

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

Taksonomia 22

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

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

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

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

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 2014

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Eugeniusz Gatnar , Balance of payments statistics and external competitiveness of Poland.....	15
Andrzej Sokolowski, Magdalena Czaja , Efektywność metody k -średnich w zależności od separowalności grup.....	23
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw	30
Elżbieta Gołata , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
Aleksandra Łuczak, Feliks Wysocki , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów	49
Marek Walesiak , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej	60
Paweł Lula , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i>	69
Mariusz Kubus , Propozycja modyfikacji metody złagodzonego LASSO.....	77
Andrzej Bąk, Tomasz Bartłomowicz , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
Justyna Brzezińska , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki	104
Barbara Batóg, Jacek Batóg , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010	113
Małgorzata Markowska, Danuta Strahl , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
Kamila Migdał-Najman, Krzysztof Najman , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	131
Kamila Migdał-Najman, Krzysztof Najman , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	139
Beata Basiura, Anna Czapkiewicz , Badanie jakości klasyfikacji szeregów czasowych	148
Michał Trzęsiok , Wybrane metody identyfikacji obserwacji oddalonych.....	157

Grażyna Dehnel, Tomasz Klimanek , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
Michał Bernard Pietrzak, Justyna Wilk , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
Maciej Beręsewicz , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena	186
Marcin Szymkowiak, Tomasz Józefowski , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji	195
Marcin Pelka , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym	202
Małgorzata Machowska-Szewczyk , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
Justyna Wilk , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
Andrzej Dudek , Metody analizy skupień w klasyfikacji markerów map Google	229
Ewa Roszkowska , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW	237
Marcin Szymkowiak, Marek Witkowski , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
Bartłomiej Jefmański , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
Karolina Bartos , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych	266
Joanna Trzęsiok , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych	275
Beata Bal-Domańska , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
Beata Bieszk-Stolorz, Iwona Markowicz , Wpływ zasiłku na proces poszukiwania pracy	294
Marta Dziechciarz-Duda, Klaudia Przybysz , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
Tomasz Klimanek , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Wybrane metody analizy danych wzdluznych.....	321
Artur Zaborski , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych	330
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

Katarzyna Wawrzyniak , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego	346
---	-----

Summaries

Eugeniusz Gatnar , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski	22
Andrzej Sokółowski, Magdalena Czaja , Cluster separability and the effectiveness of k -means method	29
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
Elżbieta Golata , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011	48
Aleksandra Łuczak, Feliks Wysocki , Determination of weights for features in problems of linear ordering of objects	59
Marek Walesiak , Reinforcing measurement scale for ordinal data in multivariate statistical analysis	68
Paweł Lula , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
Mariusz Kubus , The proposition of modification of the relaxed LASSO method.....	84
Andrzej Bąk, Tomasz Bartłomowicz , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
Justyna Brzezińska , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models	103
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
Barbara Batóg, Jacek Batóg , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity	120
Małgorzata Markowska, Danuta Strahl , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
Kamila Migdał Najman, Krzysztof Najman , Formal quality assessment of group structure mapping on the Kohonen's map	138
Kamila Migdał Najman, Krzysztof Najman , Graphical quality assessment of group structure mapping on the Kohonen's map	147
Beata Basiura, Anna Czapkiewicz , Validation of time series clustering	156
Michał Trzęsiok , Selected methods for outlier detection.....	166

Grażyna Dehnel, Tomasz Klimanek , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics	176
Michał Bernard Pietrzak, Justyna Wilk , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
Maciej Beręsewicz , An attempt to use different distance measures in the Generalized Petersen estimator	194
Marcin Szymkowiak, Tomasz Józefowski , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
Marcin Pelka , The ensemble conceptual clustering for symbolic data.....	209
Małgorzata Machowska-Szewczyk , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
Justyna Wilk , Problem of determining the number of clusters in taxonomic analysis of symbolic data	228
Andrzej Dudek , Clustering techniques for Google maps markers.....	236
Ewa Roszkowska , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure	247
Marcin Szymkowiak, Marek Witkowski , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
Bartłomiej Jefmański , The construction of fuzzy customer satisfaction indexes using R program.....	265
Karolina Bartos , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
Joanna Trzęsiok , Cluster analysis of countries with respect to fertility rate and other demographic factors	284
Beata Bal-Domańska , An attempt to identify major regional clusters and their convergence	293
Beata Bieszk-Stolorz, Iwona Markowicz , The influence of benefit on the job finding process	302
Marta Dziechciarz-Duda, Klaudia Przybysz , Education and labor market needs. Classification of university graduates	312
Tomasz Klimanek , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Selected methods for an analysis of longitudinal data.....	329
Artur Zaborski , The application of distance measures for ordinal data for aggregation individual preferences	337
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market	345
Katarzyna Wawrzyniak , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows	355

