

**Justyna Wilk**

## **GRAFICZNA PREZENTACJA OBIEKTÓW SYMBOLICZNYCH**

### **1. Wstęp**

Koncepcja obiektów symbolicznych została zapoczątkowana w pracach Michalskiego [10] oraz Michalskiego, Steppa i Didaya [11] nad klasyfikacją pojęciową (*conceptual clustering*) [12, s. 81-82]. Tematyka obiektów symbolicznych została rozwinięta w ramach projektu SODAS<sup>1</sup>. Celem projektu było wykorzystanie danych symbolicznych do analizy dużych zbiorów danych, opisywanych przez zmienne o złożonej, nietypowej strukturze, szczególnie w instytucjach przeprowadzających analizy statystyczne i badania marketingowe.

Wśród metod statystycznej analizy wielowymiarowej, rozszerzonych na obiekty symboliczne, można wskazać m.in. symboliczną analizę dyskryminacyjną, metody klasyfikacji, drzewa klasyfikacyjne i analizę czynnikową [5, s. 17-18; 2, s. 33-39].

W artykule zaprezentowano podstawowe pojęcia analizy danych symbolicznych (SDA), charakterystykę zmiennych symbolicznych oraz sposoby graficznej prezentacji obiektów symbolicznych.

### **2. Podstawowe pojęcia analizy danych symbolicznych**

Obiekt, w rozumieniu tradycyjnym, jest elementarną jednostką badawczą, opisaną za pomocą zmiennych „klasycznych”, tj. zmiennych mierzonych na mocnych (ilorazowa lub przedziałowa) lub słabych (porządkowa lub nominalna) skalach pomiaru. Jeśli zbiór obiektów oznaczmy jako  $A = \{A_i\}_1^n = \{A_1, A_2, \dots, A_n\}$ , to zmienna „klasyczna”  $M_j$  stanowi odwzorowanie:  $M_j: A \rightarrow R$  ( $j = 1, 2, \dots, m$ ),

---

<sup>1</sup> SODAS (Symbolic Official Data Analysis System), project no. 20281 of the European Commission, General Directory III, Industrial rtd, Eurostat, DOSES program.

gdzie  $R$  oznacza zbiór liczb rzeczywistych, a  $m$  – liczbę zmiennych. Obiekty opisane tego rodzaju zmiennymi są zapisane w macierzy danych, w której wartość  $j$ -tej zmiennej zaobserwowana w  $i$ -tym obiekcie oznaczona jest przez  $x_{ij}$  ( $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ ) [8, s. 17-18].

Przykładowe obiekty i zmienne „klasyczne”, tj. bankowe konta osobiste oraz ich charakterystyki, czyli „miesięczną opłatę za prowadzenie rachunku” i „oprocentowanie w skali roku” przedstawiono w tab. 1.

Tabela 1. Obiekty i zmienne „klasyczne” – bankowe konta osobiste

| Obiekty „klasyczne”    | Zmienne „klasyczne”                         |                                 |
|------------------------|---------------------------------------------|---------------------------------|
|                        | miesięczna opłata za prowadzenie konta (zł) | oprocentowanie w skali roku (%) |
| Konto 24 Prestiż       | 11                                          | 1,25                            |
| Konto 24               | 6                                           | 0                               |
| Konto 24 X-tra Student | 0                                           | 0                               |

Źródło: opracowanie na podstawie danych BZ WBK SA z 20.11.2004.

W analizie danych symbolicznych (SDA) obiektami symbolicznymi są obiekty opisane za pomocą zmiennych symbolicznych [13, s. 299; 12, s. 83-84; 7, s. 67-70].

Jeśli  $\Omega$  to zbiór obiektów symbolicznych, a  $V$  – zbiór zmiennych opisujących obiekty symboliczne, to zmienna  $V_j$  jest odwzorowaniem  $\Omega \rightarrow O$ , gdzie  $O$  jest zbiorem realizacji zmiennej, tj. ( $V_j \subset O$ ) [4, s. 7; 5, s. 9].

Zmienne symboliczne mogą mieć charakter [1, s. 471-478; 3, s. 2, 32-37; 5, s. 7; 14]:

1) przedziałów klasowych (*interval-valued variable*) rozłącznych lub spójnych (nakładających się);

2) listy kategorii lub wartości (*multivalued variable*) – to zmienne dopuszczające występowanie wielu kategorii lub wartości (w tym przedziałów klasowych) tej samej zmiennej dla jednego obiektu.

