

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

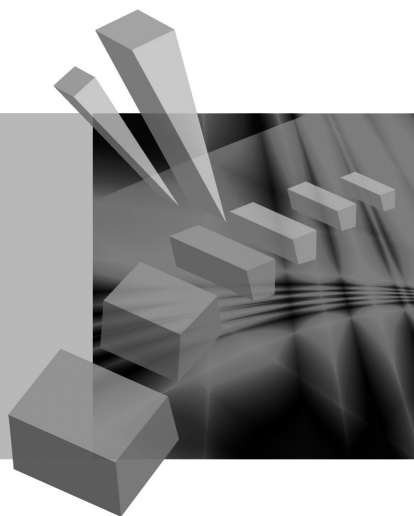
RESEARCH PAPERS

of Wrocław University of Economics

278

Taksonomia 20

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

| | |
|--|-----|
| Wstęp | 9 |
| Józef Pocięcha: Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm | 15 |
| Eugeniusz Gatnar: Analiza miar adekwatności rezerw walutowych | 23 |
| Marek Walesiak: Zagadnienie doboru liczby klas w klasyfikacji spektralnej | 33 |
| Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska: Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight | 44 |
| Andrzej Bąk: Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code> | 54 |
| Aleksandra Łuczak, Feliks Wysocki: Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia | 63 |
| Ewa Roszkowska: Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych | 74 |
| Jacek Batóg: Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych | 85 |
| Jerzy Korzeniewski: Modyfikacja metody HINoV selekcji zmiennych w analizie skupień | 93 |
| Małgorzata Markowska, Danuta Strahl: Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony | 101 |
| Elżbieta Sobczak: Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej..... | 111 |
| Elżbieta Gołata, Grażyna Dehnel: Rozbieżności szacunków NSP 2011 i BAEL..... | 120 |
| Iwona Foryś: Wykorzystanie analizy historii zdarzeń do badania powtórnych sprzedaży na lokalnym rynku mieszkaniowym | 131 |
| Hanna Dudek, Joanna Landmesser: Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów..... | 142 |
| Grażyna Łaska: Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych | 151 |
| Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach | 161 |

| | |
|---|-----|
| Andrzej Bąk, Tomasz Bartłomowicz: Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R | 169 |
| Andrzej Dudek, Bartosz Kwaśniewski: Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA | 180 |
| Michał Trzęsiok: Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej | 188 |
| Joanna Trzęsiok: Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji..... | 197 |
| Artur Mikulec: Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji | 206 |
| Artur Zaborski: Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji | 216 |
| Justyna Wilk: Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego | 225 |
| Karolina Bartos: Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP | 236 |
| Ewa Genge: Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych | 246 |
| Izabela Kurzawa: Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych | 254 |
| Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego..... | 262 |
| Aleksandra Łuczak: Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych | 271 |
| Marcin Pelka: Rozmyta klasyfikacja spektralna <i>c</i> -średnich dla danych symbolicznych interwałowych | 282 |
| Małgorzata Machowska-Szewczyk: Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne | 290 |
| Ewa Chodakowska: Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania | 300 |
| Beata Bieszk-Stolorz, Iwona Markowicz: Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia | 311 |
| Marcin Salamaga: Weryfikacja teorii poziomu rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej | 321 |
| Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce. | 330 |
| Hanna Gruchociak: Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem | 343 |

| | |
|---|-----|
| Radosław Pietrzyk: Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych | 351 |
| Sabina Denkowska: Procedury testowań wielokrotnych | 362 |