Aleksandra Łuczak, Feliks Wysocki

Uniwersytet Przyrodniczy w Poznaniu

USTALANIE SYSTEMU WAG DLA CECH W ZAGADNIENIACH PORZĄDKOWANIA LINIOWEGO OBIEKTÓW

Streszczenie: Celem pracy jest przedstawienie propozycji ustalania systemu współczynników wagowych dla cech. W pracy przedstawiono trzy podejścia. Pierwsze podejście polega na wykorzystaniu procedur statystycznych (podejście obiektywne). Drugie – merytoryczne – oparte jest na opiniach ekspertów (podejście subiektywne). Trzecie – zintegrowane – łączy podejście subiektywne i obiektywne. Proponowane zintegrowane podejście zostało zastosowane do badania poziomu rozwoju społeczno-gospodarczego województw w Polsce w 2011 roku przy wykorzystaniu metody TOPSIS.

Słowa kluczowe: wagi cech, metoda CCSD, analityczny proces hierarchiczny (AHP), metoda TOPSIS.

1. Wstęp

Poglądy dotyczące ustalania systemu współczynników wagowych dla cech w literaturze przedmiotu nie są w pełni wykrystalizowane [zob. np. Hellwig 1981; Grabiński 1988; Wysocki 2010]. W przypadku rozpoznawania typów struktury złożonej, gdy kryteria¹ i cechy nie są jednakowo ważne, można wykorzystać dwa podejścia do ustalania tego systemu. Pierwsze podejście polega na wykorzystaniu procedur statystycznych (podejście obiektywne). Drugie – merytoryczne – oparte jest na opiniach ekspertów (podejście subiektywne) [Ma i in. 1999; Kukuła 2000; Wysocki 2010]. W pracy w podejściu statystycznym do ustalania systemu wag dla cech została wykorzystana metoda CCSD, która uwzględnia odchylenia standardowe z wartości cech (*SD* – *standard deviation*) i stopień skorelowania znormalizowanych wartości danej cechy z wartościami miernika syntetycznego obliczonymi z wyłączeniem tej cechy (*CC* – *correlation coefficient*) [Wang, Luo 2010]. Natomiast

¹ Kryterium oceny jest to najczęściej niemierzalna bezpośrednio właściwość obiektów badanej zbiorowości, według której mamy zamiar je uporządkować od najlepszego do najgorszego [Pociecha i in. 1988; Wysocki 2010].

w podejściu merytorycznym wykorzystano metodę analitycznego procesu hierarchicznego (AHP – *Analytic Hierarchy Process*), która pozwala zagregować oceny ważności cech dokonane przez ekspertów [Saaty 1980]. W pracy zaproponowano zintegrowane podejście wykorzystujące procedury statystyczne i opinie ekspertów. Zostało ono zastosowane do badania poziomu rozwoju społeczno-gospodarczego województw w Polsce w 2011 roku przy wykorzystaniu metody TOPSIS (*Technique for Order of Preference by Similarity to Ideal Solution*) [Hwang, Yoon 1981; Wysocki 2010].

2. Metodyka badań

W procesie tworzenia cechy syntetycznej metodą TOPSIS można wyróżnić sześć podstawowych etapów postępowania (zob. tab. 1).

Pierwszym etapem jest wybór cech opisujących wybrane obiekty. Istnieją dwa podstawowe podejścia do wyboru cech: statystyczne i merytoryczne [zob. Wysocki 2010]. Następnie należy ustalić kierunek preferencji cech prostych w stosunku do rozpatrywanego kryterium ogólnego.

Tabela 1. Etapy konstrukcji cechy syntetycznej metodą TOPSIS

Etapy postępowania	Opis etapów / metody
I. Wybór cech	Weryfikacja cech pod względem merytorycznym i/lub statystycznym
II. Normalizacja	Unitaryzacja zerowana
III. Ustalenie systemu wag	Metoda CCSD lub/i AHP
IV. Obliczenie odległości każdego obiektu od wzorca i antywzorca rozwoju	Obliczenie oddalenia każdego ocenianego obiektu wielocechowego od wzorca i antywzorca rozwoju
V. Obliczenie wartości syntetycznego miernika rozwoju	Obliczenie wartości cechy syntetycznej (syntetycznego miernika rozwoju) za pomocą metody TOPSIS
VI. Uporządkowanie liniowe obiektów i identyfikacja typów rozwojowych	Wyodrębnienie klas typologicznych dla całego obszaru zmienności cechy syntetycznej metodami statystycznymi lub w sposób arbitralny

Źródło: opracowanie własne na podstawie [Wysocki 2010].

