

Tomasz Pełech-Pilichowski\*, Jan T. Duda\*

## **Analiza podobieństwa sygnałów diagnostycznych z wykorzystaniem metod odległościowych\*\***

### **1. Wprowadzenie**

Dostępność systemów komputerowych o dużych możliwościach obliczeniowych, w tym systemów automatyki przemysłowej, pozwala na recepcję, przetwarzanie oraz monitorowanie pozyskanych zestawów sygnałów diagnostycznych. Powoduje to konieczność algorytmizacji selekcji informacji usprawniających lub wręcz umożliwiających dalsze przetwarzanie danych na potrzeby określonych procedur numerycznych oraz podejmowanie decyzji operatorskich (eksperskich).

Zapewnienie efektywności takich decyzji uzależnione jest od algorytmicznej selekcji informacji istotnej, co pozwala na sygnalizowanie zdarzeń wymagających szczególnej koncentracji uwagi decydenta i tym samym zainicjowanie procedur diagnostycznych lub podjęcia decyzji w trybie pilnym. Istotny wpływ wczesnej selekcji takich zdarzeń na jakość dalszego przetwarzania danych powoduje konieczność szybkiej estymacji krótko bądź długoterminowych parametrów statystycznych przetwarzanych sygnałów diagnostycznych, a także minimalizację opóźnienia detekcji.

Wykorzystanie dedykowanych algorytmów wykrywania zmian w szeregach czasowych może być ukierunkowane m.in. na wykrywanie gwałtownych zmian poziomu sygnału oraz nietypowych przebiegów rejestrowanych szeregów oraz sekwencji (wzorców). Takie analizy mogą być prowadzone z wykorzystaniem metod odległościowych, gdzie na podstawie dostępnego zestawu sygnałów diagnostycznych analizowane jest krótkoterminowe podobieństwo pomiędzy obiektami (podobieństwo ich krótkoterminowych charakterystyk). Do tego celu implementowane mogą być procedury klasyczne [5] oraz miary dedykowane [8, 10, 3]. Na podstawie uzyskiwanych wyników analizy podobieństwa możliwa jest atrybutyzacja oraz klasyfikacja badanego w ruchomym oknie podciągu, identyfikacja

---

\* AGH Akademia Górniczo-Hutnicza, Wydział Zarządzania, Katedra Informatyki Stosowanej

\*\* Badania prowadzono w ramach projektu InStePro - POIG, w ramach osi priorytetowej 1. Badania i rozwój nowoczesnych technologii, Działanie 1.3. Wsparcie projektów B+R na rzecz przedsiębiorców realizowanych przez jednostki naukowe, Poddziałanie 1.3.1 „Projekty rozwojowe”, Grant nr UDA-POIG.01.03.01-12-171/08/00

typowych i nietypowych współzależności pomiędzy obiektami i tym samym predykcja kolejnych wartości analizowanego sygnału diagnostycznego na podstawie chwilowych charakterystyk statystycznych i częstotliwościowych dostępnego zestawu sygnałów.

W artykule opisano problematykę algorytmicznego przetwarzania sygnałów diagnostycznych pozyskiwanych z urządzeń technicznych (zestawów zmiennych procesowych) z wykorzystaniem odległościowych metod analizy podobieństwa szeregów czasowych w celu docelowej algorytmicznej identyfikacji i klasyfikacji podobnych, w sensie odległościowym, sytuacji procesowych. Podstawą algorytmicznego przetwarzania są analizy retrospektywne sygnałów diagnostycznych oraz ich transformacji w ruchomym oknie. Pozwalają one na klasyfikowanie typowych oraz nietypowych sytuacji procesowych i tym samym predykcję przyszłych wartości dla stanów przejściowych. Do badania podobieństwa sygnałów diagnostycznych zaproponowano wykorzystanie oryginalnej miary odległości ( $S_d$  – miary spektralnej determinizowanej), dedykowanej do analizy podobieństwa kształtu szeregów. Dyskusji poddano przykładowe wyniki analiz przeprowadzonych dla danych rzeczywistych, oryginalnych oraz poddanych dwóm transformacjom szeregów wejściowych.

