

**Justyna Wilk**

Uniwersytet Ekonomiczny we Wrocławiu

## **TECHNIKI GRAFICZNE W PROCESIE SEGMENTACJI RYNKU NA PODSTAWIE DANYCH SYMBOLICZNYCH**

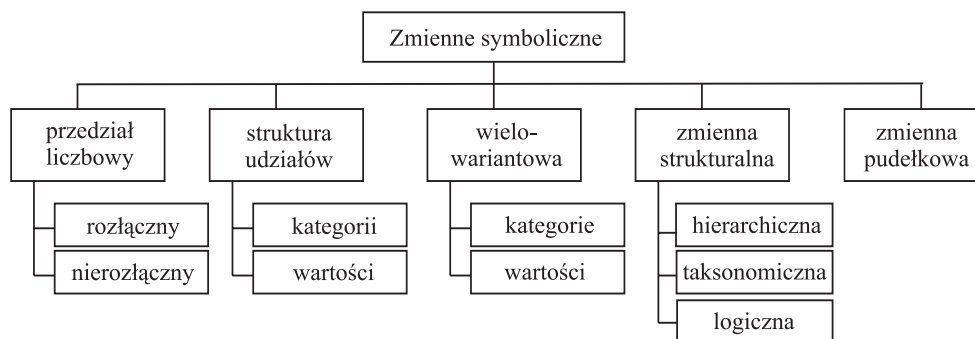
### **1. Wstęp**

Segmentacja rynku jest kluczowym elementem strategii marketingowej firmy. Proces segmentacji polega na wyborze kryteriów segmentacji, klasyfikacji obiektów (np. konsumentów) i profilowaniu segmentów. Dane, na podstawie których przeprowadzana jest segmentacja, pozyskiwane są najczęściej w drodze badań ankietowych lub z baz danych. Złożoność uzyskanych danych sprawia, że zmienne w ujęciu klasycznym (np. ilorazowe) stają się często niewystarczające do opisu obiektów. Dane symboliczne pozwalają pełniej wykorzystać informacje o konsumentach. Wiele aspektów dotyczących pojedynczego obiektu cechuje przedział wartości, wielowariantowość, struktura udziałowa i powiązania różnego rodzaju relacjami. Właściwa interpretacja wyników segmentacji rynku ułatwia poznanie potrzeb konsumentów i dostosowanie instrumentów marketingu-mix do typologii segmentów. Ułatwieniem interpretacji wyników jest zastosowanie metod graficznych. Złożona struktura danych symbolicznych powoduje konieczność zastosowania specyficznych rozwiązań. Celem artykułu jest przedstawienie technik graficznych, jakie można stosować w prezentacji wyników badań segmentacyjnych na podstawie danych symbolicznych.

### **2. Specyfika danych symbolicznych**

Bock, Diday [2000, s. 2] wśród podstawowych typów zmiennych symbolicznych wymieniają zmienne, których realizacją jest przedział liczbowy (*interval-valued variable*), lista wariantów (tzw. zmienna wielowariantowa, *multi-valued variable*) i struktura udziałowa (*modal* lub *multi-valued variable with associated weights*) – rys. 1.

Realizacją zmiennej w postaci przedziału liczbowego może być przedział rozłączny, np. miesięczny dochód netto respondenta w zł (np. (0, 3000), [3000, 6000), [6000, 9000), [9000,  $\infty$ )) lub nierozłączny, np. planowana wysokość wydatków na



Rys. 1. Rodzaje zmiennych symbolicznych

Źródło: por. [Bock, Diday 2000; Arroyo, Maté, Muñoz-San Roque 2006].

zakup samochodu w tys. zł (np. respondent 1: [10, 30], respondent 2: [30, 60], respondent 3: [50, 70]).

Zmienna wielowariantowa dopuszcza występowanie wielu kategorii lub wartości dla pojedynczego obiektu, chociażby znajomość języków obcych (np. respondent 1 włada językiem angielskim, francuskim i niemieckim, natomiast respondent 2 językiem francuskim i włoskim). Zmienne w ujęciu klasycznym są szczególnym przypadkiem zmiennej wielowariantowej. Podobną zmienną jest struktura udziałów kategorii lub wartości, z tym że wariantom nadane są stopnie ważności, prawdopodobieństwa, częstości lub udziały. Przykładem jest struktura wydatków konsumenta na zakupy w hipermarketach, np. respondent 1: {Auchan (50%), Real (30%), Tesco (20%)}, a respondent 2: {Tesco (80%) i Carrefour (20%)}

W pracy [Arroyo, Maté, Muñoz-San Roque 2006] został zaproponowany nowy typ zmiennej symbolicznej, tzw. zmienna pudełkowa (*boxplot variable*). Realizacją tej zmiennej dla obiektu są wartości pięciu miar położenia: wartość minimalna i maksymalna, kwartyle pierwszy i trzeci oraz mediana. Zmienną można traktować jako kompromis między szczegółowym (np. wysokość wydatków konsumenta na wyżywienie w ciągu roku), uśrednionym (np. średnie miesięczne wydatki na wyżywienie) a uogólnionym (np. przedział obejmujący minimalne i maksymalne miesięczne wydatki) opisem zjawiska.

