Analysis of some approaches and methods for problem solving in Data Mining technologies.
Mahmudzadeh Zarifa

Abstract
A brief overview of Data Mining methods and algorithms is carried out, their advantages and disadvantages are analyzed. In the article , it is noted that various Data Mining methods are characterized by certain features that can be decisive when choosing data analysis methods. It is noted that in order to compare them with each other, it is important to evaluate their properties such as accuracy, scalability, interpretability, testability, productivity, flexibility, speed and popularity.
Keywords: Data Mining , Cross tabulation , neural networks, etc.
The rapid development of information technology, especially the advancement in data collection, storage and processing methods, has enabled many organizations to collect large amounts of data that need to be analyzed. The volume of these data is so large that the ability of experts (specialists) to analyze them is no longer sufficient.
Today, the direction of intellectualization of data processing and analysis methods is intensively developing. Intelligent data analysis systems are designed to minimize the efforts of the decision maker in the process of data analysis, as well as in the construction of analysis algorithms. Many Data intelligent analysis systems allow not only to solve classic decision-making problems, but also to achieve cause-and-effect relationships and the discovery of hidden knowledge in the analyzed systems.
Intelligent analysis of data. Data Mining (obtaining useful knowledge, intellectual analysis of data, deep analysis of data) is a general name used to identify previously unknown, non-obvious, practically useful, interpretable and necessary knowledge in various fields of human activity. The term was coined by Grigory Piatetsky-Shapiro in 1989.
Data Mining methods include all types of classification, modeling and prediction methods. Data Mining methods include statistical methods (descriptive analysis, correlation and regression analysis, factor analysis, variation analysis, component analysis, discriminant analysis, time series mean analysis). One of the most important features of Data Mining methods is the visual presentation of the results of calculations, which allows people without special mathematical training to use Data Mining tools. Knowledge obtained by data mining methods is usually presented in the form of models .
Data Mining knowledge representation models. 1. Associative rules 2. Decision tree 3. Clusters 4. Mathematical functions
Data Mining methods and algorithms include:
• artificial neural networks
• decision trees, symbolic rules
• nearest neighbor and k-nearest neighbor methods
• method of support vectors
• Bayesian networks
• linear regression
• correlation - regression analysis
• hierarchical methods of cluster analysis
• non-hierarchical methods of cluster analysis, including k-means and k-median algorithms
• evolutionary programming and genetic algorithms
• limited search method
• different methods of data visualization, etc.
Most of the analytical methods used in Data Mining technology are based on well-known mathematical algorithms and methods. What is new in their application is the possibility of using them to solve certain specific problems due to the emerging capabilities of hardware and software. It should be noted that most of the Data Mining methods were developed within the framework of the theory of artificial intelligence. The

method itself is a certain way, norm or system of rules for solving theoretical, practical, mental, management problems.

Characteristics of Data Mining Methods. Different Data Mining methods are characterized by certain features that can be decisive when choosing data analysis methods. In order to compare the methods with each other, it is important to evaluate the characteristics of their properties. The main properties and characteristics of Data Mining methods are determined by the following: accuracy, scalability, interpretability, verifiability, productivity, flexibility, speed and popularity.

Classification of methods. All Data Mining methods can be divided into two large groups according to the principle of learning with primary data. The upper level in this classification is based on whether the data is stored after Data Mining or distilled for later use.

In the case of direct data use or data storage, the raw data is stored in a transparently detailed form and used directly in the prognostic modeling and/or exception analysis stages. The problem with this group of methods may be that difficulties may arise when using them when analyzing very large databases.

Cluster analysis, Nearest neighbor method, K -nearest neighbor method. With the identification and use of formalized regularities or template distillation technology, an example of the data (template) is extracted from the original data and transformed into certain formal structures depending on the type of Data Mining method used. This process is carried out in the stage of free search, this stage is generally absent in this group of methods. The results of the free search phase are used in the prediction and exception analysis phases.

Logical methods , Visualization methods, Cross tabulation methods, Equation based methods.

• Logical methods or methods of logical induction include: fuzzy queries and analyses; symbolic rules; decision trees; genetic algorithms . The methods of this group are perhaps the most interpretable - they convey the regularities found, in most cases, in a fairly transparent form from the user's point of view. The generated rules can include continuous and discrete variables. It should be noted that decision trees can easily be converted into a set of symbolic rules by creating a single rule along the curve from the root of the tree to its terminal node. Decision trees and rules are actually different ways of solving the same problem and differ only in their capabilities. In addition, the implementation of rules is performed by slower algorithms than the induction of decision trees.

• Cross tabulation methods: agents, Bayesian (trust) networks, crosstab visualization. The latter method does not fully respond to one of the characteristics of Data Mining - independent search for regularities by the analytical system. However, the presentation of data in the form of cross-tabulations ensures the implementation of the main task of Data Mining - the search for regularities, so this method can also be considered one of the Data Mining methods.

Conclusion

Equation-based methods. The methods of this group express the determined regularities in the form of mathematical expressions - equations. Therefore, they can only work with numeric variables, and other types of variables must be coded accordingly. This somewhat limits the use of methods in this group, but they are widely used in solving various problems, especially forecasting problems.

References

[1] Dyakonov A. G. Some problems of discrete mathematics arising in modern applications during data analysis // Spectral and Evolution Problems, 2012. т. 22. С. 66–75.

[2] Heckerman D. Bayesian Networks for Data Mining Data Mining and Knowledge Discovery. 1997. No. 1. Pp. 79–119.

[3] Chubukova I. Data Mining [Електронный ресурс] // НОУ ИНТУИТ [NOU INTUIT], URL https://loginom.ru/blog/decision-tree-p1 .www . kdnuggets . Com