## **2. Monitorowanie sygnałów diagnostycznych**

### **2.1. Metody statystyczne i odległościowe**

Zmienne procesowe są automatycznie rejestrowane przez systemy pomiarowe. Mogą być generowane przez zaimplementowane procedury obliczeniowe, urządzenia alarmowe oraz dane wpisywane przez operatorów. Zmienne kodujące te informacje (ciągłe i dyskretne), bądź ich transformacje, tworzą przestrzeń sygnałów diagnostycznych [1].

Konstrukcja algorytmów monitorowania właściwości statystycznych sygnałów diagnostycznych opiera się na wykorzystaniu klasycznych technik analizy sygnałów (analiz statystycznych, częstotliwościowych), a także metod rozpoznawania obrazów, inteligencji obliczeniowej czy analizy wielowymiarowej szeregów czasowych [7]. Wyniki takiego przetwarzania stanowią dane wejściowe dla algorytmów predykcji bądź wykrywania anomalii (w tym retrospektywnych analiz właściwości szeregów), które mają rozległe zastosowanie w technice (np. w diagnostyce technicznej [4]).

Podjęcie statystyczne stosowane w analizach sygnałów diagnostycznych wymaga przede wszystkim selekcji podciągów odpowiadającym sobie w sensie dynamicznym. Ciągi wyjściowe muszą być opóźnione w stosunku do wejściowych (dane odpowiadające sobie w sensie domniemanego stanu stacjonarnego). Do wykonania takiej selekcji potrzebna jest znajomość czasu ustalania odpowiedzi skokowej – zastępczych stałych czasowych dla wszystkich kanałów wejścia-wyjścia, które można oszacować przez eksperckie przyporządkowanie wejść i wyjść bądź z wykorzystaniem algorytmów estymacji opóźnień (np. z wykorzystaniem algorytmu rozpoznawania zmian przez analizę istotności trendu). Mając przyporządkowane odpowiadające sobie ciągi wejść i wyjść możliwa jest charakteryzacja punktu pracy, która pozwala na wykonanie atrybutyzacji i tym samym klasyfikacji sy-

tuacji procesowych. Typowo, jako atrybuty wykorzystać można, przykładowo, wartości średnie zmiennych procesowych, ich zmienność (np. odchylenie standardowe), jakość punktu pracy (np. osiągnięcie założonej jakości produkcji). Mogą dotyczyć także poziomu wartości zmiennych procesowych (ujęcie jakościowe), który powinien być odniesiony do wartości nominalnych.

Dostępne sygnały diagnostyczne poddawane mogą być algorytmicznemu przetwarzaniu, którego jednym z zadań jest identyfikacja zmian poziomu wartości zmiennych procesowych bądź detekcja sytuacji alarmowych dla danego szeregu czasowego. Wprawdzie wykorzystywane algorytmy przetwarzające pojedyncze szeregi czasowe, oparte zazwyczaj na procedurach statystycznych, pozwalają uzyskać zadowalające rezultaty [2, 9], niemniej jednak analiza współzależności pomiędzy dostępnymi sygnałami diagnostycznymi może stanowić źródło dodatkowej informacji, wspomagającej podejmowanie decyzji operatorskich, w szczególności dla stanów przejściowych, oraz zwiększające wiarygodność dalszych obliczeń.

Dodatkowo, wśród atrybutów uwzględnić można kompleksową charakterystykę typu odległości szeregów czasowych od szeregów wzorcowych jako charakterystykę stopnia przynależności do grupy procesów, co winno umożliwić detekcję sytuacji nietypowych (tzw. wczesne ostrzeganie). W odróżnieniu od podejścia klasycznego, analizowane jest podobieństwo między przetwarzanymi sygnałami diagnostycznymi (wyjściami procesu), pozwalające na sprawdzanie podobieństwa (w sensie odległościowym) pomiędzy czasem reakcji procesu na określone zmiany procesu, występujące retrospektywnie dla podobnych wejść i wyjść. Wysokie podobieństwo, tj. przynależność do grupy, może być podstawą do zakwalifikowania punktu pracy do identyfikacji modeli statyki i dynamiki.

