

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 426

**Taksonomia 26**

**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>Jacek Batóg:</b> Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis .....	13
<b>Andrzej Bąk:</b> Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
<b>Grażyna Dehnel:</b> <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
<b>Andrzej Dudek:</b> <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
<b>Iwona Foryś:</b> Analiza dyskryminacyjna w wyborze obiektów podobnych w procesie szacowania nieruchomości / The discriminant analysis in selection of similar objects in the real estate valuation process .....	51
<b>Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz:</b> Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
<b>Iwona Konarzewska:</b> Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria .....	69
<b>Anna Król, Marta Targaszewska:</b> Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
<b>Marek Lubicz:</b> Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
<b>Aleksandra Łuczak:</b> Wpływ różnych sposobów agregacji opinii ekspertów w FAHP na oceny priorytetowych czynników rozwoju / Influence of different methods of the expert judgments aggregation on assessment of priorities for evaluation of development factors in FAHP.....	99
<b>Iwona Markowicz:</b> Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity .....	108

<b>Małgorzata Markowska, Danuta Strahl:</b> Filary inteligentnego rozwoju a wrażliwość unijnych regionów szczebla NUTS 2 na kryzys ekonomiczny – analiza wielowymiarowa / Smart development pillars and NUTS 2 European regions vulnerability to economic crisis – a multidimensional analysis.....	118
<b>Kamila Migdał-Najman, Krzysztof Najman:</b> Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis .....	130
<b>Kamila Migdał-Najman, Krzysztof Najman:</b> Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis .....	139
<b>Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta:</b> Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce) / Problem of outliers in corporate bankruptcy prediction (case of manufacturing companies in Poland) .....	148
<b>Wojciech Roszka:</b> Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
<b>Małgorzata Rószkiewicz:</b> Czynniki różnicujące efektywność pracy ankietera w wywiadach <i>face-to-face</i> w środowisku polskich gospodarstw domowych / Factors affecting the efficiency of face-to-face interviews with Polish households.....	166
<b>Adam Sagan, Marcin Pelka:</b> Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data .....	174
<b>Marcin Salamaga:</b> Zastosowanie drzew dyskryminacyjnych w identyfikacji czynników wspomagających wybór kraju alokacji bezpośrednich inwestycji zagranicznych na przykładzie polskich firm / The use of classification trees in the identification of factors supporting the choice of FDI destination on the example of Polish companies.....	185
<b>Agnieszka Stanimir:</b> Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
<b>Mirosława Sztemberg-Lewandowska:</b> Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge .....	206
<b>Tadeusz Trzaskalik:</b> Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature .....	214

---

<b>Joanna Trzęsiok:</b> Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions .....	226
<b>Hanna Wdowicka:</b> Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
<b>Artur Zaborski:</b> Zastosowanie skalowania dynamicznego oraz metody wektorów dryfu do badania zmian w preferencjach / The use of dynamic scaling and the drift vector method for studying changes in the preferences.....	245

## 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. Przewodniczącymi Komitetu Organizacyjnego konferencji byli prof. dr hab. Mirosław Szreder oraz dr hab. Krzysztof Najman, prof. nadzw. UG, sekretarzami naukowymi dr hab. Kamila Migdał-Najman, prof. nadzw. UG oraz dr hab. Anna Zamojska, prof. nadzw. UG, a sekretarzem organizacyjnym Anna Nowicka z Fundacji Rozwoju Uniwersytetu Gdańskiego.

Konferencja Naukowa została dofinansowana ze środków Narodowego Banku Polskiego.

Zakres tematyczny konferencji obejmował takie zagadnienia, jak:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczymi celami konferencji SKAD były prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 81 osób. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: AGH w Krakowie, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwer-

sytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, a także przedstawiciele NBP i PBS Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz trzynastu 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. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Józef Pociecha, Eugeniusz Gatnar, Tadeusz Trzaskalik, Krzysztof Jajuga, Marek Walesiak, Barbara Pawełek, Feliks Wysocki, Ewa Roszkowska, Andrzej Sokołowski, Andrzej Bąk, Tadeusz Kufel, Mirosław Krzyśko, Krzysztof Najman, Małgorzata Rószkiewicz, Mirosław Szreder.

Teksty 25 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii „Taksonomia” nr 26. Pozostałe recenzowane artykuły znajdują się w „Taksonomii” nr 27.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pociecha. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

- A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
- B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
- C. Organizacja konferencji SKAD PTS w latach 2016 i 2017.
- D. Wybór przedstawiciela Rady Sekcji SKAD PTS do IFCS.
- E. Dyskusja nad kierunkami rozwoju działalności Sekcji.

Prof. dr hab. Józef Pociecha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajdują się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii „Taksonomia” nr 24 i 25 (PN UE we Wrocławiu nr 384 i 385). W „Przeglądzie Statystycznym” (zeszyt 4/2014) ukazało się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Międzyzdrojach, w dniach 8–10 września 2014 r. Prof. Barbara Pawełek przedstawiła także informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków i sympatyków SKAD.

