

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
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
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Jacek Batóg: Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis	13
Andrzej Bąk: Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
Grażyna Dehnel: <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
Andrzej Dudek: <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
Iwona Foryś: 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
Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz: Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
Iwona Konarzewska: Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria	69
Anna Król, Marta Targaszewska: Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
Marek Lubicz: Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
Aleksandra Łuczak: 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
Iwona Markowicz: Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity	108

Małgorzata Markowska, Danuta Strahl: 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
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis	130
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis	139
Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta: 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
Wojciech Roszka: Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
Małgorzata Rószkiewicz: 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
Adam Sagan, Marcin Pelka: Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data	174
Marcin Salamaga: 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
Agnieszka Stanimir: Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
Mirosława Sztemberg-Lewandowska: Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge	206
Tadeusz Trzaskalik: Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature	214

Joanna Trzęsiok: Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions	226
Hanna Wdowicka: Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
Artur Zaborski: 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 Pocięcha, 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 Pocięcha. 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 Pocięcha 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 Pocięcha 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

Artur Zaborski

Uniwersytet Ekonomiczny we Wrocławiu
e-mail: artur.zaborski@ue.wroc.pl

**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**

DOI: 10.15611/pn.2016.426.25

Streszczenie: W artykule zaprezentowano dwie metody zaliczane do skalowania wielowymiarowego, które mogą znaleźć zastosowanie w analizie zmian zachodzących w preferencjach konsumentów. Pierwsza z nich to skalowanie dynamiczne wykorzystujące w swojej konstrukcji analizę Procrustes. Z kolei metoda wektorów dryfu umożliwia równoczesną ilustrację na mapie percepcyjnej konfiguracji punktów reprezentujących analizowane obiekty, jak również wektorów wskazujących kierunek i siłę zmian zachodzących w preferencjach respondentów. Prezentacja metod zostanie zilustrowana przykładem empirycznym, w którym obliczenia przeprowadzono z wykorzystaniem pakietów PS IMAGO i NewMDSX.

Słowa kluczowe: badanie preferencji, skalowanie dynamiczne, analiza Procrustes, wektory dryfu.

Summary: The article presents two multidimensional scaling methods which can be used in the analysis of changes in consumer preferences. The first one is the dynamic scaling which uses in its construction the Procrustes analysis. The second one, the drift vector method, enables to present simultaneously on the perceptual map both the configuration of points representing the analyzed objects and the vectors indicating the direction and the strength of changes in the respondents' preferences. The presentation of the methods is illustrated by an empirical example in which calculations were performed with use of SPSS and NewMDSX packages.

Keywords: preference analysis, dynamic scaling, Procrustes analysis, drift vectors.

1. Wstęp

Analiza zmian zachodzących w preferencjach konsumentów z wykorzystaniem metod skalowania wielowymiarowego przeprowadzana jest na podstawie wielu macierzy preferencji uzyskanych w różnych okresach. Macierze te nie zawsze spełniają warunek symetryczności (zob. m.in. [Holyoak, Gordon 1983; Tversky, Gati 1982; Harshman i in. 1982; Chino 1978]). Przykładem może być macierz, w której element w i -tym wierszu j -tej kolumnie przedstawia liczbę osób, które w okresie $t - 1$ wybierały markę i , a w okresie t wybierają markę j .

W artykule zaprezentowano dwie metody, które można wykorzystać do prezentacji zmian zachodzących w preferencjach respondentów. Celem pierwszej z metod, tj. skalowania dynamicznego, jest przedstawienie na mapie percepcyjnej konfiguracji nT punktów, gdzie każdy z n punktów reprezentujących obiekty jest prezentowany T razy, oddzielnie dla każdego z okresów. Metoda wektorów dryfu umożliwi równoczesną prezentację konfiguracji punktów reprezentujących obiekty otrzymanej w wyniku niemetrycznego skalowania wielowymiarowego oraz wektorów wskazujących kierunek i siłę zmian w preferencjach. Prezentacja metod zostanie zilustrowana przykładem empirycznym, w którym obliczenia przeprowadzono z wykorzystaniem pakietów PS IMAGO i NewMDSX.

