

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

**Taksonomia 27**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego  
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania  
znajdują się na stronach internetowych  
[www.pracnaukowe.ue.wroc.pl](http://www.pracnaukowe.ue.wroc.pl)  
[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Publikacja udostępniona na licencji Creative Commons  
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska  
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2016

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**e-ISSN 2392-0041**

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:  
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
ul. Komandorska 118/120, 53-345 Wrocław  
tel./fax 71 36 80 602; e-mail:[econbook@ue.wroc.pl](mailto:econbook@ue.wroc.pl)  
[www.ksiegarnia.ue.wroc.pl](http://www.ksiegarnia.ue.wroc.pl)

Druk i oprawa: TOTEM

## Spis treści

|   |    |
|---|----|
| <b>Wstęp</b> .....  | 9  |
| <b>Beata Bal-Domańska:</b> Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....   | 11 |
| <b>Tomasz Bartłomowicz:</b> Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....   | 20 |
| <b>Maciej Beręsewicz, Marcin Szymkowiak:</b> Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....  | 30 |
| <b>Beata Bieszk-Stolorz:</b> Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....  | 40 |
| <b>Justyna Brzezińska:</b> Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....   | 49 |
| <b>Mariola Chrzanowska, Nina Drejerska:</b> Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union..... | 58 |
| <b>Sabina Denkowska:</b> Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....   | 66 |
| <b>Adam Depta:</b> Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jękających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....   | 76 |
| <b>Mariusz Doszyń, Sebastian Gnat:</b> Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....  | 84 |

|  |     |
|--|-----|
| <b>Marta Dziechciarz-Duda, Anna Król:</b> Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences .....   | 94  |
| <b>Ewa Genge:</b> Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour .....   | 103 |
| <b>Joanna Górna, Karolina Górna:</b> Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools .....  | 112 |
| <b>Alicja Grześkowiak:</b> Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population .....  | 122 |
| <b>Agnieszka Kozera, Feliks Wysocki:</b> Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods .....   | 131 |
| <b>Mariusz Kubus:</b> Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....   | 143 |
| <b>Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski:</b> Analiza wydzźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....  | 153 |
| <b>Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki:</b> Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R ..... | 165 |
| <b>Dorota Rozmus:</b> Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method .....  | 176 |
| <b>Adam Sagan, Aneta Rybicka, Justyna Brzezińska:</b> <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....  | 184 |
| <b>Michał Stachura:</b> O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case .....  | 195 |
| <b>Michał Stachura, Barbara Wodecka:</b> Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events .....  | 205 |
| <b>Iwona Staniec, Jan Żółtowski:</b> Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości  |     |

---

|  |     |
|--|-----|
| technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....  | 215 |
| <b>Marcin Szymkowiak, Wojciech Roszka:</b> Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....        | 224 |
| <b>Lucyna Wojcieszka:</b> Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy..... | 234 |

## **Wstęp**

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

*Krzysztof Jajuga, Marek Walesiak*

**Dorota Rozmus**

Uniwersytet Ekonomiczny w Katowicach  
e-mail: dorota.rozmus@ue.katowice.pl

---

**BADANIE STABILNOŚCI  
TAKSONOMICZNEJ CZYNNIKOWEJ  
METODY ODLEGŁOŚCI PROBABILISTYCZNEJ**  
**STABILITY OF THE FACTOR PROBABILITY  
DISTANCE CLUSTERING METHOD**

---

DOI: 10.15611/pn.2016.427.18

**Streszczenie:** Czynnikiowa metoda odległości probabilistycznej (*Factor Probability Distance Clustering*), tworząc liniową kombinację oryginalnych zmiennych, prowadzi do mniejszej liczby ortogonalnych czynników, a następnie dokonuje podziału obiektów przy zastosowaniu metody odległości probabilistycznej (*Probability Distance Clustering*). Metoda ta może w znaczący sposób poprawić dokładność rozpoznawania rzeczywistej struktury grup; ponadto w przypadku dużych zbiorów danych może się przyczynić do podniesienia stabilności rozwiązania. Głównym celem artykułu będzie zbadanie stabilności czynnikiowej metody odległości probabilistycznej. Przyjęte w pracy kryterium stabilności można sformułować następująco: metodę grupowania uznaje się za stabilną, jeżeli uzyskane przy wielokrotnym jej zastosowaniu wyniki grupowania nie różnią się między sobą.

