

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 384

Taksonomia 24

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

Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

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 stronie internetowej Wydawnictwa
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 2015

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
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
Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: 25 lat SKAD.....	15
Beata Basiura, Anna Czapkiewicz: Symulacyjne badanie wykorzystania entropii do badania jakości klasyfikacji.....	25
Andrzej Bąk: Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie <code>pllord</code>	33
Justyna Brzezińska: Analiza klas ukrytych w badaniach sondażowych.....	42
Grażyna Dehnel: Rejestr podatkowy oraz rejestr ZUS jako źródło informacji dodatkowej dla statystyki gospodarczej – możliwości i ograniczenia ..	51
Sabina Denkowska: Wybrane metody oceny jakości dopasowania w <i>Propensity Score Matching</i>	60
Marta Dziechciarz-Duda, Klaudia Przybysz: Zastosowanie teorii zbiorów rozmytych do identyfikacji pozafiskalnych czynników ubóstwa.....	75
Iwona Foryś: Potencjał rynku mieszkaniowego w Polsce w latach dekonjunktury gospodarczej.....	84
Eugeniusz Gatnar: Statystyczna analiza konwergencji krajów Europy Środkowej i Wschodniej po 10 latach członkostwa w Unii Europejskiej.....	93
Ewa Genge: Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa.....	100
Alicja Grześkowiak: Wielowymiarowa analiza uwarunkowań zaangażowania Polaków w kształcenie ustawiczne o charakterze pozaformalnym.....	108
Monika Hamerska: Wykorzystanie metod porządkowania liniowego do tworzenia rankingu jednostek naukowych.....	117
Bartłomiej Jefmański: Zastosowanie modeli IRT w konstrukcji rozmytego systemu wag dla zmiennych w zagadnieniu porządkowania liniowego – na przykładzie metody TOPSIS.....	126
Tomasz Józefowski, Marcin Szymkowiak: Wykorzystanie uogólnionej miary odległości do porządkowania liniowego powiatów województwa podkarpackiego w świetle funkcjonowania specjalnej strefy ekonomicznej Euro-Park Mielec.....	135
Krzysztof Kompa: Zastosowanie testów parametrycznych i nieparametrycznych do oceny sytuacji na światowym rynku kapitałowym przed kryzysem i po jego wystąpieniu.....	144
Mariusz Kubus: Rekurencyjna eliminacja cech w metodach dyskryminacji....	154

Marta Kuc: Wpływ sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego państw Unii Europejskiej pod względem poziomu życia ludności	163
Paweł Lula: Kontekstowy pomiar podobieństwa semantycznego	171
Iwona Markowicz: Model regresji Feldsteina-Horioki – wyniki badań dla Polski	182
Kamila Migdał-Najman: Ocena wpływu wartości stałej Minkowskiego na możliwość identyfikacji struktury grupowej danych o wysokim wymiarze	191
Małgorzata Misztal: O zastosowaniu kanonicznej analizy korespondencji w badaniach ekonomicznych.....	200
Krzysztof Najman: Zastosowanie przetwarzania równoległego w analizie skupień	209
Edward Nowak: Klasyfikacja danych a rachunkowość. Rozważania o relacjach	218
Marcin Pelka: Adaptacja metody <i>bagging</i> z zastosowaniem klasyfikacji pojęciowej danych symbolicznych.....	227
Józef Pocięcha, Mateusz Baryła, Barbara Pawelek: Porównanie skuteczności klasyfikacyjnej wybranych metod prognozowania bankructwa przedsiębiorstw przy losowym i nielosowym doborze prób	236
Agnieszka Przedborska, Małgorzata Misztal: Wybrane metody statystyki wielowymiarowej w ocenie jakości życia słuchaczy uniwersytetu trzeciego wieku	246
Wojciech Roszka: Konstrukcja syntetycznych zbiorów danych na potrzeby estymacji dla małych domen	254
Aneta Rybicka: Połączenie danych o preferencjach ujawnionych i wyrażonych	262
Elżbieta Sobczak: Poziom specjalizacji w sektorach intensywności technologicznej a efekty zmian liczby pracujących w województwach Polski	271
Andrzej Sokołowski, Grzegorz Harańczyk: Modyfikacja wykresu radarowego	280
Marcin Szymkowiak, Marek Witkowski: Wykorzystanie mediany do klasyfikacji banków spółdzielczych według stanu ich kondycji finansowej ..	287
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych – zastosowanie testu <i>join-count</i>	296
Dorota Witkowska: Wykorzystanie drzew klasyfikacyjnych do analizy zróżnicowania płac w Niemczech	305
Artur Zaborski: Analiza niesymetrycznych danych preferencji z wykorzystaniem modelu punktu dominującego i modelu grawitacji.....	315