2. Klasyfikacja metod analizy danych niesymetrycznych

Dla każdej niesymetrycznej macierzy kwadratowej \mathbf{P} można dokonać jej dekompozycji poprzez przedstawienie w postaci: $\mathbf{P} = \mathbf{A} + \mathbf{B}$, gdzie \mathbf{A} jest macierzą symetryczną, a \mathbf{B} jest macierzą skośnosymetryczną. Oznacza to, że $\mathbf{A} = \mathbf{A}^T$, $\mathbf{B} = -\mathbf{B}^T$, $\mathbf{A} = (\mathbf{P} + \mathbf{P}^T)/2$ i $\mathbf{B} = (\mathbf{P} - \mathbf{P}^T)/2$. Poniżej przedstawiono macierz podobieństw między literami od A do F nadawanymi alfabetem Morse'a (zob. [Rothkopf 1957]). \mathbf{P} należy traktować jako macierz podobieństw ponieważ element w i -tym wierszu i j -tej kolumnie oznacza odsetek odpowiedzi, że dwa sygnały były postrzegane jako takie same jeżeli sygnał j był nadawany po sygnale i .

$$\mathbf{A} = \begin{bmatrix} 92 & 4 & 6 & 13 & 3 & 14 \\ 5 & 84 & 37 & 31 & 5 & 28 \\ 4 & 38 & 87 & 17 & 4 & 29 \\ 8 & 62 & 17 & 88 & 7 & 23 \\ 6 & 13 & 14 & 6 & 97 & 2 \\ 4 & 51 & 33 & 19 & 2 & 90 \end{bmatrix} = \begin{bmatrix} 92 & 4,5 & 5 & 10,5 & 4,5 & 9 \\ 4,5 & 84 & 37,5 & 46,5 & 9 & 39,5 \\ 5 & 37,5 & 87 & 17 & 9 & 31 \\ 10,5 & 46,5 & 17 & 88 & 6,5 & 21 \\ 4,5 & 9 & 9 & 6,5 & 97 & 2 \\ 9 & 39,5 & 31 & 21 & 2 & 90 \end{bmatrix} + \begin{bmatrix} 0 & -0,5 & 1 & 2,5 & -1,5 & 5 \\ 0,5 & 0 & -0,5 & -15,5 & -4 & -11,5 \\ -1 & 0,5 & 0 & 0 & -5 & -2 \\ -2,5 & 15,5 & 0 & 0 & 0,5 & 2 \\ 1,5 & 4 & 5 & -0,5 & 0 & 0 \\ -5 & 11,5 & 2 & -2 & 0 & 0 \end{bmatrix}$$

Łatwo jest wykazać (zob. [Borg, Groenen 2005, s. 497]), że dla tak przeprowadzonej dekompozycji $\sum_{ij} p_{ij}^2 = \sum_{ij} a_{ij}^2 + \sum_{ij} b_{ij}^2$. Sugeruje to, że analiza danych zawartych w macierzy niesymetrycznej może składać się z dwóch elementów: analizy

części symetrycznej oraz analizy części skośnosymetrycznej. Dla przedstawionego przykładu suma kwadratów podobieństw z wyłączeniem elementów diagonalnych wynosi 15483, z czego 94,1% wyjaśniona jest przez symetryczną macierz **A**, zaś 5,9% przez skośnosymetryczną macierz **B**. Wynika z tego, że symetryczna część podobieństw stanowi rolę dominującą w macierzy **P**.

Z możliwości dekompozycji dowolnej macierzy podobieństw wynika klasyfikacja metod dla niesymetrycznych danych w badaniach preferencji. Metody te można podzielić na trzy grupy:

- metody oparte na analizie macierzy **A** (np. metody skalowania dynamicznego),
- metody oparte na analizie macierzy **P** (np. *unfolding*, *slide vector model*, *hill-climbing model*, *dominance point model*),
- metody oparte na analizie macierzy **B** lub **A** oraz **B** (np. *signed-distance model*, *radius-distance model*, *drift vectors model*).

W dalszej części artykułu zostaną zaprezentowane dwie metody: jedna z metod skalowania dynamicznego oraz metoda wektorów dryfu należąca do trzeciej z wymienionych grup.