## **2.2. Klasyczne i dedykowane odległościowe miary badania podobieństwa szeregów czasowych**

### **2.2.1. Monitorowanie podobieństwa sygnałów diagnostycznych**

Kompleksowe przetwarzanie danego (bazowego) sygnału diagnostycznego winno opierać się zarówno na wykorzystaniu algorytmów detekcji zdarzeń w tym sygnale, jaki i procedur monitorowania badanego sygnału z uwzględnieniem powiązanych zmiennych procesowych. Algorytmy te winny być ukierunkowane na analizę podobieństwa kształtów przetwarzanych szeregów oraz konfiguracji pojawiających się odchyłek (wzorców).

Analizy podobieństwa dostępnego zbioru zmiennych procesowych mogą być wykorzystane jako mechanizmy identyfikacji bieżącej sytuacji procesowej w odniesieniu do obserwowanych wcześniej charakterystyk badanych sygnałów diagnostycznych, w oparciu o retrospektywne analizy odległościowe, prowadzone asynchronicznie w przypadku stwierdzenia sytuacji nietypowej czy wcześniej nieobserwowanej (stanów przejściowych). Takie procedury mogą być uruchamiane automatycznie lub przez operatora w przypadku stwierdzenia nieprawidłowości. Efektem będzie identyfikacja danej sytuacji procesowej w zadanym (bieżącym) przedziale czasu, tj. odległościowa atrybutyzacja wejściowych sygnałów

diagnostycznych, a następnie dalsze ich przyporządkowanie do wyodrębnionych grup, zidentyfikowanych dla danych historycznych. Na podstawie takiej atrybutyzacji przy danej klasie oddziaływań i ich skutków możliwe jest generowanie prognoz (przykładowo, analizy podobieństwa mogą być wykorzystane także dla potrzeb komputerowej diagnostyki technicznej, do algorytmicznej klasyfikacji bieżącego stanu urządzeń do jednej z wyspecyfikowanych wcześniej klas sprawności reprezentujących wzorce diagnostyczne [1]).

Wyniki obliczeń z wykorzystaniem odległościowych metod badania podobieństwa szeregów czasowych w oknie ruchomym zależne od charakterystyki statystycznej i częstotliwościowej ciągów wejściowych. Przetwarzane szeregi czasowe mogą być poddane transformacjom prostym (wartości względne, przyrosty, logarytmy zmiennych, wartości uśrednione, ekstrakcja trendu, ekstrakcja średniej, potęgowanie z wykładnikiem parzystym lub nieparzystym) lub złożonym, w tym przekształceniom wielokrotnym. Transformacje te umożliwiają efektywne wykorzystanie metod klasycznych (np. transformacje widmowe, falkowe, ortogonalizujące [1]), a także innych algorytmów, w tym odległościowych metod analizy podobieństwa badanych obiektów [5].

Zastosowanie algorytmicznych analiz podobieństwa wymaga implementacji danej metody odległościowej (wraz z ustaleniem podstawowych parametrów metody) lub zaprojektowania miary dedykowanej w relacji do typu przetwarzanego sygnału diagnostycznego, charakterystyki analizowanych (poszukiwanych) zmian oraz przesunięć czasowych pomiędzy zdarzeniami (sytuacjami procesowymi) występującymi w dostępnym zestawie sygnałów diagnostycznych. Na końcowe obliczenia istotny wpływ ma dobór szerokości przesuwanego okna analizy, co jest nietrywialnym problemem badawczym (szerokość ruchomego okna jest kompromisem pomiędzy czułością danego algorytmu, a rozmyciem uzyskiwanych wyników).

Do analizy podobieństwa szeregów czasowych wykorzystać można miary klasyczne (m.in. metrykę euklidesową, miarę Minkowskiego, odległość miejską (City), Mahalanobisa, Czebyszewa [5]), miary korelacyjne (korelacje rang Spearmana, współczynnik  $\tau$  Kendalla [10]) oraz miary dedykowane (np. współczynnik Dice'a [5], miarę DTW (*Dynamic Time Warping*) [8] i inne).

Miary klasyczne nie pozwalają na detekcję realnej różnicy między badanymi szeregami w przypadku występowania przesunięć czasowych pomiędzy zdarzeniami lub wzorcami. Z kolei miary dedykowane w większości przypadków mogą być stosowane (są efektywne) dla określonych założeń (np. przetwarzanie danych binarnych, sortowanych, o określonych charakterystykach itp.), niemniej jednak przyjmuje się, że w ogólnym przypadku odległość euklidesowa daje najlepsze rezultaty [159].

