

ABSTRACT

Thesis: 115 p., 10 tables, 33 fig., 3 add. 25 references.

HEURISTIC METHOD, DATA MINING, PROBABILISTIC OUTPUT, CREDITWORTHINESS, MACHINE LEARNING, BAYESIAN NETWORK, TRAINING DATA.

Relevance of the topic - technology of data mining appeared in 1978. Today, data is becoming widespread. A significant amount of such data is constantly coming from different sources and remains unprocessed. These can be medical devices, economic indicators, images, etc. The processing of such information is essential, since an uninterpreted information does not have anything useful. Therefore, there is currently a need for tools that allow to analyze such data in order to obtain useful information that is commercially available. So, in previous years, significant amounts of data on the experience of individuals crediting in Ukraine have been accumulated. Bayesian Network is an effective tool for working with them. The result, presented as a graph, gives an opportunity to understand the process itself, which resulted in the collected data.

The purpose of the study is to create a software product for building a Bayesian discrete network structure, based on learning data, using a heuristic method.

Achieving the goal requires solving the following tasks:

1. An overview of the theoretical foundations of the Bayes networks
2. Analysis of the principles of constructing MB for learning data.
3. Analysis of methods for constructing a probabilistic conclusion in MB.
4. Implementation of the heuristic algorithm for constructing MB in the form of a software product.
5. Testing the software product, using the training data, which is a sample of personal data of individuals in their credit history.

6. Analysis of the effectiveness of created models, comparing the structures obtained with the results of academic programs available for the construction of MB with learning data (GeNIe and BayesiaLab).

Objective of the study is the experimental and statistical data on the credit history of individuals who need effective analytical processing in order to identify an unknown, practically necessary knowledge and interrelations between variables, on the basis of which decisions will be made on the creditworthiness of individuals.

The subject of the research are the methods of creating Bayesian probabilistic networks on the basis of the initial data and the formation of a probabilistic conclusion in them.

Research methods are based on concepts of probability theory and mathematical statistics, graph theory and the theory of probabilistic conclusion formation in MB.

Scientific updates of the received results: a heuristic method of building a Bayesian network based on learning data is implemented. A comparative analysis of the performance and computing costs of a software product with other analogs implementing the work of the networks of Bayes.