Summaries

| | |
|--|-----|
| Józef Pocięcha: Financial ratios and classification models of bankruptcy prediction | 22 |
| Eugeniusz Gatnar: Analysis of FX reserve adequacy measures | 32 |
| Marek Walesiak: Automatic determination of the number of clusters using spectral clustering | 43 |
| Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska: Structural analysis as a method of data classification in foresight research | 53 |
| Andrzej Bąk: Linear ordering methods in Polish taxonomy – pllord package | 62 |
| Aleksandra Łuczak, Feliks Wysocki: The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living | 73 |
| Ewa Roszkowska: Application of the fuzzy TOPSIS method to the estimation of negotiation offers..... | 84 |
| Jacek Batóg: Sensitivity analysis of ELECTRE III method for outliers and change of thresholds | 92 |
| Jerzy Korzeniewski: Modification of the HINoV method of selecting variables in cluster analysis | 100 |
| Małgorzata Markowska, Danuta Strahl: Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions | 110 |
| Elżbieta Sobczak: Smart workforce structures versus structural effects of employment changes in the European Union countries | 119 |
| Elżbieta Gołata, Grażyna Dehnel: Divergence in National Census 2011 and LFS estimates..... | 130 |
| Iwona Foryś: Event history analysis in the resale study on the local housing market | 141 |
| Hanna Dudek, Joanna Landmesser: Impact of the relative deprivation on subjective income satisfaction | 150 |
| Grażyna Łaska: Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities | 160 |
| Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analysis of relations between fundamental processes and capital market in China..... | 166 |
| Andrzej Bąk, Tomasz Bartłomowicz: Microeconomic polynomial models and their application in the analysis of preferences using R program..... | 179 |

| | |
|---|-----|
| Andrzej Dudek, Bartosz Kwaśniewski: Parallel processing of clustering algorithms in CUDA technology | 187 |
| Michał Trzęsiok: Real estate market value estimation based on multivariate statistical analysis | 196 |
| Joanna Trzęsiok: On some simulative procedures for comparing nonparametric methods of regression..... | 205 |
| Artur Mikulec: Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices | 215 |
| Artur Zaborski: Unfolding analysis by using gravity model | 224 |
| Justyna Wilk: Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital..... | 235 |
| Karolina Bartos: Risk analysis of bachelor students' university abandonment – the use of MLP networks | 245 |
| Ewa Genge: Clustering of industrial holiday participants with the use of latent class analysis..... | 253 |
| Izabela Kurzawa: Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households..... | 261 |
| Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modelling class imbalance problems: comparing classification approaches for surgical risk analysis | 270 |
| Aleksandra Łuczak: The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts..... | 281 |
| Marcin Pełka: A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data | 289 |
| Małgorzata Machowska-Szewczyk: Clustering algorithms for mixed-feature symbolic objects | 299 |
| Ewa Chodakowska: Malmquist index in enterprises classification on the basis of relative productivity changes | 310 |
| Beata Bieszk-Stolorz, Iwona Markowicz: Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment | 320 |
| Marcin Salamaga: Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries | 329 |
| Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Socio-economic situation as a determinant of internal migration in Poland | 342 |
| Hanna Gruchociak: Delimitation of local labor markets in Poland on the basis of the employment-related population flows research..... | 350 |
| Radosław Pietrzyk: Selectivity and timing in Polish mutual funds performance measurement | 361 |
| Sabina Denkowska: Multiple testing procedures..... | 369 |

Jacek Batóg

Uniwersytet Szczeciński

ANALIZA WRAŻLIWOŚCI METODY ELECTRE III NA OBSERWACJE NIETYPOWE I ZMIANĘ WARTOŚCI PROGOWYCH

Streszczenie: W artykule przeprowadzono analizę wrażliwości dwóch wybranych metod porządkowania obiektów, Taksonomicznej Miary Rozwoju i ELECTRE III, na wprowadzenie obserwacji nietypowych. Uzyskane wyniki pozwalają stwierdzić, że w przypadku rozważanego zbioru obserwacji niskim poziomem wrażliwości na wpływ obserwacji nietypowych charakteryzuje się wyłącznie ta druga metoda. Badanie obejmowało również ocenę wpływu zmiany wartości progowych stosowanych w metodzie ELECTRE III na rezultaty porządkowania obiektów. Potwierdzony został pogląd o dużym znaczeniu określanych poziomów wartości progowych na otrzymywane wyniki klasyfikacji.

Słowa kluczowe: klasyfikacja, metody porządkowania, obserwacje nietypowe.