Drugi etap polega na normalizacji cech. Jej celem jest uwolnienie cech od miar lub ich ujednoczenie pod względem zakresów liczbowych. Istnieją rozmaite procedury normalizacji cech. W badaniach wykorzystano metodę unitaryzacyjną, noszącą nazwę unitaryzacji zerowanej [zob. np. Kukuła 2000; Wysocki 2010].

Ustalenie systemu wag dla cech prostych może być przeprowadzone na podstawie analizy statystycznej lub merytorycznej (etap III). Pierwsze podejście wykorzystuje procedury statystyczne, bazując na informacjach o cechach tkwiących tylko w samej macierzy danych, a w szczególności wykorzystuje analizę zmienności cech i korelacji między cechami albo tylko jedną z tych analiz [Wysocki 2010]. Drugie podejście opiera się na opiniach ekspertów i AHP.

W pracy w podejściu statystycznym do ustalania sytemu wag dla cech proponuje się wykorzystanie metody CCSD [Wang, Luo 2010]. Wychodząc od jednego z najprostszych mierników syntetycznych – średniej ważonej: $d_i = \sum_{k=1}^K z_{ik} \cdot w_k$,

gdzie: z_{ik} jest znormalizowaną wartością k -tej cechy w i -tym obiekcie ($z_{ik} \in \langle 0,1 \rangle$), w_k jest wagą k -tej cechy (wstępnie można przyjąć, że $w_k = 1/K$), należy zauważyć, że d_i jest funkcją liniową wag cech. A zatem im większa jest wartość d_i , tym lepsza jest ocena badanego obiektu. W tym miejscu należy rozważyć, jaki ma związek k -ta cecha z wartościami syntetycznego miernika utworzonego z pozostałych cech. Można to przeanalizować, usuwając ze zbioru badanych cech k -tą cechę. W takim przypadku wartości miernika syntetycznego ulegną zmianie:

$d_{ik} = \sum_{j=1, j \neq k}^K z_{ij} \cdot w_j$. Związek usuniętej cechy z wartościami skorygowanego miernika (pozostałe cechy) można ocenić, wykorzystując współczynnik korelacji:

$$R_k = \frac{\sum_{i=1}^N (z_{ik} - \bar{z}_k) \cdot (d_{ik} - \bar{d}_k)}{\sqrt{\sum_{i=1}^N (z_{ik} - \bar{z}_k)^2 \cdot \sum_{i=1}^N (d_{ik} - \bar{d}_k)^2}}, \quad -1 \leq R_k \leq 1,$$

gdzie: $\bar{z}_k = \frac{1}{N} \sum_{i=1}^N z_{ik}$, $\bar{d}_k = \frac{1}{N} \sum_{i=1}^N d_{ik} = \sum_{j=1, j \neq k}^K \bar{z}_j \cdot w_j$.

Jeśli R_k ma wartość bliską jeden, wtedy usunięcie danej (k -tej) cechy będzie miało niewielki wpływ na wartości miernika syntetycznego utworzonego z pozostałych cech, co wiąże się z przypisaniem k -tej cesze odpowiednio małej wagi. Jeśli R_k ma małą wartość, w skrajnym przypadku bliską minus jeden, dodanie k -tej cechy do zestawu pozostałych cech będzie miało istotny wpływ na wartości syntetycznego miernika rozwoju i w konsekwencji na ranking obiektów. W tym przypadku k -ta cecha powinna uzyskać odpowiednio większą wagę. Ponadto im większe jest odchylenie standardowe (SD) danej cechy, tym większą wagę powinno się jej przypisać [Wang, Luo 2010].

Na podstawie powyższych rozważań można zdefiniować współczynniki wagowe k -tej cechy jako [Wang, Luo 2010]:

$$w_k = s_k \sqrt{1 - R_k} / \sum_{j=1}^K s_j \sqrt{1 - R_j}, \quad (k = 1, \dots, K),$$

gdzie: $s_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{ik} - \bar{z}_k)^2}$ jest odchyleniem standardowym z wartości k -tej cechy.