Summaries

Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: XXV years of SKAD	24
Beata Basiura, Anna Czapkiewicz: Simulation study of the use of entropy to validation of clustering.....	32
Andrzej Bąk: Problem of choosing the optimal linear ordering procedure in the p_llord package.....	41
Justyna Brzezińska-Grabowska: Latent class analysis in survey research...	50
Grażyna Dehnel: Tax register and social security register as a source of additional information for business statistics – possibilities and limitations.....	59
Sabina Denkowska: Selected methods of assessing the quality of matching in Propensity Score Matching	74
Marta Dziechciarz-Duda, Klaudia Przybysz: Applying the fuzzy set theory to identify the non-monetary factors of poverty.....	83
Iwona Foryś: The potential of the housing market in Poland in the years of economic recessions.....	92
Eugeniusz Gatnar: Statistical analysis of the convergence of CEE countries after 10 years of their membership in the European Union.....	99
Ewa Genge: Trust to the public and financial institutions in the Polish society – an application of latent Markov models.....	107
Alicja Grześkowiak: Multivariate analysis of the determinants of Poles' involvement in non-formal lifelong learning	116
Monika Hamerska: The use of the methods of linear ordering for the creating of scientific units ranking.....	125
Bartłomiej Jefmański: The application of IRT models in the construction of a fuzzy system of weights for variables in the issue of linear ordering – on the basis of TOPSIS method	134
Tomasz Józefowski, Marcin Szymkowiak: GDM as a method of finding a linear ordering of districts of Podkarpackie Voivodeship in the light of the operation of the Euro-Park Mielec special economic zone	143
Krzysztof Kompa: Application of parametric and nonparametric tests to the evaluation of the situation on the world financial market in the pre- and post-crisis period.....	153
Mariusz Kubus: Recursive feature elimination in discrimination methods ...	162
Marta Kuc: The impact of the spatial weights matrix on the final shape of the European Union countries ranking due to the standard of living.....	170
Paweł Lula: The impact of context on semantic similarity.....	181
Iwona Markowicz: Feldstein-Horioka regression model – the results for Poland.....	190

Kamila Migdal-Najman: The assessment of impact value of Minkowski's constant for the possibility of group structure identification in high dimensional data.....	199
Małgorzata Misztal: On the use of canonical correspondence analysis in economic research.....	208
Krzysztof Najman: The application of the parallel computing in cluster analysis.....	217
Edward Nowak: Data classification and accounting. A study of correlations	226
Marcin Pelka: The adaptation of bagging with the application of conceptual clustering of symbolic data.....	235
Józef Pociecha, Mateusz Baryła, Barbara Pawelek: Comparison of classification accuracy of selected bankruptcy prediction methods in the case of random and non-random sampling technique.....	244
Agnieszka Przedborska, Małgorzata Misztal: Selected multivariate statistical analysis methods in the evaluation of the quality of life of the members of the University of the Third Age.....	253
Wojciech Roszka: Construction of synthetic data sets for small area estimation.....	261
Aneta Rybicka: Combining revealed and stated preference data.....	270
Elżbieta Sobczak: Specialization in sectors of technical advancement vs. effects of workforce number changes in Poland's voivodships.....	279
Andrzej Sokółowski, Grzegorz Harańczyk: Modification of radar plot.....	286
Marcin Szymkowiak, Marek Witkowski: Classification of cooperative banks according to their financial situation using the median.....	295
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: The influence of classification method selection on the identification of spatial dependence – an application of join-count test.....	304
Dorota Witkowska: Application of classification trees to analyze wages disparities in Germany.....	314
Artur Zaborski: Asymmetric preference data analysis by using the dominance point model and the gravity model.....	323

