmierzających do określenia metod wyznaczania wartości granicznych ilościowych atrybutów diagnostycznych, jak również kontynuowania badań nad możliwościami ograniczania liczności zbiorów tych atrybutów. Zaproponowany sposób oceny metod redukcji i selekcji informacji może być zastosowany nie tylko w dziedzinie diagnostyki maszyn, ale również w innych dziedzinach zastosowania metod sztucznej inteligencji.

Wyniki przeprowadzonych badań potwierdziły potrzebę kontynuowania prac związanych z omawianym zagadnieniem. W szczególności, jak wspomniano wyżej, dotyczy to:

- rozwoju metod wyznaczania wartości granicznych atrybutów ilościowych, a w tym przede wszystkim sposobów określania i minimalizacji liczby wartości granicznych,
- rozwoju metod selekcji atrybutów relewantnych,
- rozwoju oprogramowania pozwalającego na zastosowanie zarówno metod wyznaczania wartości granicznych, jak i metod selekcji atrybutów relewantnych oraz badania przydatności tych metod.

Bibliografia

- [1] Adamczyk J.: *Klasyfikacja istotnych cech sygnałów w czynnym i biernym eksperymencie diagnostycznym*. VII Szkoła Diagnostyki, 17–30, Poznań-Rydzyzna, 1985.
- [2] Adamczyk Z., Jemielniak K., Kosmol J., Sokołowski A.: *Automatyzacja wytwarzania. Monitorowanie ostrza skrawającego. Metody konwencjonalne i sieci neuronowe*. WNT, Warszawa, 1996.
- [3] Arciszewski T., Dybala T., Wnek J.: *A method for evaluation of learning systems*. Heuristics, The Journal of Knowledge Engineering, Vol. 5, No. 4, 1992.
- [4] Auda G., Kamel M.: *CMNN: Cooperative Modular Neural Networks for pattern recognition*. Pattern Recognition Letters, Vol. 18, No. 15., 1391–1398, 1997.
- [5] Bazan J., Skowron A., Synak P.: *Dynamic reducts as a tool for extracting laws from decision tables*. Symposium on Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence, Vol. 869, Springer-Verlag, 346–355, Charlotte, NC, 1994.
- [6] Bazan J., Szczuka M.: *RSES Homepage*, <http://alfa.mimuw.edu.pl/~rses>.
- [7] Będkowski L.: *Elementy diagnostyki technicznej*. WAT, Warszawa, 1991.
- [8] Belue L., Bauer K.: *Determining input features for multilayer perceptron*. Neurocomputing, Vol. 7, No. 2, 111-122, 1995.
- [9] Bently Nevada: *Rotor Kit. Instruction manual No. TW8029319*. Minden NV, 1996.

-
- [10] Cempel C.: *Redukcja zbioru danych w diagnostyce maszyn*. Zagadnienia Eksploatacji Maszyn, Z. 4 (44) 1980, 571–585.
- [11] Cempel C.: *Podstawy wibroakustycznej diagnostyki maszyn*. WNT, Warszawa, 1982.
- [12] Cempel C.: *The limit value in vibration condition monitoring of machinery*. Materiały VII Szkoły Diagnostyki „Wnioskowanie Diagnostyczne”, 151–158, Poznań-Rydzyna, 1985.
- [13] Cempel C.: *Diagnostyka wibroakustyczna maszyn*. PWN, Warszawa, 1989.
- [14] Cempel C.: *Limit value in the practice of machine vibration diagnostics*. Mechanical Systems and Signal Processing, Vol. 4, No. 6, 483–493, 1990.
- [15] Chmielewski M., Grzymała-Busse J.: *Global discretization of continuous attributes as preprocessing for machine learning*. 3rd International Workshop on Rough Sets and Soft Computing, 294–301, 1994.
- [16] Cholewa W.: *Metoda diagnozowania maszyn z zastosowaniem zbiorów rozmytych*. ZN Politechniki Śląskiej nr 764, seria: Mechanika Z. 79, Gliwice, 1983.
- [17] Cholewa W.: *Systemy doradcze w diagnostyce technicznej*. IX Konferencja Naukowo-Techniczna „Diagnostyka Maszyn Roboczych i Pojazdów”, 63–81, Bydgoszcz, 1994.
- [18] Cholewa W.: *Dynamiczne systemy doradcze w diagnostyce technicznej*. Kongres Diagnostyki Technicznej, tom I, 57–84, KPKM Pol. Śl. i IMP PAN, Gdańsk, 1996.
- [19] Cholewa W.: *Real-time expert systems for technical diagnostics*. 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, Vol. 6, 715–720, Berlin, 1997.
- [20] Cholewa W.: *Bazy danych w diagnostyce technicznej*. II Międzynarodowy Kongres Diagnostyki Technicznej „Diagnostyka 2000”, 115–122, Warszawa, 2000.