Powyższa formuła przedstawia układ K -równań nieliniowych, którego rozwiązanie pozwala jednoznacznie określić K zróżnicowanych współczynników wagowych dla cech. W celu jego rozwiązania można go przekształcić w następujące zagadnienie programowania nieliniowego ($J, w_k, k = 1, \dots, K$) [Wang, Luo 2010]:

$$J = \sum_{k=1}^K \left(w_k - s_k \sqrt{1 - R_k} / \sum_{j=1}^K s_j \sqrt{1 - R_j} \right)^2 = \text{minimum}$$

przy ograniczeniach $\sum_{k=1}^K w_k = 1, w_k \geq 0, k = 1, \dots, K$.

W podejściu merytorycznym wykorzystano metodę analitycznego procesu hierarchicznego (AHP), która pozwala zagregować oceny ważności cech dokonane przez ekspertów [Saaty 1980]. W AHP dokonuje się porównań parami ważności kryteriów pośrednich w odniesieniu do kryterium głównego oraz porównań parami cech w obrębie każdego kryterium. Wyniki porównań zestawia się w macierze porównań. W przypadku porównań cech mamy [Saaty 1980; Łuczak, Wysocki 2005]:

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \dots & a_{1K} \\ 1/a_{12} & 1 & \dots & a_{2K} \\ \dots & \dots & \dots & \dots \\ 1/a_{1K} & 1/a_{2K} & \dots & 1 \end{bmatrix},$$

gdzie: a_{kl} – ważność cechy (kryterium) k -tej w stosunku do cechy (kryterium) l -tej uzyskana z porównań parami według dziewięciostopniowej skali Saaty'ego [1980].

Macierz \mathbf{A} tworzona jest według następujących zasad:

- przechodniości ocen: jeżeli ważność k -tej cechy nad l -tą jest $a_{kl} = \alpha$, wtedy $a_{lk} = 1/\alpha, \alpha \neq 0$,
- równoważności ocen: jeżeli k -ta cecha jest równie relatywnie ważna jak l -ta, wtedy $a_{kl} = a_{lk} = 1$.

Oblicza się składowe wektora własnego rozpatrywanych cech (kryteriów) $\mathbf{w} = [w_1, w_2, \dots, w_K]^T$ z macierzy porównań parami \mathbf{A} [Saaty 1980; Łuczak, Wysocki 2005]. Zalecany sposób polega na wyznaczeniu średniej geometrycznej z ocen porównań parami w każdym wierszu macierzy \mathbf{A} . Otrzymane liczby normalizuje się do jedności poprzez podzielenie każdej z nich przez ich sumę. Wartości znormalizowane tzw. priorytety lokalne wyrażają udział danej cechy (kryterium) w kryterium na poziomie bezpośrednio wyższym i mogą być wykorzystane do obliczenia priorytetów globalnych, które reprezentują udział każdej cechy (kryterium) w kryterium głównym. Priorytet globalny uzyskuje się przez przemnożenie wartości priorytetu lokalnego tego poziomu przez wartość priorytetu globalnego poziomu bezpośrednio wyższego [Saaty 1980; Łuczak, Wysocki 2005].

Wartości znormalizowane cech zostają przemnożone przez uzyskane współczynniki wagiowe:

a) $z_{ik(s)} = z_{ik} \cdot w_{k(s)}$, gdzie: $w_{k(s)}$ oznaczają współczynniki wagowe otrzymane za pomocą procedur statystycznych (CCSD),

b) $z_{ik(m)} = z_{ik} \cdot w_{k(m)}$, gdzie: $w_{k(m)}$ oznaczają współczynniki wagowe bazujące na przesłankach merytorycznych (AHP).

Wprowadzenie jednego systemu wag może następować po etapie normalizacji cech. W przypadku konieczności wprowadzenia dwóch systemów wag przy obliczaniu wartości cechy syntetycznej można system wag ustalonych za pomocą procedur statystycznych wprowadzić po etapie normalizacji cech, a system wag ustalonych na podstawie przesłanek merytorycznych w etapie obliczania odległości od wzorca i antywzorca rozwoju. Nie jest wskazane wprowadzenie obu systemów łącznie po etapie normalizacji cech. Z założenia wagi cech sumują się do jedności, a przemnożenie dwóch systemów wag cech przez siebie nie pozwoli na spełnienie tego warunku.

W etapie IV ustalone zostają współrzędne obiektów modelowych – wzorca, np. według zasady [Wysocki 2010]:

$$A^+ = \left(\max_i (z_{i1(\bullet)}) , \max_i (z_{i2(\bullet)}) , \dots , \max_i (z_{iK(\bullet)}) \right) = (z_1^+ , z_2^+ , \dots , z_K^+)$$

i antywzorca rozwoju:

$$A^- = \left(\min_i (z_{i1(\bullet)}) , \min_i (z_{i2(\bullet)}) , \dots , \min_i (z_{iK(\bullet)}) \right) = (z_1^- , z_2^- , \dots , z_K^-),$$

gdzie (\bullet) oznacza (s) lub (m) .

