

Artur Zaborski

Uniwersytet Ekonomiczny we Wrocławiu

WIELOWYMIAROWE SKALOWANIE PREFERENCJI Z WYKORZYSTANIEM MODELU PREFSCAL

1. Wstęp

Wśród metod, których wyniki są prezentowane w formie graficznej, w analizie preferencji duże znaczenie ma analiza *unfolding*, będąca szczególnym przypadkiem skalowania wielowymiarowego. Zakłada ona, że respondenci postrzegają wybrane obiekty w tej samej przestrzeni, ale różnią się w ocenie ważności ich poszczególnych atrybutów. Celem analizy jest odkrycie, na podstawie danych preferencji, wspólnej przestrzeni punktów reprezentujących respondentów i badane obiekty, która pozwala na ocenę zależności występujących pomiędzy respondentami, obiektami oraz pomiędzy respondentami a obiektami (w zależności od celu badania możliwa jest również demonstracja punktów reprezentujących respondentów i zmienne lub obiekty i zmienne).

Dla danych preferencji mierzonych na skali porządkowej wyniki skalowania wielowymiarowego często są nieinterpretowalne. Prawdziwa struktura danych wyjściowych pozostaje ukryta, mimo że bliska zeru wartość funkcji dopasowania STRESS wskazuje na doskonałe dopasowanie konfiguracji punktów do danych wejściowych. Prezentowany model PREFSCAL umożliwia uniknięcie rozwiązań trywialnych przez wykorzystanie w konstrukcji funkcji dopasowania STRESS współczynnika zmienności Pearsona.

Celem artykułu jest przedstawienie problemu pojawiania się trywialnych rozwiązań w analizie *unfolding*, możliwości ich uniknięcia przez modyfikację funkcji dopasowania oraz wielowariantowego wykorzystania modelu PREFSCAL w badaniach preferencji.

2. Zasady analizy *unfolding*

Analiza *unfolding* jest metodą, w której na podstawie indywidualnych ocen preferencji dokonuje się rozmieszczenia na jednej mapie percepcyjnej punktów reprezentujących respondentów oraz badane obiekty. Punkty oznaczające respondentów

lub grupy respondentów oznaczają tzw. punkty idealne, czyli punkty reprezentujące obiekt mający najbardziej preferowaną przez respondenta kombinację cech. Pozycja punktów reprezentujących badane obiekty w stosunku do punktu idealnego określa względne preferencje respondentów, tzn. ich odległości od punktu idealnego mają takie samo uporządkowanie jak rangowe uporządkowanie preferencji.

W skalowaniu wielowymiarowym wyznaczanie punktów idealnych może odbywać się w sposób bezpośredni lub pośredni. W pierwszym podejściu prosi się respondentów o uwzględnienie w badanych obiektach hipotetycznego obiektu idealnego ze względu na badane cechy, a następnie o porównanie pozostałych obiektów z obiektem idealnym. Bezpośrednie wyznaczanie punktów idealnych może sprawiać pewne problemy, ponieważ respondent często wyobraża sobie obiekt idealny jako nieistniejący obiekt abstrakcyjny, mający maksymalne realizacje wymienionych cech lub jako zbliżony do najbardziej preferowanego ze zbioru badanych obiektów. Należy również pamiętać, że respondent powinien oceniać obiekty w kategoriach preferencji, a nie podobieństwa, co jest utrudnione w wypadku uwzględniania nieznanymi obiektów. Powyższe trudności sprawiają, że badacze do wyznaczania punktów idealnych chętniej wykorzystują metodę pośrednią.

W pośrednim podejściu możliwe są dwa warianty wyznaczania punktu idealnego: zewnętrzny (*external*) i wewnętrzny (*internal*). W pierwszym skalowanie wielowymiarowe obejmuje dwa etapy. Najpierw wyznacza się konfigurację punktów reprezentujących obiekty na podstawie danych podobieństwa. Następnie, wykorzystując dane preferencji od jednego respondenta lub kilku respondentów, dokonuje się takiego rozmieszczenia punktów idealnych na mapie percepcyjnej, aby uporządkowanie odległości punktu idealnego od pozostałych punktów odpowiadało uporządkowaniu preferencji respondentów.