-
- [21] Cholewa W.: *Dynamiczne systemy doradcze*. Symposium on Methods of Artificial Intelligence in Mechanics and Mechanical Engineering „AI-Mech 2000”, 37–54, Gliwice, 2000.
- [22] Cholewa W., Kaźmierczak J.: *Data processing and reasoning in technical diagnostics*. WNT, Warszawa, 1995.
- [23] Cholewa W., Kiciński J.: *Metody identyfikacji reguł dla diagnostycznych systemów doradczych*. XXXV Sympozjon „Modelowanie w Mechanice”, 57–66, Gliwice, 1995.
- [24] Cholewa W., Kiciński J. (red.): *Diagnostyka techniczna. Odwrotne modele diagnostyczne*. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [25] Cholewa W., Moczulski W.: *Diagnostyka techniczna maszyn. Pomiary i analiza sygnałów*. Skrypt Politechniki Śląskiej, Seria: Mechanika, nr 1758, Gliwice, 1993.
- [26] Cholewa W., Pedrycz W.: *Systemy doradcze*. Skrypt uczelniany Politechniki Śląskiej nr 1447, Gliwice, 1987.
- [27] Cholewa W., Solipiwo A.: *Optymalizacja przestrzeni cech w wibroakustycznych badaniach diagnostycznych*. VII Szkoła Diagnostyki, 185–192, Politechnika Poznańska, Poznań-Rydzyna, 1985.
- [28] Cholewa W., White M.: *Application of knowledge-based systems to machine diagnosis*. Trondheim, 1991.
- [29] Chrzanowski P.: *Metody definiowania klas wartości cech sygnałów diagnostycznych*. II Krajowa Konferencja Naukowo-Techniczna „Diagnostyka Procesów Przemysłowych”, 239–244, Politechnika Zielonogórska, Zielona Góra-Łagów, 1997.
- [30] Cichosz P.: *Systemy uczące się*. WNT, Warszawa, 2000.
- [31] Ciupke K.: *Akwizycja wiedzy diagnostycznej z zastosowaniem klasyfikatora przybliżonego*. Kongres Diagnostyki Technicznej, T. 2, 181–184, KPKM Politechnika Śląska i IMP PAN, Gdańsk, 1996.

- [32] Ciupke K.: *Przykład zastosowania teorii zbiorów przybliżonych do akwizycji wiedzy diagnostycznej*. II Krajowa Konferencja Naukowo-Techniczna „Diagnostyka Procesów Przemysłowych”, Politechnika Zielonogórska, 99–102, Zielona Góra-Łagów, 1997.
- [33] Ciupke K.: *An example of an application of the rough set theory to diagnostic knowledge acquisition*. Intelligent Information Systems, 41–46, IPI PAN, Zakopane, 1997.
- [34] Ciupke K.: *Discretization of real-valued diagnostic symptoms*. Intelligent Information Systems, 21–25, IPI PAN, Bystra, 2000.
- [35] Ciupke K.: *Kwantowanie wartości ilościowych cech diagnostycznych*. II Międzynarodowy Kongres Diagnostyki Technicznej „Diagnostyka 2000”, Warszawa, 2000.
- [36] Ciupke K.: *Selekcja i redukcja informacji diagnostycznej*. Symposium on Methods of Artificial Intelligence in Mechanics and Mechanical Engineering „AI-Mech 2000”, 131–136, Gliwice, 2000.
- [37] Ciupke K., Maniak P.: *Application of machine learning methods to diagnostics of machinery*. Intelligent Information Systems, 132–135, IPI PAN, Malbork, 1998.
- [38] Deb K., Agrawal S.: *Understanding interactions among genetic algorithm parameters*. W: Banzhaf W., Reeves C. (eds.): *Foundations of Genetic Algorithms*, 5, Morgan Kaufmann, 1998.
- [39] Dietrych J.: *System i konstrukcja*. WNT, Warszawa, 1985.
- [40] Dougherty J., Kohavi R., Sahami M.: *Supervised and unsupervised discretization of continuous features*. 12th International Conference „Machine Learning”, 194–202, San Francisco, 1995.
- [41] Draper N., Smith H.: *Analiza regresji stosowana*. PWN, Warszawa, 1973.
- [42] Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (eds): *Sieci neuronowe*. Biocybernetyka i Inżynieria Biomedyczna 2000, T. 6., Akademicka Oficyna Wydawnicza, Warszawa, 2000.

-
- [43] Duda R., Hart P.: *Pattern classification and scene analysis*. J. Willey and Sons, New York, 1973.
- [44] Fayyad U., Irani K.: *Multi interval discretization of continuous valued attributes for classification learning*. Proceedings of the 13th International Joint Conference on Artificial Intelligence, 1022–1027, Morgan Kaufmann, 1993.
- [45] Friswell M., Penny J., Garvey S.: *Parameter subset selection in damage location*. Inverse Problems in Engineering, Vol. 5, No. 3, 189–215, 1997.
- [46] Grzymała-Busse J.: *LEERS - a system for learning from examples based on rough sets*. W: R. Słowinski, (ed.) Intelligent Decision Support, 3–18, Kluwer Academic Publishers, 1992.
- [47] Grzymała-Busse J.: *Managing uncertainty in machine learning from examples*. Intelligent Information Systems, 70–84, IPI PAN, Wigry, 1994.
- [48] Grzymała-Busse J., Stefanowski J.: *Discretization of numerical attributes by direct use of the LEM2 induction algorithm with interval extension*. Intelligent Information Systems, 149–158, IPI PAN, Zakopane, 1997.
- [49] Hand D.J.: *Discriminant and classification*. J. Wiley and Sons. Chichester, 1981.
- [50] Holland J.: *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI, 1975.
- [51] Jelonek J., Krawiec K., Stefanowski J.: *Comparative study of feature subset selection techniques for machine learning tasks*. Intelligent Information Systems, 68–77, IPI PAN, Malbork, 1998.
- [52] John G., Kohavi R., Pfleger K.: *Irrelevant features and the subset selection problem*. 11th International Conference on Machine Learning, 121–129, New Brunswick, NJ, Morgan Kaufman, 1994.
- [53] Johnson D.: *Approximation algorithms for combinatorial problems*. Journal of Computer and System Sciences, No. 9, 256–278, 1974.
- [54] Kaźmierczak J.: *Odwzorowanie działania elektrostalowniczego pieca łukowego w przestrzeni sygnału diagnostycznego*. Archiwum Akustyki, 19.3, 201–218, 1984.