Po ustaleniu wzorca i antywzorca rozwoju oblicza się odległości euklidesowe każdej ocenianej jednostki od wzorca z^+ i antywzorca rozwoju z^- :

(1) jeden system wag

$$d_i^+ = \sqrt{\sum_{k=1}^K (z_{ik(\bullet)} - z_k^+)^2} \quad \text{oraz} \quad d_i^- = \sqrt{\sum_{k=1}^K (z_{ik(\bullet)} - z_k^-)^2},$$

(2) dwa systemy wag

$$d_i^+ = \sqrt{\sum_{k=1}^K w_{k(m)} (z_{ik(s)} - z_k^+)^2} \quad \text{oraz} \quad d_i^- = \sqrt{\sum_{k=1}^K w_{k(m)} (z_{ik(s)} - z_k^-)^2}.$$

Do konstrukcji miernika syntetycznego zastosowano metodę TOPSIS za pomocą miernika (etap V) [Hwang, Yoon 1981; Wysocki 2010]:

$$S_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad (i = 1, 2, \dots, N),$$

przy czym $0 \leq S_i \leq 1$.

Im mniejsza jest odległość danego obiektu od obiektu modelowego – wzorca rozwoju, a tym samym większa od drugiego bieguna – antywzorca rozwoju, tym wartość miernika syntetycznego jest bliższa 1.

Wyznaczone wartości syntetycznego miernika S_i wykorzystuje się do liniowego porządkowania obiektów oraz – na tej podstawie – do wyodrębnienia ich klas typologicznych (etap VI). Wyodrębnienie klas dla całego obszaru zmienności cechy syntetycznej może zostać przeprowadzone metodami statystycznymi z zastosowaniem średniej i odchylenia standardowego z wartości cechy syntetycznej lub w sposób arbitralny [Wysocki 2010].

3. Wyniki badań

W badaniach poziomu rozwoju społeczno-gospodarczego wykorzystano dane statystyczne z Głównego Urzędu Statystycznego [Bank Danych Lokalnych 2011]. W pierwszym etapie dokonano wyboru 10 cech² opisujących poziom rozwoju według województw w Polsce: saldo migracji wewnętrznych i zagranicznych na 1000 ludności (X_1), pracujący w przemyśle i budownictwie w procentach ogółu pracujących (X_2), stopa bezrobocia rejestrowanego (%) (X_3), przeciętne miesięczne wynagrodzenia brutto w złotych (X_4), nakłady na B+R na 1 mieszkańca w złotych (X_5), podmioty gospodarcze na 100 osób w wieku produkcyjnym (X_6), wskaźniki zagrożenia ubóstwem według granic ubóstwa – odsetek osób w gospodarstwach domowych poniżej relatywnej granicy ubóstwa (%) (X_7), korzystający z instalacji kanalizacyjnej w procentach ogółu ludności (X_8), zużycie NPK na 1 ha użytków rolnych w kilogramach (X_9), dochody własne gmin w dochodach ogółem (%) (X_{10}).

Wyboru cech dokonano na podstawie analizy merytorycznej i statystycznej. W analizie statystycznej wykorzystano macierz odwrotną do macierzy korelacji pomiędzy przyjętymi cechami [Malina, Zeliaś 1997]. Elementy na głównej przekątnej odwrotnej macierzy do macierzy korelacji R (współczynników korelacji liniowej między cechami prostymi X_1, \dots, X_{10}) nie były większe niż 20. Wybrane cechy diagnostyczne mogą być zatem nośnikami istotnych informacji dotyczących poziomu rozwoju społeczno-gospodarczego.

W drugim etapie przyjęto, że dwie cechy mają charakter destymulant (X_2 i X_6), a pozostałe – stymulant (etap II). Cechy zostały znormalizowane za pomocą metody unitaryzacji zerowanej. Stanowiły one podstawę do wyznaczenia współczynników wagowych cech metodą CCSD (tab. 2, 3). Należy zauważyć, że największa wartość $CC_{opt}(R_{k(opt)})$ została osiągnięta dla cechy X_{10} , a najmniejsza dla cechy X_9 , co przekłada się na najmniejszą wartość wagi dla cechy X_{10} i największą – dla X_9 (tab. 3).

² Ze względu na bardzo silne skorelowanie wartości cechy dotyczącej PKB na 1 mieszkańca w zł z wartościami pozostałych cech została ona usunięta ze zbioru cech opisujących poziom rozwoju społeczno-gospodarczego według województw w Polsce.

