

Justyna Wilk

Uniwersytet Ekonomiczny we Wrocławiu

METODY ANALIZY DANYCH SYMBOLICZNYCH*

Streszczenie: Analiza danych symbolicznych stanowi interesujący kierunek rozwoju metod eksploracyjnej analizy danych i reprezentacji danych ekonomicznych. Znajduje zastosowanie w badaniach dotyczących różnorodnych zagadnień społeczno-gospodarczych. Celem artykułu jest usystematyzowanie wiedzy na temat metod analizy danych symbolicznych. W artykule zaprezentowano specyfikę danych symbolicznych i ich rolę w opisie zjawisk ekonomicznych. Przedstawiono klasyfikację metod analizy danych symbolicznych, wymogi odnośnie do rodzaju zmiennych symbolicznych oraz problemy związane z wyborem metod analizy danych symbolicznych. Dokonano przeglądu zastosowań metod analizy danych symbolicznych w badaniach ekonomicznych.

Słowa kluczowe: obiekty symboliczne, zmienne symboliczne, statystyczna analiza wielowymiarowa, analiza danych symbolicznych.

1. Wstęp

Analiza danych symbolicznych (*Symbolic Data Analysis* – SDA) jest rozwijana od lat 80. XX wieku. Wśród podstawowych publikacji na ten temat należy wymienić prace pod redakcją Bocka i Didaya [2000], Billard i Didaya [2006] oraz Didaya i Noirhomme-Fraiture [2008].

Założenia SDA bazują na ułatwieniu analizowania dużych zbiorów danych i prawidłowym opisie badanych zjawisk. Złożoność problemów empirycznych sprawia, że dane w ujęciu klasycznym, reprezentowane przez pojedynczą kategorię lub liczbę rzeczywistą, stają się niewystarczające w opisie badanej rzeczywistości. Możliwość pełniejszego opisu daje wykorzystanie danych symbolicznych.

Podstawy teoretyczne SDA są związane z badaniami z zakresu psychologii poznawczej człowieka, rozwojem klasyfikacji pojęciowej, statystycznej teorii rozpoznawania obrazów, badaniami nad „sztuczną inteligencją” i „uczeniem się maszyn” oraz eksploracyjną analizą danych (*data mining*).

* Artykuł sfinansowano częściowo ze środków na działalność statutową Katedry Ekonometrii i Informatyki oraz ze środków na naukę w latach 2009-2012 jako projekt badawczy nr N N111 446037.

Celem artykułu jest zaprezentowanie wielowymiarowych metod analizy danych symbolicznych. W pierwszej części omówiona zostanie specyfika danych symbolicznych i ich znaczenie w opisie zjawisk ekonomicznych. W części drugiej zaprezentowana zostanie klasyfikacja metod analizy danych symbolicznych, wraz ze wskazaniem wymogów odnośnie do zmiennych symbolicznych oraz problemów związanych z wyborem metod SDA. W trzeciej części przedstawiony zostanie przegląd zastosowań metod SDA w badaniach empirycznych.

2. Specyfika danych symbolicznych

Bock, Diday [2000, s. 2] wśród podstawowych typów zmiennych symbolicznych wymieniają zmienne, których realizacją są:

- przedziały liczbowe (*interval-valued variable*). Szczególnym przypadkiem tej zmiennej jest zmienna metryczna, której realizacją są liczby rzeczywiste;
- listy kategorii, tzw. zmienna wielowariantowa (*multi-valued variable*). Szczególnym jej przypadkiem jest zmienna niemetryczna, o realizacjach w postaci pojedynczych kategorii;
- listy kategorii z wagami (*modal, multi-valued variable with associated weights*). Szczególnym przypadkiem zmiennej jest lista kategorii, w której wszystkim wariantom nadano jednakowe wagi.

Własności tych zmiennych zaprezentowano w tab. 1.

Ponadto SDA daje możliwość uwzględnienia logicznych powiązań między zmiennymi. Można wskazać struktury taksonomiczne, hierarchiczne i logiczne, różniące się charakterem relacji, jaka zachodzi między ich poziomami (tab. 2).

W SDA wyróżnia się dwa rodzaje obiektów symbolicznych, tj. obiekty w ujęciu klasycznym (elementarne jednostki badania) oraz obiekty złożone (zagregowane) – tab. 3.