**Słowa kluczowe:** taksonomia, grupowanie, stabilność, czynnikiowa metoda odległości probabilistycznej.

**Summary:** Factorial clustering methods have been developed in recent years thanks to the improving of computational power. They have been proposed in order to cluster large datasets, where large is referred to the number of variables. The factor probability distance clustering method performs a linear transformation of original variables into a reduced number of orthogonal ones and in the next step it clusters transformed data by means of probability distance method. This method can significantly improve the algorithm performance; moreover large datasets can be partitioned into clusters with increasing stability of the results. The main aim of this article is a study of stability of the factor probability distance clustering method.

**Keywords:** taxonomy, stability, factor probability distance clustering method.

## 1. Wstęp

Stosowanie metod taksonomicznych w jakimkolwiek zagadnieniu grupowania wymaga zapewnienia wysokiej zgodności między strukturą grup obecną w zbiorze danych, a strukturą uzyskaną w wyniku grupowania. Ona bowiem warunkuje skuteczność wszelkich decyzji podjętych na podstawie uzyskanych rezultatów. Przez pojęcie dokładność grupowania należy rozumieć zdolność metody do rozpoznawania rzeczywistej struktury grup. Jednocześnie pojawia się coraz więcej licznych zbiorów danych, charakteryzowanych przez dużą liczbę zmiennych, co sprawia, że klasyczne algorytmy taksonomiczne mogą się okazać niestabilne [Vichi, Kiers 2001]. Przez stabilność należy rozumieć niezmienność rezultatów podziału przy wielokrotnym stosowaniu metody. Dlatego też w literaturze wciąż proponowane są nowe rozwiązania, których zadaniem jest poprawa zarówno dokładności grupowania, jak i stabilności rozwiązań w stosunku do tradycyjnie stosowanych metod (np.  $k$ -średnich). Przykładami mogą tu być metody polegające na zastosowaniu podejścia zagregowanego [Dudoit, Fridlyand 2003; Hornik 2005; Leisch 1999], metoda propagacji podobieństwa [Frey, Dueck 2007], metoda odległości probabilistycznej [Israel, Iyigun 2008] czy też czynnikowa metoda odległości probabilistycznej [Gettler Summa, Palumbo, Tortora 2011]<sup>1</sup>.

Celem artykułu jest przedstawienie wyników dotyczących badania stabilności czynnikowej metody odległości probabilistycznej. Badanie to przeprowadzone zostanie w porównaniu do stabilności klasycznej metody  $k$ -średnich.

## 2. Taksonomiczna czynnikowa metoda odległości probabilistycznej

Czynnikowe metody grupowania pojawiły się w celu podziału zbiorów danych charakteryzujących się dużą liczbą cech. Ogólnie rzecz ujmując, tego rodzaju metody składają się z dwóch kroków:

- 1) liniowej transformacji wyjściowych zmiennych do mniejszej liczby czynników;
- 2) grupowania obiektów w przekształconej przestrzeni cech.

W literaturze po raz pierwszy takie dwuetapowe podejście zostało zaproponowane w 1984 r. [Lebart, Morineau, Warwick 1984].

Metoda czynnikowej odległości probabilistycznej (*Factor Probabilistic Distance Clustering* – FPDC) to metoda nieparametryczna, będąca uogólnieniem metody odległości probabilistycznej (*Probabilistic Distance Clustering* – PDclustering). W metodzie tej wspomniane wcześniej dwa kroki można bardziej szczegółowo przedstawić jako:

---

<sup>1</sup> Wyniki badań dotyczących dokładności i stabilności metod zagregowanych można znaleźć np. w pracach: [Rozmus 2010, 2012, 2013].



1) dekompozycję macierzy odległości między obiektami a załączkami skupień za pomocą metody Tuckera 3 (*Tucker3 decomposition, three-mode factor analysis – 3MFA, three-mode principal component analysis – 3MPCA*);

2) grupowanie obiektów w przekształconej przestrzeni cech za pomocą metody odległości probabilistycznej.

Według autorów metody połączenie transformacji czynnikowej oraz metody odległości probabilistycznej powoduje, że grupowanie staje się bardziej stabilne, odporne na występowanie obserwacji odstających; możliwe jest także grupowanie obiektów ze zbiorów o dużej liczbie zmiennych, czy też charakteryzujących się skupieniami o różnej liczbie obiektów, bądź też o nieeliptycznych kształtach.