W drugim sposobie do skalowania wielowymiarowego wykorzystuje się tylko dane preferencji. Na ich podstawie na jednej mapie percepcyjnej dokonuje się rozmieszczenia zarówno punktów reprezentujących badane obiekty, jak i punktów idealnych.

Celem analizy *unfolding* jest znalezienie dla ustalonych ocen preferencji δ_{ij} takiego odwzorowania φ , dla którego:

$$d_{ij} \approx \hat{d}_{ij} = f(\delta_{ij}), \quad (1)$$

gdzie: $d_{ij} = \sqrt{\sum_{a=1}^r (x_{ia} - y_{ja})^2}$ – odległość między \mathbf{x}_i a \mathbf{y}_j ,

x_{ia} – współrzędne punktów idealnych,

y_{ja} – współrzędne punktów reprezentujących obiekty,

$i = 1, \dots, n$ (n – liczba punktów idealnych), $j = 1, \dots, m$ (m – liczba obiektów),

$a = 1, \dots, r$ (r – wymiar przestrzeni, w której prezentowane są wyniki analizy),

\hat{d}_{ij} – monotoniczna funkcja regresji między d_{ij} a δ_{ij} .

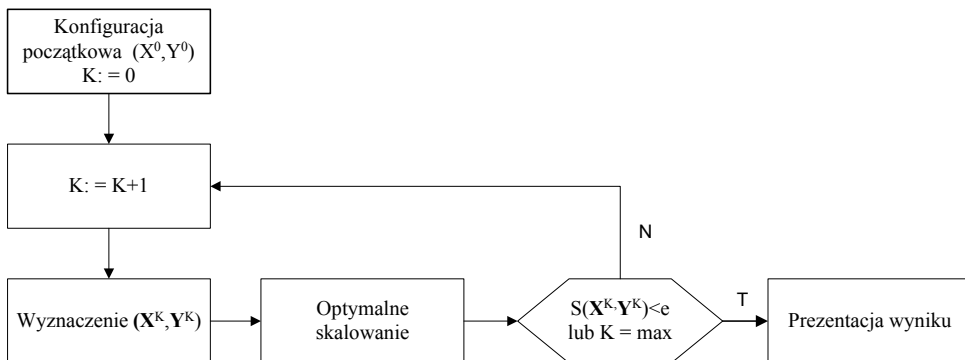
Dla ocen preferencji mierzonych na skali porządkowej funkcja φ spełnia zależność [Takane, Young, de Leeuw 1977]:

$$\delta_{ij} < \delta_{i'j'} \Rightarrow \hat{d}_{ij} \leq \hat{d}_{i'j'}. \quad (2)$$

W analizie *unfolding* współrzędne punktów \mathbf{X} i \mathbf{Y} wyznaczone są tak, aby minimalizowały wartość funkcji dopasowania STRESS¹:

$$S = \sqrt{\frac{\sum_{i,j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j} d_{ij}^2}}. \quad (3)$$

Na algorytm skalowania składają się cztery fazy (zob. rys. 1): dwie fazy podstawowe (faza optymalnego skalowania i faza estymacji) oraz dwie fazy pomocnicze (faza wstępna i faza końcowa).



Rys. 1. Algorytm skalowania w analizie *unfolding*

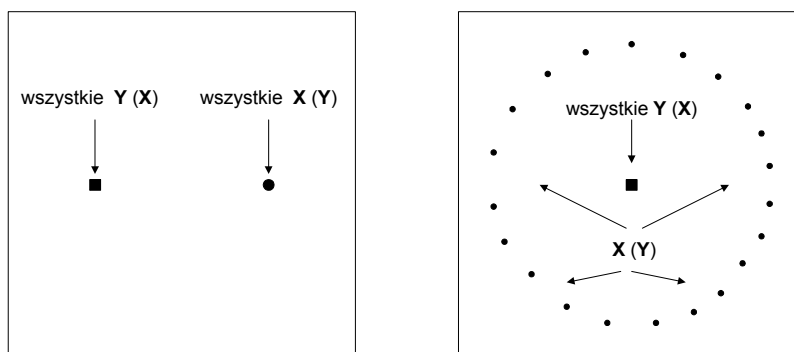
Źródło: opracowano na podstawie pracy [Zaborski 2001, s. 76].

