PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1151 — 2007
Ekonometria 18

Justyna Wilk

INTERPRETACJA WYNIKÓW KLASYFIKACJI OBIEKTÓW SYMBOLICZNYCH

1. Wstęp

Wiele problemów ekonomicznych ze względu na swoją złożoność jest charakteryzowanych za pomocą zmiennych symbolicznych (*symbolic variables*). Mogą one przyjmować więcej niż jeden wariant zmiennej dla pojedynczego obiektu, dodatkowo każdemu z wariantów można przypisać stopień ważności. Mogą one mieć postać przedziałów wartości bądź być powiązane zależnościami.

Specyficzny charakter tych zmiennych i nietypowa struktura implikują odmienny sposób interpretacji wyników klasyfikacji niż w przypadku obiektów opisanych zmiennymi mierzonymi na mocnych bądź słabych skalach pomiaru. Dla zmiennych ilościowych wyznacza się zwykle środki ciężkości poszczególnych klas oraz odchylenia standardowe w poszczególnych klasach, a dla zmiennych jakościowych – frakcje i odsetki występowania w danej klasie poszczególnych kategorii zmiennych [11, s. 344]. Odrębna jest również konstrukcja wskaźników dotyczących udziału skupienia w ogólnej zmienności próby, ważności zmiennej dla zmienności skupienia oraz udziału skupienia w zmienności zmiennej.

Artykuł ma na celu:

- scharakteryzowanie danych symbolicznych,
- omówienie, na przykładach, sposobu interpretacji klas obiektów symbolicznych (*symbolic objects*),
- prezentację wskaźników udziałowych (udział skupienia w zmienności zmiennej, ważność zmiennej dla zmienności skupienia oraz udział skupienia w zmienności zmiennej),
- prezentację modułu „Clint" w aplikacji *Sodas v. 2.0*.

2. Specyfika danych symbolicznych

Jeśli przyjmiemy, że Y_k to zmienna symboliczna (k – numer zmiennej, $k = 1, 2, \ldots, p$) ze zbioru Y (gdzie V_k zbiór realizacji zmiennej Y_k) o wariantach y_{fk} ($f = 1, 2, \ldots, z$), to v_{ik} jest wartością zmiennej Y_k dla obiektu o_i. O jest zbiorem obiektów symbolicznych $O = \{o_1, o_2, \ldots, o_n\}$, gdzie i to numer obiektu, $i = 1, 2, \ldots, n$.

Zmienne symboliczne mogą mieć charakter [5, s. 471-478; 1, s. 2, 32-37; 9, s. 7, 13; 7, s. 13]:

1. Przedziałów klasowych (*interval-valued variable*) rozłącznych (*disjoint interval*) lub nierozłącznych (*non-disjoint interval*), gdzie $v_{ik} = [a_{fk}; b_{fk}]$.
2. Listy kategorii, wartości (*multivalued variable*). To zmienne dopuszczające występowanie wielu kategorii lub wartości tej samej zmiennej dla jednego obiektu, gdzie $v_{ik} = \{y_{fk}, \ldots, y_{zk}\}$.
3. Listy kategorii (wartości) z wagami, prawdopodobieństwami czy częstościami (*multivalued modal variable*). To podobne zmienne do listy kategorii lub wartości, z tym że wariantom przyporządkowane są stopnie ważności. Zmienna symboliczna z wagami jest odwzorowaniem $Y_k = (y_{fk}, p(y_{fk}))$ [8, s. 79; 10, s. 326], gdzie $p(y_{fk})$ to waga f-tego wariantu zmiennej Y_k dla obiektu o_i, gdzie $p(y_{fk}) \in [0; 1]$ i $\sum_{f=1}^{z} p(y_{fk}) = 1$, zatem $v_{ik} = \{y_{1k}(p(y_{1k})), y_{2k}(p(y_{2k})), \ldots, y_{zk}(p(y_{zk}))\}$.
4. Zmiennych w postaci drzewa wariantów (*dependent variable*), tj. zmiennych o strukturze:
   a) taksonomicznej (*taxonomic dependent*), gdzie wariant zmiennej z najniższego poziomu nie determinuje wariantu z poziomu wyższego,
   b) hierarchicznej (*hierarchical dependent, mother-doughter*), gdy dla jednego z wariantów nie można określić podwariantu („brakujący" podwariant w analizie symbolicznej ma wartość *NA- non-applicable*),
   c) logicznej (*logical dependence*); występuje między zmiennymi, gdy wartość drugiej zmiennej zależy logicznie lub funkcyjnie od wartości pierwszej zmiennej,
   d) stochastycznej (*stochastic dependence*), gdzie szczeble hierarchii są ze sobą skorelowane (współczynnik korelacji i kowariancji jest różny od zera).

