

## REDUKCJA DANYCH W DIAGNOSTYCZNYCH BAZACH DANYCH

Paweł CHRZANOWSKI

Katedra Podstaw Konstrukcji Maszyn, Politechnika Śląska  
ul. Konarskiego 18a, 44-100 Gliwice, tel. (32) 237 14 67, fax. (32) 237 13 10, email: p.ch@polsl.pl

### Streszczenie

W artykule przedstawiono problematykę związaną ze wspomaganie procesu wnioskowania o stanie obiektu rzeczywistego. Głównym zagadnieniem jest redukcja olbrzymiej ilości danych dostarczanych do systemu monitorowania. Wyróżniono trzy grupy metod: ograniczania liczby rozpatrywanych cech, ograniczania w zbiorze rozróżnianych wartości oraz ograniczania liczby elementów wykresu wartości. W wyniku przeprowadzonej analizy danych zaproponowano metodę kwantowania z histerezą w celu redukcji liczby rozróżnianych wartości. Ponadto zaproponowano metodę optymalizacji szerokości pasma kwantowania z histerezą z użyciem testu statystycznego.

Słowa kluczowe: systemy doradcze, systemy monitorowania, diagnostyczne bazy danych, redukcja danych

### DATA REDUCTION IN DIAGNOSTIC DATABASES

#### Summary

The article present problems about computer aided machinery state reasoning. The main task of this issue was reduction of huge quantity data sets provides to monitoring system. The methods can be divided into methods for reduction of features, methods for reduction of measured data sets and time-domain methods. On the basis of data analysis, was proposed data set reduction by sampling with hysteresis to reduction of measured data sets. Moreover the method for a tolerable degree fit was proposed and analyzed. The method is based upon statistical analysis.

Keywords: expert systems, monitoring systems, diagnostic databases

## 1. WSTĘP

Współczesna technika daje olbrzymie możliwości w zakresie monitorowania stanu złożonych obiektów technicznych. Niezaprzeczalny wpływ ma na to między innymi rozwój techniki pomiarowej, coraz większe możliwości sprzętu komputerowego, duża moc obliczeniowa komputerów, nowoczesne sieci komputerowe (Internet, Intranet). Dzięki tym nowym możliwościom, wspomaganie procesu obserwacji stanu np. turbospółów, może być prowadzone w czasie rzeczywistym. Proces ten może być wspomagany przez systemy doradcze (pakiet programów komputerowych), wykorzystujące wiedzę i procedury rozumowania eksperta danej dziedziny [7]. Spośród wielu zastosowań, systemy doradcze służą m.in. do [3]:

- obserwacji badanego obiektu,
- numerycznego wyznaczania cech obserwowanych sygnałów,
- gromadzenia danych,
- wnioskowania o stanie technicznym obserwowanego obiektu na podstawie wyznaczonych cech sygnałów diagnostycznych.

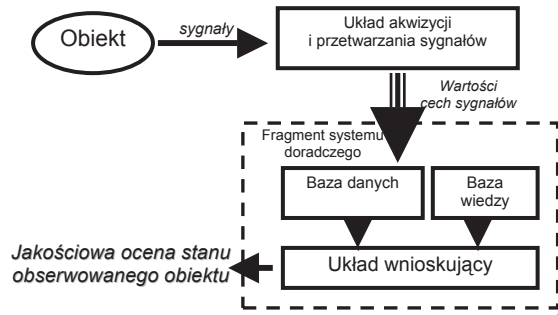
Głównymi elementami systemu doradczego są m.in. [2]:

- baza wiedzy,
- baza danych,
- układ wnioskujący,
- układ objaśniający,
- układ sterowania dialogiem.

Ze względu na specyfikę dostarczanych danych, systemy doradcze dzielą się na systemy statyczne, które wspomagają poszukiwanie rozwiązania w stałym otoczeniu oraz systemy dynamiczne działające w zmieniającym się otoczeniu.