W fazie wstępnej na podstawie danych preferencji wyznacza się (np. za pomocą klasycznego skalowania wielowymiarowego) początkową konfigurację punktów.

W fazie optymalnego skalowania dla konfiguracji początkowej lub dla konfiguracji punktów wyznaczonych w poprzednim cyklu iteracyjnym oblicza się odległości między punktami, a następnie wyznacza funkcję regresji tak, aby odległości między punktami oraz odpowiednie niepodobieństwa spełniały warunek monotoniczności, gwarantując jednocześnie minimalizację wartości funkcji dopasowania (por. [Zaborski 2001, s. 78-79]).

¹ Inne postacie funkcji dopasowania prezentują m.in. [Borg, Groenen 2005, s. 251-253; Zaborski 2001, s. 53].

W fazie końcowej obliczana jest wartość funkcji STRESS. Jeżeli jest ona mniejsza od założonej z góry wartości progowej (np. 0,1), następuje prezentacja wyników. Jeżeli nie – przechodzi się do fazy estymacji, w której wyznaczana jest nowa konfiguracja punktów minimalizująca wartość funkcji dopasowania (zob. [Borg, Groenen 2005, s. 298; Zaborski 2001]).



Rys. 2. Zdegenerowane rozwiązania w analizie *unfolding*

Źródło: opracowano na podstawie pracy [Borg, Groenen 2005, s. 301].

W analizie *unfolding* bardzo często pojawia się problem zdegenerowanych rozwiązań. Oznacza to, że uzyskujemy wyniki analiz, w których wartość funkcji STRESS jest równa zero, a struktura danych wejściowych jest nadal ukryta i nieinterpretowalna. Rozwiązanie (\mathbf{X}, \mathbf{Y}) nazywamy całkowicie zdegenerowanym, jeżeli dla każdego i, j $d_{ij} = d$ (d – pewna dodatnia stała). Tym samym wszystkie, spełniające warunek (2), przekształcone odległości $\hat{d}_{ij} = d$ (por. np. [Cox, Cox 2001, s. 65; Zaborski 2001, s. 79]). Przyjmuje się jednocześnie, że nie wszystkie oceny preferencji są jednakowe, tzn. istnieją takie i, j oraz i', j' , dla których $\delta_{ij} \neq \delta_{i'j'}$. Przykłady całkowicie zdegenerowanych rozwiązań w analizie *unfolding* prezentuje rys. 2.

3. Model PREFSCAL

Model PREFSCAL, opracowany przez Data Theory Group Uniwersytetu w Leiden, jest rozwinięciem algorytmu PROXSCAL, który do minimalizacji funkcji dopasowania w kolejnych cyklach iteracyjnych wykorzystuje metodę majoryzacji. W modelu PREFSCAL uniknięcie równych wartości \hat{d}_{ij} , a tym samym zdegenerowanego rozwiązania, osiągnięto przez modyfikację funkcji dopasowania STRESS.

W konstrukcji miary dopasowania [Busing, Groenen, Heiser 2005] zaproponowano wykorzystanie współczynnika zmienności Pearsona:

$$v(\hat{d}) = \frac{\sqrt{K^{-1} \sum_{k=1}^K (\hat{d}_k - \bar{\hat{d}})^2}}{\bar{\hat{d}}}, \quad (4)$$

gdzie $\bar{\hat{d}} = K^{-1} \sum_{k=1}^K \hat{d}_k$ (K – liczba wszystkich \hat{d}_k).

Współczynnik zmienności traktowany jest jako diagnostyka służąca identyfikacji rozwiązań o równych wartościach \hat{d}_{ij} .