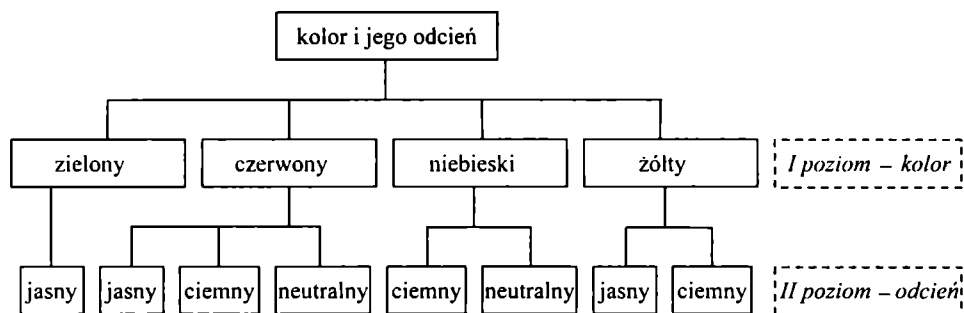
3) listy kategorii lub wartości z wagami, prawdopodobieństwami czy częstościami (*multivalued modal variable*) – to zmienne podobne do listy kategorii lub wartości, z tym że wariantom przyporządkowane są stopnie ważności;

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 (brak wewnętrznej zależności logicznej lub stochastycznej – rys. 1),

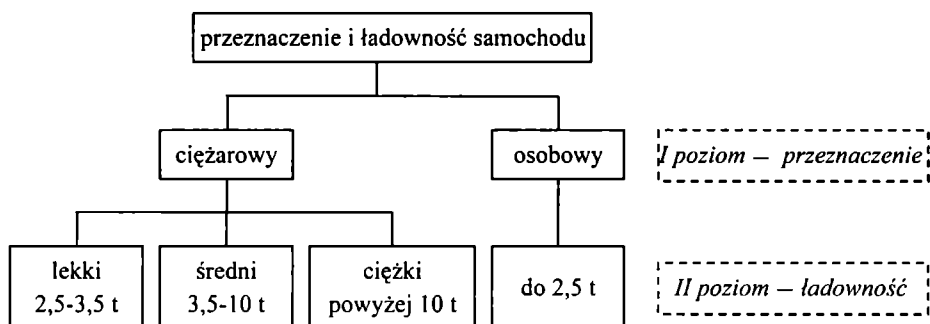
b) hierarchicznej (*hierarchical dependent, mother-daughter*), gdy dla jednego z wariantów nie można określić podwariantu („brakujący” podwariant w analizie symbolicznej ma wartość *NA* – *non-applicable*),

c) logicznej (*logical dependence*); występuje między zmiennymi  $Y$  i  $Z$ , gdy wartość drugiej zmiennej  $Z$  zależy logicznie lub funkcyjnie od wartości pierwszej zmiennej  $Y$  (rys. 2),



Rys. 1. Przykład zmiennej o strukturze taksonomicznej – kolor i jego odcień

Źródło: opracowanie własne.



Rys. 2. Przykład zmiennej o strukturze logicznej – przeznaczenie i ładowność samochodu

Źródło: opracowanie własne.

d) stochastycznej (*stochastic dependence*), gdzie szczeble hierarchii są ze sobą skorelowane (współczynnik korelacji i kowariancji jest różny od zera).

W zestawie zmiennych charakteryzujących obiekty symboliczne mogą się znaleźć również zmienne „klasyczne”, tj. zmienne mierzone na:

a) mocnych skalach pomiaru (ilorazowe, przedziałowe); wartość zmiennej  $y = a$  w analizie danych symbolicznych jest traktowana jak przedział klasowy  $\xi = [a, a]$ ;

b) słabych skalach pomiaru (porządkowe, nominalne); są traktowane w SDA jako zmienne w postaci listy wartości z wagami, z tym że dla poszczególnego obiektu można określić wyłącznie jeden wariant zmiennej o wadze równej 1,0.

Można wyróżnić dwa podstawowe typy obiektów symbolicznych, a mianowicie obiekty symboliczne pierwszego i drugiego rzędu [4, s. 9-10].