W konferencji Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS – International Federation of Classification Societies) w dniach 6–8 lipca 2015 r. w Bolonii, zorganizowanej przez Università di Bologna, udział wzięło 19 osób z Polski (w tym 17 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 79,0%). Ponadto prof. Józef Pociecha był członkiem Komitetu Naukowego Konferencji z ramienia SKAD, członkiem Międzynarodowego Komitetu Nagród IFCS oraz organizatorem i przewodniczącym sesji nt. „Classification models for forecasting of economic processes”.

W konferencji „European Conference on Data Analysis” (Colchester, 2–4 września 2015 r.) zorganizowanej przez The German Classification Society (GfKI) we współpracy z The British Classification Society (BCS) i Sekcją Klasyfikacji i Analizy Danych PTS (SKAD) udział wzięło 18 osób z Polski (w tym 14 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 66,0%). Ponadto profesorowie Krzysztof Jajuga oraz Józef Pociecha byli członkami Komitetu Naukowego konferencji, prof. Andrzej Dudek został poproszony przez organizatorów o przygotowanie referatu i wygłoszenie na Sesji Plenarnej „Cluster analysis in XXI century, new methods and tendencies”, prof. Krzysztof Jajuga był przewodniczącym sesji plenarnej, przewodniczącym sesji nt. „Finance and economics II” oraz organizatorem i przewodniczącym sesji nt. „Data analysis in finance”, prof. Józef Pociecha był organizatorem i przewodniczącym sesji nt. „Outliers in classification procedures – theory and practice”, prof. Andrzej Dudek był przewodniczącym sesji nt. „Machine learning and knowledge discovery II”.

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pociecha poinformował o dwóch wybranych konferencjach krajowych (były to XXXIV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015”, Łódź, 16–18 listopada 2015 r. i X Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 10–13 maja 2016 r.) oraz o trzech wybranych konferencjach zagranicznych. Konferencja „European Conference on Data Analysis” odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu w dniach 26–28 września 2017 r. W przeddzień tej konferencji, tj. 25.09.2017 r., odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2017”. Następna konferencja Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS) odbędzie się w 2017 r. w Tokio. W 2019 r. Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2019” organizuje prof. Andreas Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2016 zorganizuje Katedra Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

W kolejnej części zebrania dokonano wyboru przedstawiciela Rady Sekcji SKAD PTS do IFCS na kadencję 2016–2019. Powołano Komisję Skrutacyjną, której przewodniczącym został prof. Tadeusz Kufel, a członkami dr hab. Iwona Konarzewska i dr Dominik Rozkrut. Profesor Józef Pociecha poprosił zebranych o proponowanie kandydatur zgłaszając jednocześnie prof. Andrzeja Sokołowskiego. Wobec braku następnych kandydatur listę zamknięto. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 41 członków Sekcji. Profesor Andrzej Sokołowski został przedstawicielem Rady Sekcji SKAD PTS do



IFCS na kadencję 2016–2019, uzyskując następujący wynik: 39 głosów na „tak”, 1 głos na „nie”, 1 głos był nieważny.

W ostatnim punkcie zebrania dyskutowano nad kierunkami rozwoju działalności Sekcji obejmującymi następujące problemy: udział w międzynarodowym ruchu naukowym (wspólne granty, publikacje), umiędzynarodowienie konferencji SKAD (uczestnicy zagraniczni, dwujęzyczność konferencji), wydawanie własnego czasopisma.

Profesor Józef Pociecha zamknął posiedzenie Sekcji SKAD.

*Krzysztof Jajuga, Marek Walesiak*

**Iwona Konarzewska**

Uniwersytet Łódzki  
e-mail: i\_konarzewska@uni.lodz.pl

---

**RANKINGI WIELOKRYTERIOWE  
A WSPÓLZALEŻNOŚĆ LINIOWA KRYTERIÓW**  
**MULTI-CRITERIA RANKINGS  
AND LINEAR RELATIONSHIPS AMONG CRITERIA**

---

DOI: 10.15611/pn.2016.426.07

**Streszczenie:** W pracy podjęto temat konstrukcji wskaźników syntetycznych przy prowadzeniu badań z wykorzystaniem metod dyskretnej wielokryteriowej analizy decyzyjnej (np. SAW, TOPSIS, VIKOR) czy też metod wielowymiarowej analizy porównawczej. Metody te znajdują coraz częstsze zastosowanie w praktyce gospodarczej i polityce społeczno-ekonomicznej. Wymaga się przyjęcia założenia o niezależności kryteriów, lecz jego weryfikacja jest w praktyce często pomijana. Kluczowym etapem jest normalizacja kryteriów i nadanie im priorytetów. Wykorzystuje się tu, oprócz preferencji merytorycznych, mierniki statystyczne zmienności oraz mierniki siły informacyjnej. Zaproponowano zastosowanie metody tworzenia zmiennej syntetycznej z wykorzystaniem głównych składowych macierzy kowariancji kryteriów. Propozycje teoretyczne zostały zilustrowane wynikami przeprowadzonego badania empirycznego dla województw Polski pod względem kryteriów zrównoważonego rozwoju, na podstawie danych GUS (Bank Danych Lokalnych) z 2014 r.