Model PREFSCAL wykorzystuje współczynnik zmienności do „karania” funkcji dopasowania dla równych wartości \hat{d}_{ij} . W omawianym modelu funkcja dopasowania przyjmuje postać:

$$S_{pref}(\hat{\mathbf{d}}, \mathbf{X}, \mathbf{Y}) = S^\lambda(\hat{\mathbf{d}}, \mathbf{X}, \mathbf{Y}) \left(1 + \frac{\omega}{v^2(\hat{\mathbf{d}})} \right), \quad (5)$$

gdzie: $S(\hat{\mathbf{d}}, \mathbf{X}, \mathbf{Y})$ – znormalizowana funkcja STRESS zdefiniowana równaniem (3),

$\left. \begin{array}{l} 0 < \lambda < 1 \\ \omega > 0 \end{array} \right\}$ – parametry karania,

λ – siła kary,

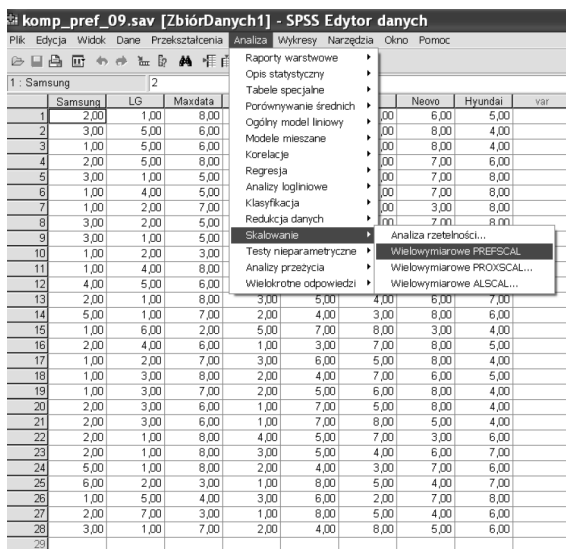
ω – zakres działania siły kary.

4. PREFSCAL w pakiecie statystycznym SPSS

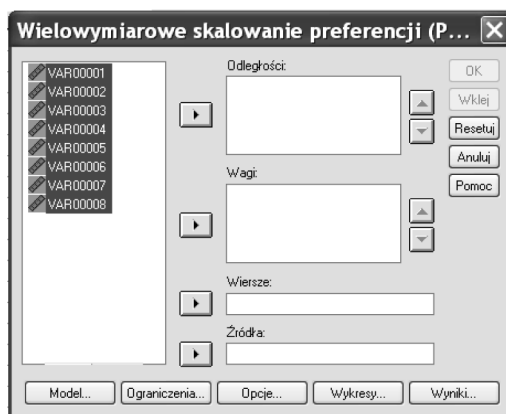
PREFSCAL jest dostępny w pakiecie statystycznym SPSS od wersji 14.0. Sposób wywoływania procedury skalowania wielowymiarowego za pomocą modelu PREFSCAL w trybie menu i okien dialogowych przedstawia rys. 3.

Po wybraniu opcji *Analiza/Skalowanie/Wielowymiarowe PREFSCAL* zostaje wyświetlone okno dialogowe *Wielowymiarowe skalowanie preferencji (PREFSCAL)*, przedstawione na rys. 4, gdzie widoczna jest źródłowa lista zmiennych (obiektów) otwartego pliku danych. Z listy należy wybrać te zmienne (obiekty), które będą uwzględnione w procedurze skalowania. Z tego okna możliwy jest również dostęp do kolejnych okien dialogowych.

W oknie dialogowym *Model* określamy, czy dane mają charakter podobieństw (niepodobieństw), liczbę wymiarów przestrzeni, w której będą prezentowane wyniki, oraz rodzaj przekształceń odległości dokonywanych w etapie optymalnego skalowania.

Rys. 3. Usytuowanie modelu PREFSCAL w menu programu *SPSS for Windows*

Źródło: pakiet statystyczny *SPSS 14.0PL for Windows*.

Rys. 4. Okno dialogowe *Wielowymiarowe skalowanie preferencji (PREFSCAL)*

Źródło: pakiet statystyczny *SPSS 14.0PL for Windows*.