- [55] Kiciński J. i inni: *Pozyskiwanie relacji diagnostycznych metodą badań symulacyjnych*. Raport IMP PAN w ramach projektu KBN 8T10B05708, Gdańsk, 1995.
- [56] Klimek A.: *Ograniczanie zbioru cech opisujących drgania wału turbozespołu*. III Krajowa Konferencja „Diagnostyka Techniczna Urządzeń i Systemów”, T. 3, 125–132, Instytut Techniczny Wojsk Lotniczych, Warszawa-Szczyrk, 1995.
- [57] Klimek A.: *Metody doskonalenia odwrotnych modeli diagnostycznych*. Zeszyty Naukowe Politechniki Śląskiej, Z. 134, seria: Mechanika, Gliwice, 1999.
- [58] Klösgen W., Żytkow J.: *Knowledge discovery and data mining*. Tutorial notes, PKDD, Trondheim, 1997.
- [59] Kohavi R., John G.: *The wrapper approach*. W: Liu H., Motoda H. (eds.): *Feature extraction, construction and selection: A data mining perspective*. Kluwer Academic Publisher, 1998.
- [60] Kohonen T.: *Self-organizing maps*. Springer Verlag, Heidelberg Berlin, 1995.
- [61] Komorowski J., Løken T.: *Diagnosing rotating machinery using rough sets*. Intelligent Information Systems, 137–145, IPPI PAN, Ustroń, 1999.
- [62] Kowalczyk W., Slisser F.: *Modeling customer retention with rough data models*. First European Symposium, PKDD, 145–155, Trondheim, 1997.
- [63] Kostka P.: *Optimization of representation space for knowledge acquisition on rotating machinery using machine learning methods*. Intelligent Information Systems, 235–238, IPI PAN, Malbork, 1998.
- [64] Kostka P.: *Badanie wrażliwości cech sygnałów diagnostycznych*. III Krajowa Konferencja Naukowo-Techniczna: „Diagnostyka Procesów Przemysłowych”, 85–90, Jurata, 1998.
- [65] Kościelny J.M.: *Analiza przydatności zbiorów przybliżonych do identyfikacji uszkodzeń*. III Krajowa Konferencja Naukowo-Techniczna „Diagnostyka Procesów Przemysłowych”, 55–60, Jurata, 1998.

-
- [66] Kościelny J.M., Ostasz A.: *Zastosowanie rozmytych sieci neuronowych do detekcji uszkodzeń urządzeń wykonawczych*. III Krajowa Konferencja Naukowo-Techniczna „Diagnostyka Procesów Przemysłowych”, 119–124, Jurata, 1998.
- [67] Kościelny J.M., Szczepaniak P.: *Terminologia oraz klasyfikacja metod detekcji i diagnostyki procesów przemysłowych*. II Krajowa Konferencja Naukowo-Techniczna „Diagnostyka Procesów Przemysłowych”, 57–67, Zielona Góra-Łagów, 1997.
- [68] Kryszkiewicz M., Rybiński H.: *Finding reducts in composed information systems*. International Workshop on Rough Sets and Knowledge Discovery, 261–273, Banff, Alberta, 1993.
- [69] Kubat M., Bratko I., Michalski R.S.: *A review of machine learning methods*. W: [80].
- [70] Langley P.: *Selection of relevant features in machine learning*. AAAI Fall Symposium, 140–144, New Orleans, LA, AAAI Press, 1994.
- [71] Liu H., Setiono R.: *Feature selection and classification — a probabilistic wrapper approach*. 9th International Conference on Industrial and Engineering Applications of AI and ES. 419–424, 1996.
- [72] Liu H., Setiono R.: *Incremental feature selection*. Journal of Applied Intelligence, Vol. 9, No. 3, 217–230, Kluwer Academic Publishers, 1998.
- [73] Łączkowski R.: *Wibroakustyka maszyn i urządzeń*. WNT, Warszawa, 1983.
- [74] Maniak P.: *Badanie przydatności metod indukcyjnych do pozyskiwania wiedzy projektowej*. Zeszyty Katedry Podstaw Konstrukcji Maszyn Politechniki Śląskiej, Z. 116, Gliwice, 2000.
- [75] Maniak P., Moczulski W.: *Przykład indukcji reguł dotyczących postaci niewyrównowazenia wirnika*. I Krajowa Konferencja „Metody i Systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim”, 159–164, AGH, Kraków, 1997.

- [76] Marszałek-Paszek B., Paszek P.: *Searching for attributes which „well” determinate decision in the decision table*. Intelligent Information Systems, 146–148, IPI PAN, Ustroń, 1999.
- [77] Michalewicz Z.: *Algorytmy genetyczne + struktury danych = programy ewolucyjne*. WNT, Warszawa, 1999.
- [78] Michalski R.S.: *A theory and methodology of inductive learning*. Artificial Intelligence, 20, 111–161, 1983.
- [79] Michalski R.S.: *Machine learning, data mining and knowledge discovery. Principles and applications*. Tutorial, Intelligent Information Systems, IPI PAN. Zakopane, 1997.
- [80] Michalski R.S., Bratko I., Kubat M. (eds.): *Machine learning and data mining: Methods and applications*. John Wiley and Sons, 1996.
- [81] Michalski R.S., Wnek J.: *Constructive induction: an automated improvement of knowledge representation spaces for machine learning*. Proceedings of the 2nd Conference on Practical Aspects of Artificial Intelligence, 188–236, IPI PAN, Augustów, 1993.
- [82] Michie D.: *Machine learning in the next five years*. EWSL-88 Proc. 3rd European Working Session on Learning, 107–122, Glasgow, London: Pittman, 1988.
- [83] MIMOSA. <http://www.mimosa.org>.
- [84] Mitchell M.: *An introduction to genetic algorithms*. MIT Press, Cambridge, MA, 1996.
- [85] Moczulski W.: *Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn*. Monografia. Zeszyty Naukowe Politechniki Śląskiej, seria: Mechanika, Z. 130, Gliwice, 1997.
- [86] Moczulski W.: *Methodology of knowledge acquisition for machinery diagnostics*. Computer Assisted Mechanics and Engineering Sciences, 6, 163–175, 1999.