Analiza danych symbolicznych daje ponadto możliwość uwzględnienia logicznych powiązań między zmiennymi. Zmienne pozostające w relacji tworzą tzw. zmienną strukturalną (*dependent variable*). Można wskazać zmienne o strukturze taksonomicznej, hierarchicznej i logicznej, różniące się charakterem relacji, jaka zachodzi między ich poziomami.

Systematyka zmiennej taksonomicznej (*taxonomic variable*) jest ogólnie przyjęta (znana) lub ustalana *a priori* przez badacza, np. samochód można scharakteryzować za pomocą marki, wersji i pojemności silnika (np. Honda Civic 1.8, Renault Laguna 2.0). W zmiennej hierarchicznej (*mother-daughter variable*) wariant z po-

ziomu wyższego determinuje warianty z poziomu niższego. Przykładem zmiennej są intencje zakupu samochodu, np. respondent 1: Tak, planuję kupić nowy samochód {Honda Civic lub Renault Laguna}; respondent 2: Nie, nie zamierzam nabywać samochodu {NA}. Kategoria NA (*Not Applicable*) występuje dlatego, że respondenta 2 nie dotyczy pytanie o markę samochodu. Zależność logiczna występuje z kolei wtedy, gdy wartość jednego poziomu zależy logicznie lub funkcyjnie od wartości drugiego poziomu zmiennej, np. waga i przeznaczenie auta: jeżeli ciężar samochodu przekracza 3,5 tony, to jest on klasyfikowany jako samochód ciężarowy.

Zasadność stosowania danych symbolicznych w badaniach segmentacyjnych wynika z:

- specyfiki zjawiska, np. gdy realizacją zmiennej nie jest pojedyncza wartość liczbową (dane nieprecyzyjne, trudne do oszacowania) lub kategoria (np. składniki produktu),
- konstrukcji instrumentu pomiarowego (kwestionariusza ankiety), m.in. kafeteria koniunktywna (np. preferowane marki produktu), informacje osobiste (np. miesięczny dochód konsumenta), pytania złożone, np. miejscowość zamieszkania respondenta (miasto {poniżej 100, 100-200, powyżej 200 tys. mieszkańców} lub wieś),
- braku dostępu do szczegółowych danych, np. w wyniku ich utajnienia,
- intencji badacza – gromadzenie danych w zagregowanej formie, np. struktura wydatków konsumenta na różne grupy towarowe, np. żywność, odzież, kosmetyki itd.,
- interpretacji badacza – szczególnie w sytuacji, gdy badacz dysponuje dużym zbiorem danych (np. bazą danych) i wstępnym krokiem analizy jest agregacja danych,
- opisu wyodrębnionych klas – dane symboliczne mogą posłużyć jako forma opisu klas obiektów charakteryzowanych zmiennymi symbolicznymi lub „klasycznymi”.

### 3. Techniki graficzne w procesie segmentacji rynku

Podstawowym założeniem w badaniach segmentacyjnych jest heterogeniczność rynku, która oznacza zróżnicowane potrzeby i wymagania konsumentów [Smith 1956]. Naturalnym etapem w procesie segmentacji jest badanie poziomu jednorodności obiektów. Wśród technik graficznych stosowanych w tym celu można wymienić metody skalowania wielowymiarowego. Umożliwiają one odwzorowanie relacji między obiektami wielowymiarowymi w zredukowanej przestrzeni, z zachowaniem pierwotnej struktury danych.

W dalszym etapie procedury segmentacji konieczne jest zastosowanie metod służących klasyfikacji w celu wyodrębnienia jednorodnych klas obiektów. Wśród najpopularniejszych metod segmentacji należy wskazać metody analizy skupień [Wedel, Kamakura 1998]. Wynikiem zastosowania tych metod jest podział zbioru

obiektów na klasy. Badacz dysponuje informacjami o strukturze zbioru obiektów, tj. liczbie i liczebności klas oraz przynależności obiektów do klas. W segmentacji istotne staje się ustalenie, z czego wynika podobieństwo obiektów w klasach. Podstawą opisu klas są zmienne, które brały udział w klasyfikacji obiektów. Charakterystykę klas można zilustrować na rysunkach i wykresach.