Zmienne strukturalne mogą przyjmować warianty zmiennych z punktów od 1 do 3.

3. Opis klas obiektów symbolicznych

Złożony charakter zmiennych symbolicznych uniemożliwia zastosowanie technik analitycznych adresowanych do zmiennych mierzonych na mocnych bądź słabych skalach pomiaru. Autorzy analizy danych symbolicznych zaproponowali więc odrębny sposób opisu klas oraz wyznaczania wskaźników udziałowych [6, s. 209, 211-213; 4, s. 31-33; 7, s. 15-16; 12, s. 3-4].

Skupienia obiektów symbolicznych można opisywać w wąskim lub szerokim zakresie [6, s. 209, 211]. Charakterystyka skupienia w szerokim zakresie (*long description*) polega na tym, że konstruuje się „nowy" obiekt symboliczny, obejmujący wszystkie elementy klasy. Jego konstrukcja przebiega w różny sposób, w zależności od rodzaju analizowanych zmiennych.

Dla przedziałów klasowych należy znaleźć najmniejszy przedział wartości, obejmujący wszystkie przedziały wartości zmiennej charakteryzujące obiekty badanego skupienia.

Na przykład jeśli wszystkie obiekty z tab. 1 należą do jednej klasy, to charakterystyka skupienia wygląda następująco: *Skupienie* : = [Y_1 *wiek* $\subseteq$ [20; 50] $\wedge$ *dochód* $\subseteq$ [1000; 4000]].

Tabela 1. Przykładowe obiekty symboliczne

Nr OS	Wiek	Miesięczny dochód
1	[20; 45]	[1000; 3000]
2	[35; 40]	[1200; 3500]
3	[25; 45]	[2000; 4000]
4	[30; 50]	[2000; 3200]

Źródło: opracowanie własne.

Tabela 2. Struktura kadry naukowej uczelni X

Nr OS	Płeć	Znajomość języków obcych
1	{mężczyzna}	{francuski, angielski}
2	{mężczyzna, kobieta}	{niemiecki, angielski}
3	{kobieta}	{francuski, włoski}
4	{mężczyzna,kobieta}	{francuski, hiszpański}

Źródło: opracowanie własne.

W przypadku list kategorii (wartości) wyznacza się zbiór wariantów, w skład którego wchodzą wszystkie warianty charakteryzujące poszczególne obiekty skupienia (jest to wynik tzw. operatora połączenia $v_{ik} \oplus v_{jk}$ (zob. [1, s. 170-171]).

Rozważając np. obiekty symboliczne z tab. 2, obserwujemy, że charakterystyka skupienia wygląda następująco *Skupienie*: = [Y_1 *płeć* $\subseteq$ {mężczyzna, kobieta} $\wedge$ znajomość języków obcych $\subseteq$ {francuski, włoski, niemiecki, angielski, hiszpański}].

Dla listy kategorii (wartości) z wagami charakterystykę skupienia można otrzymać na dwa sposoby:

- Uogólnianie za pomocą kryterium maksimum. Dla poszczególnego wariantu y_f zmiennej Y_k, opisującej obiekty tworzące skupienie, ustala się maksymalną wagę spośród wszystkich. Tak wyznaczony rozkład wag nie jest już rozkładem prawdopodobieństwa i suma wag może być większa niż 1, tj. $\sum_{f=1}^{z} p(y_f) \geq 1$.

Na przykład jeśli wszystkie obiekty z tab. 3 należą do jednej klasy, to charakterystyka skupienia wygląda następująco: *Skupienie* : = [Y_1 *kadra naukowa* ≤ {profesor (16%), doktor hab. (30%), doktor (55%), doktorant (10%), magister (10%)}]. Można zatem stwierdzić, że maksymalnie 16% pracowników naukowych wydziałów A, B i C ma tytuł naukowy profesora, natomiast co najwyżej 10% to osoby z tytułem magistra.