Artur Zaborski

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

ANALIZA NIESYMETRYCZNYCH DANYCH PREFERENCJI Z WYKORZYSTANIEM MODELU PUNKTU DOMINUJĄCEGO I MODELU GRAWITACJI

Streszczenie: W artykule zaprezentowano dwa modele analizy niesymetrycznych danych preferencji: model punktu dominującego i model grawitacyjny. W modelu punktu dominującego na jednej mapie percepcyjnej przedstawia się punkty reprezentujące badane obiekty oraz punkt reprezentujący hipotetyczny obiekt posiadający największą przewagę konkurencyjną. Model grawitacyjny został wykorzystany do graficznej prezentacji zmian preferencji w czasie. Jego zaletą jest to, że może być stosowany do porównania w różnych okresach dwóch, niekoniecznie dokładnie takich samych, zbiorów obiektów.

Słowa kluczowe: dane niesymetryczne, badanie preferencji, skalowanie wielowymiarowe, model punktu dominującego, model grawitacyjny.

DOI: 10.15611/pn.2015.384.34

1. Wstęp

Większość badań wykorzystujących metody skalowania wielowymiarowego przeprowadzana jest na podstawie symetrycznej macierzy niepodobieństw δ_{ij} obrazujących odległości między obiektami O_i oraz O_j , gdzie $i, j = 1, 2, \dots, n$ (n – liczba obiektów)¹. W badaniach preferencji macierz niepodobieństw często nie jest symetryczna. Jeżeli niesymetryczność wynika np. z błędu pomiaru lub z tendencji ocen respondentów, najbardziej popularnym zabiegiem jest uśrednienie niediagonalnych elementów macierzy.

Jednak istnieje szeroka gama problemów badawczych, dla których dane zawarte w macierzy niepodobieństw z założenia nie spełniają warunku symetryczności, tzn.

¹ Macierz ta jest nazywana macierzą odległości, jeżeli niepodobieństwo obiektów mierzy się za pomocą miar odległości, natomiast macierzą bliskości, gdy niepodobieństwa mierzy się za pomocą miar bliskości.

$\delta_{ij} \neq \delta_{ji}$ dla $i \neq j$ (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 i j -tej kolumnie przedstawia liczbę osób, którzy w okresie $t-1$ wybierali markę i , a w okresie t wybierają markę j . Przedstawioną macierz należy traktować jako macierz podobieństw, ponieważ wysoka wartość elementu macierzy oznacza, że konsumenci chętnie dokonują zamiany marki, a tym samym, zgodnie ze swoimi preferencjami, traktują je jako podobne. Dla takich danych symetryzacja macierzy poprzez uśrednienie odpowiednich wielkości niepodobieństw prowadzi do utraty cennych informacji dotyczących analizowanego zjawiska. Tym samym konieczne jest zastosowanie metod właściwych dla danych niesymetrycznych.

W artykule zaprezentowano dwa modele analizy niesymetrycznych danych preferencji: model punktu dominującego i model grawitacyjny. Zastosowanie modelu punktu dominującego pozwala na wskazanie na mapie percepcyjnej hipotetycznej marki posiadającej przewagę konkurencyjną na rynku. Model grawitacyjny oparty na analizie *unfolding* zostanie wykorzystany do badania zmian preferencji w czasie. Jego zaletą jest to, że może być stosowany do porównania w różnych okresach dwóch, niekoniecznie dokładnie takich samych, zbiorów obiektów.