Złożoność obiektu, jakim jest turbospół wymaga analizy wielu danych pozyskiwanych w czasie obserwacji obiektu, a także, kiedy ocenę stanu należy prowadzić w zmieniającym się otoczeniu [4]. Do tego celu budowane są systemy komputerowe, których jednym z elementów są dynamiczne systemy doradcze. Proces wnioskowania w tych systemach bazuje między innymi na danych gromadzonych w bazie danych. Dane te to w głównej mierze wartości cech sygnałów diagnostycznych, które wyznaczane są przez specjalistyczne układy akwizycji i przetwarzania sygnałów dostarczanych z obserwowanego obiektu. W przypadku

turbozespołów, w głównej mierze obserwowane są sygnały drganiowe i sygnały procesowe.

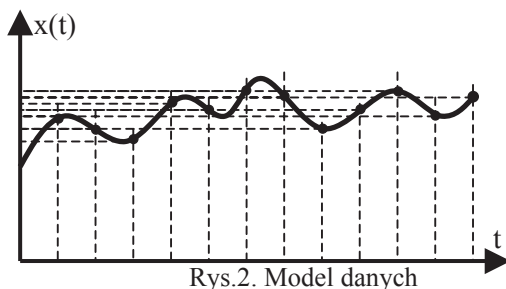


Rys.1 Ocena stanu obserwowanego obiektu

Dla potrzeb dynamicznego gromadzenia wartości cech sygnałów w bazach danych wymagane jest przyjęcie odpowiedniego modelu danych w postaci ich wykresów.

$$x = \{ \langle t, x(t) \rangle \} \quad (1)$$

Wykresem możemy nazywać zarówno rysunek pokazujący przebieg zmiennej zależnej w funkcji zmiennej niezależnej jak również zbiór par złożonych z wartości zmiennej zależnej i odpowiadającej jej zmiennej niezależnej (rys.2.).



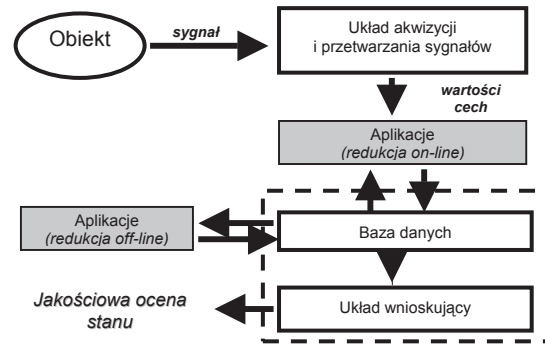
Rys.2. Model danych

Dla obserwowanego obiektu technicznego, jakim jest turbozespół liczba wyznaczanych wykresów wartości cech sygnałów diagnostycznych sięga kilku tysięcy. Z tego powodu zaistniała potrzeba ograniczania wykresów wartości cech sygnałów.

## 2. REDUKCJA DANCH

Działania zmierzające do ograniczania wykresów wartości cech sygnałów można rozpatrywać jako działania na danych bieżących (*on-line*) i działania na danych historycznych (*off-line*).

W przypadku ograniczania danych historycznych dysponujemy statycznym zbiorem danych, z kolei metody ograniczania *on-line* bazują na danych bieżących i danych historycznych, przez co analizowany zbiór danych jest zbiorem ciągle zmieniającym się.



Rys. 3. Dynamiczne ograniczanie wykresów wartości cech

Przyjęto, że rozpatrywane będą metody umożliwiające redukcję danych bieżących oraz danych historycznych. Poważnym wyzwaniem jest opracowanie i zastosowanie metody ograniczania wykresów wartości cech sygnałów diagnostycznych sporządzanych i gromadzonych w czasie rzeczywistym.

Można wyróżnić następujące grupy metod ograniczania wykresów wartości cech sygnałów diagnostycznych:

- ograniczanie liczby cech,
- ograniczanie liczby elementów zbioru rozróżnianych wartości,
- ograniczanie liczby elementów wykresów wartości cech.