-
- [87] Moczulski W.: *Methodology of acquisition of diagnostic knowledge on machinery*. II Międzynarodowy Kongres Diagnostyki Technicznej „Diagnostyka 2000”, 201–210, Warszawa, 2000.
- [88] Moczulski W.: *Pozyskiwanie wiedzy deklaratywnej i proceduralnej*. Symposium on Methods of Artificial Intelligence in Mechanics and Mechanical Engineering „AI-Mech 2000”, 55–69, Gliwice, 2000.
- [89] Moczulski W., Maniak P.: *Pozyskiwanie reguł metodą selektywnej indukcji z wyników obliczeń symulacyjnych dotyczących różnych postaci niewyrównowazenia*. Raport RMT6357, Katedra Podstaw Konstrukcji Maszyn, Politechnika Śląska, Gliwice, 1997.
- [90] Moczulski W., Żytkow J.: *Automated search for knowledge on machinery diagnostics*. Intelligent Information Systems, 194–203, IPI PAN, Zakopane, 1997.
- [91] Morel J.: *Drgania maszyn i diagnostyka ich stanu technicznego*. PTDT, Warszawa, 1992.
- [92] Mrózek A.: *Rough sets and dependency analysis among attributes in computer implementations of expert's inference models*. International Journal on Man-Machine Studies 30, 457–473, 1989.
- [93] Munakata T.: *Fundamentals of the new artificial intelligence. Beyond traditional paradigms*. Springer-Verlag, New York, 1998.
- [94] Negoita C.V.: *Expert systems and fuzzy systems*. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA, 1985.
- [95] Nguyen H.: *Discretization of real value attributes: A boolean approach*. Praca doktorska, Uniwersytet Warszawski, Warszawa, 1997.
- [96] Nguyen S., Nguyen H.: *Some efficient algorithms for rough set methods*. Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU, 1451–1456, Granada, Spain, 1996.
- [97] Nguyen S., Nguyen H.: *Discretization methods in data mining*. W: Polkowski L., Skowron A. (eds.): *Rough Sets in Knowledge Discovery*, 451–482, Physica-Verlag, Heidelberg, 1998.

-
- [98] Nguyen H., Skowron A.: *Quantization of real value attributes*. Proc. of the Second Joint Annual Conference on Information Sciences, 34–37, Wrightsville Beach, NC, 1995.
- [99] Nix A., Vose M.: *Modelling genetic algorithms with Markov chains*. Annals of Mathematics and Artificial Intelligence, 5, 79–88, 1992.
- [100] Nowicki R., Słowiński R., Stefanowski J.: *Evaluation of vibroacoustic diagnostic symptoms by means of the rough set theory*. Computers in Industry No. 20, 141–152, 1992.
- [101] Nowicki R., Słowiński R., Stefanowski J.: *Rough sets analysis of diagnostic capacity of vibroacoustic symptoms*. Computers Math. Applic. Vol. 24, No. 7, 109–123, 1992.
- [102] Øhrn A., Komorowski J., Skowron A., Synak P.: *The design and implementation of a knowledge discovery toolkit based on rough sets — the ROSETTA system*. Rough Sets in Knowledge Discovery: Methodology and Application, 376–399. Physica-Verlag, 1998.
- [103] Pawlak Z.: *Rough sets*. International Journal of Information and Computer Sciences, 11, No. 5, 341–356, 1982.
- [104] Pawlak Z.: *Rough sets: Theoretical aspects of reasoning about data*. Kluwer, Dordrecht, 1991.
- [105] Pham D., Xing L.: *Neural networks for identification, prediction and control*. Springer Verlag, London, 1995.
- [106] Puzoń J., Moczulski W.: *Akwizycja wiedzy diagnostycznej od ekspertów dziedzinowych z zastosowaniem formularzy*. Kongres Diagnostyki Technicznej, T. III, 189–194, KPKM Pol. Śl. i IMP PAN, Gdańsk, 1996.
- [107] Quinlan J.: *Induction of decision trees*. Machine Learning, 1, 81–106, 1986.
- [108] Quinlan J.: *C4.5 Programs for machine learning*. Morgan Kaufmann, San Mateo, CA, 1993.

-
- [109] Quinlan J.: *Improved use of continuous attributes in C4.5*. Journal of Artificial Intelligence Research, nr 4, 77–90, 1996.
- [110] Rissanen J.: *Stochastic complexity and modeling*. Ann. Statist 14, 1080–1100, 1986.
- [111] Rumelhart D., Hinton G., Williams R.: *Learning integral representation by error propagation*. Parallel Distributed Processing, Vol. 1, 318–362, 1986.
- [112] Rybczyński J.: *Stanowisko do badań dynamiki wirników i łożysk ślizgowych*. Zeszyt Naukowy Instytutu Maszyn Przepływowych PAN Nr 429/1394/94, Warszawa, 1994.
- [113] *C5/See5 Home page*, <http://www.rulequest.com>.
- [114] Setiono R., Liu H.: *Neural network feature selector*. IEEE Transactions on Neural Networks 8(3), 654–659, 1997.
- [115] SigLab Version 2.13. User's Guide. DSP Technology Inc. Fremont, CA. 1996.
- [116] Skowron A.: *A synthesis of decision rules: Application of discernibility matrix properties*. Intelligent Information Systems, 30–46, IPI PAN, Augustów, 1993.
- [117] Skowron A., Rauszer C.: *The discernibility matrices and functions in information systems*. W: Intelligent Decision Support. R. Słowiński (ed.) Handbook of Application and Advances of the Rough Set Theory, 331–362, Kluwer Academic Publishers, 1992.
- [118] *Słownik języka polskiego*. PWN, Warszawa, 1996.
- [119] Sobczak W., Malina W.: *Metody selekcji i redukcji informacji*. WNT, Warszawa, 1985
- [120] Sokołowski A., Kosmol J.: *Wybrane testy metod selekcji i integracji cech sygnałów pomiarowych*. Technologia maszyn. Naukowe warsztaty profesorskie TM'98, Wydawnictwo Uczelniane Politechniki Koszalińskiej, Koszalin, 1998.