Dane symboliczne mają szczególne znaczenie w badaniach ankietowych, m.in. w badaniach preferencji. Gromadzenie danych w postaci symbolicznej:

- ułatwia pomiar, ponieważ respondenci chętniej wskazują przedział wydatków na produkt niż precyzyjną wartość. Mogą oni nie znać precyzyjnej wartości, wolą nie ujawniać niektórych informacji lub wartość jest trudna do oszacowania;
- daje możliwość pozyskania danych poufnych, jak np. przedział dochodowy czy wiekowy respondenta. Respondent poproszony o podanie dokładnego wieku może nie ujawnić takich informacji albo udzielić nieprawdziwej odpowiedzi;
- umożliwia respondentom wyrażenie ich preferencji w sposób bardziej naturalny i pełniejszy, np. podanie kilku preferowanych marek produktów, kilku istotnych cech produktu wpływających na ich decyzje zakupowe. Konsument preferujący dwie marki produktu, zmuszony do podania tylko jednej, podałyby w efekcie informację niepełną.

Tabela 1. Podstawowe rodzaje zmiennych symbolicznych

Typ zmiennej symbolicznej	Zbiór realizacji zmiennej	Rodzaj wariantów zmiennej	Relacje i operacje matematyczne	Przykłady zmiennych symbolicznych
Przedział liczbowy	przedział wartości (skokowych lub ciągłych) ze zbioru liczb rzeczywistych	przedziały rozłączne, uporządkowane	zliczanie zdarzeń (liczba relacji równości, różności, większości, mniejszości)	– przedziały dochodowe lub wiekowe respondentów – temperatura krzepnięcia produktu
		przedziały nierozłączne, nieuporządkowane	zliczanie zdarzeń (liczba relacji równości, różności)	
Lista wariantów	zbiór kategorii lub wartości	kategorie (równorzędne lub uporządkowane)	zliczanie zdarzeń (liczba relacji równości, różności)	– znajomość języków obcych – posiadane kategorie prawa jazdy
		wartości (liczby rzeczywiste lub przedziały liczbowe)	zliczanie zdarzeń (liczba relacji równości, różności, większości, mniejszości)	
Lista wariantów z wagami	zbiór kategorii lub wartości oraz macierz wag	kategorie (równorzędne lub porządkowane) z wagami	zliczanie zdarzeń (liczba relacji równości, różności, większości, mniejszości), dodawanie i odejmowanie, mnożenie i dzielenie wag wariantów	– struktura wydatków konsumentów na zakupy w hipermarketach – motywy wyboru produktu według stopnia ważności
		wartości (liczby rzeczywiste lub przedziały liczbowe) z wagami		

Źródło: opracowanie własne z wykorzystaniem prac [Bock, Diday 2000; Wilk 2010].

Tabela 2. Rodzaje zmiennych strukturalnych

Rodzaj zmiennej strukturalnej	Liczba zmiennych tworzących strukturę	Liczba poziomów struktury	Rodzaj zależności	Przykłady zmiennych strukturalnych
Hierarchiczna	jedna lub więcej	najczęściej dwa	hierarchia poziomów zmiennej	– wielkość dochodu (lub brak dochodu) respondenta – zamiar (lub brak zamiaru) zakupu produktu
Taksonomiczna	minimum dwie	minimum dwa	brak lub naturalna zależność poziomów zmiennej	– marka, wersja i typ nadwozia samochodu – poziom i rodzaj wykształcenia respondenta
Logiczna	minimum dwie	minimum dwa	związek logiczny lub funkcyjny poziomów zmiennej	– wielkość miejscowości zamieszkania (wieś, miasto) respondenta – ciężar i przeznaczenie (osobowy, ciężarowy) samochodu

Źródło: opracowanie własne.

Tabela 3. Rodzaje obiektów symbolicznych

Lp.	Rodzaj obiektów symbolicznych	Stopień agregacji	Rodzaj zmiennych opisujących obiekty	Przykładowe obiekty
1	Elementarna jednostka badania	obiekt niezagregowany	– zmienne symboliczne – zmienne metryczne i niemetryczne	– respondent – gmina – produkt
2	Obiekt złożony (zagregowany)	obiekt będący wynikiem agregacji zbioru obiektów	– zmienne symboliczne	– grupa konsumentów preferująca określoną markę produktu – obszar geograficzny (np. województwo) złożony z podregionów (np. gmin)

Źródło: opracowanie własne z wykorzystaniem pracy [Bock, Diday 2000, s. 2].