3. Idea skalowania dynamicznego

Skalowanie dynamiczne jest metodą analizy danych podobieństw między obiektami otrzymanych w T kolejnych okresach. Jej celem jest przedstawienie na mapie percepcyjnej konfiguracji nT punktów, gdzie każdy z n punktów reprezentujących obiekty jest prezentowany T razy, oddzielnie dla każdego z okresów.

Przeprowadzenie skalowania wielowymiarowego dla danych podobieństw z poszczególnych okresów, a następnie przedstawienie ich na jednej mapie percepcyjnej nie jest właściwym rozwiązaniem, ponieważ różnice w układach konfiguracji punktów na mapach percepcyjnych nie zawsze wynikają z różnic w postrzeganiu obiektów przez poszczególnych respondentów. Jedną z metod, która umożliwia doprowadzenie do porównywalności mapy percepcyjne otrzymane dla różnych okresów, jest analiza Procrustes wykorzystywana zazwyczaj w modelach różnic indywidualnych (zob. Zaborski [2007]). Polega ona na takim wzajemnym dopasowaniu konfiguracji, aby zostały zachowane proporcje odległości między punktami.

Dla dwóch wyników skalowania wielowymiarowego $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n]^T$ i $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_n]^T$ (punkty \mathbf{X} oraz \mathbf{Y} reprezentują te same obiekty otrzymane za pomocą różnych metod, w różnych okresach lub na podstawie ocen dokonanych przez różnych respondentów) dopasowanie konfiguracji jest najlepsze jeżeli suma kwadratów odległości między odpowiadającymi sobie punktami \mathbf{X} oraz \mathbf{Y} jest najmniejsza, czyli gdy funkcja dopasowania przybierająca postać:

$$R^2 = tr(\mathbf{X} - \mathbf{Y})^T (\mathbf{X} - \mathbf{Y})$$

osiąga wartość najmniejszą.

W celu dopasowania do konfiguracji \mathbf{Y} zbiór punktów \mathbf{X} poddaje się przekształceniom, którymi są translacja, rotacja i jednokładność. Dopasowanie konfiguracji za pomocą analizy Procrustes przebiega przez następujące etapy (zob. [Borg, Groenen 2005, s. 434–436; Zaborski 2007]):

- translacja \mathbf{X} oraz \mathbf{Y} tak, aby ich środki ciężkości znajdowały się w początku układu współrzędnych,
- wymnożenie macierzy \mathbf{X} przez macierz rotacji $\mathbf{E} = (\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}} (\mathbf{Y}^T \mathbf{X})^{-1}$,
- znalezienie obrazu \mathbf{X} przy jednokładności o stosunku $\hat{\phi} = \frac{\text{tr}(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}}}{\text{tr}(\mathbf{X}^T \mathbf{X})}$,
- obliczenie miary dopasowania (*statystyki Procrustes*): $R^2 = 1 - \frac{\{\text{tr}(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})^{\frac{1}{2}}\}^2}{\text{tr}(\mathbf{X}^T \mathbf{X}) \text{tr}(\mathbf{Y}^T \mathbf{Y})}$.

Większość algorytmów skalowania wielowymiarowego dokonuje takiego rozmieszczenia punktów na mapie percepcyjnej, aby ich środek ciężkości znajdował się w początku układu współrzędnych, a średnia kwadratów odległości punktów od początku układu współrzędnych wynosiła jeden, dlatego w praktyce etap pierwszy i trzeci jest pomijany.

Skalowanie dynamiczne wykorzystujące analizę Procrustes umożliwia geometryczną prezentację podobieństw między obiektami, jak również zachodzących w czasie zmian w tych podobieństwach. Algorytm metody jest następujący (zob. [Cox, Cox 2001, s. 102]):

- wyznaczenie map preferencji dla poszczególnych okresów,
- dopasowanie wszystkich map preferencji do mapy dla okresu t_1 (lub mapy dla okresu t do mapy dla okresu $t - 1$),
- przedstawienie trajektorii zmian w podobieństwach preferencji.

