

REDUKCJA WYMIARU I LICZNOŚCI PRÓBY DLA POTRZEB SYNTEZY STATYSTYCZNEGO UKŁADU WYKRYWANIA USZKODZEŃ

Piotr KULCZYCKI*, Szymon ŁUKASIK*

1. Wprowadzenie

Współczesna analiza danych dysponuje szerokim zakresem zróżnicowanej metodyki, opartej zarówno na tradycyjnych jak i nowoczesnych, nierzadko wyspecjalizowanych procedurach statystycznych, obecnie coraz częściej wspomaganych znacznymi możliwościami inteligencji obliczeniowej, poza klasycznymi – logiką rozmytą, sieciami neuronowymi i algorytmami genetycznymi, także metaheurystykami nowej generacji: metodą symulowanego wyżarzania, rojami cząstek, czy też algorytmami mrówkowymi. Właściwe połączenie i wykorzystanie zalet tych koncepcji umożliwi skuteczne rozwiązywanie zagadnień inżynierii wiedzy, w szczególności bazujących na analizie eksploracyjnej danych.

Coraz częściej proces eksploracji danych przeprowadzany jest z wykorzystaniem wielowymiarowych zbiorów o dużej liczności. Ekstrakcja wiedzy z dużych zbiorów danych jest zagadnieniem złożonym. Trudności związane są tu głównie z ograniczeniami narzucanymi przez wydajność systemów komputerowych – w przypadku prób o dużej liczności – oraz problemami metodycznymi nieodłącznymi dla analizy danych wielowymiarowych. Te drugie wynikają głównie z szeregu zjawisk występujących w tego typu zbiorach, znanych pod pojęciem „przekleństwa wielowymiarowości”. Obejmuje ono przede wszystkim eksponencjalny wzrost liczności próby potrzebnej do osiągnięcia odpowiedniej skuteczności metod analizy danych przy zwiększającym się wymiarze, tak zwany „fenomen pustej przestrzeni”, a także zanikanie różnicy między bliskimi a dalekimi punktami przy zastosowaniu typowych metryk. Z kolei, redukcję liczności zbioru przeprowadza się głównie w celu przyspieszenia lub nawet urealnienia czasu obliczeń, a w konkretnych zagadnieniach wskazana jest także eliminacja elementów niosących informację mniej wartościową lub nawet błędną, czego efektem jest wręcz polepszenie jakości otrzymywanych wyników.

W niniejszej pracy zostanie przedstawiona metoda redukcji wymiaru i liczności próby, której przeznaczeniem będzie użycie w procedurach eksploracyjnej analizy danych, określonych z

* Instytut Badań Systemowych, Polska Akademia Nauk, ul. Newelska 6, 01-447 Warszawa; e-mail: {Piotr.Kulczycki, Szymon.Lukasik}@ibspan.waw.pl .

użyciem metod estymacji nieparametrycznej. Redukcja wymiaru będzie realizowana za pomocą liniowej transformacji, przy wymaganiu aby możliwie w jak najmniejszym stopniu wpływała ona na wzajemne położenie elementów próby pierwotnej i wynikowej. W tym celu zostanie wykorzystana heurystyczna metoda symulowanego wyżarzania. Dodatkowo te elementy próby, które w wyniku transformacji zmieniają istotnie swe położenie będą podlegać eliminacji lub przypisaniu mniejszej wagi w definicji estymatora jądrowego – uzyskuje się w ten sposób poprawę jakości estymacji oraz ewentualnie redukcję liczności próby. Skuteczność prezentowanej metody zostanie zweryfikowana dla fundamentalnych procedur eksploracyjnej analizy danych: identyfikacji elementów nietypowych (odosobnionych), klasteryzacji (analizy skupień) oraz klasyfikacji, określonych z użyciem jednolitej metodyki statystycznych estymatorów jądrowych. Estymatory tego typu, których rozwój jest bezpośrednio związany ze współczesną ekspansją techniki komputerowej, stanowią obecnie wiodącą metodę estymacji nieparametrycznej. Główny cel stanowi tu wyznaczenie różnorodnych charakterystyk rozkładów probabilistycznych bez arbitralnych założeń dotyczących ich przynależności do określonej klasy. Warto zauważyć, iż użycie jednolitej metodyki do wszystkich rozważanych zadań znacząco ułatwia proces analizy i syntezy badanych zagadnień, a także późniejszą implementację.