Obiekt symboliczny pierwszego rzędu (*first order symbolic object*), podobnie jak w przypadku obiektów „klasycznych”, charakteryzuje elementarną jednostkę badania (*elementary event*), np. pojedynczego konsumenta, produkt lub przedsiębiorstwo.

Przykładowo, w badaniu zadowolenia z użytkowanego auta, obiektami symbolicznymi I rzędu będą poszczególni respondenci (zob. tab. 2). Zmienna „zadowolenie z użytkowanego auta” przyjmuje wartości na skali 5-stopniowej, gdzie 1 oznacza „niezadowolony”, a 5 – „zadowolony”. Zmienna „dochód” jest zmienną symboliczną w postaci przedziału klasowego rozłącznego.

Tabela 2. Obiekty symboliczne I rzędu

| Obiekty symboliczne | Zmienne symboliczne i zmienne klasyczne |                                 |           |               |
|---------------------|-----------------------------------------|---------------------------------|-----------|---------------|
|                     | marka                                   | zadowolenie z użytkowanego auta | pleć      | dochód (zł)   |
| Respondent 1        | Toyota                                  | 3                               | Kobieta   | [1000 : 1500] |
| Respondent 2        | Škoda                                   | 2                               | Mężczyzna | [1500 : 2000] |
| Respondent 3        | Peugeot                                 | 4                               | Mężczyzna | [1000 : 1500] |
| Respondent 4        | Toyota                                  | 4                               | Kobieta   | [1500 : 2000] |
| Respondent 5        | Peugeot                                 | 5                               | Kobieta   | [2000 : 2500] |
| Respondent 6        | Peugeot                                 | 5                               | Mężczyzna | [2000 : 2500] |
| Respondent 7        | Toyota                                  | 4                               | Mężczyzna | [1000 : 1500] |
| Respondent 8        | Škoda                                   | 1                               | Kobieta   | [500 : 1000]  |
| Respondent 9        | Peugeot                                 | 2                               | Mężczyzna | [1500 : 2000] |
| Respondent 10       | Škoda                                   | 3                               | Mężczyzna | [2000 : 2500] |

Źródło: opracowanie własne.

Obiekt symboliczny II rzędu (*second order symbolic object*) powstaje przez grupowanie (agregowanie) obiektów pierwszego rzędu w zespoły (klasy) o podobnych (wspólnych) cechach [4, s. 9-10].

Tworzy się go na podstawie informacji uzyskanych z relacyjnych baz danych, określając kwerendy różnego typu [4, s. 12-20; 6, s. 287-288].

Przykładowo, obiekty I rzędu z tab. 2 można wyrazić w zagregowanej (wg „marki”) formie, jako obiekty symboliczne II rzędu:  $o_1$ ,  $o_2$ ,  $o_3$ , np. obiekt  $o_1$  charakteryzuje poziom zadowolenia z użytkowania samochodów marki Toyota respondentów: 1, 4, 7 (zob. tab. 3). Zmienne „zadowolenie z użytkowanego auta” oraz „dochód” po agregacji mają charakter listy wartości z wagami, a zmienna „pleć” – listy kategorii z wagami.

Calvo-Garrido wskazuje trzy rodzaje obiektów symbolicznych II rzędu, tj. obiekty typu *Assertion*, *Hoard* i *Synthesis* [4, s. 8-10; 12, s. 82; 9, s. 2-4].

Obiekty symboliczne można również podzielić ze względu na typ zmiennych, które je opisują, tj. [5, s. 10; 4, s. 7; 2, s. 12-13]:

1) obiekty typu Boolean (*Boolean symbolic objects*), jeśli obiekty są opisane za pomocą zmiennych symbolicznych typu 1, 2, 4;