- [121] Sokołowski A., Kosmol J.: *Wybrane aspekty praktycznego zastosowania koncepcji inteligentnego układu diagnostycznego*. II Międzynarodowy Kongres Diagnostyki Technicznej „Diagnostyka 2000”, Warszawa, 2000.
- [122] Solipiwo A.: *Metoda optymalizacji zbioru cech sygnałów diagnostycznych w badaniach maszyn*. Praca doktorska, Instytut Mechaniki i Podstaw Konstrukcji Maszyn, Politechnika Śląska, 1986.
- [123] Solartron Instruments: *Common Vibration Symptoms for Turbomachinery in Distress*.
- [124] Stawinoga J.: *Badania porównawcze metod określania progów kwantyzacji cech ilościowych dla potrzeb diagnostyki maszyn*. Praca dyplomowa, Katedra Podstaw Konstrukcji Maszyn, Politechnika Śląska, 1997.
- [125] Steinebach C., White M.: *VIBEX (Part 1): An intelligent knowledge-based system for fault diagnosis in turbomachinery*. Vibration and Wear in High-Speed Rotating Machinery, NATO ASI Series E: Applied Sciences, Vol. 174, Kluwer Academic Publishers, Netherlands, 1990.
- [126] Stepaniuk J.: *Attribute discovery and rough sets*. First European Symposium, PKDD'97, 145–155, Trondheim, 1997.
- [127] Susmaga R., Słowiński R., Greco S., Matarazzo B.: *Computation of reducts for multi-attribute and multi-criteria classification*. Intelligent Information Systems, 154–163, IPI PAN, Ustroń, 1999.
- [128] Urbanek G.: *Clustering and classification of elements in multidimensional metric spaces*. II Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim”, 225–230, Kraków, 1999.
- [129] Ventura D.: *On discretization as a preprocessing step for supervised learning models*. Masters Thesis, Department of Computer Science of Brigham Young University, 1995.
- [130] Warwick K., Ekwue A., Aggarwal R. (eds.): *Artificial intelligence techniques in power systems*. The Institution of Electrical Engineers, London, 1997.

-
- [131] Weiss S., Kapouleas I.: *An empirical comparison of pattern recognition, neural nets and machine learning classification methods*. 11th International Joint Conference on Artificial Intelligence, 177–183, 1989.
- [132] Wróblewski J.: *Finding minimal reducts using genetic algorithm*. Second Annual Joint Conference on Information Sciences, 186–189, Wrightsville Beach, NC, 1995.
- [133] Wróblewski J.: *Genetic algorithms in decomposition and classification problems*. W: Polkowski L., Skowron A.: *Rough sets in knowledge discovery*. 471–488, Physica-Verlag, Heidelberg, 1998.
- [134] Wysogład B.: *Metody reprezentacji drgań wałów maszyn wirnikowych w diagnostycznych bazach danych*. Zeszyty Naukowe Politechniki Śląskiej, seria: Mechanika, Z. 126, Gliwice, 1996.
- [135] Wyleźoł M.: *Elektroniczny formularz-edytor do pozyskiwania reguł empirycznych dla celów diagnostycznych*. I Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim”, 179–184, Akademia Górniczo-Hutnicza, Kraków, 1997.
- [136] Wyleźoł M.: *Metody pozyskiwania procedur i relacji diagnostycznych od specjalistów w dziedzinie eksploatacji maszyn*. Zeszyty Katedry Podstaw Konstrukcji Maszyn Politechniki Śląskiej, Z. 115, Gliwice, 2000.
- [137] Yang J., Honovar V.: *Feature subset selection using genetic algorithm*. Liu, Motoda (eds.): *Feature Extraction, Construction and Selection — A Data Mining Perspective*. Kluwer Academic Publishers, Boston, 1998.
- [138] Zadeh L.: *Fuzzy sets*. Information and Control. No. 8, 330–353, 1965.
- [139] Ziarko W., Shan N.: *Machine learning through data classification and reduction*. Intelligent Information Systems, 121–132, IPI PAN, Wigry, 1994.
- [140] Żytkow J., Zembowicz R.: *Database exploration in search of regularities*. Journal of Intelligent Information Systems, 2, 39–82, 1993.

Dodatek

Elementy teorii zbiorów przybliżonych

Pojęcie *zbioru przybliżonego* (ang. *rough set*) zostało wprowadzone przez Z. Pawłaka [104]. Dla zdefiniowania pojęcia zbioru przybliżonego wygodnie jest posłużyć się reprezentacją danych w postaci tzw. *systemu informacyjnego* [104], który jest uporządkowaną parą:

$$\mathcal{S} = (U, C), \quad (\text{A.1})$$

gdzie: $U = \{x_1, x_2, \dots, x_N\}$ jest niepustym i skończonym zbiorem przykładów zwanym *uniwersum*, natomiast C jest niepustym i skończonym zbiorem atrybutów $a_i \in C$ opisujących przykłady takim, że:

$$(\forall a \in C) \quad [a: U \rightarrow \text{Dom}(a)]. \quad (\text{A.2})$$

*Tablicą decyzyjną*¹ nazywa się system informacyjny:

$$\mathcal{S} = (U, A \cup D), \quad (\text{A.3})$$

gdzie: A jest zbiorem *atrybutów warunkowych*, D jest zbiorem *atrybutów decyzyjnych* i $A \cap D = \emptyset$.

Najczęściej zbiór D jest zbiorem jednoelementowym, i wówczas tablicę decyzyjną możemy zapisać jako:

$$\mathcal{S} = (U, A \cup \{d\}). \quad (\text{A.4})$$

¹W pracy termin ten stosowany jest do określenia tablicy przykładów uczących, w odróżnieniu od tablicy reguł.