Wspomniane wyżej procedury eksploracyjnej analizy danych – identyfikacja elementów nietypowych, klasteryzacji oraz klasyfikacji – zostaną wykorzystane do syntezy statystycznego układu wykrywania uszkodzeń w oparciu o metodykę estymatorów jądrowych, przedstawionego w pracach [8, 9]. Obejmuje on swym zakresem:

- detekcję uszkodzeń, a więc stwierdzenie występowania nieprawidłowości w stanie technicznym nadzorowanego systemu;
- diagnozę uszkodzeń, czyli identyfikację owych nieprawidłowości;
- predykcję uszkodzeń, to znaczy uprzedzenie o zagrożeniu ich pojawieniem się w niedalekiej przyszłości (wraz z przypuszczalną klasyfikacją).

Redukcja wymiaru przestrzeni pozwala na zwiększenie ilości zmiennych użytych do wnioskowania o stanie technicznym nadzorowanego systemu, natomiast ograniczenie liczności wykorzystywanych prób – zmniejszenie czasu obliczeń oraz liczby błędnych wskazań.

Fundamentalną polskojęzyczną pozycję bibliograficzną w zakresie diagnostyki procesów stanowi książka [3], natomiast metodyka estymatorów jądrowych szeroko przedstawiona są w monografii [7]. Szereg praktycznych aspektów szeroko rozumianych współczesnych technik informacyjnych znaleźć można w pracy [11].

2. Preliminaria

2.1. Eksploracyjna analiza danych z zastosowaniem estymatorów jądrowych

Przedmiotem niniejszego podrozdziału będzie przedstawienie krótkiej charakterystyki procedur identyfikacji elementów odosobnionych, klasteryzacji i klasyfikacji, z użyciem estymatorów jądrowych. Estymator gęstości rozkładu prawdopodobieństwa n -wymiarowej zmiennej losowej wyznacza się podstawie jej m -elementowej próby losowej

$$x_1, x_2, \dots, x_m \tag{1}$$

i w swej podstawowej postaci jest on zdefiniowany wzorem

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \tag{2}$$

przy czym dodatni współczynnik h określa się mianem parametru wygładzania, natomiast

mierzalna funkcja $K: \mathbb{R}^n \rightarrow [0, \infty)$ symetryczna względem zera, posiadająca w tym punkcie słabe maksimum lokalne i spełniająca warunek $\int_{\mathbb{R}^n} K(x) dx = 1$ jest nazywana jądrem.

Szczegółowe informacje na temat postaci funkcji K , metod wyznaczania wartości parametru wygładzania h , a także algorytmu jego modyfikacji i ograniczenia nośnika można znaleźć w monografii [7]. W przedstawionych poniżej procedurach, rozważany zbiór danych traktowany jest jako próba losowa (1).

Identyfikację elementów odosobnionych [1] stosuje się zwykle na początku procesu analizy danych, w celu ewentualnego usunięcia z próby elementów niereprezentatywnych, które w istotny sposób różnią się od ogółu populacji. Procedura wykorzystująca w tym celu estymatory jądrowe [13] opiera się na teorii testowania hipotez statystycznych. Weryfikuje się tu hipotezę, że badany element $\tilde{x} \in \mathbb{R}^n$ jest elementem typowym, przeciwko hipotezie, że takowym nie jest (i w konsekwencji można go wtedy uznać za odosobniony). Odrzucenie testowanej hipotezy następuje, gdy wartość estymatora jądrowego wyznaczonego dla próby (1) składającej się elementów typowych, na badanym elemencie, czyli

$$\hat{f}(\tilde{x}) \tag{3}$$

jest nie większa niż kwantyl ustalonego stopnia $\alpha \in (0, 1)$ (zwykle $\alpha = 0,05$), obliczonego dla próby $\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_m)$.