Tabela 3. Obiekty symboliczne II rzędu

| Obiekty symboliczne | Zmienne symboliczne i zmienne klasyczne |                                          |                                                             |                                                                        |
|---------------------|-----------------------------------------|------------------------------------------|-------------------------------------------------------------|------------------------------------------------------------------------|
|                     | Marka                                   | Zadowolenie z użytkowanego auta          | Płeć                                                        | Dochód                                                                 |
| $o_1$               | Toyota                                  | $\{3 (^{1/3}), 4 (^{2/3})\}$             | $\{\text{mężczyzna } (^{1/3}), \text{ kobieta } (^{2/3})\}$ | $\{[1500:2000] (^{1/3}), [1000:1500] (^{2/3})\}$                       |
| $o_2$               | Škoda                                   | $\{1 (^{1/3}), 2 (^{1/3}), 3 (^{1/3})\}$ | $\{\text{kobieta } (^{1/3}), \text{ mężczyzna } (^{2/3})\}$ | $\{[500:1000] (^{1/3}), [1500:2000] (^{1/3}), [2000:2500] (^{1/3})\}$  |
| $o_3$               | Peugeot                                 | $\{2 (^{1/4}), 4 (^{1/4}), 5 (^{1/2})\}$ | $\{\text{mężczyzna } (^{1/1})\}$                            | $\{[1000:1500] (^{1/4}), [1500:2000] (^{1/4}), [2000:2500] (^{1/2})\}$ |

Źródło: opracowanie własne na podstawie danych z tab. 2.

2) obiekty modalne (*modal symbolic objects*), gdy obiekty charakteryzowane są przez zmienną symboliczną (zmienne) typu 3.

Jest to podział istotny ze względu na to, że w przypadku obiektów opisanych zmiennymi symbolicznymi z wagami stosuje się techniki analityczne (m.in. metody pomiaru odległości) inne niż w przypadku obiektów opisanych pozostałymi typami zmiennych symbolicznych.

### 3. Metody prezentacji obiektów symbolicznych

Obiekty symboliczne, w odróżnieniu od obiektów „klasycznych”, mogą być prezentowane na kilka sposobów, a mianowicie w tablicy danych symbolicznych, za pomocą wykresów 2D i 3D oraz za pomocą „SOL” [4, s. 21-25].

#### 3.1. Tablica danych symbolicznych

W tablicy danych symbolicznych (*symbolic data table*) wiersze zawierają obiekty symboliczne, a kolumny zmienne symboliczne (zob. tab. 4).

W tab. 5 zawarto modalne obiekty symboliczne II rzędu, tj. internetowe konta pocztowe (zagregowane ze względu na wysokość rocznej opłaty) opisane za pomocą zmiennych:

- w formie przedziału klasowego spójnego – „okres dostępu do konta” (w miesiącach),
  - o strukturze hierarchicznej – usługa „powiadomienia sms o otrzymanej wiadomości”, do sieci Era, Idea lub Plus GSM,
- w postaci listy kategorii – „organizery”, w skład organizera mogą wchodzić kontakty (tj. książka adresowa), zadania, notatnik, terminarz, aktówka,
- w postaci listy kategorii z wagami – „poziom ochrony antywirusowej”,
- ilorazowych – „roczna opłata za prowadzenie konta” (w zł), „pojemność skrzynki” (GB),
- nominalnych binarnych – „obsługa konta z poziomu telefonu komórkowego” i „dodawanie przez serwer reklam do wysyłanych wiadomości”.

Tabela 4. Pierwotna tabela danych (obiekty symboliczne I rzędu)