Dedykowane miary odległościowe, pozwalające na uzyskiwanie wiarygodnych rezultatów, powinny być w jak największym stopniu niezależne od zewnętrznych uwarunkowań. Konieczna jest identyfikacja i uwzględnienie przesunięć czasowych pomiędzy badanymi szeregami, co jest zjawiskiem częstym zarówno dla sygnałów technicznych (np. opóźnienia typu wejście-wyjście) oraz finansowych szeregów czasowych (zmienne opóźnienia o charakterze przyczynowo-skutkowym).

### 2.2.2. Miara spektralna determinizowana

Jedną z dedykowanych metod badania podobieństwa szeregów jest miara spektralna determinizowana  $S_d$  [7]. Miara ta została zaprojektowana do badania podobieństwa kształtu badanych szeregów czasowych w oknie ruchomym o zadanej szerokości. Istotą miary  $S_d$  jest niwelowanie informacji o wzajemnym opóźnieniu zdarzeń w dwóch badanych szeregach.

Obliczenia prowadzone są w trzech głównych etapach [7, 3]:

- a) obliczenie transformat Fouriera obu badanych szeregów w oknie analizy o długości  $M$  (metodą szybkiej transformaty Fouriera, FFT),
- b) obliczenie determinizowanych sygnałów diagnostycznych o długości  $k_d$  próbek ( $k_d$  określa się jako rząd metody,  $k_d \leq M$ ) przez zastosowanie transformaty odwrotnej dla widma amplitudowego (z wyzerowaniem przesunięcia fazowego wszystkich harmonicznych),
- c) obliczenie odległości Minkowskiego [5] dla uzyskanych w (b) sygnałów diagnostycznych.

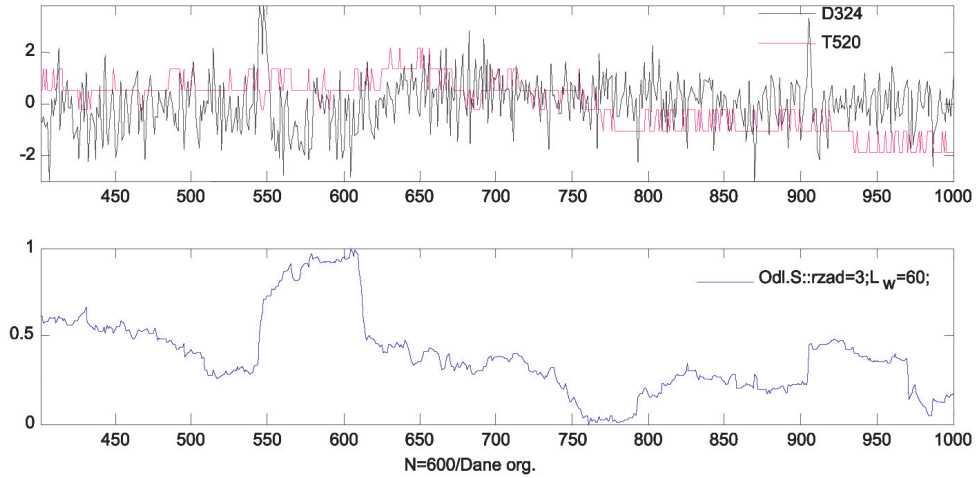
Sygnał diagnostyczny (b) agreguje informację o kształcie i mocy dominujących serii odchyłek, a przez to pozwala wychwycić ich podobieństwo w porównywanych szeregach, ze znaczną redukcją wpływu tła losowego i wzajemnych przesunięć serii. Miara  $S_d$  cechuje się selektywnością, tj. wysoką efektywnością detekcji zmian pojawiających się tylko w jednym badanym szeregu.

## 3. Przykładowe wyniki badania podobieństwa sygnałów diagnostycznych

Do badania podobieństwa zmiennych procesowych oraz zobrazowania ogólnej charakterystyki miary  $S_d$  (patrz podrozdz. 2.2.2), wybrano dwa techniczne sygnały diagnostyczne przykładowego procesu chemicznego zawierające  $L = 1000$  próbek (zastosowane oznaczenia: T520 – temperatura wody zimnej; D324 – stężenie kwasu azotowego; zestandardyzowane przebiegi zilustrowano w górnych częściach rysunków 1–3). Zakłada się, że w danej chwili czasu  $t$ , w której po otrzymaniu informacji jakościowej następuje analiza retrospektywna dostępnych zbiorów danych, badane jest  $K = 600$  bieżących próbek, co stanowi wielokrotność szerokości ruchomego okna analizy ( $L_w = 60$ ).