Ostatnim etapem procesu segmentacji jest profilowanie segmentów. Profilowanie polega na rozpoznaniu zmiennych istotnie różnicujących segmenty i sporządzeniu na tej podstawie profili segmentów. W profilowaniu biorą udział zmienne nieuczestniczące w klasyfikacji. Popularną metodą graficzną stosowaną w profilowaniu są drzewa klasyfikacyjne.

#### 4. Problem badawczy

W celu ilustracji technik wizualizacji wyników segmentacji rynku na podstawie danych symbolicznych zaprezentowany zostanie przykład empiryczny. Segmentacji poddano konsumentów dokonujących zakupów na rynku internetowym. Dane zgromadzono w badaniu ankietowym prowadzonym w okresie czerwiec-sierpień 2008 r. W przykładzie wykorzystano dane częściowe dotyczące 60 respondentów. Kryteria segmentacji zawiera tab. 1.

Tabela 1. Kryteria segmentacji „e-konsumentów”

Lp.	Zmienna	Rodzaj zmiennej symbolicznej	Zbiór realizacji zmiennej
1	2	3	4
1	Rola respondenta w procesie zakupu	zmienna wielowariantowa	{decydent, nabywca, płacący, użytkownik}
2	Przyczyny dokonywania zakupów przez Internet*	zmienna wielowariantowa	{oszczędność czasu, wysoka dostępność produktów, szeroki asortyment produktów, możliwość szybkiego porównania ofert, atrakcyjne ceny, dostawa „do domu”, dyskrecja, inne}
3	Istotność czynników wpływających na zakupy w Internecie (%)*	struktura udziałów kategorii	{marka, cena, opis produktu, promocja, szybkość dostawy, koszt przesyłki, rekomendacje innych} oraz macierz wag
4	Typ produktów najczęściej nabywanych <i>on-line</i>	zmienna wielowariantowa	{sprawdzone, nieużytkowane wcześniej, tradycyjne, nowoczesne}
5	Produkty kupowane najczęściej <i>on-line</i> *	struktura hierarchiczna (poziom II – zmienna wielowariantowa)	nie kupuję {NA}; kupuję {(książki, filmy, muzyka, gry komputerowe), (telefony, akcesoria GSM), (odzież, obuwie, galanteria), (sprzęt komputerowy, oprogramowanie), sprzęt fotograficzny, (kosmetyki pielęgnacyjne, perfumy), sprzęt sportowy, (sprzęt AGD, RTV), (artykuły dziecięce, zabawki), produkty spożywcze, inne}

1	2	3	4
6	Miejsce i częstotliwość dokonywania zakupów	struktura taksonomiczna (II poziom – zmienna porządkowa)	poziom 1: sklep internetowy lub aukcja; poziom 2: częstotliwość zakupów (kilka razy/tydzień, raz/tydzień, kilka razy/miesiąc, raz/miesiąc, kilka razy/rok, raz/rok, rzadziej niż raz/rok, nigdy)
7	Odpłatne usługi, z jakich korzysta respondent*	struktura hierarchiczna (poziom II – zmienna wielowariantowa)	nie korzystam {NA}; korzystam {porady prawne, serwisy ekonomiczne, tłumaczenia obcojęzyczne, edukacja on-line, transakcje bankowe, inne}
8	Miesięczne wydatki na produkty i usługi kupione w Internecie (zł)	przedział liczbowy nierozłączny	[0, 3000]

\* Respondent mógł wskazać maksymalnie 4 odpowiedzi.

Źródło: opracowanie własne.

	rola	przyczyny	czynniki	typ	produkty	sklep	aukcja	usługi	wydatki
20	1, 2, 4, 3	1, 5, 6, 3	1 (0.20), 2 (0.30), 5 (0.10), 3 (0.40)	2, 4, 1, 3	1, 8	[3.00 : 3.00]	[5.00 : 5.00]	7	[50.00 : 200.00]
21	1, 2, 4, 3	1, 5, 6, 3	1 (0.30), 2 (0.50), 7 (0.20)	1	8, 6, 3	[4.00 : 4.00]	[4.00 : 4.00]	2, 1	[90.00 : 100.00]
22	1, 4	1, 5, 6, 4	2 (0.50), 4 (0.30), 5 (0.20)	3	8, 6, 7	[5.00 : 5.00]	[5.00 : 5.00]	5, 3, 1	[100.00 : 300.00]
23	2, 4, 3	1, 6, 3, 4	1 (0.25), 2 (0.25), 5 (0.25), 6 (0.25)	2, 4, 1, 3	1, 8, 4, 5	[3.00 : 3.00]	[5.00 : 5.00]	5, 2	[10.00 : 200.00]
24	1, 2, 4, 3	5, 6, 3, 4	2 (0.50), 4 (0.20), 7 (0.15), 3 (0.15)	2, 4, 1, 3	9	[3.00 : 3.00]	[8.00 : 8.00]	7	[10.00 : 500.00]
25	2, 3	1, 2, 3, 4	1 (0.50), 2 (0.40), 4 (0.05), 5 (0.05)	1	8, 2, 5	[5.00 : 5.00]	[5.00 : 5.00]	5	[150.00 : 200.00]