| Obiekt            | Roczna opłata | Pojemność | Okres dostępu | Obsługa z telefonu | Reklamy | Powiadomienie sms | Idea Era Plus | Organizer                                       | Ochrona antywirusowa |
|-------------------|---------------|-----------|---------------|--------------------|---------|-------------------|---------------|-------------------------------------------------|----------------------|
| WirtualnaPolska   | 0             | 0,25      | [36 : 36]     | tak                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz          | podstawowa           |
| Intertia          | 0             | 0,25      | [36 : 36]     | tak                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz          | podstawowa           |
| YahooMail         | 0             | 0,25      | [36 : 36]     | tak                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz          | podwójna             |
| Tlen              | 0             | 0,25      | [36 : 36]     | tak                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz          | trzystopniowa        |
| Plus              | 33            | 1         | [12 : 12]     | tak                | tak     | tak               | Plus          | kontakty, zadania, notatnik, terminarz          | podstawowa           |
| PocztaPro         | 33            | 1         | [12 : 12]     | tak                | tak     | tak               | Plus          | kontakty, zadania, notatnik, terminarz          | podwójna             |
| OnetPoczta        | 33            | 1         | [12 : 12]     | tak                | tak     | tak               | Plus          | kontakty, zadania, notatnik, terminarz          | trzystopniowa        |
| 2com              | 33            | 1         | [12 : 12]     | tak                | tak     | tak               | Plus          | kontakty, zadania, notatnik, terminarz          | brak                 |
| InfocityPocztaPro | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz, aktówka | podstawowa           |
| Hoga              | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz, aktówka | podwójna             |
| Hotmail           | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz, aktówka | trzystopniowa        |
| Gazeta            | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz, aktówka | trzystopniowa        |
| Tenbit            | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty, zadania, notatnik, terminarz, aktówka | brak                 |
| Premium           | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty, zadania, notatnik, terminarz, aktówka | podstawowa           |
| PocztaBiz         | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty, zadania, notatnik, terminarz, aktówka | podstawowa           |
| PocztaPlus        | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty, zadania, notatnik, terminarz, aktówka | trzystopniowa        |
| Poczta50          | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty, zadania, notatnik, terminarz, aktówka | trzystopniowa        |
| Pocztaamax50      | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty, zadania, notatnik, terminarz, aktówka | trzystopniowa        |
| PocztaBiznes100   | 90            | 1         | [6 : 12]      | tak                | tak     | tak               | Era           | kontakty, notatnik                              | podstawowa           |
| Poczta70          | 90            | 1         | [6 : 12]      | tak                | tak     | tak               | Era           | kontakty, notatnik                              | podstawowa           |
| OnetPocztaPlus    | 97            | 2         | [12 : 36]     | tak                | tak     | tak               | Idea Plus     | kontakty, notatnik                              | podwójna             |
| YahooMailPlus     | 160           | 2         | [12 : 12]     | tak                | nie     | nie               | NA            | kontakty                                        | trzystopniowa        |

Źródło: opracowanie własne.

Tabela 5. Tablica danych symbolicznych

| Obiekt | Roczna opłata | Pojemność | Okres dostępu | Obsługa z telefonu | Reklamy | Powiadamianie sms | Idea Era Plus | Organizer                                               | Ochrona antywirusowa                                                        |
|--------|---------------|-----------|---------------|--------------------|---------|-------------------|---------------|---------------------------------------------------------|-----------------------------------------------------------------------------|
| A      | 0             | 0,25      | [33 : 36]     | tak                | nie     | nie               | NA            | kontakty<br>zadania<br>notatnik<br>terminarz            | podstawowa (0,50)<br>podwójna (0,25)<br>trzystopniowa (0,25)                |
| B      | 33            | 1         | [1 : 12]      | tak                | tak     | tak               | Plus          | kontakty<br>zadania<br>notatnik<br>terminarz            | podstawowa (0,25)<br>podwójna (0,25)<br>trzystopniowa (0,25)<br>brak (0,25) |
| C      | 61            | 2,5       | [12 : 12]     | nie                | nie     | nie               | NA            | kontakty<br>zadania<br>notatnik<br>terminarz<br>aktówka | podstawowa (0,20)<br>podwójna (0,20)<br>trzystopniowa (0,40)<br>brak (0,20) |
| D      | 66            | 2         | [1 : 12]      | nie                | nie     | tak               | Idea Era Plus | kontakty<br>zadania<br>notatnik<br>terminarz<br>aktówka | podstawowa (0,40)<br>trzystopniowa (0,60)                                   |
| E      | 90            | 1         | [6 : 12]      | tak                | tak     | tak               | Era           | kontakty<br>notatnik                                    | podstawowa (1,00)                                                           |
| F      | 97            | 2         | [12 : 36]     | tak                | tak     | tak               | Idea Plus     | kontakty<br>notatnik                                    | podwójna (1,00)                                                             |
| G      | 160           | 2         | [12 : 12]     | tak                | nie     | nie               | NA            | kontakty                                                | trzystopniowa (1,00)                                                        |

Źródło: opracowanie własne w programie *Sodas version 2.5 for Windows*.