Jeżeli dysponujemy konfiguracją punktów reprezentujących wiersze lub kolumny macierzy preferencji, czyli stosujemy zewnętrzną metodę wyznaczania punktów idealnych, wtedy w oknie *Ograniczenia* możemy wczytać te konfiguracje.

W oknie dialogowym *Opcje* definiujemy sposób wyznaczania konfiguracji początkowej, wartości parametrów karania oraz kryteria iteracji, czyli minimalną różnicę między wartościami STRESS-u w kolejnych cyklach, minimalną wartość

STRESS-u oraz maksymalną liczbę iteracji, po których proces obliczeń zostaje zakończony.

W oknie *Wykresy* i w oknie *Wyniki* możemy zadeklarować, jakie wykresy i jakie wyniki analizy zostaną zaprezentowane w końcowym raporcie.

5. Przykład empiryczny

W celu identyfikacja preferencji 28 respondentom przedstawiono 8 marek monitorów LCD. Następnie poproszono o określenie ich preferencji przez nadanie poszczególnym markom rang, przy czym rangi były kolejnymi liczbami naturalnymi od 1 do 8. Liczba 1 oznaczała ocenę najlepszą, natomiast liczba 8 – ocenę najgorszą.

Tabela 1. Macierz preferencji marek monitorów LCD

Respondent	Marka monitorów							
	Samsung	LG	Maxdata	Philips	BenQ	NEC	Neovo	Hyundai
1	2	1	8	3	4	7	6	5
2	3	5	6	1	2	7	8	4
3	1	5	6	2	7	3	8	4
4	2	5	8	3	4	1	7	6
5	3	1	5	2	4	6	7	8
6	1	4	5	3	6	2	7	8
7	1	2	7	4	5	6	3	8
8	3	2	5	1	6	4	7	8
9	3	1	5	2	4	6	7	8
10	1	2	3	6	8	4	5	7
11	1	4	8	2	3	6	7	5
12	4	5	6	3	1	2	7	8
13	2	1	8	3	5	4	6	7
14	5	1	7	2	4	3	8	6
15	1	6	2	5	7	8	3	4
16	2	4	6	1	3	7	8	5
17	1	2	7	3	6	5	8	4
18	1	3	8	2	4	7	6	5
19	1	3	7	2	5	6	8	4
20	2	3	6	1	7	5	8	4
21	2	3	6	1	7	8	5	4
22	2	1	8	4	5	7	3	6
23	2	1	8	3	5	4	6	7
24	5	1	8	2	4	3	7	6
25	6	2	3	1	8	5	4	7
26	1	5	4	3	6	2	7	8
27	2	7	3	1	8	5	4	6
28	3	1	7	2	4	8	5	6

Źródło: opracowanie własne.

Na podstawie otrzymanej w ten sposób macierzy preferencji przeprowadzono analizę *unfolding* przy wykorzystaniu algorytmu PREFSCAL. Wybrane miary dopasowania dla różnych wartości parametrów karania prezentuje tab. 2.

Tabela 2. Wybrane miary dopasowania dla różnych wartości parametrów karania

Parametry karania		$\lambda = 0,5;$ $\omega = 0,5$	$\lambda = 0,5;$ $\omega = 1,0$	$\lambda = 1,0;$ $\omega = 0,0$	$\lambda = 1,0;$ $\omega = 0,5$
Liczba iteracji		2306	2376	5000	5000
Końcowa wartość funkcji		0,3076872	0,3731754	0,0024500	0,0617637
Częstkowe wartości funkcji	STRESS	0,0387446	0,0481366	0,0024500	0,0127028
	Czynnik kary	2,4434742	2,8930154	1,0000000	4,8621973
Niedopasowanie	Znormalizowany STRESS	0,0015011	0,0023170	0,0000060	0,0001614
	STRESS-I Kruskala	0,0387438	0,0481356	0,0024498	0,0127028
Dobroć dopasowania	Rho Spearmana	0,8847641	0,8776521	0,8711058	0,8801343
	Tau-b Kendalla	0,7995368	0,7898562	0,7944053	0,8050820
Współczynnik zmienności	Zmienność przekształconych odległości	0,3843638	0,4574180	0,0021220	0,1224122
Wskaźnik degeneracji	Indeks braku degeneracji Sheparda	0,5255102	0,5497449	0,3367347	0,3456633