2. Model punktu dominującego

W modelu punktu dominującego na mapie percepcyjnej dokonuje się rozmieszczenia punktów $\mathbf{x}_i, \mathbf{x}_j$ ($i, j = 1, 2, \dots, n$) reprezentujących obiekty oraz punktu dominującego \mathbf{y} reprezentującego obiekt hipotetyczny. Punkt dominujący wyraża zjawisko asymetrii dla całej macierzy niepodobieństw (zob. rys. 1). W modelu tym przyjmuje się, że podobieństwa p_{ij} są malejącą funkcją skorygowanych odległości m_{ij} (zob. [Okada, Imaizumi 2007]):

$$p_{ij} = f(m_{ij}), \quad (1)$$

gdzie: $m_{ij} = m(\mathbf{x}_i, \mathbf{x}_j) = d_{ij} - u(d_i - d_j)$ – skorygowana odległość od \mathbf{x}_i do \mathbf{x}_j ,

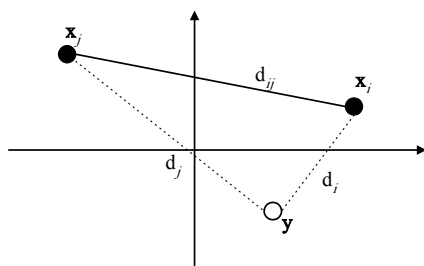
$$d_{ij} = \sqrt{\sum_{a=1}^s (x_{ia} - x_{ja})^2} \quad \text{– odległość między } \mathbf{x}_i \text{ a } \mathbf{x}_j,$$

$$d_k = \sqrt{\sum_{a=1}^s (y_a - x_{ka})^2} \quad \text{– odległość między } \mathbf{x}_k \text{ (} k = i, j \text{) a punktem dominującym } \mathbf{y},$$

gdzie:

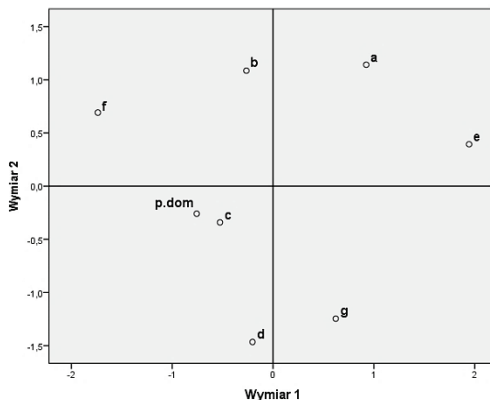
s – wymiar przestrzeni prezentacji wyników,

u ($u > 0$) – waga określająca znaczenie asymetrii w relacjach między obiektami.



Rys. 1. Wspólna konfiguracja dwóch punktów reprezentujących obiekty oraz punktu dominującego
 Źródło: opracowanie własne na podstawie [Okada, Imaizumi 2007].

Jeżeli $d_i < d_j$, to skorygowana odległość x_i od x_j jest większa niż odległość x_j od x_i ($m(x_i, x_j) > m(x_j, x_i)$), co oznacza, że j -ty obiekt jest bardziej podobny do i -tego obiektu niż obiekt i -ty do obiektu j -tego. W związku z tym, skoro $d_i < d_j$, to $m(y, x_j) > m(x_j, y)$, tym samym j -ty obiekt jest bardziej podobny do hipotetycznego obiektu reprezentowanego przez punkt dominujący niż obiekt hipotetyczny do obiektu j -tego. Ponieważ zależność taka zachodzi dla wszystkich par punktów reprezentujących obiekty dlatego punkt dominujący reprezentuje hipotetyczny obiekt, którego podobieństwo do wszystkich pozostałych obiektów jest zawsze mniejsze niż podobieństwo pozostałych punktów do punktu hipotetycznego. W badaniach marketingowych punkt dominujący będzie więc przedstawiał hipotetyczną markę posiadającą największą przewagę konkurencyjną, tzn. większa część konsumentów będzie skłonna zrezygnować z każdej marki na rzecz marki hipotetycznej. Przykład konfiguracji punktów, z której wynika, że największą przewagę konkurencyjną posiada marka c , przedstawia rys. 2.