3.1. Przykład

Zaprezentowaną metodę skalowania dynamicznego wykorzystano w badaniu, którego celem była analiza zmian zachodzących w preferencjach słuchaczy Uniwersytetu Trzeciego Wieku w Bolesławcu (woj. dolnośląskie) odnośnie chęci uczestnictwa w określonych formach zajęć w kolejnych latach od 2009 r. do 2013 r. W badaniu wzięto pod uwagę zajęcia, na które słuchacze uczęszczają zgodnie z dokonanym przez siebie wyborem. Są to: j. angielski, j. niemiecki, obsługa komputera, gimnastyka rehabilitacyjna, gimnastyka chińska i joga, siłownia, zajęcia na basenie, marsze z kijkami oraz warsztaty malarskie i rękodzieło. Badaniem objęto 109 słuchaczy Uniwersytetu, którzy niezmiennie uczestniczyli w zajęciach w analizowanym okresie. W tabeli 1 wskazano, jakie zmiany zachodziły w dokonywanych przez słuchaczy Uniwersytetu wyborach zajęć w poszczególnych latach. Wartość elementu i -tego wiersza j -tej kolumny wyraża liczbę słuchaczy, którzy na początku

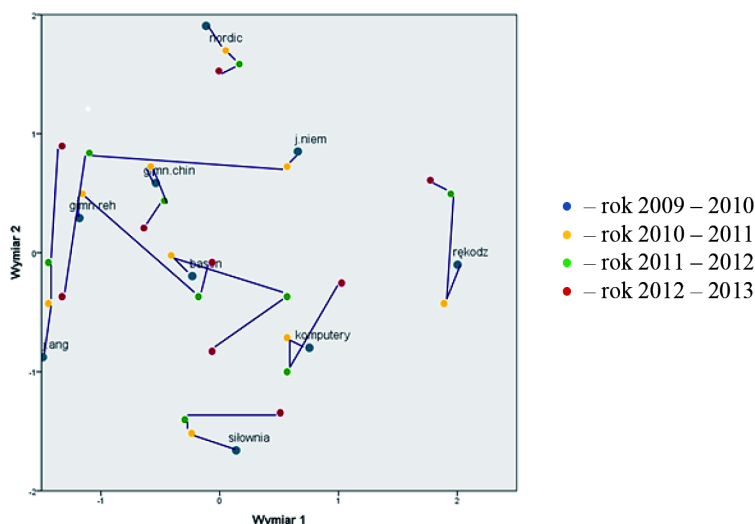
analizowanego okresu wybierali i -te zajęcia, zaś na koniec j -te zajęcia. Przedstawione macierze należy traktować jako macierze podobieństw, ponieważ duża wartość elementu macierzy oznacza, że słuchacze chętnie dokonują zamiany zajęć, a tym samym, zgodnie ze swoimi preferencjami, traktują je jako podobne.

Tabela 1. Macierze podobieństw między wybranymi zajęciami

Rodzaje zajęć		j.ang	j.niem	komp	g.reh	g.chin	basen	siłow	n.walk	w.mal
2009–2010	j. angielski	22	0	1	1	3	0	0	0	0
	j. niemiecki	1	15	2	2	4	1	1	1	0
	obsługa komputera	2	0	5	4	5	4	4	0	1
	gimnastyka rehabilitacyjna	0	0	0	50	4	2	2	1	1
	gimnastyka chińska i joga	0	0	0	0	9	3	1	0	1
	basen	0	0	0	2	0	22	2	2	0
	siłownia	0	0	0	0	0	0	5	0	0
	<i>nordic walking</i>	0	0	0	0	0	0	0	8	0
	warsztaty malarskie	0	0	2	0	0	0	0	1	7
2010–2011	j. angielski	14	0	0	3	0	2	1	0	0
	j. niemiecki	0	9	1	1	2	4	0	1	1
	obsługa komputera	0	0	11	1	0	2	1	1	0
	gimnastyka rehabilitacyjna	0	0	0	56	2	2	0	1	0
	gimnastyka chińska i joga	0	0	0	1	18	4	0	2	0
	basen	1	0	0	1	2	25	1	2	0
	siłownia	0	0	0	0	1	1	8	0	0
	<i>nordic walking</i>	0	0	0	0	0	0	0	7	0
	warsztaty malarskie	0	0	0	0	0	0	0	0	8
2011–2012	j. angielski	19	1	0	0	1	0	0	0	0
	j. niemiecki	0	8	0	1	1	0	0	0	0
	obsługa komputera	0	0	7	2	2	2	0	0	0
	gimnastyka rehabilitacyjna	0	0	1	55	4	5	3	0	0
	gimnastyka chińska i joga	1	0	0	1	17	1	0	2	0
	basen	0	0	0	0	0	41	1	0	0
	siłownia	0	0	0	1	0	0	9	0	0
	<i>nordic walking</i>	0	0	0	0	1	1	0	12	0
	warsztaty malarskie	0	0	0	0	0	0	0	0	12
2012–2013	j. angielski	20	0	0	1	0	1	0	1	0
	j. niemiecki	0	8	0	0	0	1	0	0	0
	obsługa komputera	0	0	10	1	0	0	2	0	0
	gimnastyka rehabilitacyjna	0	0	1	48	3	2	3	2	0
	gimnastyka chińska i joga	0	0	0	0	15	1	0	2	0
	basen	0	0	0	3	1	47	2	1	0
	siłownia	0	0	0	0	0	3	11	0	0
	<i>nordic walking</i>	0	0	0	1	0	0	0	11	0
	warsztaty malarskie	0	0	0	0	0	0	0	0	9