**Słowa kluczowe:** metody wielokryteriowe, zależność liniowa kryteriów, metoda głównych składowych, SAW, TOPSIS, VIKOR.

**Summary:** The paper considers the problem of the construction of synthetic indicators for multi-criteria rankings (i.e. SAW, TOPSIS, VIKOR) or multivariate comparative analysis methods. Frequent applications are observed in business and socio-economic policy. Most of multi-criteria methods require the considered criteria to be independent. Verification of this assumption is often omitted. The key elements of the multi-criteria methods are normalization and weighting procedures. We propose the procedure for criteria weighting and the method to develop the synthetic indicator which exploits the principal components of the considered criteria covariance matrix. Theoretical propositions are illustrated by the empirical research results for indicators of sustainable growth for voivodships in Poland in the year 2014.

**Keywords:** multi-criteria methods, criteria interdependence, principal components, SAW, TOPSIS, VIKOR.

## 1. Wstęp

Przedmiotem zainteresowania niniejszej pracy są metody podejmowania decyzji przy występowaniu wielu kryteriów. Najczęstsze cele stosowania metod wielokryterio- wych (metody MCDA – *Multi-Criteria Decision Analysis*) to: wybór najlepszego wariantu, klasyfikacja wariantów w jednorodne grupy, uporządkowanie wariantów (rangowanie). Przegląd metod wielokryterio- wych można znaleźć np. w pracach: [Ishizaka, Nemery 2013; Trzaskalik (red.) 2014]. Cele, jakie formułuje się dla metod MCDA, są w zasadzie tożsame z celami stawianymi metodom wielowymiarowej analizy porównawczej (WAP)<sup>1</sup>. Różnice między metodami MCDA i WAP związane są m.in. z doбором kryteriów i ustalaniem ich priorytetów, a także z celem prowadzenia badań. Obie grupy metod zakładają, że kryteria oceny wariantów wybierane są w sposób merytoryczny i że są niezależne. Metody MCDA najistotniejszą rolę w ustalaniu wag dla kryteriów przypisują subiektywnym ocenom preferencji decy- denta, dokonując, w przypadku decydentów grupowych lub przy próbie obiektywizacji wyników badań, agregacji ocen indywidualnych uzyskanych w przeprowadzo- nym przeglądowym badaniu. W metodach WAP ustalanie wag odbywa się w oparciu o statystyczną analizę zmienności wartości kryteriów w porównywanej zbiorowości wariantów, a także biorąc pod uwagę siłę informacyjną kryteriów. W przypadku metod MCDA nie wymaga się, aby ilość porównywanych wariantów była większa od ilości kryteriów, co jest zasadą w przypadku metod WAP<sup>2</sup>.

Motywacją dla niniejszej pracy jest rosnąca popularność metod wielokryterio- wych MCDA. Za ich pomocą przeprowadza się różnego rodzaju rankingi mające charakter opiniotwórczy, służący wspomaganie istotnych decyzji, np. finansowych. Brak analiz zależności stosowanych kryteriów i świadomości wpływu tych zależno- ści na uzyskiwane wyniki budzi niepokój. W artykule próbujemy odpowiedzieć na następujące pytania:

- Jak powinny być konstruowane wagi dla kryteriów?
- Jak wykorzystać informację o sile i kierunku zależności kryteriów do konstruk- cji skorygowanej zmiennej syntetycznej będącej podstawą końcowego rankin- gu obiektów/wariantów decyzyjnych?

W pracy przyjęto następujące założenia: wszystkie kryteria są mierzone na skali przedziałowej lub ilorazowej oraz liczba kryteriów  $K$  nie przekracza liczby  $N$  porów- nywanych obiektów/wariantów. Zaprezentowany został sposób ustalania wag dla kryteriów oparty na doświadczeniach metod WAP: do pomiaru zmienności wyko-

---

<sup>1</sup> Szeroko rozumianym metodom WAP (statystycznym i taksonomicznym) poświęcona jest np. praca [Panek, Zwierzchowski 2013], w której zawarto również dość obszerną bibliografię przed- miotu.