Rys. 2. Fragment tablicy danych symbolicznych

Źródło: opracowanie własne w programie Sodas v. 2.0.

Rysunek 2 prezentuje fragment tablicy danych symbolicznych charakteryzujących konsumentów.

#### 4.1. Badanie poziomu jednorodności obiektów symbolicznych

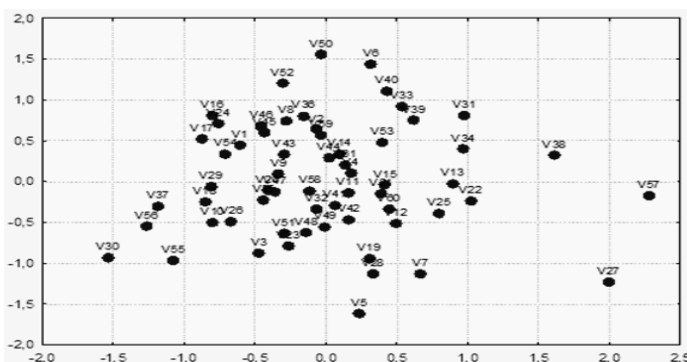
Badanie poziomu jednorodności zbioru obiektów opisanych zmiennymi symbolicznymi można przeprowadzić z wykorzystaniem metod skalowania wielowymiarowego dla danych symbolicznych (tab. 2).

Technika *Bi-Dimensional-Mapping* jest rozwinięciem nieliniowego odwzorowania Sammona (*Sammon's nonlinear mapping*) na obiekty symboliczne. W przeciwieństwie do pozostałych algorytmów bazuje na macierzy odległości obiektów, dlatego może być stosowana w sytuacji, kiedy obiekty opisane są dowolnym rodzajem zmiennych symbolicznych. Zastosowanie jej w omawianym przykładzie jest zatem uzasadnione (rys. 3).

Tabela 2. Metody skalowania wielowymiarowego dla danych symbolicznych

Algorytmy (źródło literaturowe)	Dane wejściowe	Typ zmiennych symbolicznych	Reprezentacja obiektów w przestrzeni
<i>Bi-Dimensional-Mapping</i> [Noirhomme-Fraiture 2004a, s. 337]	tablica danych symbolicznych ⇒ macierz odległości macierz odległości	dowolne	punkty w przestrzeni 2-wymiarowej ( $s = 2$ )
<i>InterScal</i> [Denoeux, Mason 2000] <i>SymScal</i> [Groenen i in. 2005] <i>I-Scal</i> [Groenen i in. 2006]	tablica danych symbolicznych ⇒ macierz odległości minimalnych i maksymalnych macierz odległości minimalnych i maksymalnych	przedziały liczbowe dowolne	hiperprostopa- dłościany, np. dla $s = 2$ – prostokąt, $s = 3$ – prostopa- dłościan

Źródło: opracowanie własne.



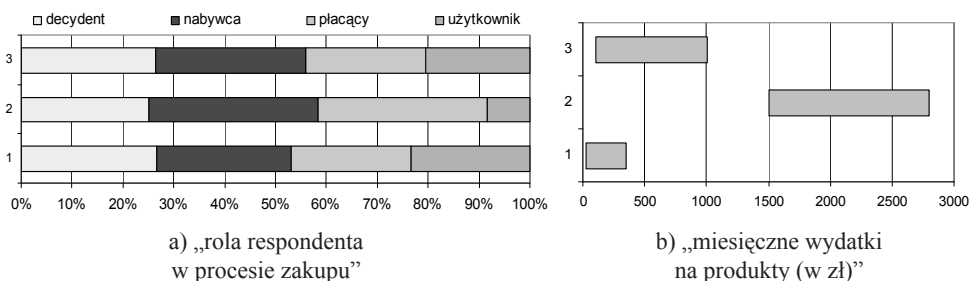
Rys. 3. Konfiguracja punktów (obiektów) otrzymana metodą *Bi-Dimensional-Mapping*

Źródło: opracowanie własne (program Sodas v. 2.0).

Wyniki analizy potwierdzają heterogeniczność zbioru obiektów, jednak na ich podstawie trudno jest jednoznacznie wyodrębnić klasy. W dalszej kolejności przeprowadzono klasyfikację obiektów. Wyznaczono odległości obiektów symbolicznych za pomocą miary Ichino-Yaguchiego (zob. [Gordon 1999, s. 138-139; Wilk 2006b]) i dokonano klasyfikacji metodami hierarchicznymi (Warda i kompletnego połączenia). Indeks sylwetkowy, służący do oceny jakości klasyfikacji, jako najlepszą wskazał strukturę 3 klas. Następnie dokonano grupowania obiektów optymalizacyjną metodą klasyfikacji dla danych symbolicznych SCLUST (zob. [Noirhomme-Fraiture 2004b, s. 159-163]).

## 4.2. Interpretacja klas obiektów symbolicznych

Do opisu klas obiektów charakteryzowanych zmiennymi symbolicznymi wykorzystano technikę CLINT (zob. [Noirhomme-Fraiture 2004b, s. 207-211; Wilk 2007]). W przykładzie zastosowano formułę „w szerokim zakresie”, która polega na konstrukcji „nowego” obiektu symbolicznego (tzw. obiektu syntetycznego) obejmującego wszystkie elementy klasy. Jego konstrukcja jest determinowana rodzajem zmiennych opisujących obiekty. Dla zmiennych symbolicznych, których realizacją jest przedział liczbowy, wyznaczono najmniejszy przedział liczbowy, obejmujący wszystkie przedziały realizacji zmiennej obserwowane dla obiektów należących do klasy. Natomiast w przypadku zmiennych wielowariantowych wyznaczono zbiór kategorii, w którym znalazły się wszystkie obserwowane w klasie kategorie zmiennej. Opis uzupełniono, wyznaczając frakcje występowania w każdej klasie poszczególnych kategorii zmiennej. Dla zmiennej symbolicznej, której realizacją jest struktura udziałów kategorii, wybrano maksymalną wagę wariantu obserwowaną dla obiektów klasy. Charakterystykę skupień uzyskanych metodą CLINT można zaprezentować graficznie za pomocą wykresów prostych (klasy-zmienna) oraz wykresów VSTAR. Wykresy proste ułatwiają porównanie klas względem każdej ze zmiennych. Rysunek 4 prezentuje charakterystykę klas w odniesieniu do zmiennej wielowariantowej „rola respondenta w procesie zakupu” oraz zmiennej w postaci przedziału liczbowego „miesięczne wydatki na produkty i usługi zakupione w Internecie”.



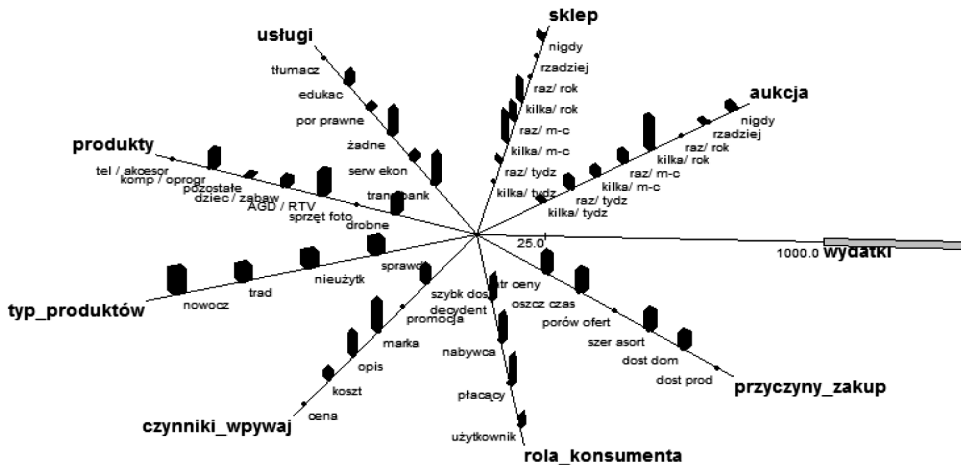
Rys. 4. Wykresy proste (klasy-zmienna)

Źródło: opracowanie własne.

W przeciwieństwie do wykresów prostych wykresy VSTAR (ZoomStar) umożliwiają prezentację pełnej charakterystyki jednej klasy bądź kilku klas (zob. np. [Calvo-Garrido 2000; Wilk 2006a]). Dają możliwość prezentacji klas obiektów opisanych dowolnym rodzajem zmiennych symbolicznych (z wyjątkiem zmiennych pudełkowych). Osie wykresów reprezentują zmienne symboliczne i ich realizacje w klasach. Na jednym wykresie można umieścić maksymalnie 24 zmienne i 15 wariantów.

tów zmiennej na jednej osi. Można wskazać dwa rodzaje wykresów VSTAR: wykresy typu 2D oraz typu 3D.

Charakterystykę klasy 2 zilustrowano na wykresie typu 3D (rys. 5). Słupki oznaczają wagi wariantów zmiennych symbolicznych w postaci struktury udziałowej, natomiast zaznaczony obszar na osi charakteryzuje „wydatki na zakupy w Internecie”.



Rys. 5. Charakterystyka segmentu 2 – wykres VSTAR typu 3D

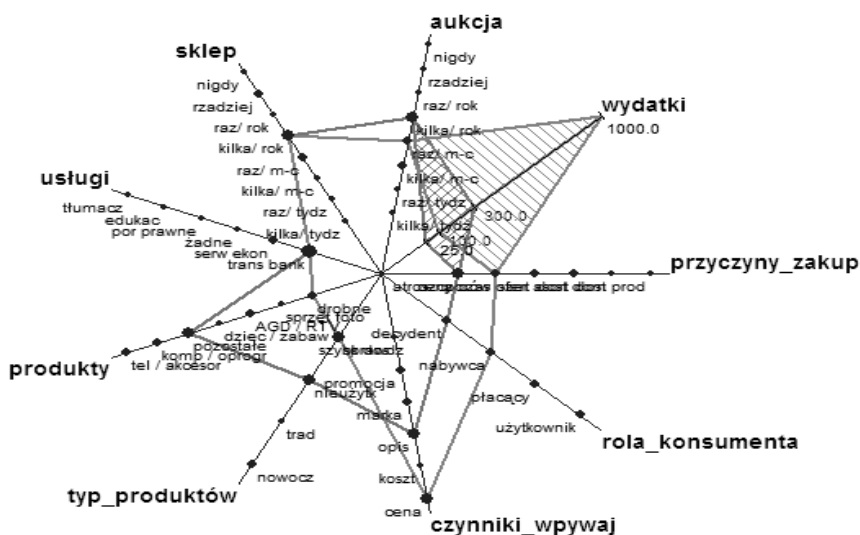
Źródło: opracowanie w programie Sodas v. 2.0.

Segment 2 cechuje wysoka częstotliwość zakupów i miesięczne wydatki (w granicach 1500-2750 zł). Powodem zakupów przez Internet jest oszczędność czasu i atrakcyjne ceny. Konsumenty korzystają głównie z oferty sklepów internetowych, a istotnym kryterium wyboru produktu jest cena. Najczęściej nabywają produkty nowoczesne (sprzęt komputerowy i AGD/RTV) i produkty drobne (książki, gry komputerowe itp.). Zwykle nie użytkują zakupionych produktów. Oprócz bankowości elektronicznej nie korzystają z usług *on-line*.

Porównanie dwóch klas jest wygodniejsze z wykorzystaniem wykresów 2D (rys. 6).

Krzywa łącząca warianty zmiennych określa charakterystykę klasy, punkty symbolizują wagi wariantów zmiennej, a obszar w paski – przedział liczbowy. Miesięczne wydatki konsumentów reprezentujących segment 1 są znacznie niższe w porównaniu z segmentem 3. W segmencie 1 głównym powodem zakupów przez Internet są atrakcyjne ceny, a w segmencie 3 – oszczędność czasu. Dominującym kryterium wyboru produktu dla segmentu 1 jest jego opis, a dla segmentu 3 – cena. Konsumenty reprezentujący segment 1 najczęściej nabywają produkty nieużytkowane wcześniej, a segment 3 – produkty sprawdzone.





Rys. 6. Porównanie charakterystyk segmentów 1 i 3 – wykres VSTAR typu 2D

Źródło: opracowanie w programie Sodas v. 2.0.

### 4.3. Profilowanie klas obiektów symbolicznych

W dyskryminacji klas obiektów opisanych zmiennymi symbolicznymi proponowane są trzy algorytmy drzew klasyfikacyjnych (tab. 3).

Tabela 3. Algorytmy drzew klasyfikacyjnych dla danych symbolicznych

Algorytmy	Sposób podziału	Zmienna zależna	Zmienne niezależne
<i>TREE</i> [Bock, Diday 2000, s. 245-261]	binarny	nominalna	przedział liczbowy, zmienna wielowariantowa, struktura udziałów
<i>SDT</i> [Noirhomme-Fraiture 2004b, s. 273-283]	dowolny	nominalna	struktura udziałów
<i>SBTREE</i> [Noirhomme-Fraiture 2004b, s. 287-294]	binarny	nominalna dwustanowa	przedział liczbowy

Źródło: opracowanie własne.

Tabela 4 zawiera zbiór potencjalnych zmiennych profilowych opisujących „e-konsumentów”.

W przykładzie wykorzystano algorytm *TREE*, ponieważ może być stosowany, gdy obiekty opisane są zmiennymi symbolicznymi, których realizacją jest przedział liczbowy, zbiór wariantów lub struktura udziałów. Algorytm nie uwzględnia zmiennych strukturalnych, ale istnieje możliwość wykorzystania części informacji przez

Tabela 4. Potencjalne zmienne profilowe

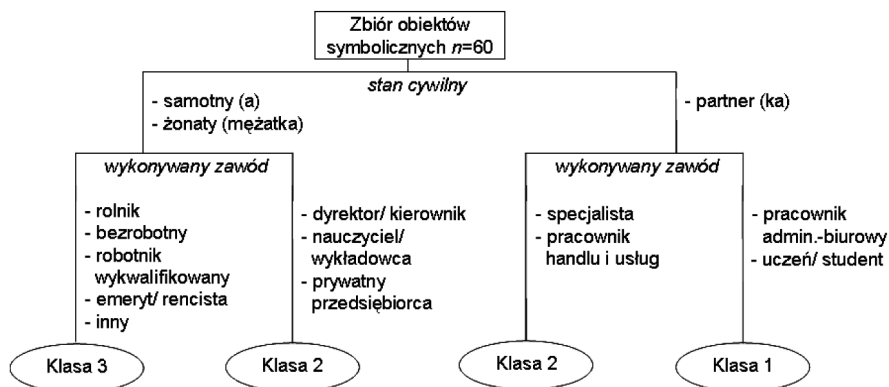
Lp.	Zmienna	Zmienna symboliczna	Zbiór realizacji zmiennej
1	Płeć	nominalna dwustanowa	kobieta, mężczyzna
2	Wiek (lata)	przedział liczbowy rozłączny	[18, 19], [20, 25], [26, 30], [31, 35], [36, 40], [41, 45], [46, 50], [51, 55], [56, 60], [61, ∞]
3	Poziom wykształcenia	struktura taksonomiczna (II poziom – porządkowa)	podstawowe, gimnazjalne, zasadnicze zawodowe, niepełne średnie (ogólnokształcące/zawodowe), średnie (ogólnokształcące/zawodowe), pomaturalne, policealne, licencjackie, wyższe (inżynierskie/ magisterskie)
4	Stan cywilny	nominalna	żonaty/mężatka, partner (ka), wdowiec/wdowa, samotny (a)
5	Liczba mieszkańców	struktura hierarchiczna (II poziom – przedział liczbowy rozłączny)	- wieś: {NA}, - miasto: (0, 50], (50, 100], (100, ∞)
6	Wykonywany zawód	zmienna wielowariantowa	{dyrektor/kierownik, specjalista, prywatny przedsiębiorca, pracownik administracyjno-biurowy, pracownik usług/handlu, nauczyciel/wykładowca, robotnik wykwalifikowany, robotnik niewykwalifikowany, rolnik, wolny zawód, gospodyni domowa, uczeń/student, emeryt/rencista, bezrobotny, inny}
7	Miesięczny dochód netto	struktura hierarchiczna (II poziom – przedział liczbowy rozłączny)	- brak dochodu {NA}, - dochód (zł): (0, 1000], [1000, 1500], [1501, 2000], [2001, 2500], [2501, 3000], [3001, 4000], [4001, ∞)
8	Źródła dochodu	zmienna wielowariantowa	{praca zarobkowa, renta, emerytura, stypendium, kieszonkowe, inne (np. zasiłek)}

Źródło: opracowanie własne.

przekształcenie zmiennych. W strukturze hierarchicznej pozostawiono II poziom zmiennej, a wariant NA zamieniono na wartość 0, natomiast w strukturze taksonomicznej warianty zmiennej potraktowano jako osobne kategorie, np. wykształcenie średnie ogólnokształcące.

Ustalono następujące parametry algorytmu TREE: liczba końcowych węzłów – 7, liczebność graniczna dla rozmiaru węzła – 5, kryterium jakości podziału – kryterium informacji. Zmiennymi najistotniej różnicującymi segmenty okazały się „stan cywilny” i „wykonywany zawód”, a w dalszej kolejności „źródła dochodów”, „liczba mieszkańców” i „miesięczny dochód netto”. Uzyskane drzewo klasyfikacyjne prezentuje rys. 7.