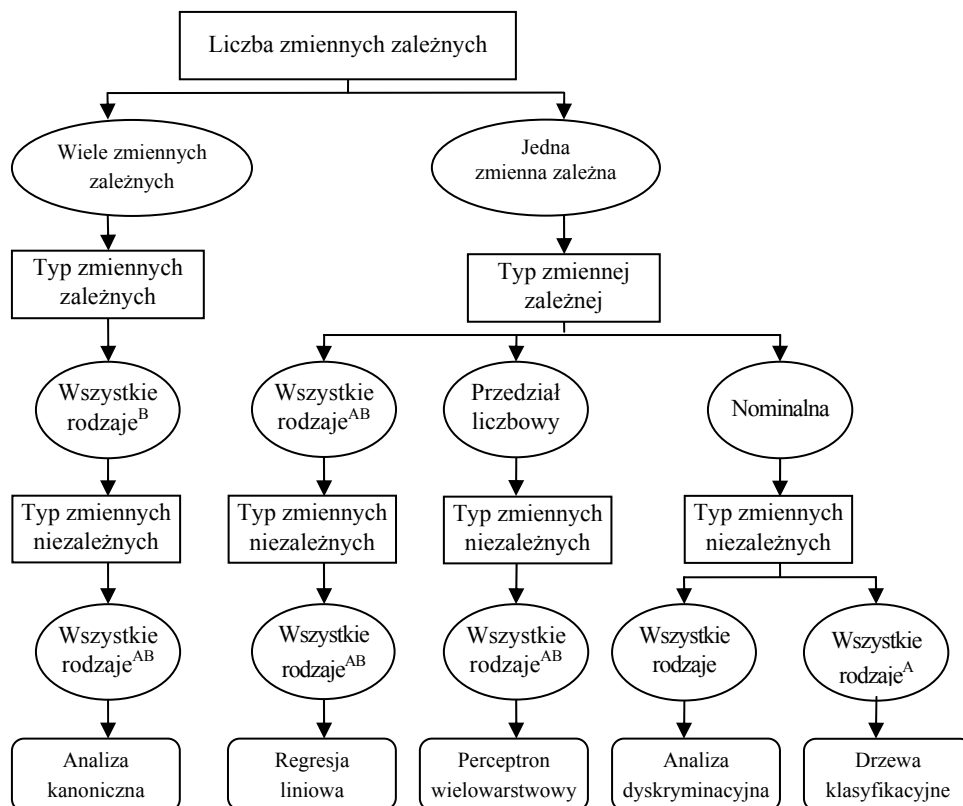
Poza tym dane symboliczne mogą być wynikiem procesu agregacji, gdy:

- efektem zabiegu jest redukcja zbioru obiektów ze względu na bardzo liczny zestaw danych,
- intencją badacza jest zebranie danych uogólnionych, np. struktura procentowa miesięcznych wydatków respondenta na żywność, opłaty, usługi itd.; przedział liczbowy planowanych wydatków na zakup sprzętu AGD/ RTV lub samochodu,
- celem analizy jest uzyskanie informacji dodatkowych (uzupełnienie posiadanych informacji). Na przykład w 2009 r. stopa bezrobocia w województwie dolnośląskim wynosiła 10,0%. W takiej formie jest to jednak informacja niepełna, ponieważ stopa bezrobocia w powiatach tego województwa kształtowała się od 3,3% (miasto Wrocław) do 23,0% (powiat złotoryjski). Zastosowanie zmiennej symbolicznej, np. przedziału liczbowego, lepiej obrazowałoby sytuację w tym regionie.

3. Klasyfikacja metod analizy danych symbolicznych

Wśród metod SDA można wyróżnić metody badania zależności i metody badania współwystępowania. W metodach badania zależności jedną lub więcej zmiennych określa się jako zmienną zależną od zbioru pozostałych zmiennych (niezależnych). Badaniu podlega wpływ zmiennych niezależnych na zmienną zależną.