Zadanie klasteryzacji [4] dla określonego zbioru danych (1) polega z kolei na jego podziale na podgrupy zawierające elementy do siebie podobne, ale istotnie różniące się pomiędzy poszczególnymi podgrupami. Idea klasteryzacji z zastosowaniem estymatorów jądrowych opiera się na metodzie gradientowej [2]. Najpierw każdy z elementów zbioru (1) jest iteracyjnie przesuwany zgodnie z kierunkiem gradientu $\nabla \hat{f}$. Dalszą część procedury klasteryzacji stanowi przyporządkowanie poszczególnych elementów do odpowiednich klastrów. W tym celu konstruuje się estymator jądrowy dla próby złożonej ze wzajemnych odległości między wszystkimi elementami zbioru przesuniętych danych. Argument, dla którego pojawia się pierwsze, większe od zera lokalne minimum tego estymatora, reprezentuje połowę odległości między dwoma najbliższymi położonymi klastrami. Przypisanie ustalonego elementu do klastra następuje wtedy, gdy jego odległość od któregośkolwiek elementu tego klastra jest nie większa niż tak wyznaczona wartość. Szerszy opis tego algorytmu znajduje się w publikacji [10].

I wreszcie, celem klasyfikacji [4] jest przypisanie badanego elementu $\tilde{x} \in \mathbb{R}^n$ do jednej z ustalonych klas ze znanymi próbami wzorcowymi, o postaci analogicznej do formuły (1). Zgodnie z klasycznym ujęciem bayesowskim, zapewniającym minimum potencjalnych strat wynikłych z błędnych klasyfikacji, przy założeniu iż liczności prób wzorcowych są proporcjonalne do prawdopodobieństwa wystąpienia w danej populacji elementu z i -tej klasy, klasyfikowany element przypisuje się do tej klasy, dla której wartość iloczynu liczności próby wzorcowej m_i oraz wartości estymatora wyznaczonego na podstawie i -tej próby losowej na badanym elemencie $\hat{f}_i(\tilde{x})$, czyli ostatecznie

$$m_i \hat{f}_i(\tilde{x}) \tag{4}$$

jest największa. Dodatkowe informacje można znaleźć w pracy [12].

Warto podkreślić jednolitość metodyki użytej do powyższych procedur. Ułatwia to ich implementację poprzez lepsze zrozumienie stosowanego aparatu oraz możliwość wielokrotnego wykorzystania tych samych podprogramów.

2.2. Algorytm symulowanego wyżarzania

Symulowane wyżarzanie jest algorytmem heurystycznym przeznaczonym do efektywnego przeszukiwania przestrzeni rozwiązań w ramach różnorodnych problemów z zakresu badań systemowych. Idea metody opiera się na analogii do procesu wyżarzania znanego z metalurgii. Algorytm symulowanego wyżarzania bazuje na technice iteracyjnego lokalnego przeszukiwania oraz na indywidualnym kryterium akceptacji rozwiązań. Kryterium to pozwala określić rozwiązanie aktualne w danym kroku dla opisywanego algorytmu, zwykle wykorzystując przy tym wartość wskaźnika jakości z poprzedniej i bieżącej iteracji oraz zmienny, malejący w czasie parametr zwany temperaturą wyżarzania. Dopuszcza się przy tym możliwość przyjęcia rozwiązania aktualnego gorszego od poprzedniego, co pozwala na uniknięcie sytuacji, w której algorytm „utyka” w lokalnym minimum. Dodatkowo zakłada się, że prawdopodobieństwo akceptacji rozwiązań gorszych powinno maleć wraz z upływem czasu. Wszystkie wyżej wymienione cechy posiada tak zwana reguła Metropolis – jest ona najczęściej stosowana jako kryterium akceptacji w algorytmach symulowanego wyżarzania.