<sup>2</sup> Metody WAP bazują na statystycznych miernikach zmienności i współzależności zmiennych losowych. Jeżeli liczba obiektów porównywanych jest mniejsza od liczby kryteriów/zmiennych dia- gnostycznych, to szacowane macierze kowariancji i korelacji zmiennych są osobliwe.

rzystano klasyczny współczynnik zmienności, do pomiaru informatywności zaś miarę VIF (*Variance Inflation Factor*), związaną ze współczynnikiem korelacji wielorakiej wybranego kryterium ze zbiorem pozostałych. W celu uodpornienia wyników rankingów na występujące zależności między kryteriami, zaproponowano modyfikację metody SAW (*Simple Additive Weighting*), o roboczej nazwie MCPC (*Multi-Criteria Principal Components*). Metoda wykorzystuje rozkład SVD (*Singular Value Decomposition*) macierzy wartości kryteriów według wartości osobliwych<sup>3</sup>. Przedstawiono przykład zastosowania metody MCPC oraz, dla porównania wyników, metod: SAW, TOPSIS, VIKOR, do przeprowadzenia rankingu województw Polski pod względem wartości wskaźników zrównoważonego rozwoju w 2014 r.<sup>4</sup>

## 2. Wybrane metody służące porównaniom wielokryteriowym

Przedstawiamy krótko metody z klasy MCDA, które zostały wykorzystane w pracy w przeprowadzonych rankingach. Przyjęto następujące oznaczenia:

- $f_{ik}$  – wartość  $k$ -tej funkcji kryterium dla  $i$ -tego wariantu,  $k = 1, \dots, K$ ,  $i = 1, \dots, N$ ,
- $w_k$  – waga dla kryterium  $k$ -tego;  $\sum_{k=1}^K w_k = 1$  oraz  $\forall k \ w_k \geq 0$ .

### 2.1. Metoda addytywnego ważenia<sup>5</sup> SAW<sup>6</sup>

Przyjmuje się, że wartości kryteriów są dodatnie. Oryginalne ujemne wartości można przekształcić, np. za pomocą podstawienia:  $f_{ik} := f_{ik} + \left| \min_i f_{ik} \right| + 1$ . Stymulację i normalizację wartości zmiennych/kryteriów można wykonać, wybierając jeden z poniżej określonych dwóch sposobów: przekształcenie ilorazowe albo unitaryzację. Konstruuje się indeks wielokryteriowy według wzoru:

$$Q_i = \sum_{k=1}^K w_k \bar{f}_{ik}, \quad (1)$$

gdzie symbol  $\bar{f}_{ik}$  oznacza znormalizowaną wartość  $k$ -tego kryterium dla  $i$ -tego obiektu porównań.

Największa wartość indeksu  $Q_i$  odpowiada najlepszemu wariantowi. Należy zauważyć, że sposób normalizacji wywiera istotny wpływ na uzyskiwane wyniki.

<sup>3</sup> Algorytm numeryczny dekompozycji SVD macierzy podają [Golub, Reinsch 1971]. Do przeprowadzenia obliczeń wykorzystano: *Matrix and Linear Algebra for Excel v.2.3.2*, <http://digilander.libero.it/foxes>, autor: L. Volpi [2006].

<sup>4</sup> Wartości wskaźników zaczerpnięto ze strony [www.stat.gov.pl](http://www.stat.gov.pl). Dostępne są tam również definicje stosowanych wskaźników.

<sup>5</sup> T. Trzaskalik [(red.) 2014] stosuje nazwę: metoda sumy ważonej.

<sup>6</sup> Zob. [Churchman, Ackoff 1954].

W przeprowadzonym badaniu wybrano przekształcenie ilorazowe, które nie powoduje zniekształcenia oryginalnych wartości statystycznych charakterystyk względnej zmienności kryteriów, a także współczynników korelacji.

## 2.2. Metoda TOPSIS

Główną ideą metody TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*)<sup>7</sup> jest przyjęcie zasady, że wybrany wariant powinien być najbliższy rozwiązaniu „idealnemu” i jednocześnie najdalszy od rozwiązania „antyidealnego”. W pierwszym kroku oblicza się elementy ważonej znormalizowanej macierzy wartości kryteriów według wzoru:

$$\omega_{ik} = \frac{w_k f_{ik}}{\sqrt{\sum_{i=1}^N f_{ik}^2}}. \quad (2)$$

Następnie, dla każdego kryterium dla  $k = 1, \dots, K$  wyznaczane jest rozwiązanie „idealne” oraz „antyidealne”<sup>8</sup>. Ranking końcowy uzyskuje się na podstawie malejących wartości wskaźnika  $C_i^*$ , wyrażającego względną „bliskość”  $i$ -tego wariantu do rozwiązania „idealnego”, zdefiniowanego jako

$$C_i^* = \frac{D_i^-}{D_i^* + D_i^-}, \quad (3)$$

gdzie  $D_i^*$  oraz  $D_i^-$  są odległościami euklidesowymi wariantu  $i$ -tego od rozwiązań: „idealnego” i „antyidealnego”.

## 2.3. Metoda VIKOR<sup>9</sup>

Kroki metody VIKOR pokazujemy dla przypadku, gdy wszystkie funkcje kryterium są stymulantami, w wersji zastosowanej w tej pracy:

1. Dla każdego kryterium  $k = 1, \dots, K$  ustal największą  $f_k^*$  i najmniejszą  $f_k^-$  wartość funkcji kryterium.