Tabela 3. Struktura kadry naukowej uczelni X

Nr OS	Wydział	Kadra naukowa
1	A	profesor (13%), doktor hab. (30%), doktor (47%), doktorant (10%)
2	B	profesor (10%), doktor hab. (20%), doktor (55%), doktorant (8%), magister (7%)
3	C	profesor (16%), doktor hab. (30%), doktor (44%), magister (10%)

Źródło: opracowanie własne.

- Uogólnianie za pomocą kryterium minimum polega na tym, że porównując obiekty względem *f*-tego wariantu zmiennej Y_k, wybiera się minimalną wagę analizowanego wariantu. W otrzymanym rozkładzie suma wag może być mniejsza niż 1, tj. $\sum_{f=1}^{z} p(y_f) \leq 1$.

Na przykład charakterystyka skupienia złożonego z obiektów w tab. 3 wygląda następująco: *Skupienie* : = [Y_1 *kadra naukowa* ≥ {profesor (10%), doktor hab. (20%), doktor (44%), doktorant (8%), magister (7%)}]. Co najmniej 10% kadry naukowej skupienia to pracownicy z tytułem profesora, a minimum 7% to osoby z tytułem magistra.

Interpretacja wyników analizy skupień w wąskim zakresie (*short description*) opiera się na znalezieniu takiego wariantu zmiennej lub takiego przedziału prawdopodobieństwa dla poszczególnych wariantów, którego nie przyjmują obiekty w pozostałych klasach [6, s. 209].

4. Wskaźniki udziałowe

Wyodrębnione skupienia obiektów symbolicznych można również analizować pod względem udziału skupienia w ogólnej zmienności próby, ważności zmiennej dla zmienności skupienia oraz udziału skupienia w zmienności zmiennej.

4.1. Udział skupienia w ogólnej zmienności próby

Udział skupienia *C* w zmienności próby *PP* (*contribution of a cluster to the total variability of the sample*) wyznacza się za pomocą wskaźnika *Cont*(*C*), *Cont*(*C*) ∈(0; 1] [6, s. 212]:

$$Cont(C) = \frac{G(s_c)}{G(s_{pp})} = \frac{\prod_{k=1}^{p} c\left(V_k^C\right)}{\sum_{i=1}^{n}\prod_{k=1}^{p} c\left(V_k^i\right)}, \tag{1}$$

gdzie: V_k^C – zbiór realizacji zmiennej Y_k w skupieniu C,

V_k^i – zbiór realizacji zmiennej Y_k dla i-tego obiektu,

$c\left(V_k^C\right) = \mu(v_{Ck})$ dla przedziału klasowego lub listy kategorii (wartości) Y_k opisującej skupienie C,

$c\left(V_k^i\right) = \mu(v_{ik})$,

$G(s)$ – *generality degree of its associated long symbolic objects*,

C – skupienie,

PP – próba, $PP \equiv \bigcup_{i=1}^{n} o_i$.

W przypadku list kategorii bądź wartości z wagami $G(s)$ wyznacza się w sposób uzależniony od rodzaju kryterium zastosowanego do opisu klasy. Jeśli opisu skupienia dokonano z wykorzystaniem kryterium maksimum, to $G(s)$ definiuje się jako [7, s. 16-17]:

$$G_{\max}(s) = \prod_{k=1}^{p} \frac{1}{\sqrt{z}} \sum_{f=1}^{z} \sqrt{p\left(y_{fk}\right)}, \tag{2}$$

własności:

$$\frac{1}{\sqrt{z}} \sum_{f=1}^{z} \sqrt{p\left(y_{fk}\right)} = \sum_{f=1}^{z} \sqrt{p\left(y_{fk}\right)\frac{1}{z}},$$

$$\prod_{k=1}^{p} \frac{1}{\sqrt{z}} \le G_{\max}(s) \le 1,$$

$$\prod_{k=1}^{p} \frac{1}{\sqrt{z}} \le G_{\max}(s) \le \prod_{k=1}^{p} \sqrt{z},$$

$$G_{\max}(s \cup s') \ge \max\left\{G_{\max}(s), G_{\max}(s')\right\}.$$

Na przykład jeśli za cechy skupienia opisującego pracowników uczelni Y, otrzymane za pomocą kryterium maksimum, przyjmiemy: *Skupienie*: [Y_1 *kadra naukowa* ≤ profesor (13%), doktor hab. (30%), doktor (47%), doktorant (10%)] ∧