Rys. 2. Konfiguracja punktów reprezentujących wybrane marki oraz punkt dominujący
 Źródło: opracowanie własne.

Wspólna konfiguracja punktów reprezentujących obiekty i punkt dominujący oraz wagi u wyznaczone są za pomocą niemetrycznego skalowania wielowymiarowego na podstawie macierzy uśrednionych ocen preferencji $C = [p_{ij}^*]$, gdzie $p_{ij}^* = \frac{p_{ij} + p_{ji}}{2}$ oraz dla $\mathbf{y}^{(0)}$ znajdującego się w początku układu współrzędnych (por. [Kruskal 1964; Okada, Imaizumi 2007]).

W kolejnych cyklach iteracyjnych wspólna konfiguracja punktów oraz wagi minimalizujące wartość funkcji dopasowania STRESS:

$$S = \sqrt{\sum_{\substack{i,j=1 \\ i \neq j}}^n (m_{ij} - \hat{m}_{ij})^2 / \sum_{\substack{i,j=1 \\ i \neq j}}^n (m_{ij} - \bar{m})^2}, \quad (2)$$

gdzie: \hat{m}_{ij} – monotoniczna funkcja regresji między m_{ij} a p_{ij} ,

$$\bar{m} = \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^n \hat{m}_{ij},$$

są wyznaczone z wykorzystaniem metody gradientowej. Po K -tym cyklu iteracyjnym wynoszą one:

$$\begin{aligned} x_{ia}^K &= x_{ia}^{K-1} - \alpha \frac{\delta S}{\delta x_{ia}}, \\ y_a^K &= y_a^{K-1} - \alpha \frac{\delta S}{\delta y_a}, \\ u^K &= u^{K-1} - \alpha \frac{\delta S}{\delta u}, \end{aligned} \quad (3)$$

przy czym α jest współczynnikiem kroku zmieniającym się w każdym cyklu iteracyjnym (zob. np. [Bąk 2004, s. 101]).

3. Model grawitacyjny

Kolejny model analizy danych niesymetrycznych to model grawitacyjny oparty na analizie *unfolding*. W modelu tym na jednej mapie percepcyjnej przedstawiane są punkty reprezentujące zarówno wiersze macierzy preferencji, jak i punkty reprezentujące jej kolumny. W przeciwieństwie do klasycznej analizy *unfolding*, w której przyjmuje się założenie, że preferencje respondentów determinowane są przez takie same wymiary, w modelu grawitacyjnym uwzględnia się dodatkowe czynniki, które mają wpływ na preferencje. Są nimi m.in. udział w rynku marek, wartość

marki, wielkość dochodów, lojalność względem marki. Te dodatkowe czynniki w modelu grawitacyjnym określa się jako masy. W zależności od tego, czy są to czynniki wynikające z własności badanych obiektów czy respondentów, nazywa się je masą obiektu bądź masą respondenta.

Model grawitacyjny określony jest równaniem (por. [DeSarbo i in. 2002]):

$$p_{ij} = \varphi_i \left(\frac{M_j^\beta S_i^\chi}{d_{ij}^2} \right), \quad (4)$$

gdzie: p_{ij} – ocena preferencji i -tego respondenta względem j -tego obiektu,

φ_i – pewna monotonicznie nie malejąca funkcja dla i -tego respondenta,

$d_{ij} = \sqrt{\sum_{a=1}^s (x_{ja} - y_{ia})^2}$ – odległość między punktem obiektu (kolumny) \mathbf{x}_j

a punktem respondenta (wiersza) \mathbf{y}_i ,

M_j – masa j -tego obiektu,

S_i – masa i -tego respondenta,

β i χ – parametry masy.