W celu scharakteryzowania wpływu prostych transformacji danych na uzyskiwane rezultaty badania podobieństwa z wykorzystaniem miary  $S_d$ , obliczenia przeprowadzono dla danych oryginalnych (rys. 1), jednokrokových przyrostów bezwzględnych (rys. 2) oraz residuów ekstrakcji trendu liniowego (rys. 3).

Jako główny parametr miary  $S_d$  przyjęto rząd  $k_d = 3$  próbki, dobrany w drodze eksperymentalnej. W celu zwiększenia czytelności uzyskanych wyników poszczególne wartości  $S_d$  znormalizowano (w efekcie wartości cech uzyskanych wektorów mieszczą się w przedziale  $<0;1>$ ; patrz dolne wykresy na rysunkach 1–3).

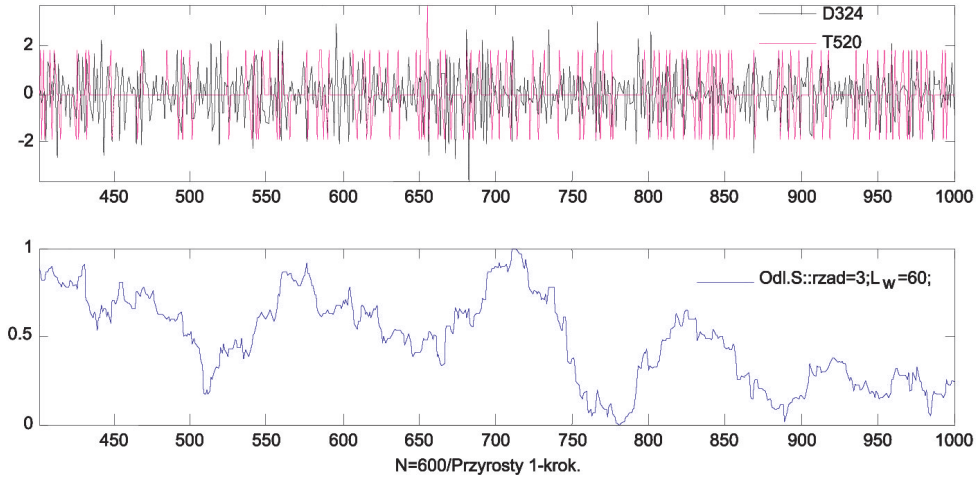


**Rys. 1.** Charakterystyka podobieństwa szeregów badanego z wykorzystaniem miary  $S_d$  (dolny wykres), liczonej w oknie o szerokości  $L_w = 60$  próbek, dla rzędu  $k_d = 3$  próbek. Obliczenia dla danych oryginalnych (zestandaryzowane przebiegi zilustrowano na górnym wykresie)

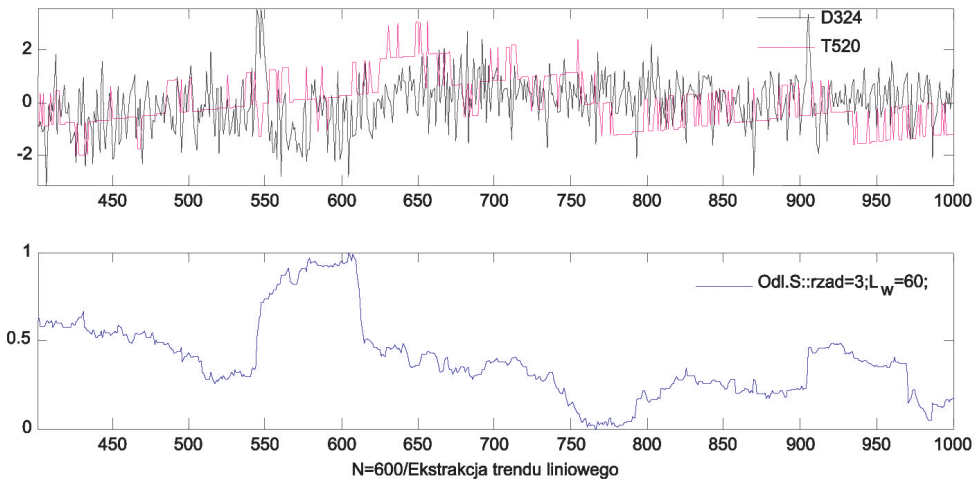
Jak pokazano na rysunkach 1–3, miara  $S_d$  jest niewrażliwa na niewielkie zmiany poziomu przetwarzanych sygnałów diagnostycznych. Obserwowane są istotne zmiany wartości  $S_d$  w przypadku pojawienia się skokowych zmian w jednym z sygnałów (patrz: rys. 1 i 3, około 550. i 920. próbki).