Przykładowo, obiekt symboliczny *D*, zapisany w formie koniunkcji cech: [opłata = 66,0]  $\wedge$  [pojemność = 2000]  $\wedge$  [okres dostępu  $\subseteq$  {1–12}]  $\wedge$  [obsługa z telefonu komórkowego  $\in$  {nie}]  $\wedge$  [dodawanie reklam  $\in$  {nie}]  $\wedge$  [powiadamianie sms  $\in$  {tak- Era, Idea, Plus}]  $\wedge$  [organizer  $\in$  { kontakty, zadania, notatnik, terminarz, aktówka}]  $\wedge$  [ochrona antywirusowa  $\in$  {podstawowa (0,4), trzystopniowa (0,6)}], charakteryzuje zbiór płatnych kont internetowych (66 zł za 12 miesięcy), o pojemności 2 GB, z brakiem możliwości obsługi konta z poziomu telefonu komórkowego. Serwer nie dodaje reklam do wysyłanych wiadomości. Istnieje możliwość powiadamiania sms o otrzymanej wiadomości do sieci Era, Idea i Plus. Organizer zawiera kontakty, zadania, notatnik, terminarz i aktówkę. 40% obiektów z tego zbioru ma podstawową wersję ochrony antywirusowej, a 60% – skaner trzystopniowy.

### 3.2. Prezentacja obiektów symbolicznych w języku SOL

Obiekty symboliczne można również przedstawiać za pomocą tzw. SOL (*symbolic objects language*). Opis obiektu rozpoczyna się od scharakteryzowania typu i wariantów poszczególnych zmiennych (zob. tab. 6). Następnie wyodrębniane są cechy poszczególnych obiektów symbolicznych (zob. tab. 7).

Tabela 6. Wybrane zmienne w języku SOL

```
variable okres_dostepu_(liczba_miesiecy)
  real [1:36]
  interval;

variable powiadamianie_sms
  nominale {"nie", "tak"}
  multiple,mode=boolean;

variable idea_era_plus
  nominale {"Era", "Idea", "NA", "Plus"}
  multiple,mode=boolean;

variable organizer
  nominale {"aktówka", "kontakty", "notatnik", "terminarz", "zadania"}
  multiple,mode=boolean;

variable ochrona_antywirusowa
  nominale {"brak", "podstawowa", "podwójna", "trzystopniowa"}
  multiple,mode=probabilist;

idea_era_plus is applicable if powiadamianie sms o otrzymanej wiadomości
in {"tak"}
```

Źródło: opracowanie własne w programie *Sodas version 2.5 for Windows*.

Tabela 7. Obiekt *D* w języku SOL

```
D =
  ROCZNA_OPLATA_ ( = [ 66.00 : 66.00 ]
AND  POJEMNOSC_(MB) = [ 2000.00 : 2000.00 ]
AND  OKRES_DOSTEPU_ ( = [ 1.00 : 12.00 ]
AND  OBSLUGA_Z_TELEF = NIE (1.00)
AND  REKLAMY = NIE (1.00)
AND  POWIADAMIANIE_S = TAK (1.00)
AND  IDEA_ERA_PLUS = ERA (0.40), IDEA (0.40), PLUS (0.20)
AND  ORGANIZER = KONTA (0.20), ZADAN (0.20), NOTAT (0.20), TERMI (0.20),
AKTÓW (0.20)
AND  OCHRONA_ANTYWIR = PODST (0.40), TRZYS (0.60)
```

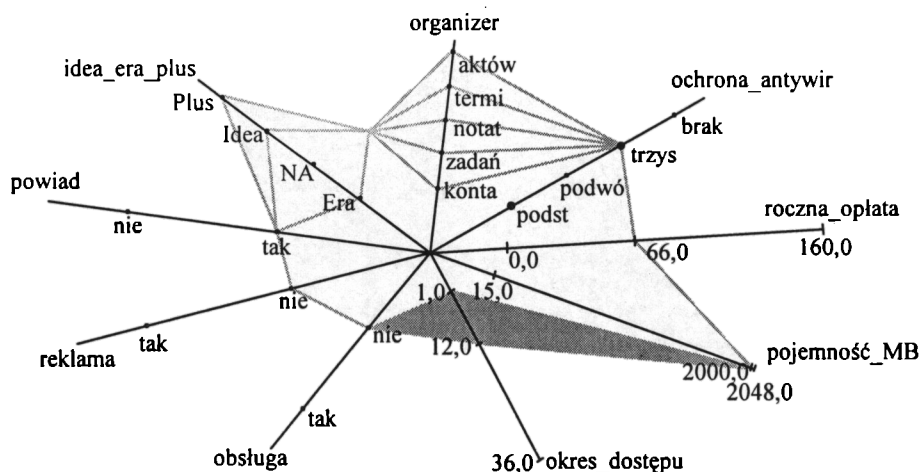
Źródło: opracowanie własne w programie *Sodas version 2.5 for Windows*.