W modelu grawitacyjnym przyjmujemy założenie, że oceny preferencji są wprost proporcjonalne do iloczynu mas, a odwrotnie proporcjonalne do odległości między punktami reprezentującymi wiersze i punktami reprezentującymi kolumny macierzy preferencji na mapie percepcyjnej.

Algorytm grawitacyjnej analizy *unfolding* ma charakter iteracyjny i jest analogiczny jak w przypadku innych metod niemetrycznego skalowania wielowymiarowego. Konfiguracje punktów w kolejnych cyklach iteracyjnych wyznaczane są za pomocą metody gradientowej (zob. [Kruskal 1964; Imaizumi 2005]). Jeżeli znane są wartości mas, nowe konfiguracje punktów można również wyznaczyć za pomocą transformacji Guttmana (zob. [Borg, Groenen 2005, s. 297-298; Zaborski 2011]).

4. Przykład empiryczny

Celem badania była analiza zmian zachodzących w preferencjach słuchaczy Uniwersytetu Trzeciego Wieku w Bolesławcu (woj. dolnośląskie) odnośnie do chęci uczestnictwa w określonych formach zajęć w okresie od 2007 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, zajęcia na basenie oraz nordic walking (marsze z kijkami). Nie uwzględniono specjalistycznych zajęć rozwijających zainteresowania i umiejętności słuchaczy, takich jak warsztaty malar-

skie, rękodzieło czy warsztaty muzyczne, oraz cyklicznych otwartych wykładów tematycznych. Badaniem objęto 109 słuchaczy uniwersytetu, którzy niezmiennie uczestniczyli w zajęciach w analizowanym okresie. Ponieważ zajęcia z gimnastyki chińskiej i jogi oraz zajęcia na basenie rozpoczęły się dopiero od roku 2009, badanie podzielono na dwa podokresy, tj. lata 2007-2009 i 2009-2013.

W tabelach 1 a i b wskazano, jakie zmiany zaszły w dokonywanych przez słuchaczy uniwersytetu wyborach zajęć. Element i -tego wiersza i j -tej kolumny przedstawia liczbę słuchaczy, którzy na początku analizowanego okresu wybierali i -te zajęcia, zaś na koniec okresu j -te zajęcia.

Tabela 1. Macierze podobieństw między wybranymi zajęciami

a) lata 2007 – 2009								
Rodzaje zajęć	j.ang	j.niem	komp	g.reh	g.chin	basen	n.walk	m_j
j. angielski	17	4	4	1	0	4	1	32
j. niemiecki	2	10	1	2	1	8	1	25
obsługa komputera	4	2	22	4	1	5	1	39
gimnastyka rehabilitacyjna	1	1	1	35	1	17	2	58
nordic walking	0	0	0	1	1	1	4	7
m_i	24	17	28	43	5	35	9	
a) lata 2009 – 2013								
Rodzaje zajęć	j.ang	j.niem	komp	g.reh	g.chin	basen	n.walk	m_j
j. angielski	17	0	0	3	4	4	3	31
j. niemiecki	0	8	0	2	7	10	3	30
obsługa komputera	1	3	6	5	4	12	1	32
gimnastyka rehabilitacyjna	0	0	0	45	4	6	2	57
gimnastyka chińska	0	0	0	0	6	3	2	11
basen	0	0	0	1	0	29	1	31
nordic walking	0	0	0	1	0	0	9	10
m_i	18	11	6	57	25	64	21	

Źródło: opracowanie własne.