### 2.1. Ograniczanie liczby cech

Względy praktyczne powodują, iż liczba obserwowanych cech powinna być ograniczona. Istnieje wiele przyczyn ograniczania liczby cech, najczęściej na skutek:

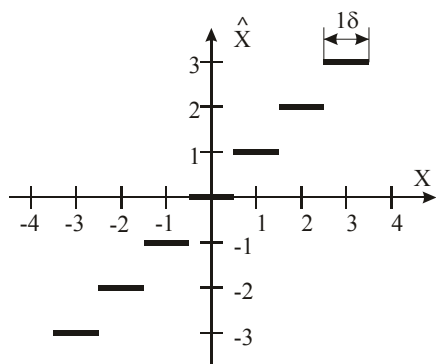
- analizy działania obiektu badań,
- braku możliwości technicznych prowadzenia obserwacji wybranych cech sygnałów diagnostycznych,
- ograniczonej liczby rozróżnianych klas stanów obiektu,

Spośród wielu metod ograniczania liczby cech, można wyróżnić redukcję cech jako wynik selekcji cech niosących znaczące informacje [8]. Inną grupą metod są metody zmierzające do wyznaczenia nowych cech reprezentujących zbiór określonych cech. Proces ograniczania liczby cech rozpoczyna się już w trakcie budowania systemu diagnostycznego. Istotne jest podjęcie decyzji przez specjalistę, wartości jakich cech mają być wyznaczone i gromadzone dla potrzeb procesu monitorowania oraz wnioskowania o stanie obserwowanego obiektu. Rozpatrując dalszą część omawianego problemu, przyjmuje się, iż zbiór obserwowanych cech jest ustalony w ramach odrębnych badań.

## 2.2. Ograniczanie liczby elementów w zbiorze rozróżnianych wartości

Do metod ograniczania wykresów wartości cech w zbiorze wartości można zaliczyć m.in. metody klasyfikacji, reprezentacji przybliżonej, reprezentacji rozmytej, kwantowania. Stosowanie metod klasyfikacji wymaga definiowania odpowiednich wzorców odpowiadających przyjętym klasom wartości.

Popularną metodą ograniczania liczności zbioru wartości cech jest kwantowanie [1], [9]. Metody kwantowania można podzielić na kwantowanie równomierne i kwantowanie nierównomierne. W wyniku kwantowania równomiernego otrzymywane są nowe wartości będące przybliżeniem dokładnych wartości chwilowych z tolerancją równą połowie szerokości pasma kwantowania. Jeżeli omawiane metody kwantowania będą stosowane do redukcji zbioru rozróżnianych wartości w kolejnych chwilach czasu (dynamicznie), to istnieje duże prawdopodobieństwo występowania wartości w kolejnych chwilach na granicach sąsiadujących pasm kwantowania. Powstały w ten sposób szum kwantowania jest zjawiskiem bardzo niepożądanym, w szczególności, gdy zredukowane wartości cech będą przekształcane do postaci jakościowej. Wrażliwość tę można zlikwidować poprzez wprowadzenie histerezy do procedury kwantowania. W wyniku takiego postępowania, bieżąca wartość cechy jest wyznaczana z uwzględnieniem poprzedzającej wartości będącej wynikiem kwantowania. Tak zmodyfikowaną metodę nazwano *kwantowaniem z histerezą*.



Rys. 4. Kwantowanie

Funkcja kwantowania z histerezą zapisywana jest następująco [5]:

**if**

$$\hat{x}_i(t-1) - n\delta \leq x_i(t) \leq \hat{x}_i(t-1) + n\delta$$

**then**

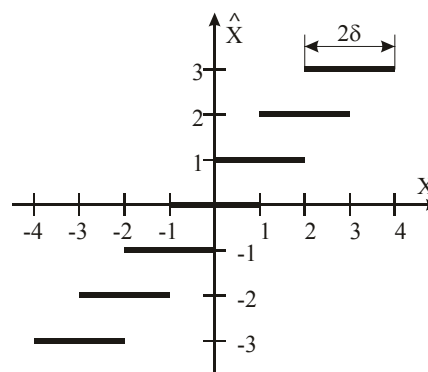
$$\hat{x}_i(t) = \hat{x}_i(t-1) \quad (2)$$

**else**

$$\hat{x}_i(t) = \text{round}\left(\frac{x_i(t)}{\delta}\right)\delta$$

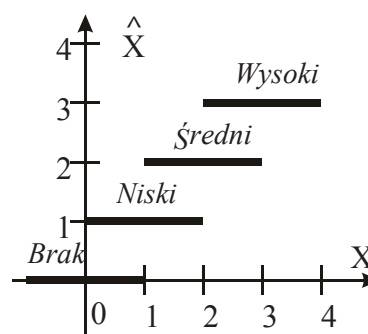
$$\text{dla } n = (0,5;1)$$

Postać tak zmodyfikowanej funkcji kwantowania z histerezą dla  $n=1$  przedstawia rys.5.