Na podstawie zilustrowanych przebiegów miary odległości (rys. 1–3) można eksperycko wyodrębnić klasy podobnych charakterystyk obu badanych szeregów, jak m.in. klasę, w której znormalizowana wartość  $S_d$  przekraczająca 0,5 obserwowana jest dla zmian sygnału wejściowego o największej amplitudzie (rysunki 1 i 3 – około 450. próbki oraz w zakresie 550–620; rysunek 2 – dla około 560 próbek), klasę dla wartości  $S_d$  z przedziału  $<0,3; 0,5$  oraz klasę, w której wartości badanej miary odległości są bliskie zeru (np. rys. 1–3, około 770. próbki).

Spośród dwóch zastosowanych transformacji danych oryginalnych, informatywny sygnał generowany jest dla jednokrokových przyrostów bezwzględnych (patrz rys. 2). Na wykresach (rys. 1 i 2) widoczne są zarówno przedziały, w których obserwowany jest podobny przebieg miary  $S_d$  w relacji do danych oryginalnych, jak i przedziały, w których widoczne są zmiany asynchroniczne (np. wartości  $S_d$  przekraczające 0,5 dla 430. i 830. próbki – rys. 2). Można także zauważyć, że dla skokowej zmiany badanej miary odległości, widocznej dla danych oryginalnych w otoczeniu 550. próbki (rysunek 1), stosunkowo wysoka wartość  $S_d$  liczona dla przyrostów bezwzględnych utrzymuje się znacznie dłużej, w porównaniu do przebiegu  $S_d$  uzyskanego dla danych surowych (wejściowych), i uwzględnia zmianę poziomu oryginalnego sygnału D324 w przedziale od 620. do 650. próbki.



**Rys. 2.** Charakterystyka podobieństwa szeregów badanego z wykorzystaniem miary Sd (dolny wykres), liczonej w oknie o szerokości  $L_w = 60$  próbek, dla rzędu  $k_d = 3$ . próbek. Obliczenia dla jednokrokových przyrostów bezwzględnych (zestandaryzowane przebiegi zilustrowano na górnym wykresie)



**Rys. 3.** Charakterystyka podobieństwa szeregów badanego z wykorzystaniem miary Sd (dolny wykres), liczonej w oknie o szerokości  $L_w = 60$  próbek, dla rzędu  $k_d = 3$ . próbek. Obliczenia dla reziduiów ekstrakcji trendu liniowego (zestandaryzowane przebiegi zilustrowano na górnym wykresie)

Ekstrakcja trendu liniowego wprawdzie powoduje wygenerowanie nowych sygnałów diagnostycznych (patrz górna część rys. 3), jednak wartości Sd liczone dla takich reziduiów są silnie skorelowane z wynikami otrzymanymi dla danych oryginalnych (patrz: rys. 1)

i tym samym nie mają wpływu na wzrost efektywności prowadzonych analiz. W badaniach wstępnych podobny efekt uzyskano dla reziduoów ekstrakcji średniej z danych oryginalnych, transformacji logarytmicznej oraz jednokrokowych przyrostów logarytmicznych.