2. Oblicz:

$$S_i = \sum_{k=1}^K w_k \frac{f_k^* - f_{ik}}{f_k^* - f_k^-} \quad \text{oraz} \quad R_i = \max_k \left[ w_k \frac{f_k^* - f_{ik}}{f_k^* - f_k^-} \right] \quad \text{dla } i = 1, \dots, N, \quad (4)$$

gdzie  $S_i$  oraz  $R_i$  to, odpowiednio, ważne znormalizowane odległości: miejska i Czebyszewa wariantu  $i$ -tego od „idealnego”.

<sup>7</sup> Metoda zaprezentowana w pracy [Chen, Hwang 1992].

<sup>8</sup> Odpowiadające maksimum/minimum wartości  $\omega_{ik}$  w przypadku stymulant/destymulant.

<sup>9</sup> Zob. [Opricovic 1998; Opricovic, Tzeng 2004].

3. Oblicz wartości indeksu kompromisowego:

$$Q_i = \alpha \frac{S_i - S^*}{S^- - S^*} + (1 - \alpha) \frac{R_i - R^*}{R^- - R^*} \quad (5)$$

dla wybranej wartości parametru  $\alpha \in [0,1]$ .

4. Utwórz 3 rankingi wariantów decyzyjnych zgodnie z malejącymi wartościami  $S, R, Q$ .

5. Zaproponuj rozwiązanie kompromisowe jako wariant  $a'$ , odpowiadający minimum  $Q_i$ , jeżeli spełnione są dwa następujące warunki:

**C1:** „Akceptowalna przewaga”, która oznacza  $Q(a'') - Q(a') \geq DQ$ , gdzie  $a''$  jest drugim z kolei wariantem na liście według  $Q$  oraz  $DQ = 1/N-1$ .

**C2:** „Akceptowalna stabilność decyzji”, która oznacza, że  $a'$  musi być także najlepszym wariantem zgodnie z  $S$  i/lub  $R$ .

W przypadku niespełnienia któregoś z warunków C1 lub C2 proponuje się zestaw rozwiązań kompromisowych (np. [Trzaskalik (red.) 2014]). W przeprowadzonym badaniu, ranking metodą VIKOR uzyskano według wartości indeksu kompromisowego  $Q$  dla  $\alpha = 0,5$ .

### 3. Zasady doboru kryteriów oraz ustalania wag

Podstawową zasadą doboru kryteriów jest dobór merytoryczny, zależny od celu przeprowadzania analiz. Można go dokonać, np. metodą punktową, wykorzystując oceny ekspertów. W metodach WAP przyjmuje się, że kryteria powinny charakteryzować się jak największą siłą dyskryminacyjną, którą utożsamiamy ze zróżnicowaniem ich wartości. W praktyce posługujemy się współczynnikami zmienności  $V(f_k)$  (klasycznymi bądź pozycyjnymi). Miernikiem siły dyskryminacyjnej  $k$ -tego kryterium jest

$$W_k^a = \frac{V(f_k)}{\sum_{k=1}^K V(f_k)}, k = 1, \dots, K. \quad (6)$$

Eliminujemy kryteria, dla których  $W_k^a < 0,1$ . Ponadto, kryteria powinny być niezależne, co oznacza, że nie powinny powielać informacji dostarczanych przez inne kryteria. Kryteria powinny być również wysoce informatywne, tzn. być dobrymi reprezentantami pozostałych, nie wybranych kryteriów zmiennych (por. [Nowak 1990, s. 28]).

Z przytoczonych zasad wynika, że ostateczny wybór kryteriów powinien być poprzedzony analizą merytoryczną ich istotności dla badanego problemu. Kryteria słabo dyskryminujące należy wyeliminować. Dla pozostałych należy przeprowadzić badanie współzależności, aby uniknąć doboru kryteriów zależnych. Proponujemy do tego celu wykorzystać współczynniki VIF (*Variance Inflation Factors*), które są diagonalnymi elementami macierzy odwrotnej do macierzy korelacji wartości kryte-

riów. Wartość  $VIF_k = 1$  oznacza ortogonalność  $k$ -tego kryterium względem zbioru pozostałych;  $VIF_k \rightarrow \infty$  oznacza liniowy związek  $k$ -tego kryterium z pozostałymi. Wzór (7) pokazuje zależność między VIF a współczynnikiem korelacji wielorakiej:

$$VIF_k \equiv r^{kk} = (1 - \delta_k^2)^{-1}, \quad (7)$$

gdzie:  $r^{kk}$  –  $k$ -ty diagonalny element macierzy odwrotnej do macierzy korelacji,  $\delta_k$  – współczynnik korelacji wielorakiej  $k$ -tego kryterium ze zbiorem pozostałych kryteriów.