Rys. 5. Kwantowanie z histerezą (dla  $n=1$ )

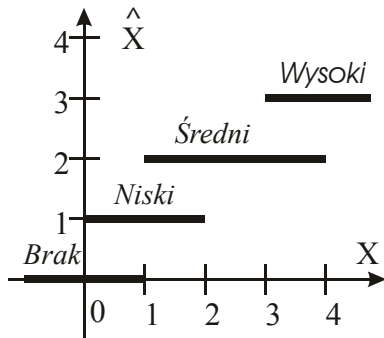
Kwantowanie z histerezą można zastosować nie tylko do redukcji wartości cech ilościowych, ale też do redukcji przekształcanych wartości cech ilościowych do postaci jakościowej. Wówczas klasom wartości jakościowych można przyporządkować wartości ilościowe uzyskiwane w wyniku procesu kwantowania.



Rys. 7. Kwantowanie z histerezą do wartości jakościowych

Przedstawione powyżej przykłady dotyczą procesu kwantowania równomiernego. Histerezą można również wprowadzić do dynamicznej redukcji cech ilościowych czy jakościowych, dla których kolejne przedziały są zdefiniowane „ostro”, lecz nierównomiernie. Dla potrzeb takiego postępowania, wymagane jest posiadanie odpowiedniej definicji wzorców klas wartości.

Możliwe jest również wprowadzenie histerezy dla klas, których przedziały są różnej długości. Przykład przedstawiono na rys.8.



Rys.8. Kwantowanie z histerezą do wartości jakościowych (klasy nierównomierne)

### 2.3. Ograniczanie liczby elementów wykresów wartości cech

Wykresy wartości cech mogą być ograniczane w zbiorze liczby elementów poprzez odpowiedni dobór częstości pobierania danych. Zbyt mała częstość pobierania danych może spowodować brak dostatecznej informacji o zjawiskach zachodzących w obiekcie (np. dziury w danych). Natomiast zbyt duża częstość pobierania danych może spowodować szybkie zapełnianie się bazy danych wartościami nadmiarowymi. Dlatego też dobór częstości zbierania danych zależy od możliwości aparatury pomiarowej, potrzeb systemu doradczego oraz dynamiki zmian obserwowanych wartości cech. Ważnym kryterium zbierania danych jest zachowanie ciągłości informacji o stanie obserwowanego obiektu. W przypadku dynamicznych systemów doradczych ważne jest rejestrowanie czasu rejestracji obserwowanej wartości cech [4].

$$x = \langle o, n(x), val(x), [t_p, t_k] \rangle \quad (3)$$

gdzie:

- o - rozpatrywany obiekt,
- n - nazwa cechy,
- val - wartość cechy
- $t_p$  - chwila określająca początek obserwacji wartości cechy x,
- $t_k$  - chwila określająca koniec obserwacji wartości cechy x.

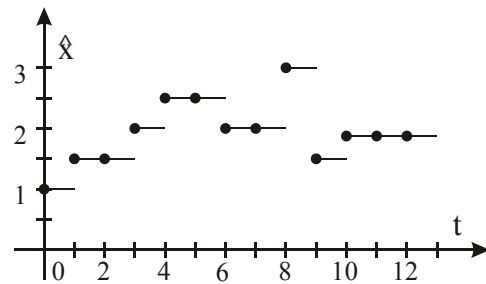
Przyjmując przedstawiony powyżej zapis, możliwe staje się ograniczanie liczności zbioru wartości cech sygnałów dziedzinie czasu poprzez sklejanie.