#### 4. Podsumowanie

Algorytmiczne metody analizy podobieństwa szeregów czasowych, implementowane dla różnych par dostępnych sygnałów diagnostycznych wraz z odpowiednimi transformacjami, pozwalają na wyodrębnienie (eksperycko lub automatycznie) klas reprezentujących charakterystyki badanych ciągów oraz relacji między nimi. W przypadku identyfikacji sytuacji nietypowej, algorytmiczna klasyfikacja bieżącej charakterystyki odległościowej zmiennych procesowych może generować istotną informację dla zapewnienia odpowiedniej jakości sterowania i w efekcie pozwoli na wczesne ostrzeżenie o sytuacjach alarmowych, eliminowanie fałszywych alarmów oraz minimalizowanie błędów predykcji, co przekłada się na szybsze podejmowanie decyzji. Problemem badawczym, warunkującym skuteczność algorytmicznej klasyfikacji sytuacji procesowych, jest w tym przypadku ustalenie odpowiedniej liczby klas. Implementacja procedur automatycznej klasyfikacji sytuacji procesowych wymaga badania dużych zestawów sygnałów diagnostycznych w celu uzyskania komplementarnej rezultatów dostępnych szeregów. Takie zestawy uzyskać można przez zastosowanie prostych i złożonych transformacji szeregów oryginalnych, liczonych w oknie przesuwanym oraz dla całego analizowanego przedziału.

W niniejszej pracy scharakteryzowano i dokonano testów miary spektralnej determinowanej, na podstawie których można stwierdzić, że dla prostych transformacji oraz stałych, dobranych arbitralnie parametrów (rzędu metody oraz szerokości okna analizy) możliwe jest uzyskanie wiarygodnych wyników detekcji zmian w badanych szeregach, tj. identyfikacji różnic kształtów przebiegów zmiennych procesowych, z jednoczesnym niwelowaniem potencjalnego opóźnienia czasowego między badanymi sygnałami.

Dalsze badania ukierunkowane będą m.in. na wdrożenie algorytmów automatycznej klasyfikacji sytuacji procesowych, adaptacji parametrów prowadzonych analiz podobieństwa (m.in. szerokości okna) oraz parametrów wykorzystywanych miar odległości szeregów.

#### Literatura

- [1] Duda J.T., *Pozyskiwanie wzorców diagnostycznych w komputerowych analizach sprawności urządzeń*. [w:] Diagnostyka procesów i systemów, red. Korbicz J., Patan K., Kowal M., Akademicka Oficyna Wydawnicza EXIT, 2007.
- [2] Duda J., Pelech T., *Wykrywanie zdarzeń w szeregach finansowych z wykorzystaniem metod statystycznych*. VI Krajowa Konferencja Inżynieria Wiedzy i Systemy Ekspertowe, Wrocław 2006.
- [3] Duda J.T., Pelech-Pilichowski T., *Miary podobieństwa szeregów czasowych w detekcji zdarzeń*. [w:] Systemy wykrywające, analizujące i tolerujące usterki, red. Kowalczyk Z., Pomorskie Wydawnictwo Naukowo-Techniczne PWNT, 2009.



- 
- [4] Geng Z., Zhu Q., *A Wavelet-Based Adaptive Mspca For Process Signal Monitoring & Diagnosis*. International Conference on Information Acquisition, 2004.
  - [5] Hand D., Mannila H., Smyth P., *Principles of Data Mining*. MIT Press, 2001.
  - [6] Mahoney M.V., Chan P.K., *Learning Rules for Time Series Anomaly Detection*. Technical Report CS-2005-04, Florida Institute of Technology, 2004.
  - [7] Pełech-Pilichowski T., *Adaptacyjne algorytmy detekcji zdarzeń w szeregach czasowych*. Rozprawa doktorska. Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie, Wydział Elektrotechniki, Automatyki Informatyki i Elektroniki, 2009.
  - [8] Shyu M., Chen S., Sarinnapakorn K., Chang L., *A Novel Anomaly Detection Scheme Based on Principal Component Classifier*. IEEE International Conference on Data Mining, 2003.
  - [9] Tarassenko L., Nairac A., Townsend N., Cowley P., *Novelty Detection in Jet Engines*. IEEE Colloquium on Condition Monitoring, Imagery, External Structures and Health, 1999.
  - [10] Taylor J.M.G., *Kendall's and Spearman's Correlation Coefficients in the Presence of a Blocking Variable*. Biometrics, vol. 43, No. 2, 1987.