Podobieństwa zawarte w tabeli 1 przekształcono w niepodobieństwa za pomocą funkcji grawitacji (zob. tab. 2):

$$\delta_{ij} = \left(\frac{m_i \cdot m_j}{p_{ij}} \right)^{\frac{1}{2}}. \quad (5)$$

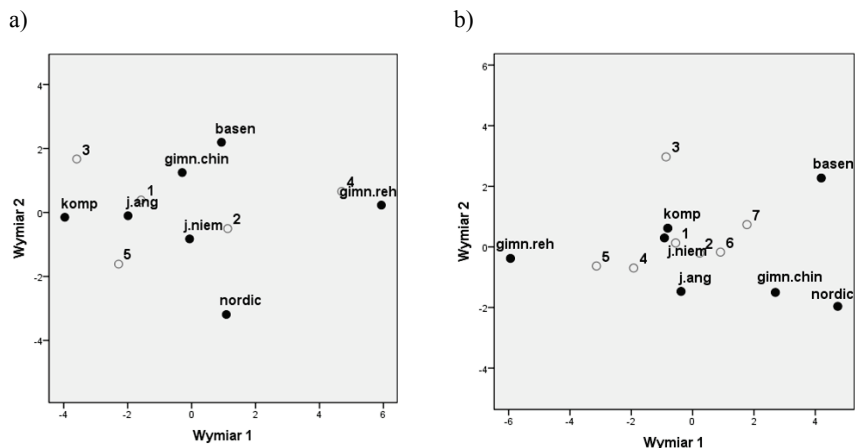
Jako masy przyjęto liczby słuchaczy uczestniczących w określonych zajęciach. Na podstawie wyznaczonych macierzy niepodobieństw przeprowadzono dla każdego okresu analizę *unfolding*, w wyniku czego otrzymano konfiguracje punktów reprezentujących wybrane rodzaje zajęć (zob. rys. 3). Ponieważ masy obiektów były znane, do wyznaczania w kolejnych cyklach iteracyjnych konfiguracji punktów w analizie *unfolding* wykorzystano transformację Guttmana.

Tabela 2. Macierze niepodobieństw między wybranymi zajęciami

a) lata 2007-2009							
Rodzaje zajęć	j. ang.	j. niem.	komp.	g. reh.	g. chin.	basen	n. walk.
j. angielski	6,72	11,66	14,97	37,09	12,65	16,73	16,97
j. niemiecki	17,32	6,52	26,46	23,18	11,18	13,29	15,00
obsługa komputera	15,30	18,21	7,04	20,48	13,96	16,52	18,73
gimnastyka rehabilitacyjna	37,31	31,40	40,30	8,44	17,03	10,93	16,16
nordic walking	0	0	0	17,35	5,92	15,65	3,97
a) lata 2009 – 2013							
Rodzaje zajęć	j. ang.	j. niem.	komp.	g. reh.	g. chin.	basen	n. walk.
j. angielski	5,73	0	0	24,27	13,92	22,27	14,73
j. niemiecki	0	6,42	0	29,24	10,35	13,86	14,49
obsługa komputera	24,00	10,83	5,66	19,10	14,14	13,06	25,92
gimnastyka rehabilitacyjna	0	0	0	8,50	18,87	24,66	24,46
gimnastyka chińska	0	0	0	0	6,77	15,32	10,75
basen	0	0	0	42,03	0	8,27	25,51
nordic walking	0	0	0	23,87	0	0	4,83

Źródło: opracowanie własne.

Rozkłady punktów na mapach percepcyjnych (rys. 3 a i b) wskazują, że w latach 2007-2009 nie zaszły znaczące zmiany w preferencjach słuchaczy. Pojawiły się nowe zajęcia (gimnastyka chińska i zajęcia na basenie), ale punkty reprezentu-



Rys. 3. Konfiguracje punktów reprezentujących zajęcia a) o w 2007 r. (1 – j. angielski, 2 – j. niemiecki, 3 – obsługa komputera, 4 – gimnastyka rehabilitacyjna, 5 – nordic walking) i ● w 2009 r.; b) o w 2009 r. (1 – j. angielski, 2 – j. niemiecki, 3 – obsługa komputera, 4 – gimnastyka rehabilitacyjna, 5 – gimnastyka chińska, 6 – basen, 7 – nordic walking) i ● w 2013 r.