[Y_2 *pracownicy administracji* ≤ księgowość (40%), dział kadr (20%), dział techniczny (40%)], to:

$$G_{\max}(s_c) = \left(\frac{\sqrt{0{,}13} + \sqrt{0{,}3} + \sqrt{0{,}1}}{\sqrt{4}}\right)\left(\frac{\sqrt{0{,}4} + \sqrt{0{,}2} + \sqrt{0{,}4}}{\sqrt{3}}\right) = 0{,}92.$$

Jeśli natomiast do uogólniania obiektów wykorzystano kryterium minimum, to $G(s)$ definiuje się jako [7, s. 16-17]:

$$G_{\min}(s) = \prod_{k=1}^{p} \frac{1}{\sqrt{z(z-1)}} \sum_{f=1}^{z} \sqrt{1 - p\left(y_{fk}\right)} \tag{3}$$

własności:

$$\prod_{k=1}^{p} \frac{\sqrt{z-1}}{\sqrt{z}} \le G_{\min}(s) \le 1,$$

$$\prod_{k=1}^{p} \frac{\sqrt{z-1}}{\sqrt{z}} \le G_{\min}(s) \le \prod_{k=1}^{p} \frac{\sqrt{z}}{\sqrt{z-1}},$$

$$G_{\min}(s \cup s') \ge \max\left\{G_{\min}(s), G_{\min}(s')\right\}.$$

Wysokie wartości *Cont*(C) oznaczają, że skupienie jest podobnie heterogeniczne jak cała próba, natomiast niskie wartości świadczą o wysokiej homogeniczności w stosunku do próby.

Jeśli obiekty skupienia były opisane za pomocą zmiennej strukturalnej, gdzie r oznacza zależność między zmiennymi, to $G(s)$ wyznacza się w odrębny sposób [7, s. 17-18]:

$$G\left(s_c|r\right) = G\left(s_c\right) - G\left(s_c|\neg r\right). \tag{4}$$

Na przykład jeśli za cechy skupienia opisującego pracowników firmy budowlanej, otrzymane za pomocą kryterium maksimum, przyjmiemy s: [Y_1 *rodzaj wykształcenia* ≤ zawodowe (0,3), średnie (0,5), wyższe (0,2)] ∧ [Y_2 *tytuł zawodowy* ≤ magister (0,1), inżynier (0,1), licencjat (0,0), *NA* (0,8)], a pomiędzy zmienną Y_1 i Y_2 wystąpi zależność hierarchiczna r: $Y_1 \in$ {zawodowe, średnie} ⇔ Y_2 = *NA*, to:

$$G_{\max}(s_c) = \left(\frac{\sqrt{0{,}3} + \sqrt{0{,}5} + \sqrt{0{,}2}}{\sqrt{3}}\right)\left(\frac{\sqrt{0{,}1} + \sqrt{0{,}0} + \sqrt{0{,}8}}{\sqrt{4}}\right) = 0{,}75.$$

Wyznaczając $G(s_c|\neg r)$, należy skonfrontować ze sobą sprzeczne warianty zmiennych, tj. dla $Y_1 \in$ {zawodowe, średnie}, Y_2 = {magister, inżynier, licencjat}, natomiast dla Y_1 = wyższe, $Y_2 = NA$, tj.:

$$G_{\max}(s_c|\neg r) = \left(\frac{\sqrt{0,3}+\sqrt{0,5}}{\sqrt{0,3}}\right)\left(\frac{\sqrt{0,1}+\sqrt{0,1}\sqrt{0,0}}{\sqrt{4}}\right)+\left(\frac{\sqrt{0,2}}{\sqrt{3}}\right)\left(\frac{\sqrt{0,8}}{\sqrt{4}}\right) = 0,34,$$