Niech zatem $Z \subset \mathbb{R}^{n^*}$ oznacza zbiór rozwiązań dopuszczalnych pewnego problemu optymalizacyjnego, natomiast funkcja $g: Z \rightarrow \mathbb{R}$ stanowi wskaźnik ich jakości. Rozpatrywany jest tu przypadek minimalizacji funkcji g na zbiorze Z . Dodatkowo niech k oznacza numer iteracji, natomiast $T^k \in \mathbb{R}$, z^k , c^k odpowiednio: temperaturę i rozwiązanie bieżące dla tejże iteracji oraz jego jakość. Przy powyższych założeniach algorytm symulowanego wyżarzania można zapisać w następujący sposób:

```

procedure Symulowane_Wyżarzanie
  begin
    k := 1
    Wygeneruj( $T^1, z^1$ )
     $c^1 = \text{Oblicz\_jakość}(z^1)$ 
     $z^* = z^1$ 
     $c^* = c^1$ 
    k := 2
    repeat
       $z^k = \text{Wygeneruj\_Sasiada}(z^{k-1})$ 
       $c^k = \text{Oblicz\_jakość}(z^k)$ 
       $\Delta c = c^k - c^{k-1}$ 
       $z^k = \text{Reguła\_Metropolis}(\Delta c, z^k, z^{k-1}, T^k)$ 
      if  $c^k < c^*$ 
         $c^* = c^k$ 
         $z^* = z^k$ 
      k := k+1
      Aktualizuj( $T^k$ )
    until warunek_stopu = TRUE
  return  $z^*$ 
end

```

przy czym reguła Metropolis jest realizowana wg schematu

```

procedure Reguła_Metropolis( $\Delta c, z^k, z^{k-1}, T^k$ )
  if  $\Delta c < 0$ 
    return  $z^k$ 
  else
    if losuj_z_przedziału(0,1) <  $\exp(-\Delta c/T^k)$ 
      return  $z^k$ 
    else
      return  $z^{k-1}$ 
  end

```

Algorytm symulowanego wyżarzania w konkretnych zastosowaniach wymaga szczegółowego określenia kilku elementów funkcjonalnych, takich jak sposób generowania rozwiązania początkowego, postać wskaźnika jakości, wartość początkowa temperatury i schemat jej zmian, metoda wyznaczenia rozwiązania sąsiedniego oraz warunków zakończenia procedury. Ogólne omówienie powyższych aspektów można znaleźć w publikacji [15].

Warto nadmienić, że procedura symulowanego wyżarzania może zostać w łatwy sposób zrównoleglona, zarówno w zakresie niezbędnych obliczeń, jak i samego schematu wyznaczania kolejnych rozwiązań. Uzyskuje się w ten sposób znacznie szybszą zbieżność algorytmu, przy efektywnym wykorzystaniu dostępnych zasobów sprzętowych.

3. REDUKCJA WYMIARU PRÓBY

Koncepcja proponowanej tu metody redukcji wymiaru przestrzeni z początkowej wartości n do założonej wartości N , przy czym $N < n$, bazuje na liniowej transformacji

$$Y = A X \quad , \quad (5)$$

gdzie A jest macierzą transformacji o wymiarach $N \times n$, co można uściślić do postaci

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{Nm} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{Nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad , \quad (6)$$

przy czym poszczególne kolumny macierzy X oznaczają rozpisane na współrzędne elementy próby losowej (1) (w przypadku procedury klasyfikacji – prób losowych stanowiących wzorce poszczególnych klas), natomiast kolumny macierzy Y – ich obrazów po przekształceniu przez transformację (5).

Do wyznaczenia postaci macierzy A wykorzystana zostanie metodyka symulowanego wyżarzania w postaci przedstawionej w poprzednim rozdziale. Rozwiązanie jest w tym przypadku reprezentowane przez wektor składający się z elementów macierzy A zapisanych kolumnowo:

$$z = [a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}, \dots, a_{N1}, a_{N2}, \dots, a_{Nn}]^T \in \mathbb{R}^{nN} \quad . \quad (7)$$

Elementy macierzy transformacji dobrane są tak, by minimalizować wskaźnik jakości

$$g(z) = \sum_{i=1}^m \sum_{j=i+1}^m \left(\|y_i - y_j\|_{\mathbb{R}^N} - \|x_i - x_j\|_{\mathbb{R}^n} \right)^2 \quad . \quad (8)$$

Powyzsza postać ma na celu zapewnienie jak najmniejszej zmiany odległości między elementami próby w pierwotnej przestrzeni euklidesowej i w przestrzeni o zredukowanym wymiarze. Taki sposób określania macierzy transformacji jest intuicyjnie zgodny z generalną ideą estymatorów jądrowych – wartość estymatora dla ustalonego argumentu jest bowiem proporcjonalna do jego odległości od poszczególnych elementów próby. Równocześnie – pomimo uniwersalności przyjętej postaci wskaźnika jakości (8) – zasadne staje się stosowanie do zagadnienia jego minimalizacji algorytmu symulowanego wyżarzania. Zarówno przeprowadzone wstępne badania, jak i analiza porównawcza dostępna w literaturze przedmiotowej, wykazują na jego wyższość nad algorytmami genetycznymi w tego typu zagadnieniach.

Rozwiązanie początkowe wyznaczane jest na podstawie algorytmu wyboru cech

przedstawionego w książce [14]. Opiera się on na koncepcji podziału współrzędnych na skupienia, zawierające cechy do siebie podobne, z kryterium zgodności dwóch współrzędnych definiowanym przez tak zwany indeks maksymalnej kompresji informacji. Podział na skupienia odbywa się na podstawie algorytmu k -najbliższych sąsiadów, przy czym przyjmuje się $k = n - N$. W rezultacie liczba uzyskanych klastrów wynosi w przybliżeniu N , co bezpośrednio ustala wymiar przestrzeni zredukowanej. Warto podkreślić, iż dzięki temu stanowi on wartość nie ściśle założoną, lecz w sposób naturalny dostosowaną do rzeczywistej struktury danych, w przybliżeniu równą oczekiwanej, ale z nią nie tożsamą.

Generowanie rozwiązania sąsiedniego odbywa się poprzez losowy wybór jednego z elementów macierzy transformacji i jego zmianę o wartość ustalonego kroku Δ lub $-\Delta$ (z prawdopodobieństwami 0,5). Przyjęto ponadto klasyczny, logarytmiczny schemat zmiany temperatury. Temperatura początkowa wyznaczana jest na podstawie wygenerowania pewnej niewielkiej ilości przejść testowych z rozwiązania początkowego, tak by zapewnić na wstępie akceptację rozwiązania gorszego z ustalonym prawdopodobieństwem. Algorytm kończy się po ustalonej liczbie iteracji, a wynikiem jego działania są macierz A minimalizująca wskaźnik jakości g i jego wartość reprezentującą jakość otrzymanego rozwiązania oraz – bezpośrednio – wynikowy zbiór danych Y o zredukowanym wymiarze.

4. REDUKCJA LICZNOŚCI PRÓBY

W przypadku przedstawionej powyżej koncepcji redukcji wymiaru niektóre elementy próby mogą ulec niepożądanemu przesunięciu względem pozostałych. W rezultacie, wynik procedur eksploracyjnej analizy danych w zredukowanej przestrzeni euklidesowej może być odmienny od oczekiwanego. By przeciwdziałać powyższemu proponuje się by dodatkowo definicja estymatora jądrowego (2) została uogólniona do postaci

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m w_i K\left(\frac{x - x_i}{h}\right), \quad (9)$$

gdzie poszczególnym elementom próby x_i przyporządkowane są nieujemne wagi w_i , znormalizowane tak aby $\sum_{i=1}^m w_i = m$. Uwzględniając uwarunkowania rozważanego tu zagadnienia, wagę tę można zdefiniować określając najpierw pomocnicze parametry

$$w_i^* = \frac{1}{\sum_{j=1, j \neq i}^m \left(\|y_i - y_j\|_{\mathbb{R}^N} - \|x_i - x_j\|_{\mathbb{R}^n} \right)^2}, \quad (10)$$

skąd normalizując obliczyć

$$w_i = \frac{m w_i^*}{\sum_{i=1}^m w_i^*} \quad \text{dla } i = 1, 2, \dots, m. \quad (11)$$

Wagi w takiej formie zawierają informację o stopniu, w jakim dany element próby zmienił swe relatywne położenie względem pozostałych, przy czym im większa waga tym jego położenie jest relatywnie „bardziej adekwatne” i staje się on istotniejszy w procesie konstrukcji estymatora jądrowego w przestrzeni o zredukowanym wymiarze.