Źródło: opracowanie własne.

Na podstawie czterech macierzy podobieństw zawartych w tab. 1 przeprowadzono skalowanie dynamiczne, przy czym wszystkie konfiguracje punktów dopasowano za pomocą analizy Procrustes do konfiguracji dla pierwszego z analizowanych okresów. Dynamiczną mapę percepcyjną ukazującą trajektorie zmian zachodzących w podobieństwach preferencji prezentuje rys. 1.



Rys. 1. Dynamiczna mapa percepcyjna wybranych rodzajów zajęć

Źródło: opracowanie własne z wykorzystaniem pakietów NewMDSX i PS IMAGO.

Odległości między punktami na rys. 1 są odwrotnie proporcjonalne do podobieństw w preferencjach dotyczących uwzględnionych w analizie zajęć, a trajektorie punktów reprezentujących zajęcia w kolejnych okresach odzwierciedlają zmiany jakie zachodziły w preferencjach respondentów. Rysunek 1 wskazuje, że w analizowanym okresie największe zmiany w pozycji na mapie podobieństw preferencji zaszły dla zajęć z języków obcych oraz gimnastyki rehabilitacyjnej. W roku 2009/2010 trudno jest zauważyć zajęcia, które dla słuchaczy uniwersytetu były traktowane pod względem preferencji jako podobne (wyjątek stanowią zajęcia z gimnastyki korekcyjnej oraz gimnastyki chińskiej). W okresie 2011–2013 grupą zajęć podobnych pod względem preferencji są zajęcia z gimnastyki rehabilitacyjnej, gimnastyki chińskiej oraz zajęcia na basenie.

4. Metoda wektorów dryfu

Metoda wektorów dryfu umożliwia równoczesną prezentację zarówno symetrycznej jak i skośnosymetrycznej części danych. Symetryczna część danych jest przedstawiona za pomocą konfiguracji punktów reprezentujących obiekty otrzymanej

w wyniku niemetrycznego skalowania wielowymiarowego. Skośnosymetryczną część obrazują wektory dryfu wskazujące kierunek i siłę zmian w preferencjach. Algorytm wyznaczania konfiguracji punktów reprezentujących obiekty (w przeprowadzonym badaniu są to poszczególne rodzaje zajęć) oraz wektorów dryfu jest następujący (zob. [Borg, Groenen 2005, s. 503]):

- dla niesymetrycznej macierzy podobieństw \mathbf{P} wyznaczyć symetryczną macierz \mathbf{A} ,

$$\text{gdzie: } a_{ij} = \frac{p_{ij} + p_{ji}}{2};$$

- dla symetrycznej macierzy \mathbf{A} przeprowadzić niemetryczne skalowanie wielowymiarowe;
- dla wszystkich punktów \mathbf{x}_i i \mathbf{x}_j (przy czym $i \neq j$), otrzymanych w wyniku skalowania wielowymiarowego wyznaczyć wektory $\mathbf{y}_{ij} = \mathbf{x}_j - \mathbf{x}_i$;
- unormować wektory \mathbf{y}_{ij} tak, aby ich długość była równa 1, tzn.:

$$\mathbf{y}_{ij}^* = \mathbf{y}_{ij} / (\mathbf{y}_{ij}^T \mathbf{y}_{ij})^{1/2};$$

