

# Chapter 130

## Data Mining Techniques Applied in the Financial Industry

Liang Huo, Tao Wang and Liu Yang

**Abstract** Data mining is the use of various techniques to discover knowledge from massive data, it has wide application and prospect. The article detailed introduction summarizes the concept of data mining, methods, and applications. And analysis, summarizes the application of data mining in the financial field, including trend forecasting, customer relationship management, financial crime detection, risk identification and risk management.

**Keywords** Data mining · Financial data · Prediction · Risk identification

### 130.1 Introduction

Bank, securities company, insurance company daily business will generate large quantities of data, using the database system can efficiently realize the data input, query, statistics and other functions, but can't find data in the presence of the relations and rules (Wang and Cao 2002), not according to the existing data to predict the future trend of development.

How can I not be overwhelmed by information from a vast expanse of water, discover useful knowledge, improve utilization of information? Therefore, we must find a effective method, automatic data analysis and processing, we quickly find valuable information. Then, the data mining technology will emerge as the times require, and is able to develop flourishingly, more and more show its

---

L. Huo · T. Wang

Department of Information Management and Engineering, Hebei Finance University,  
Baoding, Hebei, China

T. Wang (✉) · L. Yang

Experimental Center of Economics and Management,  
Hebei University, Baoding, Hebei, China  
e-mail: Wangtao731018@126.com

powerful vitality (Wang 2009). It can be a lot of data from the extraction of potentially useful information and models, to help us to make scientific decision.

At the same time, the operation of financial institutions must be financial risk. The risk management is one of the important work of financial institutions. Using data mining techniques can not only from this data was found hidden in the subsequent rule, and can well reduce financial risks (Yang 2002). Study and application of data mining technology of China's financial institutions have an important significance.

## **130.2 Methodology**

### ***130.2.1 Basic Concept***

1. Data Mining. Data mining is from a large, incomplete, noisy, fuzzy and stochastic data, in which the extraction of implicit, unknown, but potentially useful information and knowledge process. Discovery of knowledge can be used for information management, query optimization, decision support, process control and so on, can also be used for data maintenance (Yang 2002).
2. Data Warehouse. Data warehouse is wrote by W. H. Inmon in his book "Building the Data Warehouse" presented in the 1993. W. H. Inmon to the data warehouse is definition as: "a data warehouse is a Subject-Oriented, Integrated, Nonvolatile, Time-Variant data set, in order to support management decision-making process (Amihud and Mendelson 2005)." The data warehouse can be in a data model, under the guidance of the collection system, internal and external data, to ensure data consistency, accuracy and timeliness, realize data sharing and analysis using.

### ***130.2.2 Main Technical of Data Mining***

In application to a variety of financial areas in data mining technology, basically have the following kinds of methods.

1. Based on neural network method.

The neural network method is to simulate the human information processing process of a kind of intelligent information technology. Artificial neural network is the computer through a multiple input similar parallel processing structure to simulate human pattern recognition function to build up (Boyd and Hu 2005).

Neural network consists of a series of basic processing unit (also known as neurons), the processing unit according to the layer distribution. Most neural network consists of three types: the input, hidden and output layers. The mining

process is basically the data clustering and classification, calculation of weight (Brank et al. 2005). Neural network is used to store the captured knowledge and for future use, it is used to describe the association between neuron models, capture in a given collection of data between independent variable and dependent variable causal relationship.

Neural network is suitable for non-linear data and noise data, so the database analysis and modeling is widely used. Neural network provides a satisfactory accuracy, close the modes of financial tools.

## 2. Based on the Bayesian networks method.

Bayesian networks (Bayesian network), also called the probabilistic causal network, a web of trust, knowledge map, is a directed acyclic graph. Bayesian network graph to represent the variables connected probability. Nodes represent: field variables; directed edges between nodes: dependencies; for each node corresponds to a conditional probability distribution table, the table indicates the distribution of the variables and dependencies between the parent node (Zurell 2002).