Źródło: opracowanie własne.

jące pozostałe rodzaje zajęć na początku i na końcu analizowanego okresu leżą blisko siebie. Odmienna sytuacja ma miejsce w drugim z badanych okresów. Jeżeli w 2009 r. trudno jest zauważyć podobieństwa w preferencjach względem uwzględnionych w badaniu zajęć, to w 2013 r. widoczne są grupy zajęć, które pod względem preferencji oceniane są jako podobne. W pierwszej grupie znalazły się j. angielski, j. niemiecki i obsługa komputera, a w drugiej gimnastyka chińska i nordic walking. Zajęcia bardzo popularne, jednak różniące się pod względem podobieństwa preferencji, to gimnastyka rehabilitacyjna oraz zajęcia na basenie.

5. Zakończenie

W artykule zaprezentowano dwie metody skalowania wielowymiarowego dla niesymetrycznych danych preferencji. W porównaniu do innych metod analizy danych niesymetrycznych (zob. np. [Harshman i in. 1982; Chino 1978; Borg, Groenen 2005, s. 497-503]) ich zaletą jest to, że posiadają łatwą interpretację geometryczną. Zastosowanie modelu punktu dominującego jest szczególnie uzasadnione w badaniach pozycji na rynku, gdyż pozwala wskazać na mapie percepcyjnej taki obiekt, który posiada przewagę konkurencyjną. Z kolei model grawitacyjny oparty na analizie *unfolding* umożliwia analizę zmian zachodzących w preferencjach. Ponieważ analizy dokonuje się na podstawie prostokątnej macierzy podobieństw, jego dodatkowym atutem jest to, że może być wykorzystywany, gdy w różnych okresach ocenie podlegają niekoniecznie takie same zbiory obiektów.

Literatura

- Bąk A., 2004, *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Borg, I., Groenen, P., 2005, *Modern Multidimensional Scaling. Theory and Applications. Second Edition*, Springer-Verlag, New York.
- Chino N., 1978, *A graphical technique for representing the asymmetric relationship between N objects*, *Behaviormetrika*, no 5, s. 23-40.
- DeSarbo W.S., Kim J., Choi S. C., Spaluding M., 2002, *A gravity-based multidimensional scaling model for deriving spatial structures underlying consumer preference/choice judgments*, *Journal of Consumer Research*, vol. 29, s. 91-100.
- 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*, no 44, s. 881-887.
- Imaizumi T., 2005, *A Gravity-Based Multidimensional Unfolding Model for Preference Data*, [w:] Baier D., Wernecke K.D. (red.), *Innovations in Classification, Data Science, and Information Science*, Springer-Verlag, Berlin, Heidelberg, New York, s. 238-245.
- Kruskal J.B., 1964, *Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis*, *Psychometrika*, no. 29, s. 1-27.

- Okada A., Imaizumi T., 2007, *Multidimensional scaling of asymmetric proximities with a dominance point*, [w:] Decker R., Lenz H.J. (red.), *Advances in Data Analysis Studies in Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag, Berlin, Heidelberg, s. 307-318.
- Tversky A., Gati I., 1982, *Features of similarity*, *Psychological Review*, no 89, s. 123-154.
- Zaborski A., 2011, *Zastosowanie algorytmu SMACOF do badań opartych na prostokątnej macierzy preferencji*, [w:] Jajuga K., Walesiak M. (red.), *Taksonomia 18, Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 176, Wrocław, s. 262-271.

ASYMMETRIC PREFERENCE DATA ANALYSIS BY USING THE DOMINANCE POINT MODEL AND THE GRAVITY MODEL

Summary: The paper presents two models of preference asymmetric data analysis: the dominance point model and the gravity model. In the dominance point model a joint configuration consists of points representing objects and the point which represents the hypothetical object having the strongest competitive power. The gravity model was used for the graphical presentation of the changes in respondents' preferences in a definite period. Its advantage is that it can be used to compare two, not necessarily exactly the same sets of objects in different periods of time.

Keywords: asymmetric data, preference analysis, multidimensional scaling, dominance point model, gravity model.