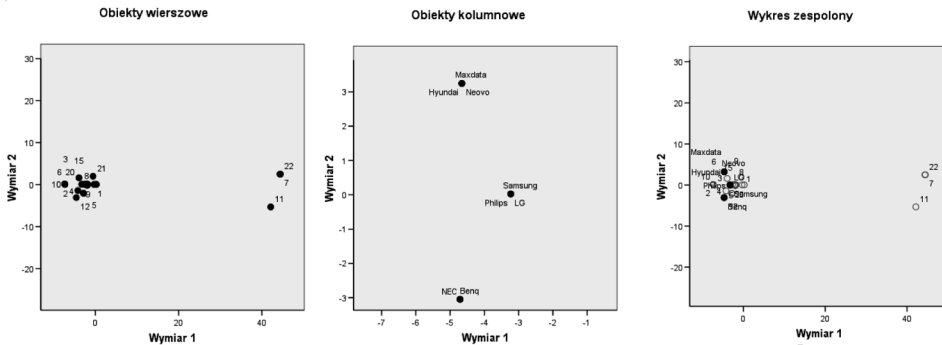
Źródło: opracowanie własne.

Najmniejsze wartości funkcji dopasowania uzyskano dla $\lambda = 1$ i $\omega = 0^2$. Wartość funkcji równa 0,0024500 może sugerować niemal idealne dopasowanie wyników skalowania do danych wejściowych, jednak niska wartość współczynnika zmienności oraz indeksu braku degeneracji Sheparda, będącego ilorzem liczby różnych odległości i liczby wszystkich odległości (różne odległości to takie, których różnica jest większa niż 0,001 przeciętnej odległości), wskazują na zdegenerowane rozwiązanie. Trywialność rozwiązania potwierdza rys. 5a. Prawie wszystkie punkty reprezentujące respondentów znajdują się w bardzo bliskim sąsiedztwie badanych marek monitorów, a punkty obrazujące marki Maxdata, Hyundai i Neovo, Samsung, Philips i LG oraz NEC i BenQ mają takie same współrzędne.

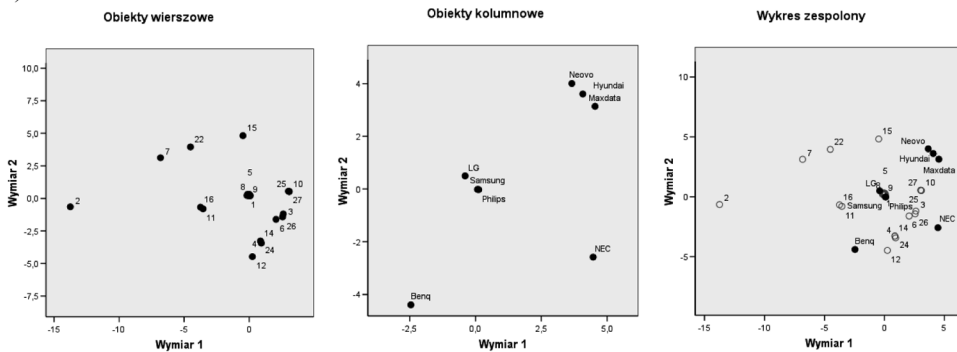
Niezdegenerowane wyniki skalowania prezentuje rys. 3b, w którym parametr $\lambda = 0,5$ i $\omega = 1,0$. Rozkład punktów na mapach percepcyjnych wskazuje na dwie grupy marek postrzeganych jako podobne. Są to Neovo, Hyundai i Maxdata oraz LG, Samsung i Philips. Jednocześnie marki LG, Samsung i Philips są najbardziej preferowane przez respondentów biorących udział w badaniu.