### 3.3. Metody wizualizacji obiektów symbolicznych

Graficzną formą prezentacji obiektów symbolicznych są wykresy 2D i 3D (*zoom star*), zróżnicowane ze względu na poziom szczegółowości. Ich zaletą jest możliwość prezentacji na jednym wykresie wszystkich analizowanych zmiennych bez względu na typ.

Każdy obiekt symboliczny jest przedstawiony na osobnym wykresie. Osie wykresów reprezentują zmienne i ich warianty<sup>2</sup>. Natomiast krzywa łącząca poszczególne warianty zmiennych zawiera charakterystykę obiektu.

Na rys. 3 przedstawiono obiekt symboliczny *D* za pomocą wykresu 2D. Linie łączące osie oznaczają wartości zmiennych dla obiektu symbolicznego *D*. Największe punkty oznaczają najwyższe wagi, a wypełniony obszar – przedział klasowy.

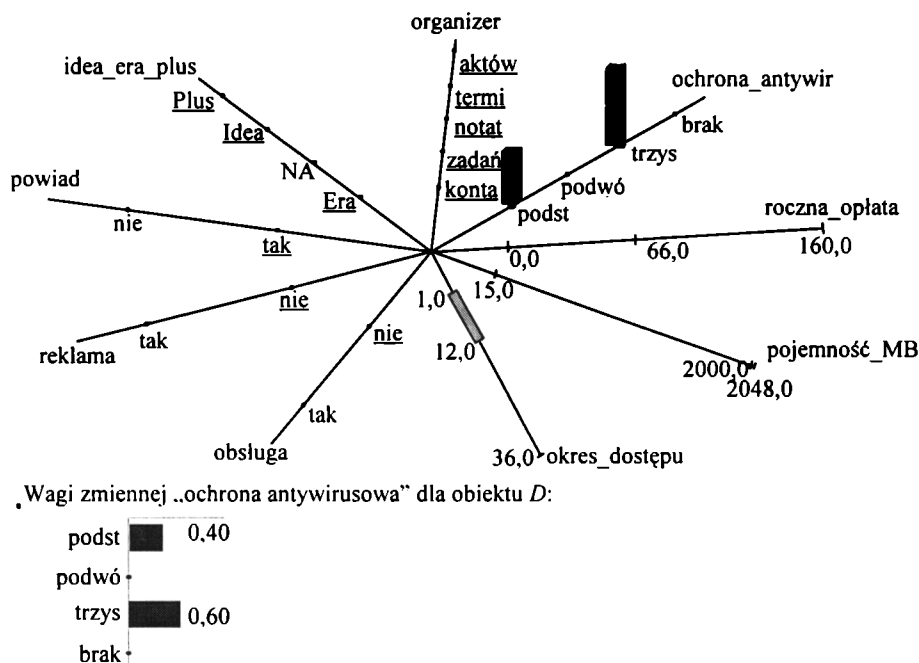


Rys. 3. Wykres 2D prezentujący wartości zmiennych dla obiektu *D*

Źródło: opracowanie własne w programie *Sodas version 2.5 for Windows*.

Na rys. 4 przedstawiono obiekt symboliczny *D* za pomocą wykresu 3D. Słupki oznaczają wagi poszczególnych wariantów zmiennej dla danego obiektu („ochrona antywirusowa”), a zaznaczony obszar na osi zmiennej „okres dostępu do konta” – przedział klasowy. Podkreślone warianty na osi zmiennej listy kategorii („organizer”), zmiennych nominalnych binarnych („obsługa z telefonu komórkowego” i „dodawanie reklam”), a także zmiennej strukturalnej („powiadamianie sms”) oznaczają warianty charakteryzujące obiekt *D*.

<sup>2</sup> Pojedynczy graf może zawierać maksymalnie 24 zmiennych (osi) i 15 wariantów zmiennej na jednej osi.



Rys. 4. Wykres 3D charakteryzujący obiekt D oraz rozkład wag zmiennej „ochrona antywirusowa”

Źródło: opracowanie własne w programie *Sodas version 2.5 for Windows*.

Wykresy 2D i 3D, oprócz uproszczonej formy prezentacji złożonych danych, umożliwiają porównanie obiektów oraz analizę zmiany wartości zmiennych dla poszczególnych obiektów w czasie [4, s. 25-29].