$$G_{\max}(s_c|r) = G(s_c) - G(s_c|\neg r) = 0,75 - 0,34 = 0,41.$$

4.2. Ważność zmiennej dla zmienności skupienia

Ważność zmiennej Y_k dla zmienności skupienia C, inaczej mówiąc wpływ zmiennej na zmienność skupienia (*the importance of each variable for each cluster*), wyznacza się za pomocą odrębnego miernika – $Cont(Y_k, c)$ [6, s. 212]:

$$Cont(Y_k, c) = \frac{c(V_k^c)}{\sum_{k=1}^{p} c(V_k^c)}, \tag{5}$$

gdzie: $c(V_k^c) = \mu(v_{Ck})$ dla przedziału klasowego lub listy kategorii (wartości) Y_k opisującej skupienie C,

$c(V_k^c) = \sum_{f=1}^{z} p(y_{fk})$ dla listy kategorii (wartości) z wagami Y_k opisującej skupienie C.

Suma wartości wskaźnika jest równa 1, tj. $\sum_{k=1}^{p} Cont(Y_k, c) = 1$. Wskaźnik przyjmuje wartości z przedziału $Cont(Y_k, c) \in \left[0; \frac{1}{p-1}\right]$. Wysokie wartości wskaźnika oznaczają, że zróżnicowanie między elementami skupienia ze względu na zmienną Y_k jest wysokie, natomiast niskie wartości oznaczają, że zmienna Y_k słabo różnicuje elementy skupienia.

Na przykład jeśli za skupienie przyjmiemy s: [Y_1 *wiek* $\subseteq$ [20 – 45] $\wedge$ Y_2 *płeć* {mężczyzna} $\wedge$ Y_3 *znajomość języków obcych* $\subseteq$ {francuski, angielski} $\wedge$ Y_4 *skład osobowy* $\leq$ {administracja (0,30), pracownicy naukowi (0,70)], to:

$$c(V_1) = 45 - 20 = 25,$$

$$c(V_2) = 1,$$

$$c(V_3) = 2,$$

$$c(V_4) = \sqrt{0,3} + \sqrt{0,7} = 4,38,$$

$$Cont(wiek, c) = \frac{25}{25+1+2+4,38} = \frac{25}{32,38} = 0,77,$$

$$Cont(płeć, c) = \frac{1}{32,38} = 0,03,$$

$$Cont(języki\ obce, c) = \frac{2}{32,38} = 0,06,$$

$$Cont(skład, c) = \frac{4,38}{32,38} = 0,135.$$

4.3. Udział skupienia w zmienności zmiennej

Udział skupienia C w zmienności zmiennej Y_k (*the contribution of a cluster to the variability of a variable*) mierzy się za pomocą $Cont(c, Y_k)$ [6, s. 212]:

$$Cont(c, Y_k) = \frac{c(V_k^c)}{c(V_k^{PP})}, \tag{6}$$

gdzie: V_k^{PP} – zbiór realizacji zmiennej Y_k dla próby PP,

$$c(V_k^{PP}) = \sum_{k=1}^{p} c(V_k). \tag{7}$$

$Cont(c, Y_k)$ przyjmuje wartości z przedziału (0; 1]. Wysokie wartości wskaźnika $Cont(c, Y_k)$ oznaczają, że dla zmiennej Y_k skupienie jest podobnie heterogeniczne jak cała zmienna, natomiast niskie wartości, oznaczają, że skupienie jest homogeniczne (jednorodne) w stosunku do zmienności zmiennej.