1. Wstęp

We wszelkiego rodzaju zjawiskach, w tym również w zjawiskach ekonomicznych, ważną rolę odgrywają zdarzenia nietypowe, które mogą i często mają znaczący wpływ zarówno na ostateczny kształt specyfikowanego modelu, jak i na uzyskiwane za jego pomocą wyniki. Wskazuje się, że obserwacje nietypowe (*outliers*) najczęściej są skutkiem błędnych danych statystycznych lub przejawem znacząco odmiennej realizacji rozpatrywanego procesu. W literaturze przytaczane są zróżnicowane definicje obserwacji nietypowych. Według D. Hawkinsa [1980], obserwacja nietypowa jest to obserwacja na tyle znacząco różniąca się od innych obserwacji, że istnieje podejrzenie, że jest generowana przez inny proces. V. Barnett i T. Lewis [1994] określają jako obserwację nietypową obserwację istotnie różniącą się od innych obserwacji w próbie. D. Pyle podkreśla, że obserwacjami nietypowymi w przypadku danych mierzonych na skali nominalnej mogą być obserwacje o niskiej częstotliwości występowania [Pyle 1999]. Różnice w definicjach obserwacji odstających mogą być również efektem założeń dotyczących charakteru (struktury) analizowanych danych, jak również metod stosowanych do identyfikacji tych obserwacji. Jako przykład można tu podać definicję przekrojowej (*spatial*) obserwacji nietypowej rozumianej jako obiekt, którego charakterystyki niezwiązane z jego przestrzennymi

atrybutami są znacząco odmienne od obiektów znajdujących się w jego sąsiedztwie, nawet wówczas, gdy nie odróżnia się on od obiektów całej populacji [Shekhar, Lu, Zhang 2003].

Oprócz obserwacji nietypowych wymieniane są w literaturze również inne rodzaje obserwacji obniżających jakość danych i wpływających na obciążenie uzyskiwanych wyników, takie jak obserwacje dźwigniowe czy obserwacje wpływowe [Kufel 2010, s. 136]. W pracach A.J. Foxa [1972] oraz C. Chena i L.M. Liu [1993] wyróżniono cztery rodzaje obserwacji mających istotny wpływ na wyniki procesu modelowania charakterystyczne dla danych w postaci szeregów czasowych: *additive*, *innovative*, *level shift* i *transitory change*. R.S. Tsay [1988] wyróżnia jeszcze jeden rodzaj obserwacji nietypowej – *variance change*.

Stosowane metody identyfikacji obserwacji nietypowych uzależnione są od charakteru tych obserwacji. Obszerny opis klasyfikacji tych metod znaleźć można w pracach [Williams i in. 2002; Lu, Chen, Kou 2003]. Warto też podkreślić, że często ważne jest nie tylko odkrycie obserwacji nietypowych, lecz również określenie stopnia ich nietypowości [Breunig i in. 2000].

Obserwacja nietypowa może powodować zniekształcenia uzyskiwanych wyników, ale często też informuje o wystąpieniu zdarzenia mającego istotne znaczenie merytoryczne. Przykładem mogą być zdarzenia ekstremalne występujące na przykład w analizie ryzyka [*Zarządzanie ryzykiem...* 2007, s. 68]. W przypadku identyfikacji obserwacji nietypowych sugerowane są zazwyczaj dwa podejścia. Pierwsze z nich sprowadza się do usunięcia tych obserwacji. Drugie natomiast polega na wykorzystywaniu w procesie modelowania metod odpornych¹.

Przesłanką przeprowadzonego badania jest stwierdzenie, że obserwacje nietypowe mają znaczący wpływ na wyniki porządkowania uzyskiwane za pomocą metod statystycznej analizy wielowymiarowej. Natomiast jego celem jest porównanie wrażliwości rezultatów dwóch wybranych metod porządkowania obiektów: ELECTRE III oraz Taksonomicznej Miary Rozwoju (TMR) na wpływ obserwacji nietypowych². Pozwoli to określić, czy metody te są wrażliwe na wprowadzenie tego typu obserwacji. Dodatkowo weryfikacji poddana zostanie zmienność wyników porządkowania obiektów uzyskiwanych za pomocą metody ELECTRE III przy dokonywaniu zmian charakterystycznych dla niej wartości progowych: równoważności, preferencji i weta.

2. Opis metody ELECTRE III

Ponieważ metoda Taksonomicznej Miary Rozwoju należąca do grupy metod porządkowania liniowego jest szeroko opisana w literaturze przedmiotu³, ograniczymy

¹ Zdarzają się również propozycje łączące te dwa podejścia, zob. [Motulsky, Brown 2006].

² Wybór tych dwóch metod pozwoli przeciwstawić sobie metodę powszechnie znaną i stosowaną (TMR) z metodą, która jest znacznie mniej popularna, a dodatkowo wykorzystuje subiektywnie określone wartości progowe (ELECTRE III).