Tabela 2. Wartości SD i CC dla analizowanych cech

Wskaźniki/ Cechy	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
SD (s_k) dla wartości cech znormalizowanych	0,270	0,266	0,266	0,245	0,228	0,284	0,316	0,274	0,260	0,294
$CC_1 (R_{k(1)})^1$	0,663	0,275	0,485	0,705	0,483	0,825	0,661	0,454	0,243	0,959
$CC_{opt}(R_{k(opt)})^2$	0,600	0,293	0,437	0,657	0,403	0,799	0,658	0,445	0,244	0,960

¹⁾ (1) oznacza pierwszą iterację przy założeniu startowych wartości współczynników wagowych $w_k = 0,1$, $k = 1, \dots, K$; ²⁾ (opt) oznacza ostatnią iterację, w której otrzymuje się optymalne wartości współczynników wagowych.

Źródło: opracowanie własne na podstawie danych z Banku Danych Lokalnych [2011].

Tabela 3. Wartości współczynników wagowych dla cech uzyskane metodami CCSD i AHP

Kryteria	Cechy	Wagi uzyskane metodami	
		CCSD	AHP
Sytuacja demograficzna	Saldo migracji wewnętrznych i zagranicznych na 1000 ludności (X_1)	0,100	0,121
Rynek pracy	Pracujący w przemyśle i budownictwie w procentach ogółu pracujących (X_2)	0,130	0,080
	Stopa bezrobocia rejestrowanego (%) (X_3)	0,117	0,080
	Przeciętne miesięczne wynagrodzenia brutto (zł) (X_4)	0,084	0,080
Gospodarka	Nakłady na B+R na 1 mieszkańca (zł) (X_5)	0,103	0,168
	Podmioty gospodarcze na 100 osób w wieku produkcyjnym (X_6)	0,074	0,168
Sytuacja gospodarstw domowych	Wskaźniki zagrożenia ubóstwem według granic ubóstwa – odsetek osób w gospodarstwach domowych poniżej relatywnej granicy ubóstwa (%) (X_7)	0,108	0,157
Infrastruktura	Korzystający z instalacji kanalizacyjnej w procentach ogółu ludności (X_8)	0,119	0,047
Rolnictwo	Zużycie NPK na 1 ha użytków rolnych w kg (X_9)	0,132	0,021
Finanse	Dochody własne gmin w dochodach ogółem (%) (X_{10})	0,034	0,078

Źródło: obliczenia własne na podstawie danych statystycznych pochodzących z Banku Danych Lokalnych [2011], <http://www.stat.gov.pl/bdl>.

Stosując metodę Saaty'ego (AHP), najpierw ustalono współczynniki wagowe w odniesieniu do kryteriów podrzędnych. Wynosiły one kolejno: sytuacja demograficzna – 0,121, rynek pracy – 0,240, gospodarka – 0,336, sytuacja gospodarstw domowych – 0,157, infrastruktura – 0,047, rolnictwo – 0,021, finanse – 0,078 (tab. 3). Następnie dla uproszczenia przyjęto, że współczynniki wagowe cech mają wartości odwrotnie proporcjonalne do liczebności danego zespołu.

Porównując otrzymane wartości wag metodami CCSD i AHP, można zauważyć istotne różnice między nimi (tab. 3). W przypadku metody CCSD większość cech miała zbliżone współczynniki wagowe, poza trzema (X_4 , X_6 , X_{10}). Zwrócić należy uwagę, że największy współczynnik wagowy uzyskała cecha dotycząca zużycia NPK na 1 ha użytków rolnych (0,132), gdyż ma ona najniższy współczynnik korelacji CC_{opt} (0,244), co oznacza niski stopień skorelowania tej cechy z

miernikiem syntetycznym skonstruowanym bez jej udziału. Taki wynik budzi jednak wątpliwości merytoryczne. Na podstawie przesłanek merytorycznych można twierdzić, że waga dla tej cechy związanej z rolnictwem w kształtowaniu poziomu rozwoju społeczno-gospodarczego nie powinna być aż tak wysoka. Potwierdza to współczynnik ważności uzyskany metodą AHP, wynoszący dla tej cechy tylko 0,021. Z kolei cecha *dochody własne gmin w dochodach ogółem* ma najwyższy – bliski 1 – współczynnik korelacji (0,960) (tab. 2), co oznacza, że cecha ta w sensie statystycznym ma niewielki udział w kształtowaniu wartości miernika syntetycznego (0,034). Natomiast największy współczynnik wagowy uzyskany metodą AHP mają dwie cechy związane z gospodarką: *nakłady na B+R na 1 mieszkańca* (zł) (0,168) oraz *podmioty gospodarcze na 100 osób w wieku produkcyjnym* (0,168).

