

2.1. An array of symbolic data, symbolic object, a symbolic variable

Real-world objects are mapped to an array of symbolic approach of symbolic data. It can be defined, as in the case of the data array classical data, as follows:

(2.1)

where: - implementation of the k-th symbolic variable in the i-th object symbolic

- Symbolic object number,

- Variable number of symbolic.

In contrast to classical data matrix of symbolic variables individual projects do not have to take a single numeric value. Symbolic data table may take the form similar to that of the tab.

2.1.

Table 2.1. Fragment of symbolic data table describing the offer investment funds

Fund

1 [12, 31] {www} {Bonds smartphone (30%), shares (70%)}

2 [-3, 17] {web, sms, smartphone} {bonds (30%), gold (12%), shares (58%)}

3 [1.1, 3.4] {sms, call_center} {Bonds (100%)}

... ..

- Annual interest rate achieved in the previous years (symbolic variable with realizations in the form of numerical ranges inseparable), - access channels (symbolic variable with realizations in the form of a list of categories), - the structure of the portfolio (variable symbolic of realizations in the form of a list of categories with weights).

Source: own.

One row in the array corresponds to an object symbolic symbolic and symbolic one column variable. A set of symbolic objects so you can ZAPI-sac as.

However, the variable is a symbolic representation:

(2.2)

Where: Ok - a collection of symbolic execution variable,

- Variable number of symbolic.

- A set of symbolic objects

2.2. Types of symbolic variables

Symbolic variable is a characteristic that describes the collective symbolic objects. The methods of the symbolic approach is defined seven types of sym-parabolic variables, which complement each other to provide a more complete presentation of mo-delowanej in the analysis of reality than traditional data. These are:

♣ Variables of realizations in the form of a single numerical value. These variables are often referred to as classical variables (numerical). Variables are measured at the measuring scale of strong, but in contrast to the classical methods of analysis of data almost no distinction between the variables measured on a scale and with a separate dressing ILOR-variables measured on the interval scale. In this approach, a symbolic variable is treated as a variable interval symbolic of the beginning and end of the range at the same point.

♣ Variables of realizations in the form of text strings. These variables are called and text variables correspond to the variables measured on the scale of the classic desired log-Central and nominal. As in the case of classical variables, where the variables rarely

distinguished text variables measured on an ordinal scale of measured variables at a nominal scale. In many methods, the text approaches the symbolic variables are treated as a list of categories and the length is always equal to one.

Variables ♣ embodiments in the form of numerical ranges (eg income <2000, 3000>) separable or inseparable. Variables of this type are called symbolic variables interval (adjective "symbolic" is used here for the on-distinction between numerical variables measured on the interval scale).

♣ realizations symbolic variables in a list of categories (eg equipment-nie_dodatkowe {abs, gps, felgi_z_metali_lekkich}). Variables of this type are in-ferred or wielokategorialnymi wielowariantowymi variables.

♣ realizations symbolic variables as a list of values with weights {eg defects {20% poor access, lack of parking 30%, 50% of the lease price}}. These are called multi-riantowych variables with weights or with weights wielokategorialnych variables.

Variables ♣ embodiments, a list of disjoint intervals of the weights (for example, The level of savings {70% <0, 1000> 20% <1000, 10000> 10% <10 000, 100 000>}). These variables are called variable histogram.

♣ boxed variables (Arrojo et al. [2006]), the variables that aggregate representations of the original list of numbers and statistics are described cumulative five-to me, the lowest value (excluding outliers), the first quartile, second-gim quartile (median), the third quartile and the highest value (excluding outliers). Boxed graphical representation of variables shown in Figure 2.1.

Figure 2.1. Projects boxed variables

Source: own study based on Walesiak, Dudek [2009], p 91st

The sixth and seventh type of the variable (variable histograms and box-variable) is as yet very little practical use in symbolic data analysis methods, so there will be further discussed in this paper.