Jeżeli wartości cechy w kolejnych chwilach nie zmieniają się, bądź zmiana obserwowanych wartości zawarta jest w określonych granicach, wówczas zbiór tych wartości chwilowych można zastąpić jedną wartością uwzględniającą łączny przedział

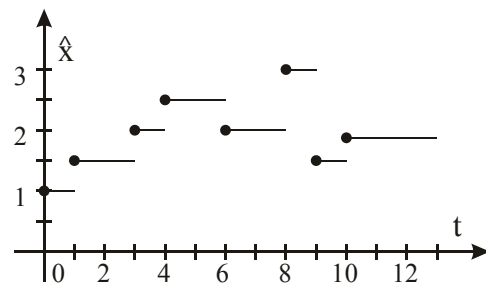
czasowy ich obserwacji. W tym przypadku  $t_p$  określa chwilę, w jakiej zaobserwowano rozpatrywaną wartość po raz pierwszy, natomiast  $t_k$  określa chwilę, w jakiej zaobserwowano ostatnią taką samą wartość. Chwila  $t_k$  jest jednocześnie czasem  $t_p$  początku obserwacji nowej wartości. Operację sklejania zapisano w postaci następującej procedury [4]:

$$\begin{aligned} &\text{if} \\ &\hat{x}_i(\overline{t_{p-1}t_{k-1}}) = \hat{x}_i(\overline{t_p t_k}) \\ &\text{then} \\ &\hat{x}_i(\overline{t_{p-1}t_k}) \\ &\text{else} \\ &\hat{x}_i(\overline{t_p t_k}) \end{aligned} \quad (4)$$

Sklejaniu mogą być poddawane zarówno dane ilościowe jak i dane jakościowe. Na rys. 8 przedstawiono przykładowe dane ilościowe w funkcji czasu. W wyniku zastosowania procedury sklejania otrzymano zredukowany zbiór danych przedstawiony na rys. 9.



Rys. 8. Przykładowe dane



Rys. 9. Przykładowe dane po sklejaniu

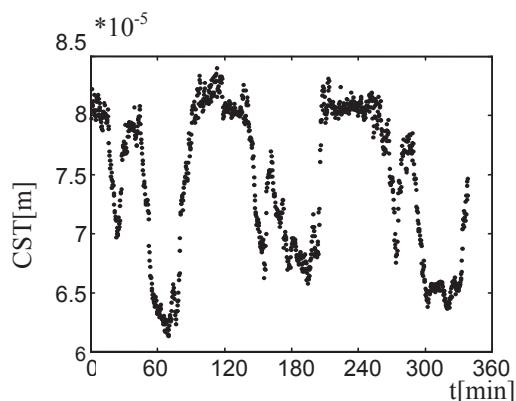
### 3. OPTIMALIZACJA SZEROKOŚCI PASMA KWANTOWANIA Z HISTEREZĄ

Dla potrzeb stosowania metody kwantowania z histerezą, należy przeprowadzić badania związane z optymalizacją szerokości pasma oraz szerokości histerezy.

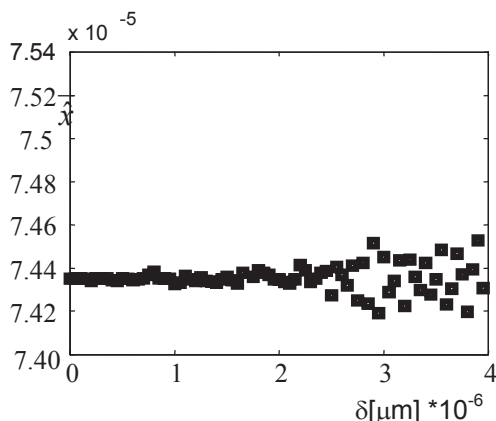
Rozpatrując zbiór wartości cech sygnałów diagnostycznych w postaci ilościowej, do rozwiązania postawionego zadania, można zastosować np. metody analizy statystycznej.