Każdemu podzbiorni atrybutów $B \in A$ można przyporządkować *relację nierozróżnialności* zdefiniowaną jako:

$$\text{IND}(B) = \{(x, y) \in U \times U : \forall a \in B, a(x) = a(y)\}. \quad (\text{A.5})$$

$\text{IND}(B)$ jest równocześnie *relacją równoważności*:

$$\text{IND}(B) = \bigcap_{a \in B} \text{IND}(A). \quad (\text{A.6})$$

Oznacza to, że dwa przykłady x, y , opisane zbiorem atrybutów B są nierozróżnialne ze względu na B , co zapisujemy $x \text{IND}(B) y$, jeżeli wartości wszystkich atrybutów $a \in B$ dla przykładów x i y są takie same. Rodzinę relacji nierozróżnialności oznaczamy będziemy $[\text{IND}(B)]$. Zbiory przykładów nierozróżnialnych nazywane są *klasami równoważności* E^B . Klasy te są oczywiście rozłączne i $U = E_1^B \cup E_2^B \cup \dots \cup E_p^B$.

Jeżeli $\mathcal{S} = (U, A)$ jest systemem informacyjnym, $B \subseteq A$ jest zbiorem atrybutów i $X \subseteq U$ jest zbiorem przykładów, wówczas zbiory:

$$\underline{B}X = \{x \in U : [x]_{\text{IND}(B)} \subseteq X\}, \quad (\text{A.7})$$

$$\overline{B}X = \{x \in U : [x]_{\text{IND}(B)} \cap X \neq \emptyset\}, \quad (\text{A.8})$$

nazywane są odpowiednio *B-górną* i *B-dolną aproksymacją* zbioru X w \mathcal{S} .

Zbiór:

$$Bn_B(X) = \underline{B}X \setminus \overline{B}X \quad (\text{A.9})$$

nazywany jest *B-brzegiem* zbioru X .

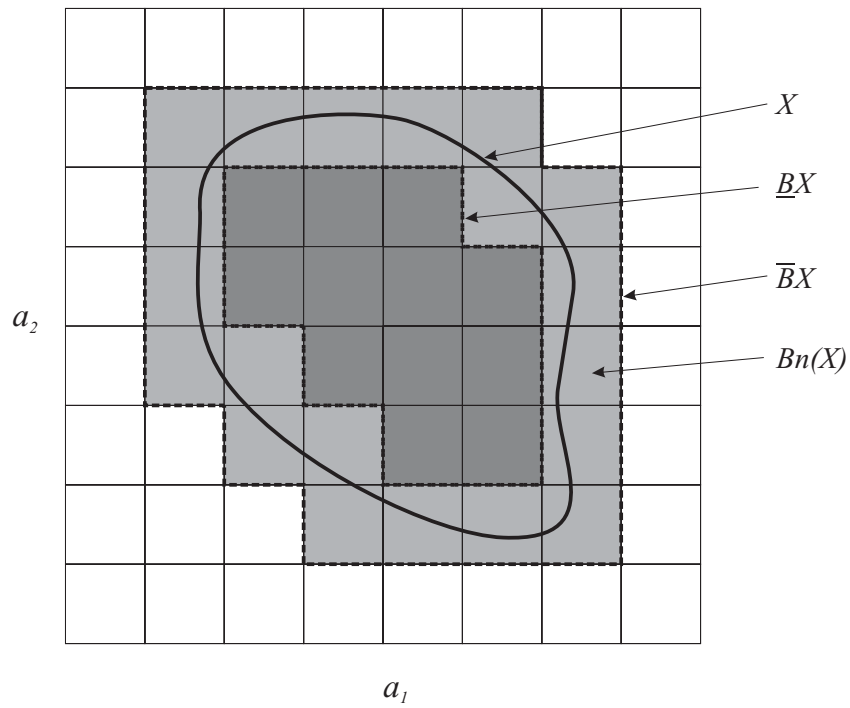
Zbiór przybliżony określony jest jako para: dolna i górna aproksymacja rozpatrywanego zbioru, przy czym aproksymacje te określone są według (A.7) i (A.8). Graficzną interpretację dolnej i górnej aproksymacji oraz brzegu zbioru przybliżonego przedstawiono na rys. A.1.

Jeżeli $\mathcal{S} = (U, A \cup \{d\})$, $\text{Dom}(d) = \{1, \dots, K\}$ oraz:

$$S_k = \{x \in U : \text{class}(x) = k\}, \quad k = 1, \dots, K \quad (\text{A.10})$$

wówczas zbiór:

$$\text{POS}_B(d) = \underline{B}S_1 \cup \dots \cup \underline{B}S_K \quad (\text{A.11})$$



Rysunek A.1: Ilustracja pojęć górnej i dolnej aproksymacji zbioru w przestrzeni dwuwymiarowej

nazywany jest *B-pozytywnym obszarem* systemu \mathcal{S} .

B-negatywnym obszarem systemu \mathcal{S} jest zbiór

$$NEG_B(d) = U \setminus [\overline{BS}_1 \cup \dots \cup \overline{BS}_K]. \quad (\text{A.12})$$

Spis rysunków

1.1	Model obiektu diagnozowania	14
1.2	Relacje w diagnostyce maszyn	15
2.1	Pozyskiwanie wiedzy metodą „uczenia maszynowego”	23
2.2	Geometryczna interpretacja wartości granicznych	31
2.3	Przykład ograniczenia liczności zbioru uwzględnianych atrybutów	33
4.1	Geometryczna interpretacja metody χ^2	46
4.2	Przykład wyznaczania wartości granicznych cechy diagnostycznej na podstawie histogramu	52
4.3	Przykład wyznaczania wartości granicznych cech diagnostycznych metodą grupowania	58
4.4	Przykład rozkładu częstości wartości atrybutu diagnostycznego AdifTD	63
5.1	Sposób selekcji atrybutów metodą bazującą na teorii algorytmów genetycznych	74
5.2	Sieć neuronalna zawierająca warstwę wejściową, jedną warstwę ukrytą i jeden węzeł w warstwie wyjściowej	77
5.3	Selekcja atrybutów z zastosowaniem algorytmu indukcyjnego „uczenia maszynowego”	79
5.4	Metoda selekcji atrybutów z zastosowaniem algorytmu indukcyjnego „uczenia maszynowego” i konwersją wartości atrybutów	80
6.1	Proces oceny i wyboru metod redukcji i selekcji informacji	92
7.1	Model maszyny wirnikowej <i>Rotor Kit</i>	97