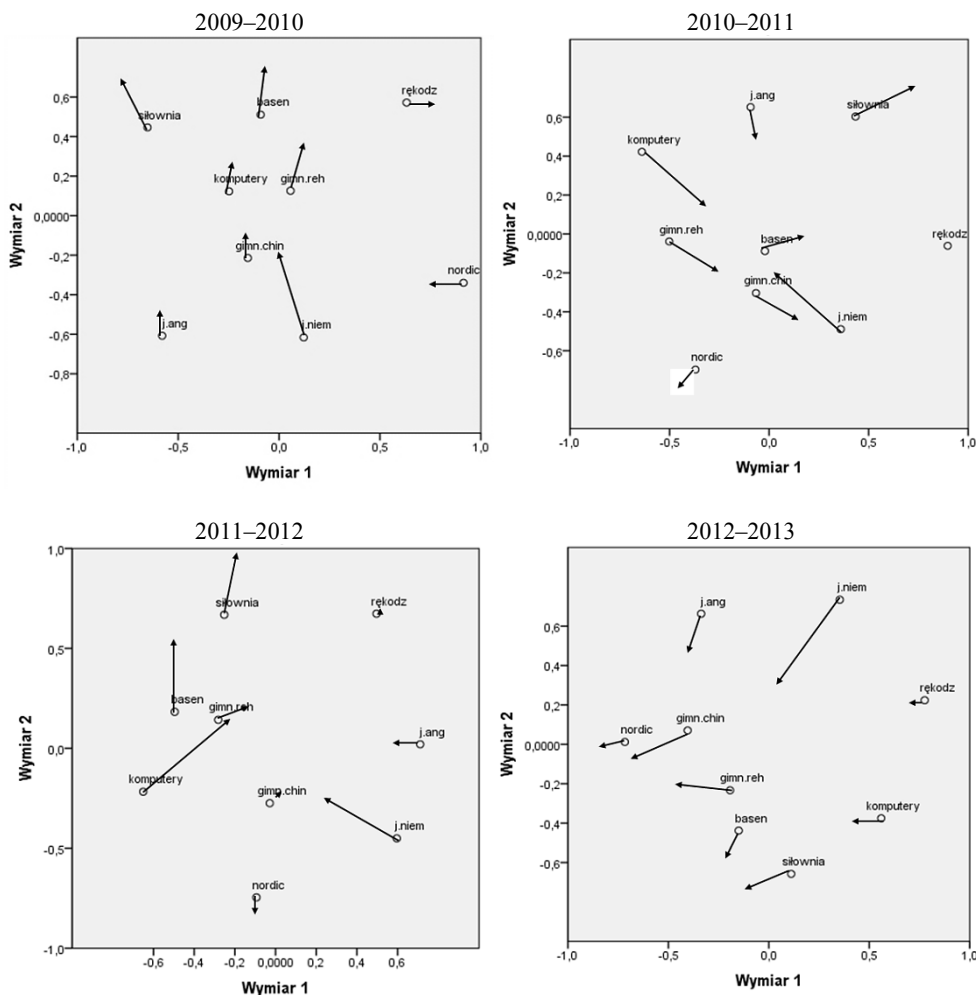
- wyznaczyć wektory $\mathbf{z}_{ij} = b_{ij} \cdot \mathbf{y}_{ij}^*$ (b_{ij} – elementy skośnosymetrycznej macierzy \mathbf{B});
- wyznaczyć wektory dryfu \mathbf{w}_i jako: $\mathbf{w}_i = n^{-1} \sum_j \mathbf{z}_{ij}$;

- wyznaczyć długości wektorów dryfu: $\|\mathbf{w}_i\| = \sqrt{w_{i1}^2 + w_{i2}^2}$;

- wyznaczyć kąty nachylenia wektorów dryfu do osi OY: $\alpha_i = \arccos(\mathbf{w}_i \mathbf{u}^T / \sqrt{\mathbf{w}_i \cdot \mathbf{w}_i^T})$, gdzie $\mathbf{w}_i = (w_{i1}, w_{i2})$, $\mathbf{u} = (0, 1)$.