Podstawowym zadaniem stosowania testu statystycznego jest przyjęcie (nie odrzucenie) hipotezy o podobieństwie dwu rozkładów (rozkład wykresu wejściowego i rozkład wykresu ograniczonego przez kwantowanie z histerezą) przy założonym poziomie istotności. Można przyjąć, iż nie odrzucenie założonej hipotezy oznaczać będzie zachowanie głównej informacji diagnostycznej. Do oceny statystycznej w procesie optymalizacji szerokości pasma kwantowania można zastosować test  $\chi^2$ . W omawianym przykładzie przyjęto, iż ograniczenie należy uznać za dopuszczalne, gdy poziom istotności jest większy od 0.9. Oznacza to, iż zgodność pomiędzy wykresem rzeczywistym, a zredukowanym mniejsza od 90% nie będzie akceptowane.

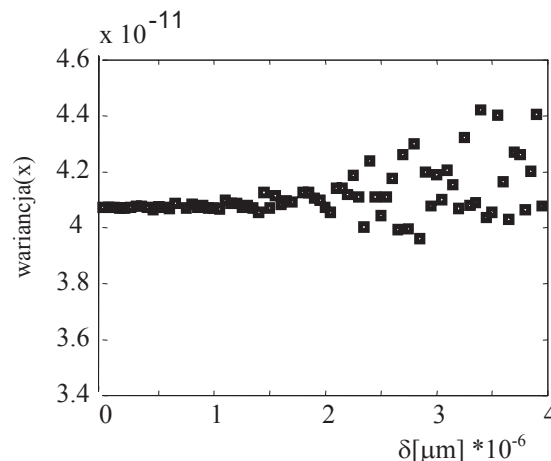
Dla przykładu optymalizację szerokości pasma kwantowania przeprowadzono dla danych pochodzących z obserwacji obiektu rzeczywistego (rys.10). Zmienność wybranych cech wykresu zredukowanego przedstawiono na rys.11 oraz rys.12, gdzie pierwsza wartość jest wartością wyznaczoną dla wykresu rzeczywistego. Dokonując obserwacji można stwierdzić, iż wyraźnie rysuje się granica, do jakiej możliwe jest przyjęcie poszukiwanej szerokości pasma kwantowania.



Rys. 10. Przykładowe dane (wykres wartości składowej stałej przemieszczeń względnych wału turbiny w łożysku nr 1)

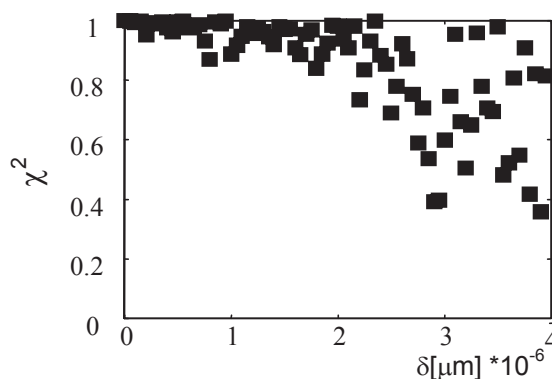


Rys.11. Zmiana średniej w funkcji szerokości pasma kwantowania



Rys. 12. Wariancja w funkcji szerokości pasma kwantowania.

Dokonując optymalizacji szerokości pasma kwantowania za pomocą proponowanego testu statystycznego  $\chi^2$  [10], wybranego przykładu uzyskano wyniki przedstawione na rys.13.



Rys. 13. Wyniki testu  $\chi^2$  przy optymalizacji pasma kwantowania.

#### 4. MIARA STOPNIA OGRANICZENIA WYKRESU WARTOŚCI

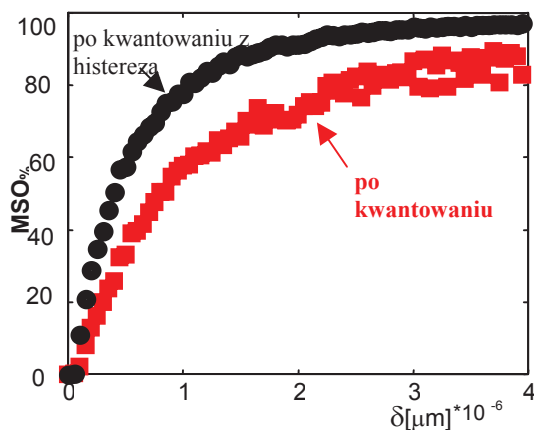
Przedstawione metody ograniczania zbioru rozróżnianych wartości przez dynamiczne kwantowanie z histerezą oraz ograniczania liczności zbioru wartości cech poprzez sklejanie można realizować jednocześnie w trakcie procesu rejestracji danych. Ocena efektywności takiego postępowania można przeprowadzić za pomocą miary określającej procentowy stopień ograniczenia (redukcji):

$$MSO_{\%} = \left(1 - \frac{n}{m}\right) * 100 \quad (5)$$

gdzie:

- n - liczba wartości ograniczonego wykresu wyjściowego (po redukcji)  
m - liczba wartości wykresu wejściowego (przed redukcją)

Przykład porównania efektywności metody kwantowania z histerezą w stosunku do metody kwantowania, przedstawiono na rys.14. Można zauważyć, iż wprowadzenie histerezy likwiduje szum powstający na granicach klas, co daje znaczną redukcję danych.



Rys.14. Ocena stopnia redukcji w wyniku sklejania wartości kwantowanych

## 7. PODSUMOWANIE

W wyniku przeprowadzonych badań stwierdzono, iż:

- zaproponowana metoda wstępnego przetwarzania przez *kwantowanie z histerezą* wraz z równoczesnym sklejaniem wartości w dziedzinie czasu, umożliwia znaczną redukcję wykresów wartości cech sygnałów diagnostycznych,
- kwantowanie z histerezą tłumi lokalne zmiany wartości, szczególnie na granicach klas,
- pomimo znacznego ograniczenia liczby gromadzonych wartości cech, możliwe jest wyznaczanie jakościowych ocen tych wartości,
- zaproponowana metoda kwantowania z histerezą może być stosowana procesie redukcji danych zapisywanych w postaci jakościowej, którym są przyporządkowane cechy ilościowe (identyfikatory).

## LITERATURA

- [1] Bendat J.S., Piersol A.G., Metody analizy i pomiaru sygnałów losowych. PWN, Warszawa 1976
- [2] Cholewa W. „Bazy danych w diagnostyce technicznej”. II Międzynarodowy kongres Diagnostyki Technicznej *Diagnostyka 2000*, str.156-164. Warszawa 2000.
- [3] Cholewa W. „Dynamiczne systemy doradcze”. Methods of Artificial Intelligence In Mechanics and Mechanical Engineering, str.37-54, Gliwice, 2000.
- [4] Cholewa W. Kiciński J., red. Pracy zbiorowej, „DT200-1. System diagnostyczny dla turbozespołów energetycznych o mocy 200MW”, KPKM, IMP PAN Gdańsk, 1998. Projekt badawczy zamawiany PBZ 038-06.
- [5] Chrzanowski P. „Metoda jakościowej oceny cech sygnałów diagnostycznych”. Rozprawa doktorska. Katedra Podstaw Konstrukcji Maszyn – Politechnika Śląska, Zeszyt 117. Gliwice 2001.
- [6] Korbicz J., Kościelny J., Kowalczyk Z., Cholewa W. „Diagnostyka procesów. Modele i metody sztucznej inteligencji. Zastosowania”. WNT, Warszawa 2002.
- [7] Moczulski W. „Metody pozyskiwania wiedzy dla potrzeb diagnostyki maszyn”, Monografia. Wydawnictwo Politechniki Śląskiej, Gliwice, 1997.
- [8] Sobczak W., Malina W. „Metody selekcji i redukcji informacji”, WNT, Warszawa, 1985.
- [9] Szabatin J. „Podstawy teorii sygnałów” WKiŁ, 1982.
- [10] Volk. W. „Statystyka stosowana dla inżynierów” Wydawnictwa Naukowo-Techniczne, Warszawa 1973



Dr inż. Paweł Chrzanowski, absolwent Wydziału Mechanicznego Technologicznego Politechniki Śląskiej, adiunkt w Katedrze Podstaw Konstrukcji Maszyn Politechniki Śląskiej. Zajmuje się głównie zagadnieniami rozwoju oraz stosowania metod i technik przetwarzania cech sygnałów diagnostycznych w zastosowaniu do dynamicznych systemów doradczych.