Tabela 4. Wartości syntetycznych mierników oraz rankingi województw w Polsce według poziomu rozwoju społeczno-gospodarczego

Lp.	Województwa ^{a)}	Wartości syntetycznego miernika TOPSIS z wagami				Rangi województw według systemu wag			
		jednakowe	CCSD	AHP	AHP + CCSD	jednakowe	CCSD	AHP	AHP i CCSD
		I	II	III	IV	I	II	III	IV
1	mazowieckie	0,685	0,611	0,808	0,691	1	2	1	1
2	dolnośląskie	0,620	0,621	0,568	0,585	2	1	2	2
3	śląskie	0,579	0,593	0,513	0,573	3	3	5	3
4	pomorskie	0,577	0,582	0,518	0,538	4	4	4	4
5	wielkopolskie	0,547	0,571	0,477	0,515	5	5	6	5
6	małopolskie	0,451	0,438	0,520	0,507	9	11	3	6
7	opolskie	0,496	0,541	0,446	0,488	6	6	9	7
8	łódzkie	0,467	0,480	0,450	0,475	8	8	8	8
9	lubuskie	0,437	0,454	0,432	0,454	10	10	10	9
10	zachodniopomorskie	0,491	0,485	0,472	0,449	7	7	7	10
11	kujawsko-pomorskie	0,411	0,458	0,323	0,384	11	9	11	11
12	podkarpackie	0,248	0,281	0,228	0,280	14	14	12	12
13	warmińsko-mazurskie	0,287	0,332	0,186	0,271	12	12	13	13
14	podlaskie	0,265	0,291	0,174	0,235	13	13	14	14
15	lubelskie	0,224	0,252	0,173	0,221	15	15	15	15
16	świętokrzyskie	0,195	0,214	0,164	0,197	16	16	16	16

a) Uporządkowanie liniowe województw według wartości syntetycznego miernika rozwoju używanego metodą IV.

Źródło: obliczenia własne na podstawie danych statystycznych Głównego Urzędu Statystycznego z Banku Danych Lokalnych [2011], <http://www.stat.gov.pl/bdl>.

Następnie obliczono wartości syntetycznego miernika rozwoju metodą TOPSIS (tab. 4). Uzyskane rankingi pokazują wysoki stopień zgodności w uporządkowaniach województw. Należy jednak zwrócić uwagę na rangi uzyskane przez woje-

wództwo małopolskie. W podejściu I i II województwo to uzyskało znacznie gorszą pozycję w rankingu (odpowiednio 9. i 11. miejsce), aniżeli w podejściu III i IV (3. i 6. miejsce). Polepszenie przeciętnej sytuacji województwa małopolskiego pod względem poziomu rozwoju społeczno-gospodarczego w podejściu IV, a zwłaszcza III, wynika ze znacznego zróżnicowania wag dla cechy X_9 (*zużycie NPK na 1 ha użytków rolnych w kg*) uzyskanych metodami CCSD ($w_{10(s)} = 0,132$) i AHP ($w_{10(m)} = 0,021$). Poziom nawożenia mineralnego w województwie małopolskim był w 2011 roku najniższy w kraju i wynosił tylko 55,8 kg/ha UR, podczas gdy przeciętnie w Polsce był ponad dwukrotnie wyższy – 126,6 kg/ha UR, a w województwie opolskim nawet ponad trzykrotnie wyższy – 181,9 kg/ha UR. W związku z tym przypisanie wysokiej wagi tej cesze (0,132) mogło spowodować zmniejszenie rangi województwa małopolskiego, a z kolei jej istotne obniżenie do 0,021 przyczyniło się do znacznie wyższej oceny tego województwa ze względu na poziom rozwoju społeczno-gospodarczego. Wydaje się, że ocena województwa oparta na współczynnikach wagowych uzyskanych metodą AHP jest bardziej poprawna merytorycznie, gdyż dla zapewnienia możliwie najwyższego poziomu rozwoju społeczno-gospodarczego cechy związane z gospodarką, takie jak podmioty gospodarcze (X_6), nakłady na B+R (X_5) są znacznie ważniejsze aniżeli cechy związane z rolnictwem (poziom nawożenia mineralnego – X_9).

4. Podsumowanie

Na podstawie przeprowadzonych obliczeń i analiz można sformułować następujące stwierdzenia i wnioski.