Metodę wektorów dryfu zastosowano do danych zawartych w tab. 1, oddzielnie dla każdego z okresów. Wyniki analizy prezentuje rys. 2.

Rozkład punktów na mapach percepcyjnych wskazuje, że w latach 2009–2012 trudno jest dostrzec podobieństwa w preferencjach względem uwzględnionych w badaniu zajęć. Jednak na podstawie wektorów dryfu można stwierdzić, że preferencje słuchaczy uniwersytetu kierują się w stronę zajęć ruchowych. W miejsce zajęć rozwijających umiejętności językowe oraz obsługę komputera coraz bardziej preferowane są zajęcia poprawiające sprawność fizyczną. Jest to szczególnie widoczne w okresie 2009–2010, kiedy to zauważalny jest trend w zmianach preferencji w kierunku zajęć ruchowych, a zwłaszcza zajęć na basenie oraz gimnastyki rehabilitacyjnej. Długość wektorów dryfu wskazuje na siłę tych zmian. Z rysunku 2 wynika, że w analizowanym okresie największe zmiany w preferencjach widoczne są wśród słuchaczy biorących udział w zajęciach z j. niemieckiego, a w okresie 2010–2012 również w zajęciach z obsługi komputera. Z kolei prawie niezmiennie były preferencje odnośnie warsztatów malarskich i rękodzieła oraz marszów z kijkami. W okresie 2012–2013 można już wyraźnie wyodrębnić grupę zajęć traktowanych jako podobne pod względem preferencji, do której należą wszystkie zajęcia ruchowe.



Rys. 2. Konfiguracje punktów reprezentujących zajęcia oraz wektory dryfu

Źródło: opracowanie własne z wykorzystaniem pakietu PS IMAGO.

5. Zakończenie

W artykule przedstawiono dwie metody skalowania wielowymiarowego, które wykorzystano do badania zmian zachodzących w preferencjach. Podstawową wartością wynikającą z zastosowania obu metod jest możliwość geometrycznej prezentacji na mapie percepcyjnej podobieństw w preferencjach respondentów względem obiektów (w przedstawionym przykładzie były to różne formy zajęć). Skalowanie dynamiczne, dzięki dopasowaniu za pomocą analizy Procrustes konfiguracji wyznaczonych dla różnych okresów, zostało wykorzystane do przedstawienia

trajektorii zmian zachodzących w podobieństwach preferencji. Analiza niesymetrycznej części danych w metodzie wektorów dryfu umożliwiła dodatkowo wskazanie kierunku i siły zmian w preferencjach. Zastosowanie obu metod potwierdziło, że na początku badanego okresu trudno było zauważyć podobieństwa w preferencjach słuchaczy Uniwersytetu Trzeciego Wieku względem wybranych form zajęć, jednak z upływem czasu słuchacze mniej byli zainteresowani nauką języków obcych i obsługą komputera, a ich preferencje kierowały się w stronę zajęć poprawiających sprawność ruchową.

Literatura

- Borg I., Groenen P., 2005, *Modern Multidimensional Scaling. Theory and Applications*, Springer-Verlag, New York.
- Chino N., 1978, *A graphical technique for representing the asymmetric relationship between N objects*, *Behaviormetrika*, no. 5, s. 23–40.
- Cox T.F., Cox M.A.A., 2001, *Multidimensional Scaling*, Chapman and Hall, London.
- Harshman R.A., Green P.E., Wind Y., Lundy M.E., 1982, *A model for the analysis of asymmetric data in marketing research*, *Marketing Science*, vol. 1, no. 2, s. 205–242.
- Holyoak K.J., Gordon P.C., 1983, *Social reference points*, *Journal of Personality and Social Psychology*, vol. 44, no. 5, s. 881–887.
- Rothkopf E.Z., 1957, *A measure of stimulus similarity and errors in some paired-associate learning*, *Journal of Experimental Psychology*, vol. 53, no. 2, s. 94–101.
- Tversky A., Gati I., 1982, *Similarity, separability and the triangle inequality*, *Psychological Review*, vol. 89, no. 2, s. 123–154.
- Zaborski A., 2007, *Przegląd wybranych modeli różnic indywidualnych w skalowaniu wielowymiarowym*, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 1151, s. 9–18.