Wprowadzenie do definicji estymatora jądrowego wag w_i pozwala na poprawę jego jakości

w przestrzeni o zredukowanym wymiarze. Ponadto można usunąć z próby te elementy x_i , dla których przyporządkowane im wagi spełniają warunek $w_i < W$, przy czym $W \in [0,1]$, a dla pozostałych albo pozostawić wagę bez zmian albo ustalić $w_i = 1$, po czym – w obu przypadkach – znormalizować. Prowadzi to do jednoczesnego zredukowania wymiaru i liczności próby, z implikowanym wartością parametru W stopniem kompresji, przy naturalnym uwzględnieniu specyfiki związanej z zastosowaniem estymatorów jądrowych. Co więcej, odpowiednie usunięcie elementów próby pozwala na istotne polepszenie wyników rozważanych procedur eksploracyjnej analizy danych.

5. SYNTEZA UKŁADU WYKRYWANIA USZKODZEŃ

Rozważane powyżej procedury identyfikacji elementów odosobnionych, klasteryzacji oraz klasyfikacji dostarczają pełnego aparatu matematycznego do skonstruowania efektywnego układu wykrywania uszkodzeń w systemach dynamicznych pracujących w czasie rzeczywistym, w zakresie detekcji, diagnozy i predykcji. Jego koncepcja została przedstawiona w pracach [8, 9].

Wnioskowanie o występowaniu uszkodzenia dokonywane będzie na podstawie tak zwanego wektora symptomów, czyli skończonego zbioru wielkości pomiarowych charakteryzujących stan techniczny (stan technicznej sprawności) nadzorowanego obiektu. Nazwę tę można zinterpretować zauważając, iż objawy (symptomy) pojawiających się nieprawidłowości powinny znajdować stosowne odzwierciedlenie w aktualnych wartościach lub relacjach pomiędzy poszczególnymi współrzędnymi tak określonego wektora.

Najpierw rozważona zostanie detekcja uszkodzeń. Załóżmy zatem, iż próba losowa reprezentuje wartości wektora symptomów uznane za typowe – świadczące o poprawnej pracy urządzenia, natomiast \tilde{x} jego aktualny stan. Stosując procedurę identyfikacji elementów odosobnionych można stwierdzić czy stan też należy uznać za typowy czy raczej za odosobniony – świadczący o wystąpieniu nieprawidłowości.

W przypadku diagnozy, jeśli dysponuje się próbami wzorcowymi charakteryzującymi poszczególne rodzaje typowych uszkodzeń, to po wykryciu nieprawidłowości można – stosując procedurę klasyfikacji – stwierdzić, z którym z nich ma się do czynienia. W przypadku gdy podział elementów reprezentujących różne rodzaje uszkodzeń nie jest znany, odpowiedni podział na klasy można uzyskać stosując opisaną procedurę klasteryzacji.

I wreszcie, jeżeli dostępne są kolejne wartości wektora symptomów uzyskiwane sukcesywnie w trakcie procesu nadzorowania, to możliwe jest zaprojektowanie predykcji uszkodzeń, realizowanej poprzez odrębne prognozy wartości funkcji \hat{f} występującej we wzorze (3) oraz $m_1 \hat{f}_1, m_2 \hat{f}_2, \dots, m_J \hat{f}_J$ uwidocznionych w formule (4), i wnioskowanie w zakresie detekcji i diagnozy zgodnie z regułami przedstawionymi w dwóch poprzednich akapitach. Do zadania prognozowania rekomenduje się stosowanie klasycznej metody regresji liniowej, aczkolwiek w wersji umożliwiającej łatwe uaktualnienie modelu w trakcie sukcesywnego pozyskiwania kolejnych wartości wektora symptomów. Odpowiednie wzory można znaleźć w rozdziale 4 publikacji [5].

Działanie powyższego układu zostało sprawdzone empirycznie. Nadzorowanym obiektem był system pozycyjny, poddany działaniu sterowania odpornego [6]. Uzyskane wyniki uwiarygodniły prezentowaną koncepcję i potwierdziły prawidłowe funkcjonowanie proponowanego tu układu wykrywania uszkodzeń, w zakresie detekcji, diagnozy i związanej z nimi predykcji. W przypadku gwałtownego charakteru pojawiających się symptomów, niesprawność urządzenia była niezwłocznie wykrywana i poprawnie rozpoznawana w ramach detekcji i diagnozy. Jeżeli natomiast uszkodzenie związane było z powolnym narastaniem jego objawów, to było ono prognozowane z prawidłowym wskazaniem na typ powstającego uszkodzenia (predykcja), a w stosownym czasie wykrywane i identyfikowane w ramach detekcji

i diagnozy. Predykcja okazała się bardzo skuteczna w przypadku wystąpienia wolno narastających ich objawów, wykrywając i rozpoznając uszkodzenie zanim charakterystyki obiektu wykroczyły poza zakres przynależny poprawnym warunkom pracy systemu, dzięki prawidłowej identyfikacji samego trendu zmian wartości wektora symptomów wskazującego na niekorzystny kierunek jego ewolucji.

Zastosowanie przedstawionych w sekcji 3 procedury redukcji wymiaru w stosunku do wektora symptomów umożliwia uwzględnienie większej ilości składników, na podstawie których następuje wnioskowanie o stanie technicznym nadzorowanego urządzenia. Z kolei odpowiednie zmniejszenie liczności próby, opisane w sekcji 4, zwiększa szybkość obliczeń – zarówno we wstępnej fazie, gdy wyznaczane są parametry statystycznego układu wykrywania uszkodzeń, jak i w trakcie późniejszej pracy w czasie rzeczywistym – a także efektywnie zmniejszy liczbę błędnych wskazań, tak w zakresie detekcji (czyli niewykrytych uszkodzeń oraz fałszywych alarmów) jak również diagnozy (w postaci błędnej klasyfikacji uszkodzenia).

Bibliografia

- [1] Barnett V., Lewis T.: *Outliers in Statistical Data*. Wiley, 1994.
- [2] Fukunaga K., Hostetler L.: The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, vol. 21, ss. 32-40, 1975.
- [3] Korbicz J., Kościelny J.M., Kowalczyk Z., Cholewa W. (red.): *Diagnostyka procesów*, WNT, 2002.
- [4] Krzyśko M., Wołyński W., Górecki T., Skorzybut M.: *Systemy uczące się*. WNT, 2008.
- [5] Kulczycki P.: *Wykrywanie uszkodzeń w systemach zautomatyzowanych metodami statystycznymi*. Alfa, 1998.
- [6] Kulczycki P.: Fuzzy Controller for Mechanical Systems. *IEEE Transactions on Fuzzy Systems*, vol. 8, ss. 645-652, 2000.
- [7] Kulczycki P.: *Estymatory jądrowe w analizie systemowej*. WNT, 2005.
- [8] Kulczycki P.: Analiza danych z użyciem estymatorów jądrowych w zastosowaniu do diagnostyki systemów. Korbicz J., Patan K., Kowal M. (red.), *Diagnostyka procesów i systemów*, ss. 231-238, EXIT, 2007.
- [9] Kulczycki P.: Statistical Kernel Estimators for Design of a Fault Detection, Diagnosis and Prognosis System. *The Open Cybernetics and Systemics Journal*, vol. 2, ss. 180-184, 2008, open access: <http://www.bentham.org/open/tocsj/openaccess2.htm>.
- [10] Kulczycki P., Charytanowicz M.: Kompletny algorytm gradientowej klasteryzacji. Malinowski K., Rutkowski L. (red.), *Sterowanie i automatyzacja: aktualne problemy i ich rozwiązania*, ss. 312-321, EXIT, 2008.
- [11] Kulczycki P., Hryniewicz O., Kacprzyk J. (red.): *Techniki informacyjne w badaniach systemowych*. WNT, 2007.
- [12] Kulczycki P., Kowalski P.: Klasyfikacja informacji niedokładnej typu przedziałowego ze zredukowanymi próbami wzorcowymi. Hryniewicz O., Straszak A., Studziński J. (red.), *Badania operacyjne i systemowe: środowisko naturalne, przestrzeń, optymalizacja*, ss. 305-314, IBS PAN (ser. Badania Systemowe, tom 63), 2008.
- [13] Kulczycki P., Prochot C.: Wykrywanie elementów odosobnionych za pomocą metod estymacji nieparametrycznej. Kulikowski R., Kacprzyk J., Słowiński R. (red.), *Badania operacyjne i systemowe: podejmowanie decyzji – podstawy teoretyczne i zastosowania*, ss. 313-328, EXIT, 2004.
- [14] Pal S.K., Mitra P.: *Pattern Recognition Algorithms for Data Mining*. Chapman and Hall, 2004.
- [15] Vidal R.V.V., Nahorski Z. (red.): Simulated Annealing Applied to Combinatorial Optimization (Special Issue), *Control and Cybernetics*, vol. 25(1), 1996.

Redukcja wymiaru i liczności próby dla potrzeb syntezy statystycznego układu wykrywania uszkodzeń

P. Kulczycki*, S. Łukasik**

* Instytut Badań Systemowych, Polska Akademia Nauk,
e-mail: Piotr.Kulczycki@ibspan.waw.pl

** Instytut Badań Systemowych, Polska Akademia Nauk,
e-mail: Szymon.Lukasik@ibspan.waw.pl

Przedmiotem niniejszej pracy jest zagadnienie redukcji wymiaru i liczności próby losowej z przeznaczeniem do procedur eksploracyjnej analizy danych, określonych przy użyciu metodyki statystycznych estymatorów jądrowych. Koncepcja opiera się na liniowej transformacji przestrzeni, przy czym współczynniki macierzy wyznaczane są z zastosowaniem metaheurystyki symulowanego wyżarzania. Ponadto stosuje się eliminację i/lub redukcję znaczenia tych elementów próby, których położenie uległo istotnej zmianie. Przedstawiona metoda została pozytywnie zweryfikowana w zagadnieniach wykrywania elementów nietypowych, klasteryzacji oraz klasyfikacji, których odpowiednie stosowanie umożliwia skonstruowanie efektywnego układu statystycznego wykrywania uszkodzeń w systemach dynamicznych w zakresie detekcji, diagnozy oraz związanej z nimi predykcji. Redukcja wymiaru i liczności wykorzystywanych prób pozwala na zwiększenie ilości zmiennych użytych do wnioskowania o stanie technicznym nadzorowanego systemu, jak również ograniczenie liczby błędnych wskazań oraz czasu obliczeń.

Reduction of dimensionality and size of a sample for synthesis of a statistical fault detection system

P. Kulczycki*, S. Łukasik**

* Systems Research Institute, Polish Academy of Science,
e-mail: Piotr.Kulczycki@ibspan.waw.pl

** Systems Research Institute, Polish Academy of Science,
e-mail: Szymon.Lukasik@ibspan.waw.pl

The subject of this paper is the task of reduction of dimensionality and size of a random sample allocated for exploratory data mining procedures, developed using the statistical kernel estimators methodology. The concept is based on linear transformation, whereby the elements of its matrix are calculated by a simulated annealing metaheuristic. Moreover, the elimination and/or reduction in importance of those sample elements with substantial changes in positions, is worked out. The presented method was successfully verified for the tasks of identification of atypical elements (outliers), clustering and classification, where appropriate usage allows the design of an effective statistical fault detection system in dynamical objects, in the scope of detection, diagnosis, and prediction connected with both of them. The reduction of dimensionality and size of samples enables an increase in the number of variables used for inference regarding the technical state of a supervised system, as well as allowing a decrease in the number of incorrect indications and calculation time.