Unormowany miernik informatywności kryterium definiujemy za pomocą wzoru (8):

$$W_k^b = \left( \frac{\min_k VIF_k}{VIF_k} \right) \cdot \left( \sum_{k=1}^K \left( \frac{\min_k VIF_k}{VIF_k} \right) \right)^{-1}. \quad (8)$$

Łączny miernik zdolności dyskryminacyjnej i informatywności kryterium skonstruowany jest jako iloczyn:  $W_k = W_k^a \cdot W_k^b$ . Unormowane wagi dla kryteriów, przy założeniu, że oceny ich istotności merytorycznej są jednakowe, oblicza się jako:

$$w_k = W_k \left( \sum_{k=1}^K W_k \right)^{-1}. \quad (9)$$

W przypadku zróżnicowanych ocen istotności merytorycznej kryteriów proponujemy nadane im przez ekspertów wagi (po ich unormowaniu) uwzględnić jako trzeci czynnik przy konstrukcji ostatecznych wag.

#### 4. MCPC – propozycja uodpornienia analiz wielokryteriowych na współzależność kryteriów

Niech  $\mathbf{A}$  oznacza macierz o wymiarach  $N \times K$  unormowanych wartości  $K$  kryteriów dla  $N$  wariantów. Celem proponowanej metody jest uodpornienie wyników rankinów na występowanie, wbrew intencjom badacza, współzależności kryteriów. Kroki proponowanej metody, o roboczej nazwie MCPC, są następujące:

1. Obliczenie wartości własnych macierzy kowariancji kryteriów po normalizacji.
2. Podjęcie decyzji o liczbie głównych składowych  $K^*$  istotnych dla objaśnienia łącznej wariancji kryteriów (np. objaśnienie co najmniej 75% wariancji łącznej).
3. Dekompozycja macierzy  $\mathbf{A}$  według wartości osobliwych

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad (10)$$

gdzie:  $\mathbf{U}$  –  $N \times K$  macierz normalizowanych wektorów własnych związanych z niezerowymi wartościami własnymi macierzy  $\mathbf{A}\mathbf{A}^T$ ,  $\mathbf{U}^T\mathbf{U} = \mathbf{I}_K$ ,  $\mathbf{D}$  – diagonalna macierz  $K \times K$  pierwiastków z wartości własnych macierzy  $\mathbf{A}^T\mathbf{A}$ ,  $\mathbf{V}$  – macierz  $K \times K$  normalizowanych wektorów własnych macierzy  $\mathbf{A}^T\mathbf{A}$ .

4. Wyznaczenie macierzy  $\mathbf{Y}^*$  o wymiarach  $N \times K^*$ , której kolumnami są wektory głównych składowych

$$\mathbf{Y}^* = \mathbf{A}\mathbf{V}^* = \mathbf{U}^*\mathbf{D}^* . \quad (11)$$

(Oznaczenie „\*” przy symbolach macierzy  $\mathbf{U}$  oraz  $\mathbf{D}$  wskazuje wybór  $K^*$  głównych składowych i odpowiedni wymiar  $K^*$  zamiast wymiaru  $K$ . Macierz  $\mathbf{V}^*$  ma wymiar  $K \times K^*$ .)

5. Wartości zmiennej syntetycznej  $S_i$   $i = 1, \dots, N$  obliczane są za pomocą wzoru

$$\mathbf{S} = \mathbf{Y}^*(\mathbf{V}^{*T}\mathbf{W})\mathbf{1} = \mathbf{Y}^*\mathbf{W}^*\mathbf{1} = \mathbf{Y}^*\mathbf{w}^* , \quad (12)$$

gdzie  $\mathbf{S} = [S_i]$ ,  $\mathbf{W} = \text{diag}(w_k)$  oznacza macierz diagonalną o wymiarach  $K \times K$  wag dla kryteriów,  $\mathbf{W}^* = \mathbf{V}^{*T}\mathbf{W}$ ,  $\mathbf{w}^* = \mathbf{W}^*\mathbf{1}$  (wektor „wag” dla głównych składowych) a symbol  $\mathbf{1}$  to kolumnowy wektor jedynek o odpowiednim wymiarze.

6. Ranking uzyskiwany jest w porządku malejących wartości  $S_i$ .

## 5. Wskaźniki zrównoważonego rozwoju dla Polski w 2014 r. – ujęcie regionalne

Przedstawione metody wielokryteriowe zostały zastosowane do wskaźników zrównoważonego rozwoju w 2014 r. opracowanych przez GUS. Wykorzystano odpowiednie informacje dla województw. Warto wspomnieć, że zrównoważony rozwój wskazany został wśród trzech priorytetów Strategii „Europa 2020”, która stanowi nowy, długofalowy program społeczno-gospodarczy Unii Europejskiej. Analizy prowadzone są w czterech obszarach.

1. Ład gospodarczy, w tym: rozwój społeczno-gospodarczy, zrównoważona konsumpcja i produkcja, zrównoważony transport.

2. Ład społeczny, w tym: zrównoważona konsumpcja i produkcja, włączenie społeczne, zmiany demograficzne.