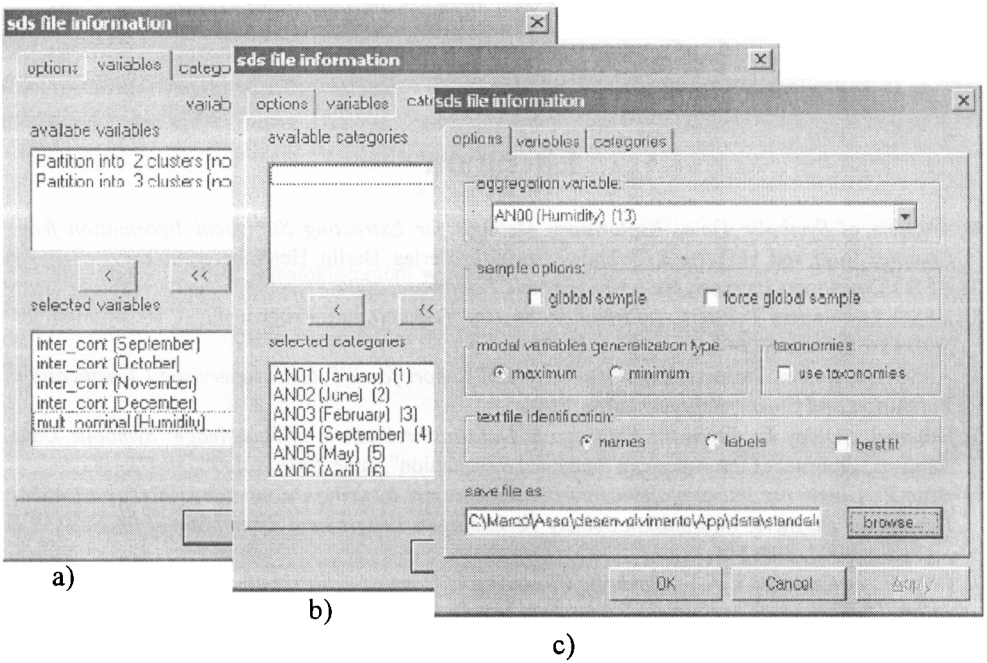
5. Moduł „Clint" aplikacji *Sodas v. 2.0*

W najnowszej wersji aplikacji *Sodas v. 2.0* (*Symbolic Official Data Analysis Software*) do opisu klas obiektów symbolicznych służy moduł „Clint" (*cluster interpretation*) (zob. [2, s. 178-181]). Plikiem wejściowym w module „Clint" jest pierwotna tablica danych symbolicznych (rozszerzenie: **.sds*), wzbogacona o dodatkową zmienną nominalną, która identyfikuje wyodrębnione skupienia za pomocą wprowadzonych oznaczeń (kategorii zmiennej).

Uruchomienie modułu „Clint" wywołuje okno z zakładkami opcje (*options*), zmienne (*variables*) oraz warianty zmiennych (*categories*) (zob. rys. 1).

Zakładka „zmienne" umożliwia wybór zmiennych podlegających interpretacji (*selected variables*) spośród zmiennych biorących udział w procesie klasyfikacji oraz zmiennych profilowych (*available variables*) (zob. rys. 1, pkt a). Moduł informuje o typie wybieranych zmiennych. „Clint" uwzględnia wszystkie rodzaje zmiennych symbolicznych.

W „opcjach" użytkownik wskazuje zmienną charakteryzującą skupienia (*aggregation variable*), domyślnie jest to pierwsza zmienna w tabeli (zob. rys. 1, pkt c). Jeśli zbiór wariantów analizowanych zmiennych dla badanych obiektów wyczerpuje cały zbiór realizacji każdej z tych zmiennych, należy zaznaczyć opcję „global sample", jeśli nie obejmuje – opcję „force global sample". Jeśli w zbiorze występują zmienne symboliczne z wagami, należy wybrać charakter uogólniania (*modal variables generalization type*). Opcję „taxonomies" wybiera się wówczas, gdy w zbiorze występują zmienne strukturalne i badacz chce je uwzględnić w opisie klas. Można również wybrać sposób oznaczania wariantów: według nazwy (*names*), numeryczny (*labels*) lub wybrać opcję najlepszego dopasowania do okna wynikowego (*best fit*). Należy również podać nazwę pliku wynikowego i wybrać ścieżkę zapisu.



Rys. 1. Okna aplikacji „Clint" programu *Sodas v. 2.0*

Źródło: [3, s. 178-181].

W oknie „kategorie" spośród dostępnych symboli klas (*available categories*) należy wybrać symbole (*selected categories*) odnoszące się do skupień podlegających interpretacji (zob. rys. 1, pkt b). Wynikają one z wcześniej wskazanej zmiennej charakteryzującej skupienia.