Metody badania zależności zaadaptowane w SDA oraz wymogi odnoszące się do zmiennych symbolicznych (zależnych i niezależnych) prezentuje rys. 1.



A – oprócz zmiennych strukturalnych, B – konieczne przekodowanie zmiennych symbolicznych.

Rys. 1. Metody badania zależności w analizie danych symbolicznych

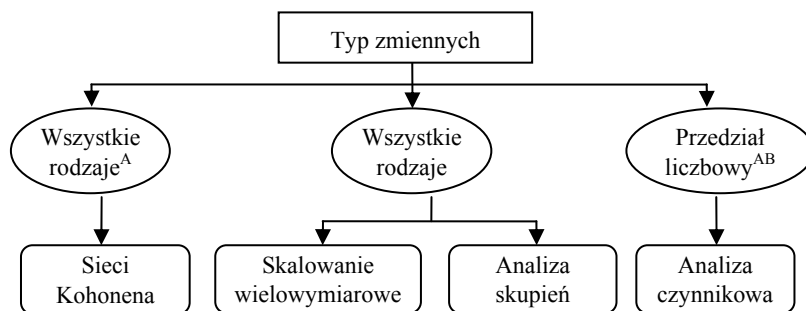
Źródło: opracowanie na podstawie [Bock, Diday 2000; Diday, Noirhomme-Fraiture 2008].

Z kolei w metodach badania współwystępowania nie dzieli się zmiennych na zależne i niezależne. Zakłada się brak tego rodzaju zależności. Analizuje się wewnętrzne relacje między zmiennymi w zbiorze zmiennych (jeśli przedmiotem badania jest zmienna) lub relacje między obiektami opisanymi tymi zmiennymi (gdy przedmiotem badania jest obiekt). Ta grupa metod ma na celu ustalenie wzajemnego położenia (usytuowania) obiektów lub zmiennych w przestrzeni wielowymiarowej.

Metody badania współwystępowania dostępne w ramach SDA oraz wymogi co do zmiennych symbolicznych prezentuje rys. 2.

Adaptacja metod statystycznej analizy wielowymiarowej w ramach SDA polegała na:

- wprowadzeniu nowych algorytmów i procedur w ramach SDA, np. niektóre metody skalowania wielowymiarowego, algorytmy drzew klasyfikacyjnych i większość algorytmów analizy skupień,



A – oprócz zmiennych strukturalnych, B – konieczne przekształcenie zmiennych symbolicznych.

Rys. 2. Metody badania współwystępowania w analizie danych symbolicznych

Źródło: opracowanie na podstawie [Bock, Diday 2000; Noirhomme-Fraiture 2004; Diday, Noirhomme-Fraiture 2008].

- modyfikacji znanych algorytmów (zob. np. [Hair i in. 2006]), jak np. niektóre algorytmy analizy skupień, algorytmy sztucznych sieci neuronowych,
- modyfikacji procedur, np. wyznaczenie macierzy odległości obiektów symbolicznych z wykorzystaniem miar opracowanych w ramach SDA (zob. np. [Bock, Diday 2000; Wilk 2005; Wilk 2006]), a następnie stosowanie metod „klasycznych” bazujących na macierzy odległości, np. metod analizy skupień, skalowania wielowymiarowego,
- wprowadzeniu systemu kodowania danych symbolicznych, m.in. w analizie regresji i analizie czynnikowej.

Wybrane metody statystycznej analizy wielowymiarowej stosowane w analizie danych symbolicznych zaprezentowano w tab. 4.

Problem wyboru metody analizy dotyczy wszystkich badań empirycznych, ale staje się bardziej pogłębiony dla danych symbolicznych. Po pierwsze, nie wszystkie metody statystycznej analizy wielowymiarowej zostały dotąd zaadaptowane na grunt SDA. Drugim problemem jest mało liczna reprezentacja grup metod. Najbardziej liczną są metody analizy skupień (taksonomii symbolicznej). Pozostałe grupy metod są znacznie uboższe, np. w ramach drzew klasyfikacyjnych dla danych symbolicznych opracowano tylko trzy algorytmy, które mają charakter drzew dyskryminacyjnych (nie zaproponowano dotąd drzew regresyjnych). Po trzecie, niektóre metody SDA są przeznaczone dla obiektów opisanych wybranym rodzajem zmiennych symbolicznych (np. wiele metod nie uwzględnia zmiennych strukturalnych), a inne wymagają przekodowania zmiennych symbolicznych, co łączy się ze zmianą struktury danych i związaną z tym zabiegiem utratą informacji. Poza tym wykonano dotąd niewiele badań weryfikujących użyteczność metod SDA w badaniach empirycznych. Jednymi z nielicznych są badania symulacyjne w zakresie analizy skupień przeprowadzone przez Dudka [2007].

Tabela 4. Wybrane metody analizy danych symbolicznych

Metody wielowymiarowej analizy statystycznej	Podział metod	Algorytmy SDA	
		zaadaptowane	opracowane w ramach SDA
Sztuczne sieci neuronowe	wzorcowe	perceptron wielowarstwowy	–
	bezwzorcowe	sieci Kohonena	–
Drzewa klasyfikacyjne	dyskryminacyjne	–	TREE, BDT, SDT
	regresyjne	–	–
Analiza skupień	hierarchiczne	m.in. metody Warda, kompletnego połączenia, średniej klasowej, giętka	m.in. metody Brito, Chavent, Gowdy i Didaya, KBG Bissona, RUMMAGE Fishera
	niehierarchiczne	k-medoidów	m.in. k-średnich Verde, SCLUST, DCLUST
Analiza czynnikowa	klasyczna analiza czynnikowa	–	–
	analiza głównych składowych	–	Centers, Vertices
Skalowanie wielowymiarowe	bazujące na macierzy odległości	bi-dimensional -mapping	–
	bazujące na tablicy danych symbolicznych	–	InterScal, SymScal, I-Scal

Źródło: opracowanie na podstawie [Bock, Diday 2000; Diday, Noirhomme-Fraiture 2008; Noirhomme-Fraiture 2004; Billard, Diday 2006; Gatnar 1998].

4. Przegląd zastosowań metod analizy danych symbolicznych

Metody SDA są stosowane w różnych dziedzinach nauki, m.in. chemii, biologii, socjologii i ekonomii (zob. np. [Wilk 2009b]). W zagadnieniach ekonomicznych mają zastosowanie szczególnie w analizach:

- marketingowych, m.in. w badaniach preferencji, struktury rynku, segmentacji rynku,
- regionalnych, m.in. w badaniach rozwoju regionalnego, jakości życia, rynku pracy,
- finansowych, m.in. w ocenie wniosków kredytowych, portfela inwestycyjnego,
- społecznych, m.in. w badaniach stylu życia, bezpieczeństwa socjalnego.

Wybrane zastosowania metod SDA w badaniach empirycznych dotyczących zagadnień ekonomicznych prezentuje tab. 5.

Tabela 5. Wybrane zastosowania metod SDA w badaniach empirycznych

Źródło literaturowe	Problem badawczy	Zmienne symboliczne*	Metoda SDA
Badania marketingowe			
[Brito 1994]	Klasyfikacja marek samochodów pod względem komfortu jazdy	przedziały liczbowe, listy kategorii	metoda Brito
[Diday, Gowda 1994]	Klasyfikacja komputerów klasy PC na podstawie parametrów technicznych	przedziały liczbowe, listy kategorii	metoda Gowdy i Didaya
[Más i in. 2002]	Klasyfikacja użytkowników Internetu w baskijskiej części Hiszpanii	listy kategorii z wagami	metoda Brito
[Duarte-Silva, Brito 2006]	Klasyfikacja marek samochodów	nominalna (przedziały liczbowe)	liniowa analiza dyskryminacyjna
[Pełka 2007]	Ocena podobieństwa monitorów ciekłokrystalicznych	przedziały liczbowe	InterScal
[Pełka, Jefmański 2008]	Profilowanie klas stacji narciarskich	nominalna (przedziały liczbowe, porządkowe, nominalne)	TREE
[Wilk 2009a]	Badanie poziomu jednorodności zbioru obiektów w badaniach segmentacyjnych	listy wariantów, listy kategorii z wagami, przedziały liczbowe, zmienne strukturalne	bi-dimensional-mapping
[Wilk 2010]	Segmentacja rynku <i>e-commerce</i>	przedziały liczbowe, listy kategorii, listy kategorii z wagami, zmienne strukturalne	metoda Warda, metoda Brito, SCLUST
[Wilk 2010]	Profilowanie segmentów rynku <i>e-commerce</i>	nominalna (przedziały liczbowe, listy kategorii, listy kategorii z wagami)	TREE
Badania regionalne			
[Silva i in. 2001]	Wyodrębnienie czynników decydujących o bezrobociu w regionach Portugalii klasy NUTS II	przedziały liczbowe	analiza głównych składowych
[Wilk 2005]	Klasyfikacja województw pod względem bezrobocia	listy kategorii z wagami	metoda Brito
[Bock, Diday 2000]	Podział osób ze względu na sposób przebiegu kariery zawodowej	nominalna (przedziały liczbowe)	jądrowa analiza dyskryminacyjna
[Bock, Diday 2000]	Wyodrębnienie czynników decydujących o przebiegu kariery zawodowej	przedziały liczbowe	analiza głównych składowych
[Bravo 2001]	Klasyfikacja mieszkańców Włoch pod względem aktywności zawodowej	nominalna (listy kategorii z wagami)	SDT
Badania finansowe			
[Goupil-Testu i in. 2000]	Wyodrębnienie czynników giełdowych decydujących o efektywności portfela inwestycyjnego	przedziały liczbowe	analiza głównych składowych
[Pełka 2007]	Ocena wniosków kredytowych klientów banku	nominalna (przedziały liczbowe, listy kategorii)	jądrowa analiza dyskryminacyjna
Badania społeczne			
[Mustjärvi 2006]	Klasyfikacja państw w odniesieniu do poziomu zabezpieczeń socjalnych i rozwoju ekonomicznego	przedziały liczbowe, listy kategorii z wagami	metoda Chavent
[Laaksonen 2008]	Badanie zaufania obywateli UE do różnych aspektów życia społecznego	przedziały liczbowe, listy kategorii z wagami	metoda Chavent

* W metodach badania zależności poza nawiasem podano rodzaj zmiennej zależnej, a w nawiasie rodzaj zmiennych niezależnych.

Źródło: opracowanie własne.

5. Podsumowanie

SDA stanowi interesujący kierunek rozwoju metod eksploracyjnej analizy danych i reprezentacji danych ekonomicznych. Znajduje zastosowanie w badaniach dotyczących różnorodnych zagadnień społeczno-gospodarczych. W analizie danych symbolicznych stosuje się metody opracowane w ramach SDA. Można również wykorzystać niektóre metody statystycznej analizy wielowymiarowej przeznaczone dla danych w ujęciu klasycznym. Jedną z możliwości jest wybór metod bazujących na macierzy odległości, z tym że do pomiaru odległości należy zastosować miary podobieństwa obiektów zaproponowane na gruncie SDA.

Literatura

- Billard L., Diday E., *Symbolic Data Analysis. Conceptual Statistics and Data Mining*, Wiley, Chichester 2006.
- Bock H.H., Diday E. (eds.) *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin-Heidelberg 2000.
- Bravo M.C., *Análisis de segmentación en el análisis de datos simbólicos*, Universidad Complutense de Madrid, *Departamento de Estadística e Investigación Operativa I* (doctoral thesis) 2001.
- Brito P., *Use of pyramids in symbolic data analysis*, [w:] E. Diday, Y. Lechevallier, P.B. Schader, B. Burtschy (eds.), *New Approaches in Classification and Data Analysis*, Springer-Verlag, Berlin-Heidelberg 1994, s. 378-386.
- Diday E., Gowda K.C. (1994), *Symbolic clustering algorithm using similarity and dissimilarity measures*, [w:] E. Diday, Y. Lechevallier, P.B. Schader, B. Burtschy (eds.), *New Approaches in Classification and Data Analysis*, Springer-Verlag, Berlin-Heidelberg 1994, s. 414-422.
- Diday E., Noirhomme-Fraiture M., *Symbolic Data Analysis and the Sodas Software*, John Wiley & Sons, Chichester 2008.
- Duarte-Silva A., Brito P., *Linear discriminant analysis for interval data*, „Computational Statistics”, vol. 21, issue 2, Physica-Verlag, Heidelberg 2006, s. 289-308.
- Dudek A., *Metody klasyfikacji dla danych symbolicznych – symulacja porównawcza*, [w:] J. Dziechciarz (red.), *Ekometria 18. Zastosowania metod ilościowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1151, Wydawnictwo AE, Wrocław 2007, s. 19-26.
- Gatnar E., *Symboliczne metody klasyfikacji danych*, PWN, Warszawa 1998.
- Goupil-Testu F., Touati M., Diday E., van der Veen H., *Analyse symbolique de données financières*, XXXIIèmes Journées de Statistiques, ASU 2000, Fes-MAROC, Mai 2000.
- Hair J.F., Black W.C., Babin B.J., Anderson R.E., Tatham R.L., *Multivariate Data Analysis*, Pearson Prentice Hall, New Jersey 2006.
- Laaksonen S., *People's life values and trust components in Europe – Symbolic data analysis for 20-22 countries*, [w:] E. Diday, M. Noirhomme-Fraiture (eds.), *Symbolic Data Analysis and the Sodas Software*, John Wiley & Sons, Chichester 2008.
- Más M., Olaeta H., Ayestarán M., *Symbolic characterization of internet users in Basque country*, *EUSTAT Technical Notebook 1*, Instituto Vasco de Estadística (EUSTAT), Donostia-San Sebastián 2002.

