



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

**Piotr Kulczycki  
Szymon Łukasik**

**Instytut Badań Systemowych  
Polska Akademia Nauk**

# Wprowadzenie

W ramach wcześniejszych badań opracowano **statystyczny układ wykrywania uszkodzeń** opierający się o metodykę estymatorów jądrowych (Kulczycki, 2007 & Kulczycki, 2008).

Obejmuje on swym zakresem:

- detekcję uszkodzeń
- diagnozę uszkodzeń
- predykcję uszkodzeń

Wnioskowanie odbywa się na podstawie wartości bądź relacji pomiędzy poszczególnymi współrzędnymi **wektora symptomów** czyli skończonego zbioru wielkości pomiarowych reprezentujących stan techniczny nadzorowanego obiektu.

# Synteza układu wykrywania uszkodzeń

**Detekcja uszkodzeń** – próba losowa reprezentuje wartości wektora symptomów uznane za typowe, natomiast  $\tilde{x}$  – aktualny stan. Stosując procedurę wykrywania elementów odosobnionych (nietypowych) można sprawdzić czy stan ten można uznać za odosobniony (tj. świadczący o wystąpieniu nieprawidłowości).

**Diagnoza uszkodzeń** – dostępne są próby wzorcowe reprezentujące poszczególne rodzaje uszkodzeń. Po wykryciu nieprawidłowości, stosując procedurę klasyfikacji można stwierdzić z którym z nich ma się do czynienia. Gdy podział elementów reprezentujących różne rodzaje uszkodzeń nie jest znany można go uzyskać poprzez procedurę klasteryzacji.

# Tematyka badań – uzasadnienie

Ekstrakcja wiedzy z dużych zbiorów danych jest zagadnieniem wysoce 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.

W przypadku rozważanego statystycznego układu wykrywania uszkodzeń 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.

# Redukcja wymiaru

Realizowana poprzez liniową transformację postaci:

$$Y = AX$$

$X$  – zbiór pierwotny ( $n \times m$ )

$A$  – macierz transformacji ( $N \times n$ )

$Y$  – wynikowy zbiór danych ( $N \times m$ ,  $N \ll n$ )

Wyznaczanie postaci macierzy – metoda symulowanego wyżarzania.

Wskaźnik jakości (*ang. raw stress*):

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

# Szczegółowe komentarze

**Rozwiązanie początkowe** wyznaczane jest na podstawie algorytmu wyboru cech (Pal&Mitra, 2004). Opiera się on na koncepcji podziału współrzędnych na skupienia, zawierające cechy do siebie podobne, z kryterium zgodności dwóch współrzędnych definiowanym przez indeks maksymalnej kompresji informacji. Podział na skupienia odbywa się na podstawie algorytmu k-najbliższych sąsiadów, przy czym przyjmuje się  $k = n - N$ .

**Generowanie rozwiązania sąsiedniego** odbywa się poprzez losowy wybór jednego z elementów macierzy transformacji i jego zmianę o wartość ustalonego kroku  $\Delta$  lub  $-\Delta$  (w obu przypadkach z prawdopodobieństwem **0,5**).

Ponadto przyjęto klasyczny, dobrze uzasadniony teoretycznie schemat zmiany temperatury – schemat logarytmiczny.

# Redukcja liczności próby

Poszczególnym elementom próby  $x_i$  przyporządkowane są nieujemne wagi  $w_i$ , które 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.

Wagi mogą być użyte:

- aby uzyskać poprawę jakości estymacji w przestrzeni o zredukowanym wymiarze
- aby zredukować licznosc próby (można usunąć z próby te elementy  $x_i$ , dla których przyporządkowane im wagi spełniają warunek  $w_i < W$ , przy czym  $W \in [0, 1]$ )

# Eksperymenty obliczeniowe - wstęp

W zakresie redukcji wymiaru, zaproponowana koncepcja została porównana z metodami głównych składowych oraz selekcji cech Pala i Mitry, a także pełnym przeszukiwaniem przestrzeni rozwiązań minimalizujących przedstawiony wcześniej wskaźnik jakości. Testy przeprowadzono dla procedur wykrywania elementów odosobnionych (stosowanej w celu detekcji ewentualnego uszkodzenia), analizy skupień (podziału próby reprezentującej uszkodzenia na klasy) oraz klasyfikacji (diagnozy uszkodzenia), z użyciem estymatorów jądrowych, w każdym z przypadków porównując wyniki otrzymane przed i po redukcji wymiaru.



# Przykład I: wykrywanie el. nietypowych

Wygenerowano 250 elementowy zbiór danych z rozkładów:

$$N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \text{ – z wagą } 0,95$$

$$N\left(\begin{bmatrix} -10 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), N\left(\begin{bmatrix} 10 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), N\left(\begin{bmatrix} 0 \\ -10 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$N\left(\begin{bmatrix} 0 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \text{ – każdy z wagą } 0,0125$$

Wynikiem działania procedury wykrywania elementów odosobnionych było **15** elementów uznanych za nietypowe. Po zredukowaniu wymiaru do  $N = 1$  otrzymano średni wynik wykrywania zgodny w **98,5%** (met. pełnego przeszukiwania – **98,5%**, met. głównych składowych – **95,2%**, selekcji cech – **95,2%**).

# Przykład II: klasyfikacja

Z trzech rozkładów normalnych wygenerowano trzy wzorce po 50 elementów oraz zbiór klasyfikowanych elementów o liczności 1500, po 500 z każdej klasy:

$$N\left(\begin{bmatrix} -10 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), N\left(\begin{bmatrix} 10 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), N\left(\begin{bmatrix} 0 \\ -10 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

W przestrzeni dwuwymiarowej otrzymano **23,9%** błędnych klasyfikacji. Po zredukowaniu wymiaru ilość błędnych klasyfikacji wzrosła do **33,0%** (met. pełnego przeszukiwania – **33,0%**, met. głównych składowych – **33,1%**, selekcji cech – **43,5%**). Po usunięciu tych elementów dla których  $w_i < 0,5$  i ustaleniu pozostałych wag na **1**, ilość błędnych klasyfikacji zmalała do **32,8%**, z równoczesną redukcją próby wzorcowej o **30** elementów. Dla próby pozbawionej elementów dla których  $w_i < 1$  osiąga się **33,1%** błędnych klasyfikacji, jednak przy eliminacji **72** elementów wzorcowych (a zatem niemal o 50%).

# Podsumowanie

## **Badania:**

Wstępna weryfikacja dla relatywnie niewymagających obliczeniowo zagadnień analizy i eksploracji danych dwuwymiarowych przyniosła bezsprzecznie pozytywne rezultaty. Metoda zapewnia skuteczność porównywalną do techniki głównych składowych (PCA), z dodatkową możliwością przeprowadzenia redukcji liczności próby.

## **Dalsze prace/problemy badawcze:**

- poprawa efektywności zaimplementowanego algorytmu symulowanego wyżarzania + jego zrównoleglenie
- eksperymenty obliczeniowe dla dużych zbiorów danych reprezentujących rzeczywiste zagadnienia detekcji i diagnozy uszkodzeń
- analiza porównawcza



**Dziękuję Państwu za uwagę!**