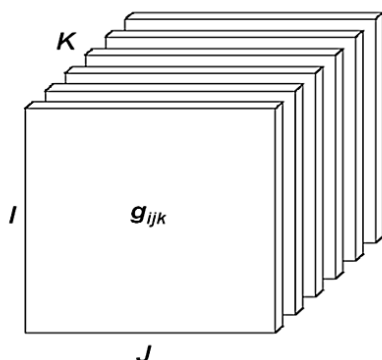
Algorytm omawianej metody można przedstawić w następujących krokach:

1. Losowa inicjalizacja macierzy załączków skupień  $C$  o elementach  $c_{kj}$  ( $k = 1, 2, \dots, K$  to liczba skupień;  $j = 1, 2, \dots, J$  – liczba zmiennych).

2. Obliczenie elementarnej 3-wymiarowej macierzy odległości między obserwacjami a załączkami skupień (rys. 1):

$$g_{ijk} = |x_{ij} - c_{kj}|,$$

gdzie  $i = 1, 2, \dots, I$  to liczba obserwacji.



**Rys. 1.** Trójwymiarowa macierz odległości między obiektami a załączkami skupień

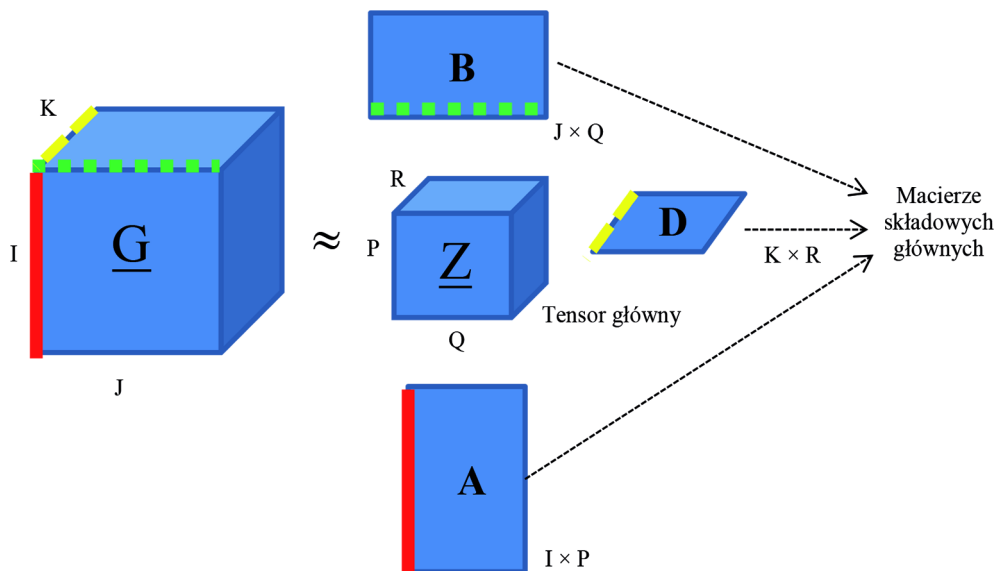
Źródło: opracowanie własne na podstawie [Kolda, Bader 2008].

3. Dekompozycja macierzy odległości za pomocą metody Tuckera 3 (rys. 2) [Tucker 1966]<sup>2</sup>:

<sup>2</sup> Dekompozycja Tuckera jest odpowiednikiem analizy głównych składowych dla danych 3-wymiarowych. W metodzie tej 3-wymiarowa macierz dekomponowana jest na tzw. tensor główny (*core tensor*)  $\underline{\mathbf{Z}}$  oraz macierze składowych (*component matrices*)  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$ . Tensor główny ma wymiary  $P \times Q \times R$ , gdzie  $P$ ,  $Q$  i  $R$  to liczba składowych głównych, które określają także liczbę kolumn w macierzach składowych  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$  [Tucker 1966].

$$g_{ijk} \approx \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R z_{pqr} a_{ip} b_{jq} d_{kr},$$

gdzie:  $p = 1, 2, \dots, P$ ;  $q = 1, \dots, Q$ ;  $r = 1, 2, \dots, R$  – liczba składowych w dekompozycji;  $a_{ip}$ ,  $b_{jq}$ ,  $d_{kr}$  – elementy macierzy składowych **A**, **B**, **D**;  $z_{pqr}$  – elementy tensora głównego **Z**.



Rys. 2. Dekompozycja Tuckera 3

Źródło: opracowanie własne na podstawie [Kolda, Bader 2008].