³ Począwszy od prac Z. Hellwiga – zob. na przykład [Hellwig 1968].

się do opisu metody ELECTRE III⁴. Metoda ta wykorzystuje relację przewyższania S zdefiniowaną na zbiorze Ω zawierającym obiekty Q_i ($i = 1, 2, \dots, n$), charakteryzowane przez cechy X_j , którym przypisane mogą być wagi k_j ($j = 1, 2, \dots, m$). Wartości cech X_j dla obiektów i oraz i' oznaczone są odpowiednio przez q_{ij} oraz $q_{i'j}$. Konstruowany model preferencji oparty jest na porównywaniu parami poszczególnych obiektów, a przy obliczaniu jego parametrów (współczynników) wykorzystuje się progi preferencji p_j , równoważności g_j i weta v_j , których wartości ustalane są przez tworzącego model. Procedura budowy modelu rozpoczyna się od obliczenia dwóch indeksów nazywanych odpowiednio współczynnikami zgodności $C(Q_p, Q_{i'})$:

$$C(Q_i, Q_{i'}) = \frac{1}{\sum_{j=1}^m k_j} \sum_{j=1}^m k_j C_j(Q_i, Q_{i'}), \quad (1)$$

$$\text{gdzie: } C_j(Q_i, Q_{i'}) = \begin{cases} 1 & \text{dla } q_{ij} + g_j \geq q_{i'j} \\ 0 & \text{dla } q_{ij} + p_j \geq q_{i'j} \\ (p_j + q_{ij} - q_{i'j}) / (p_j - q_j) & \text{dla pozostałych przypadków} \end{cases},$$

i niezgodności $D_j(Q_p, Q_{i'})$:

$$D_j(Q_i, Q_{i'}) = \begin{cases} 1 & \text{dla } q_{ij} + v_j \leq q_{i'j} \\ 0 & \text{dla } q_{ij} + p_j \leq q_{i'j} \\ (q_{i'j} - q_{ij} - p_j) / (v_j - p_j) & \text{dla pozostałych przypadków} \end{cases}, \quad (2)$$

na podstawie których obliczane są następnie współczynniki wiarygodności $S(Q_p, Q_{i'})$:

$$S(Q_i, Q_{i'}) = \begin{cases} C(Q_i, Q_{i'}) & \text{gdy } D_j(Q_i, Q_{i'}) \leq C(Q_i, Q_{i'}) \text{ dla } \forall j \\ C(Q_i, Q_{i'}) \prod_F \frac{1 - D_j(Q_i, Q_{i'})}{1 - C(Q_i, Q_{i'})} & \text{dla } D_j(Q_i, Q_{i'}) > C(Q_i, Q_{i'}) \end{cases} \quad (3)$$

pozwalające określić, w jakim stopniu obiekt Q_i przewyższa obiekt $Q_{i'}$. Relacja przewyższania S zostaje potwierdzona, jeśli poziom wiarygodności $S(Q_i, Q_{i'}) > \lambda - s(\lambda)$, gdzie parametr λ określanym mianem „poziom cięcia” definiowany jest przez decydenta. Wartości poziomów wiarygodności $S(Q_p, Q_{i'})$ pozwalają wyznaczyć macierz wiarygodności, której elementy $T(Q_p, Q_{i'})$ obliczane są następująco:

$$T(Q_i, Q_{i'}) = \begin{cases} 1 & \text{dla } S(Q_i, Q_{i'}) > \lambda - s(\lambda) \\ 0 & \text{dla } S(Q_i, Q_{i'}) \leq \lambda - s(\lambda) \end{cases}, \quad (4)$$

⁴ Zob. na przykład prace [Roy 1991; Figueira, Mousseau, Roy 2005].

a następnie „jakość” $M(i)$ poszczególnych obiektów będącą różnicą między sumą wartości $T(Q_p, Q_i)$ odpowiednich wierszy i kolumn macierzy wiarygodności. Stworzenie ostatecznego uporządkowania obiektów opiera się na dwóch uporządkowaniach cząstkowych: malejącym (zstępującym) i rosnącym (wstępującym).