7.2	Przykład wyznaczania wartości granicznych na podstawie histogramu dla atrybutu $1X_{Ax}$ (eksperyment RK)	100
7.3	Przykład wyznaczania wartości granicznych na podstawie histogramu dla atrybutu Ax_{Max} (eksperyment RK)	101
7.4	Graficzna prezentacja wyników redukcji informacji diagnostycznej dla przypadku eksperymentu czynnego RK	106
7.5	Wyniki oceny metod redukcji informacji diagnostycznej (eksperyment RK)	106
7.6	Graficzna prezentacja wyników selekcji atrybutów dla przypadku eksperymentu czynnego RK	113
7.7	Teoretyczny model układu wirnik–łożyska–podpory łożyskowe	115
7.8	Przykładowe rozmieszczenie niewyważ	116
7.9	Cechy trajektorii zastosowane w badaniach GD	119
7.10	Rozkład częstości wartości atrybutu $AdifDT$ (eksperyment GD)	120
7.11	Rozkład częstości wartości atrybutu $R02D2$ (eksperyment GD)	120
7.12	Wyniki oceny jakości podziału wartości atrybutu $AdifTD$ na grupy (eksperyment GD)	121
7.13	Wyniki oceny jakości podziału wartości atrybutu $R02D2$ na grupy (eksperyment GD)	122
7.14	Graficzna prezentacja wyników redukcji informacji diagnostycznej dla przypadku eksperymentu numerycznego GD	125
7.15	Wyniki oceny metod redukcji informacji diagnostycznej dla przypadku eksperymentu numerycznego GD	126
7.16	Proponowana metoda redukcji i selekcji informacji dla potrzeb diagnostyki maszyn	131
A.1	Ilustracja pojęć górnej i dolnej aproksymacji zbioru w przestrzeni dwuwymiarowej	153

Spis tablic

6.1	Określenie wagi przyjętych kryteriów oceny metod redukcji informacji .	89
6.2	Określenie wagi przyjętych kryteriów oceny metod selekcji informacji .	90
7.1	Liczba pomiarów przeprowadzonych dla danego rodzaju (klasy) niesprawności (eksperyment RK)	98
7.2	Wykaz atrybutów stosowanych w badaniach dotyczących wyników czynnego eksperymentu diagnostycznego RK	99
7.3	Sprawność klasyfikatora w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment RK)	103
7.4	Błędy klasyfikatora dla wybranych metod wyznaczania wartości granicznych (eksperyment RK)	103
7.5	Błędy ważone klasyfikatora dla wybranych metod wyznaczania wartości granicznych (eksperyment RK)	104
7.6	Liczba wartości granicznych w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment RK)	105
7.7	Liczby reguł uzyskane dla różnych metod wyznaczania wartości granicznych (eksperyment RK)	107
7.8	Ocena metod określania wartości granicznych na podstawie przyjętych kryteriów (eksperyment RK)	108
7.9	Ważona ocena metod określania wartości granicznych (eksperyment RK)	108
7.10	Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów dla redukcji metodą heurystyczną (eksperyment RK)	109

7.11 Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów dla redukcji metodą grupowania jednoatrybutowego (eksperyment RK)	109
7.12 Wyznaczone zbiory atrybutów relewantnych (eksperyment RK)	110
7.13 Liczba pozyskanych reguł w zależności od zastosowanej metody selekcji atrybutów (eksperyment RK)	112
7.14 Ocena metod selekcji atrybutów (eksperyment RK)	112
7.15 Ważona ocena metod selekcji atrybutów (eksperyment RK)	113
7.16 Liczba przykładów w poszczególnych klasach (eksperyment GD)	116
7.17 Wykaz atrybutów stosowanych w badaniach dotyczących wyników eksperymentu numerycznego GD	118
7.18 Sprawność klasyfikatora w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment GD)	123
7.19 Błędy klasyfikatora dla wybranych metod wyznaczania wartości granicznych (eksperyment GD)	123
7.20 Błędy ważone klasyfikatora dla wybranych metod wyznaczania wartości granicznych (eksperyment GD)	123
7.21 Liczba wartości granicznych w zależności od zastosowanej metody wyznaczania wartości granicznych (eksperyment GD)	124
7.22 Liczby reguł uzyskane dla różnych metod wyznaczania wartości granicznych (eksperyment GD)	124
7.23 Ocena metod redukcji informacji diagnostycznej (eksperyment GD)	124
7.24 Ważona ocena metod redukcji informacji diagnostycznej (eksperyment GD)	125
7.25 Liczba atrybutów relewantnych w zależności od zastosowanej metody selekcji atrybutów (eksperyment GD)	127
7.26 Wyniki selekcji atrybutów dla metody z zastosowaniem algorytmu indukcyjnego i konwersją wartości atrybutów (eksperyment GD)	128
7.27 Atrybuty relewantne uzyskane metodą z zastosowaniem algorytmu indukcyjnego i konwersją wartości atrybutów (eksperyment GD)	129

Spis algorytmów

4.1	Sposób wyznaczania wartości granicznych metodą równych szerokości przedziałów	44
4.2	Sposób wyznaczania wartości granicznych metodą równych częstości	45
4.3	Sposób wyznaczania wartości granicznych metodą χ^2	47
4.4	Sposób wyznaczania wartości granicznych metodą C	50
4.5	Sposób wyznaczania wartości granicznych według histogramu	51
4.6	Sposób wyznaczania wartości granicznych metodą minimum entropii	53
4.7	Znajdowanie najlepszej wartości granicznej dla metody minimum entropii	54
4.8	Sposób wyznaczania wartości granicznych jednoatrybutową metodą grupowania	56
4.9	Sposób wyznaczania wartości granicznych wieloatrybutową metodą grupowania	59
4.10	Sposób wyznaczania wartości granicznych metodą heurystyczną	62
5.1	Sposób selekcji atrybutów metodą opartą na analizie korelacji	68
5.2	Sposób obliczania reduktu minimalnego	70
5.3	Obliczanie reduktu losowego	71
5.4	Selekcja atrybutów oparta na teorii algorytmów genetycznych	73
5.5	Sposób selekcji atrybutów metodą minimum entropii	76
5.6	Losowa selekcja atrybutów	81
5.7	Selekcja atrybutów wg kryterium niesprzeczności	82
5.8	Przyrostowa selekcja atrybutów	83