Metody służące do ustalenia systemu wag można ogólnie podzielić na trzy grupy: obiektywne, subiektywne i zintegrowane. Obiektywne metody określają wagi cech za pomocą obiektywnych informacji istniejących w macierzy danych, w tym odchylenia standardowe ze znormalizowanych wartości cech i współczynniki korelacji obliczone dla wartości danej cechy z wartościami miernika syntetycznego wyznaczonymi z wyłączeniem tej cechy (metoda CCSD). Subiektywne metody pozwalają na określenie wag ważności cech przez ekspertów (metoda AHP). W pracy zaproponowano podejście zintegrowane łączące kryteria obiektywne i subiektywne, polegające na wykorzystaniu dwóch systemów wag.

W podejściu z zastosowaniem wag ustalonych metodą CCSD największe znaczenie uzyskała cecha związana z rolnictwem (*zużycie NPK na 1 ha użytków rolnych*), gdyż była ona w najniższym stopniu skorelowana z wartościami miernika syntetycznego skonstruowanego po wyeliminowaniu właśnie tej cechy. Z kolei cecha *dochody własne gmin w dochodach ogółem* miała najwyższy – bliski 1 – współczynnik korelacji, a tym samym miała ona najmniejsze znaczenie w kształtowaniu wartości miernika syntetycznego.

Natomiast w przypadku metody AHP największy współczynnik wagowy mają dwie cechy związane z gospodarką: *nakłady na B+R na 1 mieszkańca (zł)* oraz *podmioty gospodarcze na 100 osób w wieku produkcyjnym*, natomiast cecha *zużycie NPK na 1 ha użytków rolnych* uzyskała najmniejszą wagę. Ważności cech ustalone subiektywnie na podstawie przesłanek merytorycznych – metodą AHP, wydają się poprawniejsze aniżeli wartości współczynników wagowych ustalone obiektywnie – metodą CCSD.

Zastosowanie wag jednakowych (podejście I) lub ustalonych statystycznie (podejście II) może prowadzić do gorszych – z punktu widzenia merytorycznego – rankingów. Oparcie się na subiektywnych opiniach w procedurze ustalania wag i konstrukcji cechy syntetycznej (podejście III i IV) prowadzi do bardziej merytorycznie uzasadnionych uporządkowań obiektów. Można uznać, że rankingi uzyskane w podejściu III i IV lepiej odzwierciedliły różnice między województwami ze względu na poziom rozwoju społeczno-gospodarczego.

Literatura

- Bank Danych Lokalnych (2011), <http://www.stat.gov.pl/bdl>.
- Grabiński T. (1988), *Metody statystycznej analizy porównawczej*, [w:] *Metody statystyki międzynarodowej*, red. A. Zeliaś, PWE, Warszawa, s. 235-259.
- Hellwig Z. (1981), *Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych*, [w:] *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną*, red. W. Welfe, PWE, Warszawa.
- Hwang C.L., Yoon K. (1981), *Multiple attribute decision making: Methods and applications*, Springer, Berlin.
- Kukuła K. (2000), *Metoda unitaryzacji zerowanej*, Wydawnictwo Naukowe PWN, Warszawa.
- Łuczak A., Wysocki F. (2005), *Wykorzystanie metod taksonometrycznych i analitycznego procesu hierarchicznego do programowania rozwoju obszarów wiejskich*, Wydawnictwo AR w Poznaniu, Poznań.
- Ma J., Fan Z.-P., Huang L.-H. (1999), *A subjective and objective integrated approach to determine attribute weight*, „European Journal of Operational Research”, vol. 112 (2), s. 397-404.
- Malina A., Zeliaś A. (1997), *Taksonomiczna analiza przestrzennego zróżnicowania jakości życia ludności w Polsce w 1994 r.* „Przegląd Statystyczny”, z. 1, t. 44, s. 11-27.
- Pociecha J., Podolec B., Sokołowski A., Zajac K. (1988), *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa.
- Saaty T.L. (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York.
- Wang Y.-M., Luo Y. (2010), *Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making*, „Mathematical and Computer Modeling”, vol. 51, s. 1-12.
- Wysocki F. (2010), *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wyd. Uniwersytetu Przyrodniczego w Poznaniu, Poznań.

DETERMINATION OF WEIGHTS FOR FEATURES IN PROBLEMS OF LINEAR ORDERING OF OBJECTS

Summary: The aim of this study is to propose a system for determining the weighting coefficients for the characteristics. The paper presents three approaches. The first approach consists in using statistical procedures (objective approach). The second approach – the substantive one is based on expert opinions (subjective approach). The third one – integrated – combines subjective and objective approaches. The proposed integrated approach was applied to test the level of socio-economic development of regions in Poland in 2011, using the TOPSIS method (*Technique for Order of Preference by Similarity to Ideal Solution*).

Keywords: weights of features, CCSD method, analytic hierarchy process (AHP), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS).