PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 176
Taksonomia 18. Klasyfikacja i analiza danych – teoria i zastosowania 2011

Małgorzata Machowska-Szewczyk
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

PODOBIEŃSTWA I RÓŻNICE MIĘDZY OBIEKTAMI SYMBOLICZNYMI A OBIEKTAMI ROZMYTYMI

Streszczenie: Informacje zgromadzone w bazach danych często mają postać tekstu zapisanego w naturalnym języku. Informacje lingwistyczne są proste do przetwarzania dla człowieka, który posługuje się na co dzień takimi pojęciami, stanowi jednak dość poważną barierę dla komputera. Obiekty symboliczne oraz obiekty charakteryzowane przez cechy rozmyte umożliwiają przetworzenie danych wyrażonych lingwistycznie do postaci akceptowalnej przez komputer. Głównym celem artykułu jest analiza porównawcza, czyli przedstawienie podobieństw i różnic, jakie występują w opisie obiektów przedstawionych za pomocą zmiennych symbolicznych oraz zmiennych rozmytych.

Słowa kluczowe: obiekty symboliczne, obiekty rozmyte, dane lingwistyczne.

1. Wstęp

Stale rozwijająca się technologia informacyjna pozwala gromadzić olbrzymie zasoby danych, przez co ich analiza staje się bardzo czasochłonna i kosztowna. Pozyskiwanie wiedzy z baz danych stało się więc obszarem badań naukowych, czego dowodem są liczne publikacje naukowe w tej dziedzinie. Większość metod pozyskiwania wiedzy opiera się na poszukiwaniu niejawnych reguł, uogólnianiu czy też klasyfikowaniu. Często informacje zgromadzone w bazach danych mają postać tekstu zapisanego w naturalnym języku. Informacje lingwistyczne są proste do przetwarzania dla człowieka, który posługuje się na co dzień takimi pojęciami, stanowi jednak dość poważną barierę dla komputera. Zarówno obiekty symboliczne, jak i obiekty charakteryzowane przez cechy rozmyte, służą do opisu pojęć, którymi posługujemy się w języku naturalnym. Umożliwiają przetworzenie danych wyrażonych lingwistycznie do postaci akceptowalnej przez komputer. Głównym celem artykułu jest analiza porównawcza, czyli wykrycie podobieństw i różnic, jakie występują w opisie obiektów przedstawionych za pomocą zmiennych symbolicznych oraz zmiennych rozmytych. W celu wskazania podobieństw podjęto próbę przedstawienia tych samych obiektów za pomocą zmiennych symbolicznych oraz za pomocą zmiennych rozmytych.

Podstawowym walorem pracy jest próba dokonania analizy porównawczej dwóch metod przetwarzania informacji lingwistycznej, którą posługuje się człowiek, do postaci możliwej do przetworzenia przez komputer. Obie te metody łączą analogie przetwarzania danych numerycznych z przetwarzaniem wzorowanym na tym, które dokonuje się w umyśle ludzkim.

2. Obiekty symboliczne

Na początku lat osiemdziesiątych XX wieku podjęto skuteczną próbę wyodrębnienia skupień w zbiorze obiektów, które opisane zostały w sposób symboliczny. Niech dany będzie zbiór obiektów $\Omega = \{O_1, O_2, ..., O_N\}$, charakteryzowanych przez zmienne $X_1, X_2, ..., X_K$ o charakterze symbolicznym, tzn. każda zmienna $X_i, i \in \{1, 2, ..., K\}$, może przyjąć wartości ze zbioru D_i w postaci [Bock, Diday 2000; Dudek 2004]:

1) danych numerycznych;
2) łańcuchów tekstowych, czyli danych jakościowych;
3) przedziałów liczbowych rozłącznych lub nierozłącznych;
4) danych w postaci listy wartości;
5) danych w postaci listy wartości z wagami;
6) danych, reprezentujących strukturę gałęziową;
7) wartości powiązanych hierarchicznie lub logicznie.

Do opisu obiektu symbolicznego służy pojęcie kompleksu. Kompleksem [Gatnar 1998] jest wyrażenie koniunkcyjne:

$$\underset{i \in I}{\wedge} \left[X_i \;\; \# \;\; y_i \right],$$

gdzie $I \subset \{1, ..., K\}$, $y_i \in D_i$, symbol $\#$ oznacza jeden z operatorów relacyjnych: $=, \neq, <, \leq, >, \geq$. Składowe koniunkcji $\left[X_i \;\; \# \;\; y_i \right]$ nazywane są selektorami. Wewnątrz selektorów mogą wystąpić także alternatywy wartości poszczególnych jego cech. Przykładem kompleksu [Gatnar 1998] może być:

$$\text{[branża=usługi]} \wedge \text{[forma=spółka cywilna} \vee \text{spółka akcyjna]} \wedge$$

$$\wedge \text{ [obroty>150]} \wedge \text{[zatrudnienie= 40...50].}$$

Kompleks ten opisuje obiekty, które mogą być spółką akcyjną lub cywilną świadczącą usługi, zatrudniającą od 40 do 50 osób, o obrotach ponad 150 tys. zł miesięcznie. Zbiór obiektów, których opis jest zgodny z opisem obiektu symbolicznego, wyrażonym za pomocą kompleksu, nazywa się zasięgiem obiektu symbolicznego. Opis symboliczny nie jest zatem precyzyjny, jeden kompleks może opisywać także obiekty nieistniejące w początkowym zbiorze obiektów.

Ze względu na strukturę danych, reprezentowanych przez obiekty symboliczne, do oceny odległości czy podobieństwa między takimi obiektami nie można zastosować tradycyjnych miar znanych z taksonomii numerycznej, jak odległość euklidesowa, miejska, Mahalanobisa. Podobieństwo lub odległość między obiektami symbolicznymi jest mierzona za pomocą miar heurystycznych, które wykorzystują informację o tym, czy warianty danej cechy są jednakowe czy też nie; bądź moce zbiorów w odpowiednich selektorach kompleksów, opisujących obiekty lub moce zbiorów stanowiących logiczny związek kompleksów i zakresów obiektów symbolicznych. Bock i Diday [2000] wymieniają jako najważniejsze następujące miary podobieństwa obiektów symbolicznych:

1. Gowda, Krishna – miara wzajemnego sąsiedztwa,
2. Ichino i Yaguchi – miara oparta na pojęciach operatorów połączenia i przekroju, uogólnionych na wszystkie typy danych występujących w obiektach symbolicznych,
3. De Carvalho – rozszerzenie miar Ichino i Yaguchi, wykorzystujące funkcje: porównującą i agregującą.

Obiekt symboliczny opisany może być również przez zmienne symboliczne z wagami. Są to zmienne, które przyjmują różne warianty, a poszczególnym wariantom przypisane są wagi prawdopodobieństwa lub częstości występowania. Zmienna symboliczna $X_i, i \in \{1, 2, ..., K\}$ jest odwzorowaniem przyporządkowującym poszczególnym obiektom kolejne warianty zmiennej z ustalonymi wagami, czyli

$$X_i : O_k \mapsto \left((x_{i1}, p_k(x_{i1})), ..., (x_{im_i}, p_k(x_{im_i}))\right), \text{ przy czym } \sum\nolimits_{j=1}^{m_i} p_k(x_{ij}) = 1,$$

gdzie: $p_k(x_{ij})$ – waga przypisana j-temu wariantowi cechy X_i w obiekcie O_k,

m_i – liczba możliwych wariantów cechy X_i.

Podobieństwo obiektów O_k, O_l względem zmiennej symbolicznej X_i można zmierzyć, wykorzystując adaptacje miar sformułowanych dla rozkładów prawdopodobieństwa (zob. [Csiszar 1967; Wilk 2006]). Przykładowo również odległość Minkowskiego:

$$d_p^i(O_k, O_l) = \left(\sum\nolimits_{j=1}^{m} \left| p_k(x_{ij}) - p_l(x_{ij}) \right|^p \right)^{\frac{1}{p}}, \text{ gdzie } p \geq 1,$$

może służyć do mierzenia odległości obiektów względem ustalonej zmiennej. Spełnia ona warunki metryki.

Odległość dowolnych dwóch obiektów, opisanych za pomocą zmiennych symbolicznych z wagami, może być mierzona jako zagregowana wartość odległości zmiennych, np. odległość Minkowskiego [Wilk 2006]:

$$d_p(O_k,O_l) = \left[\sum_{i=1}^{K}\left(\omega_i \cdot d^i(O_k,O_l)\right)^p\right]^{\frac{1}{p}},$$

gdzie $\omega_i \in [0,1]$ jest wagą zmiennej X_i, $i \in \{1,...,K\}$, przy czym $\sum_{i=1}^{K}\omega_i = 1$.

Jest wiele procedur klasyfikacji symbolicznej, które zostały opracowane w zakresie sztucznej inteligencji lub statystyki czy teorii rozpoznawania obrazów, podobnie jak w taksonomii klasycznej można wyróżnić metody: hierarchiczne, optymalizacyjno-iteracyjne oraz tworzące skupienia nierozłączne [Michalski 1980; Fisher 1987].

3. Obiekty rozmyte

Kolejnym sposobem na matematyczne modelowanie rzeczywistości, której nie da się wyrazić za pomocą zdeterminowanych wartości, jest przedstawienie wartości cech jako zbiorów rozmytych. Zbiór rozmyty jest uogólnieniem pojęcia podzbioru, w którym przynależność elementów jest określona binarnie: 1 − element należy do zbioru, 0 − nie należy. W rozmytym podzbiorze pewnego uniwersum funkcja przynależności może przyjmować wartości z przedziału od 0 do 1, czyli przynależność elementów może być częściowa, niepełna. Wartości funkcji przynależności odzwierciedlają intuicję człowieka w postrzeganiu zjawisk, nie są one określone w sposób jednoznaczny dla konkretnego zbioru rozmytego. Do analizowania zjawisk charakteryzowanych przez cechy będące nośnikami informacji rozmytej można zastosować algorytm klasyfikacji numerycznej. Tradycyjne metody klasyfikacji nie mogą być jednak wykorzystane bezpośrednio, gdyż każdy obiekt O_l jest reprezentowany przez ciąg K zbiorów rozmytych, czyli odpowiada mu wektor

$$\tilde{\mathbf{x}}_l = [\tilde{x}_{l1},...,\tilde{x}_{lK}]^T,$$

gdzie $\tilde{x}_{li} = \{(x,\mu_{li}(x)) : x \in D_i\}$ jest zbiorem rozmytym określonym na zbiorze wartości D_i zmiennej X_i o funkcji przynależności $\mu_{li} : D_i \to \langle 0,1 \rangle, i \in \{1,...,K\}, l \in \{1,...,N\}$.

Określając odległość lub podobieństwo między obiektami podobnie jak w przypadku cech zwykłych, np. za pomocą sumy odległości między poszczególnymi wartościami cech, wystarczy w rezultacie zdefiniować odległość między dwoma zbiorami rozmytymi.

Istnieje wiele miar podobieństwa oraz odległości między zbiorami rozmytymi. Aby ocenić podobieństwo obiektów rozmytych O_k, O_l, reprezentowanych przez wektory

$$\tilde{\mathbf{x}}_k = [\tilde{x}_{k1},\tilde{x}_{k2},...,\tilde{x}_{kK}]^T, \ \tilde{\mathbf{x}}_l = [\tilde{x}_{l1},\tilde{x}_{l2},...,\tilde{x}_{lK}]^T$$

względem zmiennej X_i, można wykorzystać miary zdefiniowane następująco [Pappis, Nikos 1992; Machowska-Szewczyk 2001]:

$$P_1^i(\tilde{\mathbf{x}}_k, \tilde{\mathbf{x}}_l) = \frac{\sum_{x \in D_i} \min\{\mu_{ki}(x), \mu_{li}(x)\}}{\sum_{x \in D_i} \max\{\mu_{ki}(x), \mu_{li}(x)\}}, \quad P_2^i(\tilde{\mathbf{x}}_k, \tilde{\mathbf{x}}_l) = 1 - \frac{\sum_{x \in D_i} |\mu_{ki}(x) - \mu_{li}(x)|}{\sum_{x \in D_i} |\mu_{ki}(x) + \mu_{li}(x)|},$$

$$P_3^i(\tilde{\mathbf{x}}_k, \tilde{\mathbf{x}}_l) = \left(\sum_{x \in D_i} |\mu_{ki}(x) - \mu_{li}(x)|^p \right)^{\frac{1}{p}},$$

gdzie μ_{ki} oznacza funkcję przynależności zbioru rozmytego $\tilde{x}_{ki}$.

Opracowano wiele procedur służących do klasyfikacji obiektów rozmytych, z których większość stanowi adaptacje klasycznych metod dla obiektów o wartościach rozmytych. Opis wybranych procedur wraz z przykładami zastosowań można znaleźć w pracach [Machowska-Szewczyk 2001; Łuczak, Wysocki 2003, 2004].

4. Podobieństwa i różnice między obiektami symbolicznymi a rozmytymi

Aby łatwiej zauważyć podobieństwa oraz różnice, podjęto próbę opisu konkretnych przykładowych obiektów symbolicznych za pomocą zmiennych rozmytych i odwrotnie. W pracy przedstawiono jedynie prosty przykład takiej zamiany.

Jeżeli obiekty symboliczne są opisywane za pomocą następujących zmiennych:

X_1 – branża o możliwych wartościach ze zbioru D_1 = {usługi, produkcja, handel},

X_2 – forma, D_2 = {spółka akcyjna, spółka z ograniczoną odpowiedzialnością, spółka partnerska, spółka komandytowa, spółka jawna},

X_3 – obroty, $D_3 = \{ \langle 0, 50), \langle 50{,}100) s, \langle 100{,}150), \langle 150, +\infty) \}$

X_4 – zatrudnienie, D_4 = {1..4, 5..19, 20..39, 40..50, powyżej 50}

to przykładowy obiekt symboliczny:

[branża=usługi]∧[forma=spółka jawna]∧[obroty>150]∧[zatrudnienie= 40..50],

można przedstawić jako obiekt rozmyty w postaci wektora $[\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4]^T$, przy czym $\tilde{x}_1 = \{(\text{usługi},1); (\text{produkcja},0); (\text{handel},0)\}$;

$\tilde{x}_2$ = {(spółka akcyjna, 0), (spółka z ograniczoną odpowiedzialnością, 0), (spółka partnerska, 0), (spółka komandytowa, 0), (spółka jawna, 1)};

$$\tilde{x}_3 = \left\{ (x, \mu(x)); x \in \mathbb{R}, x \geq 0, \mu(x) = \begin{cases} 1 & x \in (150, \infty) \\ 0 & x \in \langle 0, 150 \rangle \end{cases} \right\};$$

$$\tilde{x}_4 = \left\{ (x, \mu(x)); x \in \mathbb{N}, \mu(x) = \begin{cases} 1 \text{ dla } 40 \leq x \leq 50 \\ 0 \text{ dla pozostałych } x \end{cases} \right\}.$$

Z kolei dla cech:

X_1 – temperatura powietrza o wartościach ze zbioru D_1={zimno, ciepło, bardzo ciepło},

X_2 – zachmurzenie o wartościach ze zbioru D_2={małe, średnie, duże}, można wartości lingwistyczne zapisać przykładowo jako następujące zbiory rozmyte:

$$\mu_{\text{zimno}}(x) = \begin{cases} 1 \text{ dla } x \leq 7 \\ 0 \text{ dla } x > 7 \end{cases}, \quad \mu_{\text{ciepło}}(x) = \begin{cases} 1 \text{ dla } 7 \leq x \leq 15 \\ 0 \text{ dla pozostałych} \end{cases},$$

$$\mu_{\text{bardzo ciepło}}(x) = \begin{cases} 1 \text{ dla } x \geq 15 \\ 0 \text{ dla } x < 15 \end{cases}$$

$$\mu_{\text{małe}}(x) = \begin{cases} 1 \text{ dla } x \in \langle 0, 25 \rangle \% \\ 0 \text{ dla } x \notin \langle 0, 25 \rangle \% \end{cases}, \quad \mu_{\text{średnie}}(x) = \begin{cases} 1 \text{ dla } x \in \langle 25, 75 \rangle \% \\ 0 \text{ dla } x \notin \langle 25, 75 \rangle \% \end{cases},$$

$$\mu_{\text{duże}}(x) = \begin{cases} 1 \text{ dla } x \in (75, 100 \rangle \% \\ 0 \text{ dla } x \notin (75, 100 \rangle \% \end{cases}$$

Wówczas obiekt rozmyty O_1 „brzydka pogoda” można przedstawić jako wektor $\tilde{\mathbf{x}}_1 = [\tilde{x}_{11}, \tilde{x}_{12}]^T$, gdzie $\tilde{x}_{11} = \{(x, \mu_{\text{zimno}}(x)) : x \in \mathbb{R}\}$,

$\tilde{x}_{12} = \{(x, \mu_{\text{duże}}(x)) : x \in \langle 0,100 \rangle \%\}$.

Obiekt taki można wyrazić w kategoriach symbolicznych, przyjmując:

X_1 – temperatura powietrza, D_1={zimno, ciepło, bardzo ciepło},

X_2 – zachmurzenie, D_2={ $\langle 0, 25 \rangle \%$, $\langle 25, 75 \rangle \%$, $\langle 75,100 \rangle \%$ },

„brzydka pogoda”=[temperatura=zimno] $\wedge$ [zachmurzenie >75%].

Ale obiektu rozmytego O_1 „brzydka pogoda”, reprezentowanego przez wektor $\tilde{\mathbf{x}}_1 = [\tilde{x}_{11}, \tilde{x}_{12}]^T$, gdzie $\tilde{x}_{11} = \{(x, \mu_{\text{zimno}}(x)) : x \in \mathbb{R}\}$, $\tilde{x}_{12} = \{(x, \mu_{\text{duże}}(x)) : x \in \langle 0,100 \rangle \%\}$

i funkcje przynależności są zdefiniowane następująco:

$$\mu_{\text{zimno}}(x) = \begin{cases} 1 & \text{dla } x \leq 0 \\ -0{,}1x + 1 & \text{dla } 0 < x \leq 10 \\ 0 & \text{dla } x > 10 \end{cases},$$

$$\mu_{\text{duże}}(x) = \begin{cases} 0 & \text{dla } x \in \langle 0,50)\% \\ \frac{1}{25}x - 2 & \text{dla } x \in \langle 50,75)\% \\ 1 & \text{dla } x \in \langle 75,100\rangle\% \end{cases}$$

już się nie da przedstawić za pomocą zmiennych symbolicznych z powodu zróżnicowanych wartości stopni przynależności.

Dwa nurty nauki, które zajmują się przetwarzaniem danych wyrażonych lingwistycznie do postaci akceptowalnej przez komputer, rozwijają się odrębnie, choć można zauważyć występujące w nich pewne analogie. Podobieństwa oraz różnice, jakie zaobserwowano między opisem obiektu za pomocą zmiennych symbolicznych oraz za pomocą zmiennych w postaci zbiorów rozmytych, przedstawiono w tab. 1.

Tabela 1. Analiza porównawcza obiektów rozmytych oraz symbolicznych

Własności	Zmienne/obiekty symboliczne	Zmienne/obiekty rozmyte
Przetwarzanie zmiennych lingwistycznych	tak	tak
Wartości zmiennej w postaci przedziału, listy kategorii, listy kategorii z wagami, ciągów tekstowych	tak	tak
Wartości zmiennej w postaci zmiennej numerycznej ciągłej z różnymi wagami	nie	tak
Możliwe wyznaczenie odległości między obiektami po współrzędnych	tak*	tak
Uszczegółowienie lub uogólnienie opisu	tak	nie
Opis dotyczy grupy obiektów o tych samych cechach	tak	nie**
Wartości zmiennej w postaci alternatywy wartości lingwistycznych	tak	nie***
Opis klas w wyniku klasyfikacji za pomocą koniunkcji wyrażeń relacyjnych	tak	nie
Uwzględnianie zależności hierarchicznej, logicznej, gałęziowej zmiennych w procedurach klasyfikacyjnych	tak	nie

* Procedura wyznaczania odległości po współrzędnych dotyczy jednie obiektów opisanych za pomocą zmiennych symbolicznych z wagami, wówczas nawet niektóre metryki (np. euklidesowa) są tożsame dla obiektów symbolicznych oraz rozmytych.

** Zwykle dla obiektu rozmytego nie stosuje się takiego znaczenia opisu.

*** Zmienna rozmyta może przyjąć wartość rozmytą, reprezentującą alternatywę wartości lingwistycznych, dla tej alternatywy należałoby ustalić, w jaki sposób zdefiniować funkcję przynależności.

Źródło: opracowanie własne.

Oba opisy spełniają podobne zadanie, to znaczy próbują rzeczywistość postrzeganą przez człowieka i wyrażoną lingwistycznie przetworzyć na opis schematyczny, który można zautomatyzować. Zmienna rozmyta może przyjmować dowolne wartości z różnymi stopniami przynależności, przy czym funkcja przynależności może mieć różne kształty i jest to uwzględniane w opisie, natomiast zmienna symboliczna − nie.

Łatwo można dokonać uogólnienia lub uszczegółowienia opisu obiektu symbolicznego przez dodanie kolejnych koniunkcji lub ich usunięcie, zawężając lub rozszerzając zasięg obiektu symbolicznego, co nie jest praktykowane w przypadku opisu obiektów rozmytych.

Interpretacja stopni przynależności przy wariantach zmiennych rozmytych jest zupełnie inna niż w przypadku wag wartości zmiennych obiektów symbolicznych z wagami. Stopnie przynależności interpretuje się jako możliwość wystąpienia danego wariantu albo przynależności danej wartości do konkretnego zbioru rozmytego, w przypadku wag jest to zwykle częstość względna lub prawdopodobieństwo.

W wyniku klasyfikacji obiektów symbolicznych otrzymuje się opis klas w postaci wyrażeń koniunkcyjnych, który jest łatwy do interpretacji, gdyż opiera się na zmiennych lingwistycznych. Opis klas klasyfikacji rozmytej jest w postaci ciągu zbiorów rozmytych, co dodatkowo utrudnia interpretację. W procedurach klasyfikacyjnych obiektów symbolicznych uwzględniana jest zależność hierarchiczna, logiczna czy też strukturalna (struktura gałęziowa) tych obiektów, co nie występuje w przypadku procedur dla obiektów rozmytych.

5. Podsumowanie

Mimo że teoria zbiorów rozmytych oraz analiza danych symbolicznych poszukują rozwiązań problemu wyrażenia zmiennych lingwistycznych w kategoriach akceptowalnych dla komputera i wydawałoby się, że w pewnych zagadnieniach są zbieżne, to rozwijają się w różnych kierunkach, a występujące różnice są istotne. Stosowanie zamiennie zmiennych rozmytych oraz symbolicznych wymaga zatem olbrzymiej rozwagi i ostrożności.

Taksonomia symboliczna według Gatnara [1998] „polega na odkrywaniu struktur obiektów (przedmiotów, ludzi, zdarzeń itp.), w istocie prowadzi do tworzenia pojęć. Opracowano różne metody reprezentacji pojęć oraz konkretne algorytmy taksonomii symbolicznej, które wykrywają prawidłowości w tak przedstawionych zbiorach obiektów". Opis obiektów w sposób symboliczny stosuje się wówczas, gdy same cechy są określone za pomocą wartości lingwistycznych i przekształcając je do postaci danych numerycznych, łańcuchów tekstowych, przedziałów liczbowych, list wartości, list wartości z wagami, dokonuje się klasyfikacji tych obiektów. Natomiast opis obiektów za pomocą cech o wartościach rozmytych stosuje się wówczas, gdy uwzględnia się niepewność leksykalną, to znaczy niepewność definicji wartości wyrażonych lingwistycznie [Piegat 2001].

Mimo rozwoju wielu nauk, które zajmują się opisem rzeczywistości na wzór umysłu człowieka, ciągle dostrzegane są różne ograniczenia i niedoskonałości opracowanych metod, co świadczy o tym, że bogactwo umysłu człowieka jest ogromne i tak naprawdę pewnie nigdy nie będzie możliwe w pełni opisanie sposobu postrzegania ludzi w kategoriach symboli i liczb. Z pewnością warto jednak rozwijać te kierunki, gdyż ciągle dążymy do doskonałości.

Literatura

Bock H.H., Diday E., *Analysis of Symbolic Data. Explanatory Methods for Extracting Statistical Information from Complex Data*, Springer Verlag, Berlin 2000.

Csiszar I., *Information – type measures of difference of probability distributions and indirect observations*, „Studia Sci. Math. Hung." 1967, vol. 2.

Dudek A., *Miary niepodobieństwa obiektów symbolicznych. Odległość Ichino-Yaguchiego*, Ekonometria 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1021, Wrocław 2004, s. 100-106.

Fisher D., *Knowledge acquisition via incremental conceptual clustering*, „Machine Learning" 1987, no. 2.

Gatnar E., *Symboliczne metody klasyfikacji danych*, PWN, Warszawa 1998.

Łuczak A., Wysocki F., *Rozmyta wielokryterialna metoda Hellwiga porządkowania liniowego obiektów*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1169, Wrocław 2007, s. 330-340.

Machowska-Szewczyk M., *Wykorzystanie informacji rozmytej w wybranych algorytmach taksonomicznych*, praca doktorska, Politechnika Szczecińska, 2001.

Michalski R.S., *Knowledge acquisition through conceptual clustering*: *A theoretical Framework and an algorithm for partitioning data into conjunctive concepts*, „International Journal of Policy Analysis and Information Systems" 1980, vol. 4.

Pappis C.P., Nikos I., A *comparative assessment of measures of similarity of fuzzy values*, „Fuzzy Sets and Systems" 1992, 56, s. 171-174.

Piegat A., *Fuzzy Modeling and Control*, Springer, Heidelberg − New York 2001.

Wilk J., *Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 13, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, Wrocław 2006, s. 224-236.

SIMILARITIES AND DIFFERENCES BETWEEN SYMBOLIC OBJECTS AND FUZZY OBJECTS

Summary: The information collected in databases often has the form of a text written in natural language. Linguistic information is easy to process for a man who speaks with such concepts every day. However, it is quite a serious barrier to a computer. Symbolic objects and objects which are characterized by fuzzy attributes enable to process linguistically expressed data to a form which is acceptable by a computer. The main object of this article is a comparative analysis, that is a presentation of similarities and differences that occur in the description of the objects represented by the symbolic variables and fuzzy variables.