- Mustjärvi S., *Application to the Finnish, Portuguese and Spanish data from European Social Survey 2002/2003*, [w:] L. Billard, E. Diday (eds.), *Symbolic Data Analysis: Extracting Knowledge from Complex Data*, John Wiley and Sons Ltd, Indianapolis 2006.
- Noirhomme-Fraiture M. (ed.), *User manual for SODAS 2 software*, Software Report, 2004, Analysis System of Symbolic Official Data, Project no. IST-2000-25161.
- Pełka M., *Analiza danych symbolicznych i jej wykorzystanie w badaniach marketingowych*, Akademia Ekonomiczna we Wrocławiu 2007, Rozprawa doktorska (nieopublikowana).
- Pełka M., Jefmański B., *O sposobie profilowania segmentów rynkowych z wykorzystaniem drzew decyzyjnych dla danych symbolicznych*, [w:] J. Pocięcha (red.), *Modelowanie i prognozowanie zjawisk społeczno-gospodarczych*, Wydawnictwo UE, Kraków 2008, s. 367-379.
- Silva A., Marcelo C., Rodriguez D., *Regional particularities of employment in Portugal: The contribution of multivariety analysis of symbolic objects*, *Revista de estudos regionais – região Lisboa e Vale do Tejo*, 1 Semestre de 2001, no. 2, s. 1-23.
- Wilk J., *Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami*, [w:] K. Jajuga, M. Walesiak (red.), *Taksonomia 13. Klasyfikacja i analiza danych – teoria i zastosowania*, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126*, Wydawnictwo AE, Wrocław 2005, s. 224-236.
- Wilk J., *Problemy klasyfikacji obiektów symbolicznych. Symboliczne miary odległości*, [w:] J. Garczarczyk (red.), *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, *Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu nr 71*, Wydawnictwo AE, Poznań 2006, s. 69-83.
- Wilk J., *Techniki graficzne w procesie segmentacji rynku na podstawie danych symbolicznych*, [w:] M. Walesiak (red.), *Wizualizacja wyników badań marketingowych – podejścia, metody i zastosowania*, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 86*, Wydawnictwo UE, Wrocław 2009a, s. 153-164.
- Wilk J., *Cluster analysis methods in symbolic data analysis*, [w:] J. Pocięcha (red.), *Data Analysis Methods in Economic Investigations*, Wydawnictwo UE, Kraków 2009b, s. 39-54.
- Wilk J., *Problemy segmentacji rynku z wykorzystaniem metod klasyfikacji i danych symbolicznych*, Uniwersytet Ekonomiczny we Wrocławiu 2010, Rozprawa doktorska (nieopublikowana).

SYMBOLIC DATA ANALYSIS METHODS

Summary: Symbolic data analysis is an interesting development direction of data mining methods and economic data interpretation. It can be applied to a wide range of social and economic issues.

The aim of the paper is to present a summary on symbolic data analysis methods. There was explained symbolic data concept and its importance in describing economic problems. Symbolic data analysis methods were classified, the requirements on symbolic variables were given and the problems of method selection were discussed. The economic applications of symbolic data analysis methods were reviewed.