4. Współrzędne  $x_{iq}^*$  obiektów w zredukowanej przestrzeni zmiennych pochodzących z dekompozycji Tuckera wyliczane są jako:

$$x_{iq}^* = \sum_{j=1}^J x_{ij} b_{jq}.$$

5. Grupowanie obiektów z kroku 4. za pomocą metody odległości probabilistycznej.

Metoda odległości probabilistycznej to metoda iteracyjna, w której przynależność obiektów do grup jest oparta na odległości (zazwyczaj Euklidesowej) obiektów od centrów skupień. Według A. Israela i C. Iyiguna [2008] w metodzie tej: „Mając dane skupienia, ich centra oraz odległości obserwacji od tych centrów skupień, prawdopodobieństwo przynależności obiektów do skupień jest odwrotnie proporcjonalne do odległości od centra skupienia, które jest brane pod uwagę”.

Metoda ta zatem określa przynależność obiektów do grup jako prawdopodobieństwo przynależności do skupienia  $p_k(\mathbf{x}_i)$ , zależne od odpowiedniej odległości  $d_k(\mathbf{x}_i)$ .

Jak twierdzą autorzy, metoda ta jest prosta, szybka (wymaga niewielkiej liczby iteracji), odporna (niewrażliwa na obserwacje odstające) i daje wysoką poprawność wyników grupowania.

Schematyczny opis algorytmu metody odległości probabilistycznej – dla ułatwienia – zaprezentowany jest dla przypadku dwóch klas [Israel, Iyigun 2008]:

**Inicjalizacja.** Mamy zbiór danych  $X$ , dowolne dwa punkty  $\mathbf{c}_1, \mathbf{c}_2$  stanowiące początkowe załączki skupień oraz  $\varepsilon > 0$ .

### Iteracja:

Krok 1. **Oblicz** odległości  $d_1(\mathbf{x}_i), d_2(\mathbf{x}_i)$  dla wszystkich  $\mathbf{x} \in X$ , np. według formuły euklidesowej<sup>3</sup>:

$$d_k(\mathbf{x}_i, \mathbf{c}_k) = \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \mathbf{c}_k\|, \quad k = 1, 2.$$

Krok 2. **Uaktualnij** centra skupień  $\mathbf{c}_1^+, \mathbf{c}_2^+$ :

$$\mathbf{c}_k^+ = \sum_{i=1}^N \left( \frac{u_k(\mathbf{x}_i)}{\sum_{j=1}^N u_k(\mathbf{x}_j)} \right) \mathbf{x}_i, \quad k = 1, 2,$$

gdzie:

$$u_k(\mathbf{x}_i) = \frac{p_k(\mathbf{x}_i)^2}{d_k(\mathbf{x}_i, \mathbf{c}_k)}, \quad k = 1, 2,$$

oraz:

$$p_k(\mathbf{x}_i) = \frac{\prod_{j \neq k} d_j(\mathbf{x}_i)}{\sum_{t=1}^K \prod_{j \neq t} d_j(\mathbf{x}_i)}, \quad k = 1, 2.$$

Krok 3. **Jeżeli**  $\|\mathbf{c}_1^+ - \mathbf{c}_1\| + \|\mathbf{c}_2^+ - \mathbf{c}_2\| < \varepsilon$  **zatrzymaj** działanie algorytmu, w przeciwnym przypadku **wrót** do kroku 1.

<sup>3</sup> Wybór odległości euklidesowej autorzy uzasadniają tym, że jest to powszechnie i często stosowana metryka. W swojej pracy wspominają także o możliwości zastosowania innych metryk, np. odległości Mahalanobisa [Israel, Iyigun 2008].

### 3. Wyniki badań empirycznych

Badanie stabilności czynnikowej metody odległości probabilistycznej przeprowadzono w odniesieniu do stabilności metody  $k$ -średnich. Zastosowana została miara stabilności dla par podziałów (*pairwise stability*) zaproponowana przez L.I. Kunchevą i D.P. Vetrova [2006]:

$$Stab = \frac{2}{M \cdot (M - 1)} \sum_{\substack{1 \leq m, l \leq M \\ m < l}}^M IR(P_m, P_l),$$

gdzie:  $M$  – liczba podziałów;  $IR$  – Indeks Randa [Rand 1971];  $P_m, P_l$  – grupowanie na podstawie  $m$ -tego i  $l$ -tego podziału.

W badaniach zastosowano zbiory danych udostępniane przez Uniwersytet Kalifornijski, które standardowo wykorzystywane są w badaniach porównawczych<sup>4</sup>. Ich charakterystyka zawarta jest w tab. 1.