3. Uzyskane wyniki

Ocena wrażliwości dwóch rozpatrywanych metod na występowanie obserwacji nietypowej przeprowadzona zostanie na podstawie porównania uporządkowań wybranych polskich miast według stopnia ich rozwoju. W badaniu wykorzystane zostały następujące cechy diagnostyczne charakteryzujące poziom rozwoju tych miast w roku 2010 (zob. tab. 1): dochody na 1 mieszkańca (X_1), zużycie energii elektrycznej na 1 mieszkańca (X_2), saldo migracji ogółem na pobyt stały (X_3), przeciętne miesięczne wynagrodzenie brutto (Polska = 100) (X_4), liczba samochodów osobowych na 1000 ludności (X_5), stopa bezrobocia (X_6), środki trwałe brutto w przedsiębiorstwach na 1 mieszkańca (X_7).

Tabela 1. Wartości cech diagnostycznych

| Miasto | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 |
|----------|-------|---------|--------|-------|-------|-------|---------|
| Warszawa | 6 085 | 976,7 | 3 940 | 136,7 | 543,6 | 3,5 | 141 948 |
| Kraków | 4 480 | 1 060,4 | 416 | 103,2 | 479,6 | 4,7 | 60 134 |
| Katowice | 4 330 | 878,3 | -1 078 | 132,9 | 500,7 | 3,8 | 80 003 |
| Poznań | 4 419 | 878,0 | -3 005 | 111,0 | 525,6 | 3,6 | 82 488 |
| Szczecin | 3 167 | 784,6 | -290 | 104,4 | 412,5 | 9,7 | 42 007 |
| Wrocław | 5 010 | 951,3 | 693 | 107,0 | 503,2 | 5,5 | 58 732 |
| Toruń | 3 612 | 742,5 | -642 | 93,8 | 412,2 | 8,2 | 40 916 |
| Gdańsk | 4 439 | 929,9 | -112 | 119,6 | 484,3 | 5,4 | 66 887 |

Źródło: BDL GUS.

W celu umożliwienia jednoznacznego porównywania wartości zmiennych X_3 (saldo migracji) oraz X_6 (stopa bezrobocia) z wartościami progowymi zostały one przeskalowane do wartości dodatnich, przy czym druga z nich dodatkowo została zamieniona na stymulantę⁵. W metodzie TMR w procesie normalizacji zmiennych zastosowana została metoda unitaryzacji zerowanej.

W pierwszej kolejności ocenie została poddana zgodność uporządkowań obiektów w wariancie podstawowym obejmującym 7 obserwacji, bez obserwacji nietypowej (Warszawy).

⁵ W przypadku zmiennej X_3 do każdej obserwacji dodano stałą równą 3100, a w przypadku zmiennej X_6 stałą równą 13. Stopa bezrobocia została wcześniej zamieniona na stymulantę za pomocą formuły: $X_j^S = -X_j^D$.

Tabela 2. Wyniki porządkowania w wariancie podstawowym (bez obserwacji nietypowej)

| Pozycja | TMR ($N = 7$) | ELECTRE III ($N = 7$) |
|---------|-----------------|-------------------------|
| 1 | Gdańsk | Wrocław |
| 2 | Katowice | Kraków |
| 3 | Wrocław | Katowice |
| 4 | Kraków | Gdańsk |
| 5 | Poznań | Poznań |
| 6 | Toruń | Toruń |
| 7 | Szczecin | Szczecin |

Źródło: obliczenia własne na podstawie tab. 1.

Tabela 3. Wyniki porządkowania po wprowadzeniu obserwacji nietypowej (TMR)

| Pozycja | TMR ($N = 7$) | TMR ($N = 8$) |
|---------|-----------------|-----------------|
| 1 | | <i>Warszawa</i> |
| 2 | Gdańsk | Wrocław |
| 3 | Katowice | Gdańsk |
| 4 | Wrocław | Kraków |
| 5 | Kraków | Katowice |
| 6 | Poznań | Poznań |
| 7 | Toruń | Toruń |
| 8 | Szczecin | Szczecin |

Źródło: obliczenia własne na podstawie tab. 1.

Tabela 4. Wyniki porządkowania po wprowadzeniu obserwacji nietypowej (ELECTRE III)

| Pozycja | ELECTRE III ($N = 7$) | ELECTRE III ($N = 8$) |
|---------|-------------------------|-------------------------|
| 1 | | <i>Warszawa</i> |
| 2 | Wrocław | Wrocław |
| 3 | Kraków | Kraków |
| 4 | Katowice | Katowice |
| 5 | Gdańsk | Gdańsk |
| 6 | Poznań | Poznań |
| 7 | Toruń | Toruń |
| 8 | Szczecin | Szczecin |

Źródło: obliczenia własne na podstawie tab. 1.