Plik wynikowy modułu „Clint" zawiera wykresy 2D i 3D (zob. [13]) wyodrębnionych skupień oraz arkusz obejmujący:

- listę obiektów wchodzących w skład poszczególnych skupień,
- opis klas w szerokim i wąskim zakresie,
- udziały poszczególnych skupień w ogólnej zmienności próby,
- tabelę wynikową ilustrującą udział skupień w zmienności każdej ze zmiennych oraz ważność poszczególnych zmiennych dla zmienności wyodrębnionych skupień.

6. Podsumowanie

Specyfika zmiennych symbolicznych implikuje odrębny sposób charakteryzowania skupień i konstrukcji wskaźników udziałowych niż w przypadku zmiennych mierzonych na mocnych lub słabych skalach pomiaru. Autorzy analizy danych symbolicznych zaproponowali rozwiązania adekwatne do analizy obiektów symbolicznych. Rozszerzyli również aplikację *Sodas* o moduł „Clint", który umożliwia otrzymanie interpretacji wyników klasyfikacji tego rodzaju obiektów.

Literatura

[1] *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, red. H.H. Bock, E. Diday, Springer-Verlag, Berlin, Heidelberg 2000.

[2] ASSO Developers Partners, *Help Guide Sodas 2 Software*, 2004.

[3] ASSO Developers Partners, *Information Society Technologies Programme. User Manual for SODAS 2 Software*, 2004.

[4] ASSO Developers Partners, *Scientific-Technical Report for WP 6. Unsupervised Classification, Validation and Cluster Representation*, 2001.

[5] Billard L., Diday E., *From the Statistics of Data to the Statistic of Knowledge: Symbolic Data Analysis*, „Journal of the American Statistical Association", June 2003, vol. 98, nr 470-487.

[6] Brito P., *Clustering Interpretation. Interpreting Clusters by using the module CLINT*, [w:] *ASSO Developers Partners, Information Society Technologies Programme. User Manual for SODAS 2 Software*, 2004, 207-214.

[7] Brito P., de Carvalho F.A.T., *Symbolic Clustering of Constrained Probabilistic Data*, [w:] *Exploratory Data Analysis in Empirical Research*, red. O. Opitz, M. Schwaiger, „Studies in Classification, Data Analysis and Knowledge Organization", Springer-Verlag, Heidelberg 2001, s. 12-22.

[8] Brito P., Polaillon G., *Structuring Probabilistic Data by Galois Lattices*, „Mathematicts and Social Sciences", 2005, vol. 43, nr 169, 77-104.

[9] Diday E., *An Introduction to Symbolic Data Analysis and the Sodas Software*, „En Electronic Journal of Symbolic Data Analysis" 2002, vol. 0, nr 0 (July).
[10] Hardy A., Lallemand P., *Clustering of Symbolic Objects Described by Multi-Valued and Modal Variables*, [w:] D. Banks, L. House, F.R. McMorris, P. Arabie, W. Gaul, *Clustering and Data Mining Applications*, „Studies in Classification, Data Analysis and Knowledge Organization", Springer-Verlag, Berlin-Heidelberg 2004, 325-332.
[11] *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, red. E. Gatnar, M. Walesiak, AE, Wrocław 2004.
[12] Verde R., Lechevallier Y., Chavent M., *Symbolic Clustering Interpretation and Visualization*, „The Electronic Journal of Symbolic Data Analysis" 2003, vol. 1, nr 1.
[13] Wilk J., *Graficzna prezentacja obiektów symbolicznych*, [w:] Ekonometria 16, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.
[14] Wilk J., Pełka M., *Dane symboliczne w zagadnieniu klasyfikacji*, [w:] *Identyfikacja struktur rynkowych. Pomiar – modelowanie – symulacja*, red. M. Rószkiewicz, Oficyna Wydawnicza SGH, Warszawa 2004.

INTERPRETATION OF CLUSTERS OF SYMBOLIC OBJECTS

Summary

The article characterizes types of symbolic data and a way of interpreting the clusters of symbolic objects. The author also presents how to evaluate the contribution of a cluster to the total variability of the sample, the importance of each variable for each cluster and the contribution of a cluster to the variability of a variable. Moreover the article illustrates the application of „Clint" module of *Sodas v. 2.0*.

Mgr Justyna Wilk jest doktorantem w Katedrze Ekonometrii i Informatyki Akademii Ekonomicznej we Wrocławiu.