3. Ład środowiskowy, w tym: zrównoważona konsumpcja i produkcja, zmiany klimatu i energia, zasoby naturalne.

4. Ład instytucjonalno-polityczny, w tym: otwartość i uczestnictwo, instrumenty ekonomiczne.

Przedstawimy wyniki rankingów województw uzyskane niezależnie dla wskaźników ładu gospodarczego oraz środowiskowego. W badaniu przyjęto, że merytoryczna istotność wziętych pod uwagę kryteriów jest jednakowa.

### 5.1. Wskaźniki ładu gospodarczego

W bazie danych znajduje się 11 wskaźników, z których ostatecznie, po eliminacji ze względu na niską zmienność, analizowano 6. Są to:



- Wskaźniki zatrudnienia: w wieku 15–24 lata (G2), według wykształcenia – gimnazjalne i niższe (G7) – *stymulanty*.
- Osoby fizyczne prowadzące działalność gospodarczą na 100 osób w wieku produkcyjnym ogółem (G8) – *stymulanta*.
- Zużycie nawozów mineralnych na 1 ha użytków rolnych (G9) – *destymulanta*.
- Obsada zwierząt w sztukach dużych ogółem na 1 hektar użytków (G10) – *destymulanta*.
- Przeciętna cena biletu jednorazowego normalnego za przejazd autobusem miejskim (G11) – *destymulanta*.

**Tabela 1.** Wagi dla kryteriów ładu gospodarczego (%)

Kryterium	G2	G7	G8	G9	G10	G11
$W^a$	13,00	11,22	11,96	19,48	35,98	8,36
$W^b$	9,57	24,90	12,89	22,61	13,30	9,57
$w$	7,70	17,28	9,54	27,24	29,60	8,65

Źródło: obliczenia własne.

Mierniki siły dyskryminacyjnej  $W^a$ , informatywności  $W^b$  oraz unormowane wagi  $w$  dla kryteriów ładu gospodarczego przedstawione są w tab. 1. Uzyskane rankingi przedstawia tab. 2. Największą siłą dyskryminacyjną cechował się wskaźnik G10; najbardziej informatywne były G7 oraz G9. Ostatecznie, najwyższe wagi przyjęto dla kryteriów G9 i G10, związanych z rolnictwem. Wyróżniono 3 główne składowe opisujące 86,5% łącznej wariancji układu kryteriów.

**Tabela 2.** Rankingi dla wskaźników ładu gospodarczego

Region	Województwo	SAW	MCPC	TOPSIS	VIKOR
Region centralny	łódzkie	11	12	13	10
	mazowieckie	10	13	11	9
Region południowy	małopolskie	6	6	6	5
	śląskie	16	16	12	13
Region wschodni	lubelskie	7	7	4	8
	podkarpackie	2	2	2	4
	podlaskie	12	10	15	11
	świętokrzyskie	5	5	5	2
Region pñ.-zach.	lubuskie	4	4	3	3
	wielkopolskie	15	15	16	16
	zachodniopomorskie	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
Region pd.-zach.	dolnośląskie	3	3	7	12
	opolskie	13	11	10	15
Region północny	kujawsko-pomorskie	14	14	14	14
	pomorskie	9	9	8	7
	warmińsko-mazurskie	8	8	9	6

Źródło: obliczenia własne.

Wyniki rankingów jednoznacznie wskazują na województwo zachodniopomorskie jako charakteryzujące się najwyższym stopniem osiągnięcia ładu gospodarczego. Najgorzej wypadają województwa śląskie i wielkopolskie.

## 5.2. Wskaźniki ładu środowiskowego

W badaniu wykorzystano następujące wskaźniki:

- emisja zanieczyszczeń powietrza z zakładów szczególnie uciążliwych: gazowych (S1), pyłowych (S2) – *destymulanty*,
- emisja CO<sub>2</sub> z zakładów szczególnie uciążliwych (S3) – *destymulanta*,
- udział obszarów prawnie chronionych w powierzchni ogółem (S4) – *stymulanta*,
- udział powierzchni odnowień i zalesień w powierzchni lasów ogółem (S5) – *stymulanta*,
- udział powierzchni użytków rolnych w powierzchni ogółem (S6) – *stymulanta*,
- lesistość (S7) – *stymulanta*.

**Tabela 3.** Wagi dla kryteriów ładu środowiskowego (%)

Kryterium:	S1	S2	S3	S4	S5	S6	S7
$W^a$	28,25	16,58	28,27	10,52	6,04	3,95	6,39
$W^b$	0,00	4,34	0,00	45,88	44,67	2,80	2,32
w	0,00	8,46	0,00	56,78	31,72	1,30	1,74

Źródło: obliczenia własne.