Widoczne jest w tym przypadku uzyskiwanie identycznych pozycji wyłącznie przez obiekty charakteryzujące się niskimi wartościami cech diagnostycznych. Warto jednocześnie zauważyć stosunkowo niską pozycję Poznania. Zostało to spowodowane występowaniem w przypadku tego miasta najwyższego ujemnego salda migracji⁶.

W kolejnym kroku wprowadzona zostaje obserwacja nietypowa (miasto Warszawa) charakteryzująca się w większości przypadków znacznie wyższymi wartościami rozważanych cech diagnostycznych. Wyniki porządkowania przedstawiają dla metody TMR i metody ELECTRE III odpowiednio tab. 3 oraz 4.

⁶ Warto byłoby również rozważyć wartość tej zmiennej dla Poznania jako specjalną odmianę obserwacji nietypowej, lecz wykracza to poza ramy niniejszego badania.

W przypadku metody TMR obserwujemy znaczące różnice w pozycjach zajmowanych przez cztery miasta (Gdańsk, Katowice, Wrocław i Kraków) w obu wariantach. Zmianie nie ulegają tylko pozycje trzech ostatnich miast (Poznań, Toruń i Szczecin). Natomiast metoda ELECTRE III wydaje się mniej wrażliwa na wprowadzenie obiektu nietypowego w postaci miasta Warszawy. Pozycje wszystkich poprzednio klasyfikowanych miast nie uległy zmianie.

Ocena wrażliwości metody ELECTRE III na zmiany wartości progowych wykonana została przy założeniu zmniejszenia oraz zwiększenia wszystkich tych wartości o 20% w porównaniu do ich wartości wyjściowych. Rezultaty przeprowadzonej analizy wrażliwości zaprezentowane zostały w tab. 5 (dla wariantu podstawowego) oraz w tab. 6 (z dodaną obserwacją nietypową).

Tabela 5. Wyniki porządkowania przy zmianie wartości progowych (wariant podstawowy)

| Pozycja | Wariant podstawowy ($N = 7$) | Progi < 20% | Progi > 20% |
|---------|--------------------------------|-----------------|-----------------|
| 1 | Wrocław | Wrocław | Wrocław |
| 2 | Kraków | Katowice | Kraków |
| 3 | Katowice | Kraków, Gdańsk | Gdańsk |
| 4 | Gdańsk | | Katowice |
| 5 | Poznań | Poznań | Toruń |
| 6 | Toruń | Toruń | Poznań |
| 7 | Szczecin | Szczecin | Szczecin |

Źródło: obliczenia własne.

Tabela 6. Wyniki porządkowania przy zmianie wartości progowych (z obserwacją nietypową)

| Pozycja | Wariant podstawowy ($N = 8$) | Progi < 20% | Progi > 20% |
|---------|--------------------------------|-----------------|------------------|
| 1 | Warszawa | Warszawa | Warszawa |
| 2 | Wrocław | Wrocław | Wrocław |
| 3 | Kraków | Katowice | Kraków |
| 4 | Katowice | Kraków, Gdańsk | Katowice, Gdańsk |
| 5 | Gdańsk | | |
| 6 | Poznań | Poznań | Toruń |
| 7 | Toruń | Toruń | Poznań |
| 8 | Szczecin | Szczecin | Szczecin |

Źródło: obliczenia własne.

Uzyskane wyniki pozwalają stwierdzić brak wpływu zmiany wartości progowych na pozycje uzyskiwane przez obiekty charakteryzujące się zdecydowanie wyższymi oraz niższymi wartościami zmiennych (Warszawa, Wrocław, Szczecin). Natomiast pozostałe miasta odznaczają się pewną wrażliwością na zmianę warto-

ści progowych. Przy czym nie ma w tym przypadku znaczenia, czy wartości te są zmniejszane czy zwiększane.