#### 4. Podsumowanie

Obiekty symboliczne są nową jednostką statystyczną. Są sposobem reprezentowania skomplikowanych danych, które w przeciwieństwie do danych „klasycznych” nie muszą spełniać postulatu atomowości. Zmienne mogą przyjmować więcej niż jeden wariant dla pojedynczego obiektu, postać przedziału liczbowego lub być powiązane relacjami. Mogą być opisane zmiennymi różnego typu bez konieczności ich przekodowania czy transformacji (zob. także [12, s. 82]).

Konstrukcja obiektów symbolicznych II rzędu umożliwia analizę dużych zbiorów obiektów. Nawet gdy wszystkie obiekty są opisane za pomocą klasycznych zmiennych, ich analiza ze względu na zbyt duży zbiór danych może nie być możliwa. Pogrupowanie danych znacznie ułatwia analizę, również metodami klasycznej, statystycznej analizy wielowymiarowej.

Obiekty symboliczne umożliwiają również wygodną prezentację i zapis danych o nietypowej strukturze na trzy sposoby, tj. za pomocą tablicy danych symbolicznych, wykresów 2D i 3D oraz tzw. SOL [3, s. 21-22; 9, s. 1].

## Literatura

- [1] Billard L., Diday E., *From the statistics of data to the statistics of knowledge: symbolic data analysis*, „Journal of the American Statistical Association” 2003 Vol. 98, No 462, 456-469.
- [2] Billard L., Diday E., *Symbolic data analysis: definitions and examples*, Centre de Recherche en Mathématiques de la Décision (CEREMADE), [http://www.stat.uga.edu/faculty/LYNNE/tr\\_symbolic.pdf](http://www.stat.uga.edu/faculty/LYNNE/tr_symbolic.pdf), 2002.
- [3] Bock H.-H., Diday E. (red.), *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin 2000.
- [4] Calvo-Garrido P., *Applications of symbolic objects in official statistics*, Eustat, 2000.
- [5] Diday E., *An introduction to symbolic data analysis and the Sodas software*, „Electronic Journal of Symbolic Data Analysis” 2002 (July),
- [6] Dudek A., *Obiekty symboliczne w klasyfikacji danych*, Prace Naukowe Akademii Ekonomicznej nr 1022, Wrocław 2004.
- [7] Gatnar E., *Symboliczne metody klasyfikacji danych*, PWN, Warszawa 1998.
- [8] Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
- [9] Gowda K.C., *Symbolic objects and symbolic classification*, [w:] Brito P., Noirhomme-Fraiture M. (red.), *Symbolic and Spatial Data Analysis: Mining Complex Data Structures*, „ECML/ PKD 2004 Workshop Programme”, Pisa, September 20<sup>th</sup>, 2004, 1–18.
- [10] Michalski R.S., *Revealing conceptual structure in data by inductive inference*, [w:] Hayes-Michic D., Michie D., Pao Y.H. (red.), *Machine Intelligence 10*, Halsted Press, New York 1981.
- [11] Michalski R.S., Stepp E., Diday E., *A recent advanced in data analysis: clustering objects into classes characterized by conjunctive concepts*, „Progress in Pattern Recognition” 1981, Vol. 1.
- [12] Ruiz-Shulcloper J., Chac-Kantun M.G., Martinem-Trinidad J., *Data analysis between set of objects*, [w:] *The 8th International Conference on Systems Research Informatics and Cybernetics*, Baden-Baden 1996, s. 81-85.
- [13] Verde R., *Clustering methods in symbolic data analysis*, [w:] Banks D., House L., McMorris F.R., Arabie P., Gaul W. (red.), *Classification, Clustering, and Data Mining Applications*, Springer-Verlag, Berlin 2004, s. 380-317.
- [14] Wilk J., Pelka M., *Dane symboliczne w zagadnieniu klasyfikacji*, [w:] Rószkiewicz M. (red.), *Identyfikacja struktur rynkowych. Pomiar – modelowanie – symulacja*, VIII Warsztaty Metodologiczne, SGH, Warszawa 2004.

## VISUALIZATION OF SYMBOLIC OBJECTS

### Summary

The are major concepts of symbolic data analysis explained in the article. Moreover, the visualization of symbolic objects was illustrated (Symbolic Data Table, Zoom Star and Symbolic Objects Language) on the example of Internet e-mail accounts.