² Przyjęcie wartości parametrów karania na poziomie $\lambda = 1$ i $\omega = 0$ oznacza, że konfigurację punktów wyznaczano przez minimalizację funkcji dopasowania STRESS bez czynnika kary.

a)



b)



Rys. 5. Konfiguracje punktów reprezentujące respondentów (obiekty wierszowe), marki (obiekty kolumnowe) oraz respondentów i marki (wykres zespolony) dla: a) $\lambda = 1$ i $\omega = 0$; b) $\lambda = 0,5$ i $\omega = 1,0$

Źródło: opracowanie własne z wykorzystaniem pakietu statystycznego SPSS 14.0PL for Windows.

6. Zakończenie

Skalowanie wielowymiarowe jest jedną z metod statystycznej analizy wielowymiarowej, która znalazła trwale miejsce w badaniach marketingowych. Jest ona ważnym narzędziem badawczym, o czym świadczą liczne przykłady zastosowań w rozwiązywaniu konkretnych problemów marketingowych (zob. [Zaborski 2001, s. 107-109]). W badaniu preferencji szczególne znaczenie ma analiza *unfolding*, która, oprócz ukazania preferencji konsumentów względem produktów i usług, umożliwia również określenie preferowanego przez respondentów zbioru cech, jakie powinny mieć obiekty.

Częstym zjawiskiem w analizie *unfolding* są trywialne rozwiązania skalowania. Wprawdzie niektórzy autorzy wskazują na możliwe przyczyny degeneracji (zob. np. [Kim, Rangaswamy, DeSarbo 1999; Busing, Groenen, Heiser 2005]), jednak nie są one w pełni wyjaśnione. Uniknięcie trywialnych rozwiązań można osiągnąć przez

modyfikację danych wejściowych, transformacji monotonicznej lub funkcji dopasowania (zob. [Borg, Groenen 2005, s. 317-330]). Zastosowanie współczynnika zmienności w konstrukcji funkcji dopasowania modelu PREFSCAL pozwala na uniknięcie zdegenerowanego rozwiązania, a jeżeli wartość zmodyfikowanej funkcji STRESS przyjmuje wartość zero, świadczy to o doskonałym dopasowaniu konfiguracji punktów do danych wejściowych.

Literatura

- Borg I., Groenen P., *Modern Multidimensional Scaling. Theory and Applications. Second Edition*, Springer-Verlag, New York 2005.
- Busing F.M.T.A., Groenen P.J.K., Heiser W.J., *Avoiding Degeneracy in Multidimensional Unfolding by Penalizing on the Coefficient Of Variation*, „Psychometrika” 2005 nr 1, s. 71-79.
- Cox T.F., Cox M.A.A., *Multidimensional Scaling. Second edition*, Chapman and Hall, Londyn 2001.
- Kim C., Rangaswamy A., DeSarbo W.S., *A Quasi-Metric Approach to Multidimensional Unfolding for Reducing the Occurrence of Degenerate Solutions*, „Multivariate Behavioral Research” 1999 nr 34, s. 143-180.
- Takane Y., Young F.W., de Leeuw J., *Nonmetric Individual Differences MDS: an Alternating Least Squares Method with Optimal Scaling Features*, „Psychometrika 1977 nr 42, s. 7-67.
- Zaborski A., *Skalowanie wielowymiarowe w badaniach marketingowych*, AE, Wrocław 2001.
- Zaborski A., *Unfolding jako model pomiaru preferencji w skalowaniu wielowymiarowym*, [w:] Taksonomia 9, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 942, AE, Wrocław 2002, s. 128-137.

MULTIDIMENSIONAL PREFERENCES SCALING WITH THE USE OF PREFSCAL MODEL

Summary

Unfolding analysis is a special case of multidimensional scaling. It assumes that different individuals perceive various objects of choice in the same space, but differ with respect to what they consider an ideal combination of the objects' attributes.

The article presents PREFSCAL algorithm in which avoiding degeneracy in multidimensional unfolding is possible by modification of the loss function. Finally, empirical data set was analyzed by PREFSCAL method to show the benefits of the modification of the loss function.