**Tabela 4.** Rankingi dla wskaźników ładu środowiskowego

Region	Województwo	SAW	MCPC	TOPSIS	VIKOR
Region centralny	łódzkie	16	13	15	16
	mazowieckie	14	12	11	11
Region południowy	małopolskie	5	6	2	4
	śląskie	13	16	16	14
Region wschodni	lubelskie	15	8	14	15
	podkarpackie	3	3	4	3
	podlaskie	10	5	9	10
	świętokrzyskie	1	2	1	1
Region pñ.-zach.	lubuskie	4	4	5	5
	wielkopolskie	7	10	8	7
	zachodniopomorskie	12	15	13	12
Region pd.-zach.	dolnośląskie	11	14	12	13
	opolskie	6	7	7	6
Region północny	kujawsko-pomorskie	9	9	10	9
	pomorskie	8	11	6	8
	warmińsko-mazurskie	2	1	3	2

Źródło: obliczenia własne.

Obliczone wagi dla kryteriów ładu środowiskowego przedstawiono w tab. 3. Metoda obliczania wag z wykorzystaniem VIF wyeliminowała wpływ wskaźników S1 i S3 o najwyższej mocy dyskryminacyjnej, ale silnie skorelowanych. Najwyższe wagi wyznaczono dla wskaźników S4 i S5.

Analiza macierzy kowariancji kryteriów pokazała, że trzy główne składowe objaśniają prawie 94% całkowitej zmienności. Jest to wynikiem występowania zależności liniowych (między S1 i S3, jak również między S6 i S7).

Rankingi przedstawione w tab. 4 nie są zgodne. Najwyższy poziom ładu środowiskowego osiągany jest przez województwa świętokrzyskie (według SAW, TOPSIS, VIKOR) i warmińsko-mazurskie (według MCPC). Najgorzej prezentują się łódzkie (według SAW i VIKOR) oraz śląskie (według MCPC, TOPSIS).

## 6. Zakończenie

W pracy pokazano, że do ważenia kryteriów w oparciu o pomiar pojemności informacyjnej można wykorzystać wskaźniki VIF – sposób ten ogranicza wpływ kryteriów silnie powiązanych z pozostałymi. Zaproponowana metoda MCPC, redukująca wymiar przestrzeni kryteriów do ortogonalnych głównych składowych macierzy kowariancji, wydaje się pomocna w sytuacji liniowej współzależności kryteriów. Bywają one trudne do przewidzenia, zwłaszcza gdy nie wskazuje na nie analiza merytoryczna lub mają charakter złożony.

Metody SAW oraz MCPC, zastosowane dla wskaźników ładu gospodarczego, dały najbardziej zbliżone wyniki rankingów – w tym przypadku w macierzy kowariancji kryteriów – nie zaobserwowano związków o charakterze liniowym. Metoda VIKOR dała wynik rankingu najslabiej zbliżony do wyników uzyskanych za pomocą innych metod (współczynniki korelacji rang Spearmana na poziomie poniżej 0,8).

W badaniu wskaźników ładu środowiskowego stwierdzono silne złe uwarunkowanie macierz kowariancji – występowanie związków o charakterze liniowym. Zaproponowany sposób ważenia kryteriów znacznie ograniczył wpływ współzależności, niemniej jednak problem uwidocznił się w postaci dużych różnic między wynikami zastosowanych metod. Metoda MCPC dała ranking najslabiej zgodny z wynikami uzyskanymi za pomocą innych metod wielokryteriowych, nieuwzględniających współzależności kryteriów. Potwierdziło to postawioną hipotezę o silnym wpływie zależności kryteriów na wyników rankingów. Dalsze badania nad własnościami zaproponowanej metody MCPC wydają się bardzo uzasadnione.

## Literatura

- Chen S.J., Hwang C.L., 1992, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer, Berlin.
- Churchman C.W., Ackoff R.L., 1954, *An approximate measure of value*, Journal of Operations Research Society of America, vol. 2, no. 1, s. 172–187.
- Ishizaka A., Nemery P., 2013, *Multi-Criteria Decision Analysis. Methods and Software*, Wiley, Chichester.
- Golub G.H., Reinsch C., 1971, *Singular value decomposition and least squares solutions*, [w:] J.H. Wilkinson, C. Reinsch (red.), *Handbook for Automatic Computation*, vol. II: *Linear Algebra*, Springer-Verlag, Berlin.
- Nowak E., 1990, *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa.
- Opricovic S., 1998, *Multicriteria Optimization of Civil Engineering Systems*, Rap. Tech., Faculty of Civil Engineering, Belgrade.
- Opricovic S., Tzeng G.H., 2004, *Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS*, European Journal of Operational Research, vol. 156, no. 2, s. 445–455.
- Panek T., Zwierzchowski J., 2013, *Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowania*, Oficyna Wydawnicza SGH w Warszawie, Warszawa.
- Trzaskalik T. (red.), 2014, *Wielokryterialne wspomaganie decyzji. Metody i zastosowania*, PWE, Warszawa.
- Volpi L., 2006, *Matrix and Linear Algebra for Excel v.2.3.2*, <http://digilander.libero.it/foxes> (18.01.2007).