**Tabela 1.** Zastosowane zbiory danych

| Nazwa zbioru        | Liczba obserwacji | Liczba cech | Liczba grup |
|---------------------|-------------------|-------------|-------------|
| UCI segment         | 2 310             | 19          | 7           |
| UCI sat             | 4 435             | 36          | 7           |
| UCI opt digits      | 3 823             | 64          | 10          |
| UCI EEG             | 80                | 14          | 2           |
| UCI spect           | 14 980            | 14          | 2           |
| UCI movement libras | 360               | 90          | 15          |

Źródło: opracowanie własne na podstawie [<https://archive.ics.uci.edu/ml/datasets.html>].

Każdą z metod – tj.  $k$ -średnich<sup>5</sup> i czynnikową metodę odległości probabilistycznej<sup>6</sup> – zastosowano 50 razy, uzyskując w efekcie 50 podziałów ( $M = 50$ )<sup>7</sup>, których stabilność została zbadano za pomocą miary *Stab*. By uzyskać szerszy pogląd na wartości miary *Stab* dla tych metod, całe badanie powtórzono 10 razy. Wyniki dla tych 10 powtórzeń przedstawiono za pomocą wykresów (rys. 3).

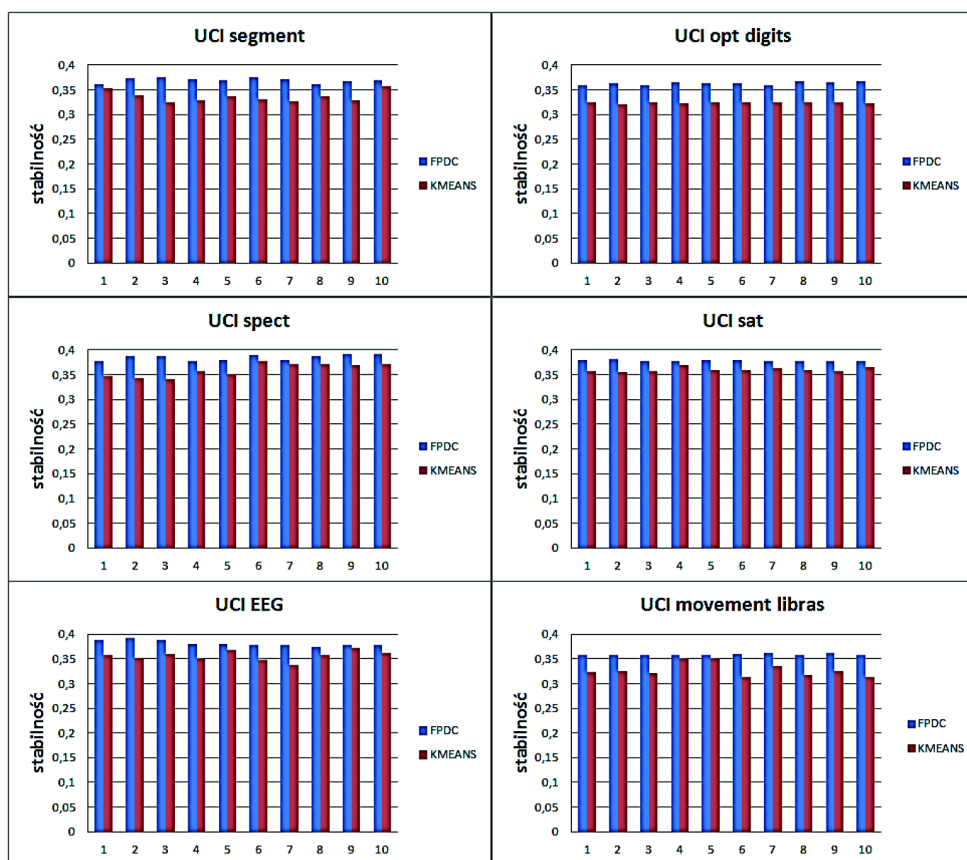
Patrząc na wyniki przedstawione na rys. 3, widzimy, że czynnikowa metoda odległości probabilistycznej (FPDC) charakteryzuje się wyższymi wartościami

<sup>4</sup> Struktura grup znana jest *a priori*.

<sup>5</sup> Do obliczeń zastosowano funkcję *kmeans* w programie R z algorytmem metody  $k$ -średnich zaproponowanym przez J.A. Hartigana i M.A. Wonga [1979].

<sup>6</sup> Czynnikowa metoda odległości probabilistycznej znajduje się w pakiecie FPDC w programie R.

<sup>7</sup> Liczba grup przyjęta w badaniu odpowiadała rzeczywistej liczbie grup w każdym zbiorze danych.



Rys. 3. Wartości miary stabilności dla czynnikowej metody odległości probabilistycznej (FPDC) i metody  $k$ -średnich (KMEANS)

Źródło: opracowanie własne.

miary stabilności  $Stab$  niż metoda  $k$ -średnich (KMEANS). Największe różnice między wartościami miary  $Stab$  dla badanych dwóch metod można zaobserwować dla zbioru UCI segment, a najmniejsze dla UCI sat. Biorąc jednak pod uwagę fakt, że miara  $Stab$  przybiera wartości z przedziału  $\langle 0, 1 \rangle$ , to uzyskane wyniki dla metody FPDC (miara  $Stab$  nie przekracza wartości 0,4), świadczą jednak o dosyć niskiej stabilności tej metody w przypadku zastosowanych zbiorów danych.

#### 4. Zakończenie

Czynnikowa metoda odległości probabilistycznej łączy ze sobą transformację czynnikową (dekompozycja Tuckera 3) oraz grupowanie obiektów za pomocą metody odległości probabilistycznej. Połączenie to pozwala m.in. na zwiększenie

stabilności wyników grupowania, grupowanie obiektów ze zbiorów o dużej liczbie zmiennych, czy też o nieeliptycznych kształtach skupień.

Celem artykułu było zbadanie stabilności czynnikowej metody odległości probabilistycznej. Badanie przeprowadzono w porównaniu do stabilności klasycznej metody  $k$ -średnich. W świetle uzyskanych wyników można stwierdzić, że metoda czynnikowa charakteryzuje się większą stabilnością niż metoda klasyczna.

## Literatura

- Dudoit S., Fridlyand J., 2003, *Bagging to improve the accuracy of a clustering procedure*, Bioinformatics, vol. 19, no. 9, s. 1090–1099.
- Frey B.J., Dueck D., 2007, *Clustering by passing messages between data points*, Science, vol. 315, no. 5814, s. 972–976.
- Gettler Summa M., Palumbo F., Tortora C., 2011, *Factor PD-clustering*, Working Paper, <http://arxiv.org/abs/1106.3830> (3.07.2012).
- Hartigan J.A., Wong M.A., 1979, *A K-means clustering algorithm*, Applied Statistics vol. 28, no. 1, s. 100–108.
- Hornik K., 2005, *A CLUE for CLUster ensembles*, Journal of Statistical Software, vol. 14, no. 12, s. 65–72.  
<https://archive.ics.uci.edu/ml/datasets.html>.
- Israel A., Iyigun C., 2008, *Probabilistic d-clustering*, Journal of Classification, vol. 25, no. 1, s. 5–26.
- Kolda T.G., Bader B.W., 2008, *Tensor decompositions and applications*, SIAM Review, vol. 51, no. 3, s. 455–500.
- Kuncheva L.I., Vetrov D.P., 2006, *Evaluation of stability of k-means cluster ensembles with respect to random initialization*, IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 28, no. 11, s. 1798–1808.
- Lebart A., Morineau A., Warwick K., 1984, *Multivariate Statistical Descriptive Analysis*, John Wiley and Sons, New York.
- Leisch F., 1999, *Bagged clustering*, Adaptive Information Systems and Modeling in Economics and Management Science, Working Papers, SFB, no. 51.
- Rand W.M., 1971, *Objective criteria for the evaluation of clustering methods*, Journal of the American Statistical Association, vol. 66 (336), s. 846–850.
- Rozmus D., 2010, *Porównanie stabilności zagregowanych algorytmów taksonomicznych opartych na macierzy współwystąpień*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 176, Taksonomia 18: *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 212–220.
- Rozmus D., 2012, *Porównanie dokładności spektralnej oraz zagregowanych algorytmów opartych na idei metody bagging*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 242, Taksonomia 19: *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 352–360.
- Rozmus D., 2013, *Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody bagging*, Prace Naukowe, Uniwersytetu Ekonomicznego we Wrocławiu, nr 279, Taksonomia 21: *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 140–152.
- Tucker L.R., 1966, *Some mathematical notes on three-mode factor analysis*, Psychometrika, vol. 31, no. 3, s. 279–311.
- Vichi M., Kiers H., 2001, *Factorial k-means analysis for two way data*, Computational Statistics and Data Analysis, vol. 37, no. 1, s. 29–64.