Segment 1 reprezentują osoby żyjące w związku partnerskim, najczęściej uczniowie lub studenci oraz pracownicy administracyjno-biurowi. Do segmentu 2 należą prywatni przedsiębiorcy, dyrektorzy i kierownicy, nauczyciele i wykładowcy, osoby



Rys. 7. Drzewo klasyfikacyjne uzyskane z wykorzystaniem algorytmu TREE

Źródło: wykres sporządzony na podstawie wyników uzyskanych w programie Sodas v. 2.0.

pracujące w handlu i usługach oraz specjaliści. Segment 3 skupia osoby pracujące (robotnicy lub rolnicy) i bezrobotne, żyjące w związku małżeńskim lub samotne, emerytów i rencistów.

## 5. Podsumowanie

Analiza danych symbolicznych otwiera nowe możliwości postrzegania, formułowania i analizowania problemów segmentacyjnych. Złożoność danych symbolicznych implikuje konieczność stosowania specyficznych narzędzi wizualizacji wyników badań segmentacyjnych. Wśród podstawowych można wskazać technikę *Bi-Dimensional-Mapping*, wykresy proste (klasy-zmienna), wykresy VSTAR oraz algorytm TREE z rodziny drzew klasyfikacyjnych.

## Literatura

- Arroyo J., Maté C., Muñoz-San Roque A., *Hierarchical Clustering for Boxplot Variables*, [w:] *Data Science and Classification, Studies in Classification, Data Analysis and Knowledge Organization*, V. Batagelj, H.H. Bock, A. Ferligoj, A. Žiberna (red.), Springer-Verlag, Berlin-Heidelberg 2006, s. 59-66.
- Bock H.H., Diday E. (red.), *Analysis of Symbolic Data. Explanatory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin-Heidelberg 2000.
- Calvo-Garrido P., *Applications of Symbolic Objects In Official Statistics*, EUSTAT Technical Notebook, Instituto Vasco de Estadística, Donostia – San Sebastian 2000.
- Denoeux T., Masson M., *Multidimensional Scaling of Interval-Valued Dissimilarity Data*, „Pattern Recognition Letters” 2000 vol. 21, s. 83-92.
- Gordon A. D., *Classification*, Chapman and Hall/CRC, London 1999.

- Groenen P. Winsberg S., Rodriguez O., Diday E., *I-Scal: Multidimensional Scaling of Interval Dissimilarities*, „Computational Statistics and Data Analysis” 2006 vol. 51, s. 360-378.
- Groenen P.J.F., Winsberg S., Rodriguez O., Diday E., *SymScal: Symbolic Multidimensional Scaling of Interval Dissimilarities*, „Econometric Report” 2005-15, Rotterdam 2005.
- Noirhomme-Fraiture M. (red.), *Help Guide for SODAS 2 Software*, Software Report, Analysis System of Symbolic Official Data, Project Number IST-2000-25161, 2004a.
- Noirhomme-Fraiture M. (red.), *User Manual for SODAS 2 Software*, Software Report, Analysis System of Symbolic Official Data, Project Number IST-2000-25161, 2004b.
- Smith W.R., *Product Differentiation and Market Segmentation as Alternative Marketing Strategies*, „Journal of Marketing”, July 1956, vol. 21, Issue 1.
- Wedel M., Kamakura W.A., *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publisher, Boston-Dordrecht-London 1998.
- Wilk J., *Graficzna prezentacja obiektów symbolicznych*, [w:] *Ekonometria 16, Zastosowania metod ilościowych*, J. Dziechciarz (red.), Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1100, AE, Wrocław 2006a, s. 86-96.
- Wilk J., *Problemy klasyfikacji obiektów symbolicznych. Symboliczne miary odległości*, [w:] *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, J. Garczarczyk (red.), ZN AE w Poznaniu nr 71, Poznań 2006b, s. 69-83.
- Wilk J., *Interpretacja i profilowanie klas w analizie danych symbolicznych*, [w:] *Taksonomia 14, Klasyfikacja i analiza danych – teoria i zastosowania*, K. Jajuga, M. Walesiak (red.), Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1169, AE, Wrocław 2007, s. 149-159.

## VISUALIZATION TECHNIQUES IN MARKET SEGMENTATION PROCEDURE BASED ON SYMBOLIC DATA

### Summary

Market segmentation is a key strategic management tool. It enables to meet the customers' needs more precisely, gain a higher market share and profits. Marketing information is collected through questionnaire surveys or from data bases. The data complexity prevents application of classical data to describe customers. Symbolic data application allows us to describe customers in an explanatory way.

Implementing of the market segmentation strategy requires a thorough understanding of the research results. Visualization techniques application simplifies communicating the findings to practitioners. The aim of the paper is to introduce visualization techniques that can be applied in market segmentation research based on symbolic data.