MULTIDIMENSIONAL DATA ANALYSIS WITH NOVOSPARK VISUALIZER

Nikitina I.E., Marukhina O.V. Tomsk Polytechnic University Lenina Avenue, 30, 634050, Tomsk, Russia E-mail: nikitina ira@docsis.ru

Introduction

Objects around us are difficult by their nature. Such objects often have a set of attributes that allows to characterize them and also surrounding them space most fully. However it becomes hard to get information at excessive multidimensionality of data. Multidimensional data contain information about two or more attributes for each studied object. In addition to information which can be taken from one-dimensional sets, multidimensional data can be used to detect whether there is a simple dependence between these variables as far as they are interconnected. To reveal dependence between variables it is necessary to carry out the analysis.

The purpose of this work is identification of expressiveness degree of objects' personal qualities.

There are the following tasks for achieving the main goal:

• Examine procedure of recognition individual observations.

• Indication data anomalies.

• Research of neurodynamic and psychodynamic factors.

• Assess statistical characteristics of research.

Visualization and Analysis

Data visualization is a task which any researcher faces in the work. Problem of visual form representation of experiment data or theoretical research results is reduced to a problem of data visualization. Traditional tools in this area – schedules and charts – badly cope with a problem of visualization when there is a need to represent more than three interconnected variables. The visual analysis can play an important role in any scientific experiment, especially where the fast assessment of the main data set aspects without using the methods of quantitative data analysis is necessary.

NovoSpark Visualizer is an advanced visualization tool that combines the intuitiveness of visual analysis with the descriptive power of statistical methods, delivering the most comprehensive information about multidimensional data [1]. The product allows making easy manipulation of loaded datasets and recalculates the underlying model and its graphical representation.

It is possible to see review statistical information and perform analysis while manipulating data. The results can be automatically recalculated and viewed in a separate window allowing to see how changes in data affect the resulting model. Conversely, we can invoke a number of functions that will automatically change the data based on our current model. By using combinations of powerful statistical methods like discriminant analysis, factor analysis and others we can quickly configure the optimal structure and contents of datasets and use the results in our research.

The product comes with a set of helper windows that show the most relevant information about image and refer to the application features that can be used to improve its look and feel.

For participation in research 103 persons were interviewed. Data are obtained by means of the formal dynamic properties of a personality Questionnaire of V. M. Rusalov[2], which relies on achievements of modern functional and system approach in neuro- and psychophysiology. The questionnaire is used for diagnostics "subject-activity" properties (psychomotor and intellectual spheres) and "communicative" aspects of temperament. The objects were described by the following basic properties of the nervous system: factors of Motor Activity (Motor Ergonicity, Motor Plasticity and Motor Tempo), Social Activity (Social Ergonicity, Social Plasticity and Social Tempo) and Intellectual Activity (Intellectual Ergonicity, Intellectual Plasticity and Intellectual Tempo). These personality properties can have different degrees of expressiveness. It very strongly influences behavior and the relations. For example, low sensitivity is indifference, and high is shown in continuous exaggeration of the events importance.

We have loaded one dataset and our 103 observations were shown on the image.



Pic. 1. Comparison of the original data set

Each curve on the picture represented one observation. The more different they look on the image the more different their observations are in the dataset. It is possible to select a record in the data grid to see the corresponding curve highlighted on the image or click on a curve to get its record selected in the data grid.

Data anomalies are determined as by a form of their images (visually significant distinctions in images indicate differences in the original data), and through "cloud" of a multidimensional interval. If the image of observation completely fits into an interval cloud, parameters values corresponding to this image are between the minimum and maximum borders of the set interval in all dimensions.

We needed to adjust the picture by using the tools in NovoSpark to find out which observation deviates norm. For example, selected view is Front View (2D), curve filter is strong and then curve color (palette) is chosen as color of a rainbow. The selected palette has 9 colors. The colors red and blue may identify data anomalies (maximum and minimum). So we can see (pic. 2) that 63 observations are marked on the Z-axis, indicating that their values are outside of the selected interval (confidence interval at 1.50 sigma).



Pic. 2. Dataset with anomalies

Let's pass to concrete observation. We have selected a marker to highlight an observation on the image and see the corresponding record in the data grid (pic. 3).



Pic. 3. One object with anomalies

Observation chosen by us leaving out an interval had indicators of a Motor Ergonicity, Motor Plasticityand Motor Tempo. Motor Ergonicity means degree of tendency to physical activity, Motor Plasticity is degree of motor programs variety and flexibility at their choice, Motor Tempo is rate of psychomotor behavior [2]. From that we draw a conclusion about deviations of observed object in the psychomotor sphere, decrease in vitality and activity.

Also NovoSpark has the tab "Descriptive Statistics", where elements of descriptive statistics are presented. Let is analyze the most important.

Standard Deviation shows how much variation or dispersion exists from the average (mean) [3]. The greatest variation at participants is observed in a Motor Ergonicity that speaks about essential distinguishing in physical activity of the observed. Objects showed the smallest dispersion in Social Tempo, and we can consider that observed tempo of speech is about the same.

Variance indicators confirm these conclusions. The greatest spread out of numbers is also visible at Motor Ergonicity and the smallest at Social Tempo, findings are the same.

Conclusion

The subject of this research was identification of expressiveness degree of objects' personal qualities. Having examined neurodynamic and psychodynamic factors, note that the majority of objects have normal indicators. However 63 observations have data anomalies. Based on statistical characteristics, we may conclude that it is connected with a difference in rates of objects physical activity, social activity is normal.

Using the method of multidimensional data visualization is possible to carry out the procedure of recognition, an indication of data anomalies, to determine their statistical characteristics. The visualization tool «NovoSpark Visualizer» enables qualitative analysis of multidimensional data on a graphical image.

References

1. NovoSpark Visualizer. [Electronic resource] – Retrieved April 24, 2013. Access mode: http://novospark.com.

2. Rusalov V., Trofimova I. Structure of Temperament and its measurement. Toronto, Canada: Psychological Services Press, 2007.

3. Standard deviation. Wikipedia. [Electronic resource] - Retrieved April 24, 2013. Access mode: http://en.wikipedia.org/wiki/Standard_deviation.

4. Sterlau J. "Individual differences in temperament: An international perspective". [Electronic resource] - Retrieved April 24, 2013. Access mode: http://e-

book.lib.sjtu.edu.cn/iupsys/Proc/mont1/mpv1ch02.ht ml.