Metoda selekcji i redukcji informacji w diagnostyce maszyn

Streszczenie

Głównym problemem związanym z zastosowaniem systemów doradczych w dziedzinie diagnostyki maszyn jest problem pozyskiwania wiedzy. Jednym ze stosowanych sposobów jest pozyskiwanie wiedzy z baz danych z zastosowaniem indukcyjnych metod „uczenia maszynowego”. Dane stanowią zwykle wyniki obserwacji i pomiarów dokonywanych na badanym obiekcie (maszynie), reprezentowanych jako wartości cech sygnałów diagnostycznych. Rejestrowane wartości cech sygnałów diagnostycznych są najczęściej wartościami ilościowymi. Stosowanie większości metod uczenia maszynowego związane jest z koniecznością konwersji wartości ilościowych na wartości jakościowe. Istotne jest również ograniczenie liczby cech uwzględnianych w procesie pozyskiwania wiedzy.

Celem rozprawy było porównanie metod selekcji i metod redukcji informacji diagnostycznej oraz opracowanie metody redukcji i selekcji informacji przydatnej do zastosowania w dziedzinie diagnostyki maszyn. Tak określony cel umożliwił sformułowanie tezy, że możliwe jest przyjęcie metody redukcji i selekcji informacji diagnostycznej, której zastosowanie w procesie pozyskiwania wiedzy dotyczącej relacji diagnostycznych powoduje wzrost efektywności tego procesu oraz zwiększa skuteczność wnioskowania.

W pracy opisano wybrane metody wyznaczania wartości granicznych atrybutów, pozwalające na konwersję wartości ilościowych na wartości jakościowe — czyli redukcję informacji diagnostycznej. Przedstawiono zarówno metody jednoatrybutowe wyznaczania wartości granicznych — dla każdego atrybutu oddzielnie, jak również metody wieloatrybutowe — wyznaczanie wartości granicznych dla wszystkich atrybutów jednocześnie. Opisano także wybrane metody selekcji atrybutów. Przedstawiono podstawy teoretyczne oraz szczegółowe algorytmy opisanych metod.

Podjęto próbę sformalizowanej oceny metod redukcji i selekcji informacji diagnostycznej. Przedstawiono kryteria oceny, a także sposoby oceny tych metod. Zastosowano pośrednią ocenę metod poprzez ocenę wiedzy pozyskanej na podstawie przykładów, dla których badanymi metodami dokonano redukcji i selekcji informacji.

Przeprowadzono weryfikację metod redukcji i selekcji informacji diagnostycznej biorąc pod uwagę wyniki dotyczące dwóch eksperymentów diagnostycznych: czynnego eksperymentu diagnostycznego na modelu maszyny wirnikowej oraz eksperymentu numerycznego. Wyniki weryfikacji potwierdzają słuszność sformułowanej tezy, dowodząc, że zastosowanie opracowanej metody redukcji i selekcji informacji zwiększa skuteczność klasyfikacji.

Słowa kluczowe: diagnostyka maszyn, uczenie maszynowe, konwersja wartości atrybutów ilościowych, selekcja atrybutów, ocena metod redukcji i selekcji informacji

Dziedziny: diagnostyka maszyn, eksploatacja maszyn

Selection and reduction of information in machinery diagnostics

Summary

The problem of knowledge acquisition is the main problem connected with applications of expert systems in the domain of machinery diagnostics. Knowledge acquisition from databases using machine learning methods is one of the methods of knowledge acquisition. Collected data are usually the results of observations and measurements carried out on an investigated object. In most cases, collected values of diagnostic attributes (diagnostic symptoms) are quantitative. The usage of machine learning methods is connected with the necessity of conversion of quantitative values to qualitative ones. Limitation of attributes number used in knowledge acquisition process is also crucial.

The aim of the dissertation was to compare methods of knowledge reduction and selection and to work out a method of knowledge reduction and selection which will be useful in the machinery diagnostic domain. Such an aim allowed to formulate a thesis of the dissertation, that it is possible to elaborate a method of diagnostic knowledge reduction and selection that, if applied to the process of diagnostic knowledge acquisition, increases the efficiency of the process and efficiency of the inference.

In the dissertation selected methods of boundary values estimation are presented. These values make the conversion of attributes values (also called reduction of information) possible. There are presented one-attribute as well as multi-attribute methods. Using one-attribute methods estimation of boundary values is carried out separately for each attribute, while multi-attribute methods allow to estimate boundary values of all attributes simultaneously. The methods of attributes selection are also presented. For these methods theoretical preliminaries and detailed algorithms are presented.

The author tried to assess (in formalized way) the methods of information reduction and selection. Some assessment criteria are presented as well as procedures of the methods assessment. An indirect procedure of the methods assessment by assessment of the acquired knowledge was applied.

Verification of diagnostic information reduction and selection methods was conducted for the results of two diagnostic experiments: active diagnostic experiment on the rotating machine model and numerical experiment. The results confirm the thesis that has been put forward. The conversion of quantitative diagnostic symptom values as well as the selection of relevant attributes improves the efficiency of the inference.

Key words: machinery diagnostics, machine learning, conversion of quantitative attributes values, selection of attributes, evaluation of methods of information reduction and selection

Domains: diagnostics of machinery, exploitation of machinery