4. Wnioski

Przeprowadzone analizy wykazały, że wyniki porządkowania obiektów dla obu wykorzystywanych metod pomimo braku obserwacji nietypowych są zbieżne wyłącznie w zakresie obiektów charakteryzujących się niskimi wartościami cech diagnostycznych. Metoda ELECTRE III charakteryzowała się w porównaniu do metody TMR, w przypadku rozważanego przykładu, mniejszą wrażliwością na wprowadzenie obserwacji nietypowej. Warto jednak zauważyć, że przeprowadzone porównanie ograniczało się do stosunkowo mało liczego zbioru obiektów, co spowodowane było wysoką pracochłonnością obliczeń. W przypadku zmiany wartości progowych g_j , p_j oraz v_j obserwowana była niezmiennosc zajmowanych pozycji wyłącznie w zakresie obiektów odznaczających się znacznie wyższymi i znacznie niższymi wartościami cech diagnostycznych w porównaniu do pozostałych obiektów. Warto jednocześnie wskazać, co było widoczne w uzyskiwanych wynikach, na ograniczenie metody ELECTRE III, polegające na trudności w ocenie, który obiekt jest lepszy w przypadku zajmowania tej samej pozycji przez co najmniej dwa obiekty. Kolejne zagadnienia, z którymi musi się liczyć użytkownik tej metody, to brak możliwości określenia „rzeczywistej” odległości między poszczególnymi obiektami oraz subiektywny wybór wartości progowych. Dalsze możliwości badawcze sprowadzić się mogą do porównania wrażliwości metody ELECTRE III na występowanie obserwacji nietypowych z innymi metodami porządkowania obiektów, oceny tej wrażliwości przy wykorzystaniu liczniejszych zbiorów danych oraz oceny wrażliwości uzyskiwanych wyników na zmianę wag.

Literatura

- Barnett V., Lewis T., *Outliers in Statistical Data*, John Wiley, 1994.
- Breunig M.M., Kriegel H-P., Ng R.T., Sander J., *LOF: Identifying Density-Based Local Outliers*, Proceedings ACM SIGMOD 2000 International Conference on Management of Data, Dallas, TX, 2000.
- Chen C., Liu L.M., *Joint estimation of model parameters and outlier effects in time series*, “Journal of the American Statistical Association”, 88, 1993, s. 284-297.
- Davidson R., MacKinnon J.G., *Econometric Theory and Methods*, Oxford University Press, New York 2004.
- Figueira J., Mousseau V., Roy B., *ELECTRE methods*, [w:] J. Figueira, S. Greco, M. Ehrgott (red.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Verlag, Boston, Dordrecht, London 2005, s. 133-162.
- Fox A.J., *Outliers in time series*, “Journal of Royal Statistical Society”, Series B, 1972, 34, s. 350-363.
- Hawkins D. *Identification of Outliers*, Chapman and Hall, 1980.

- Hellwig Z., *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, „Przegląd Statystyczny” 1968, 4, s. 307-327.
- Kufel T., *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości występowania*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2010.
- Lu C., Chen D., Kou Y., *Algorithms for spatial outlier detection*, Proceedings of the 3rd IEEE International Conference on Data-mining (ICDM'03), Melbourne, FL 2003.
- Motulsky H.J., Brown R.E., *Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate*, “Bioinformatics” 2006, 7.
- Pyle D., *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, San Francisco 1999.
- Roy B., *The outranking approach and the foundations of ELECTRE methods*, “Theory and Decision” 1991, 31, s. 49-73.
- Shekhar S., Lu C.T., Zhang P., *A unified approach to detection spatial outliers*, “Geoinformatica” 2003, s. 139-166.
- Tsay R.S., *Outliers, level shifts, and variance changes in time series*, “Journal of Forecasting” 1988, 7, 1-20.
- Williams G.J., Baxter R.A., He H.X., Hawkins S., Gu L., *A Comparative Study of RNN for Outlier Detection in Data Mining*, IEEE International Conference on Data-mining (ICDM'02), Maebashi City, Japan, CSIRO Technical Report CMIS-02/102, 2002.
- Zarządzanie ryzykiem*, red. K. Jajuga, Wydawnictwo Naukowe PWN, Warszawa 2007.

SENSITIVITY ANALYSIS OF ELECTRE III METHOD FOR OUTLIERS AND CHANGE OF THRESHOLDS

Summary: The paper carries out the sensitivity analysis for the introduction of outliers of two ranking methods Taxonomic Measure of Development and ELECTRE III. The received results show that, in the case of considered objects, only the latest method is robust. Using the same data significant sensitivity of ELECTRE III method for thresholds change was proved.

Keywords: classification, ranking methods, outliers.