## 3. Genetic algorithm.

The basic principle of genetic algorithm is given a question, in a unique population genetic groups potentially containing a solution or a better solution. Based on genetic and evolutionary principles, genetic algorithm repeatedly revised population artificial structures, an operator through a preset, select, crossover and change to gradually form a solution.

## 4. Based on the rules and decision tree tools.

Most data mining tools use rule discovery and decision tree classification techniques to find data patterns and rules its core is an inductive algorithm. Tools of this kind are usually on the database data mining production rule and decision tree, and then to new data analysis and prediction. Tools of this kind are the main advantages of rules and decision tree are readable (Han 2005).

The classification tree method as a kind of data mining based on statistic theory non parameter identification technique, not only to maintain multiple parameters, nonparametric statistics, some of the advantages, but also overcomes the shortcomings, mainly displays in: automatic variable selection of reducing dimensions, make full use of prior information processing data between the non homogeneous, and can effectively for the classification of the data.

## 5. Based on the fuzzy logic tool.

The method is the application of fuzzy logic for data query, sort. The tool uses fuzzy concept and “recently” search techniques of data query tool, which allows users to specify the target, then the database search, find out close to the goal of all records, and evaluate results.

## 6. Rough set method.

Rough set method is proposed by Z. Pawlak in 1982 Poland mathematician. It is used, lower approximation set to deal with uncertain problem. It does not require prior knowledge of data obtained, using only the data itself provides information, fully contained in the data mining. The specific application of the rough set method is usually used in conjunction with other methods, such as the use of rough sets to obtain the initial rule set and then construct the corresponding neural network model (Han et al. 2004).

### *130.2.3 The Main Steps of Data Mining*

1. Problem definition. The goal of a clear, clear definition, is also identified the need to solve the problem, the goal should be feasible, capable of operating and evaluation.
2. Data collection. A full rich data mining is premise, no data, data mining will not be. Therefore, data collection is the first step in data mining. Data can be derived from the existing transaction processing system, can also be obtained from the data warehouse (Basel Committee 2000).
3. Data processing. Data collection is the necessary link of data mining. By the data collection phase of data may have a “pollution”, manifested in the data may have its own inconsistency, or the existence of missing data, so data collection is a must. At the same time, through the data collection, data can be a simple generalization process, thus the data in the original basis to obtain more abundant data information, and then for the next step of data mining is carried out smoothly (Han and Kamber 2001, 2006).
4. Data mining. The use of artificial intelligence, mathematical statistics and other data mining methods to analyze the data, to discover useful knowledge and mode. The core of the whole process steps.
5. Data mining result assessment. Data mining of some of the results are meaningful, but some are not practical, or with actual condition is violated, it will need to be evaluated. Assessment can be based on user experience, but also can be directly used actual data to verify the correctness of the model, and then adjusting the mining model, repeated data mining (Berson et al. 2000).
6. Analysis and decision. Data mining is the ultimate goal of aided decision. Decision makers can according to the result of data mining, combined with the actual situation, adjust the competitive strategy.

In conclusion, the data mining process requires a lot of problems, adjustment, a reassessment of modified model, test cycle is repeated, will it be possible to achieve the desired effect.

### ***130.2.4 The Data Mining Application in Financial Industry***

Data mining in the financial field, including: financial market analysis and prediction, classification of accounts, bank guarantee and credit evaluation. These financial business needs to collect and process large amounts of data, it is difficult to manually or using the one or two small software for analysis and prediction. Data mining can be based on the existing data, find the data characteristics of the object and the relations between objects, and can be observed in the financial market trends. And then learning to use the model of rational analysis and prediction, and then find a customer, consumption group or organizational financial and commercial interests.

1. In customer relationship management. Data mining can be customer behavior analysis to discover customer behavior. Including the overall behavior and group behavior mode, market department can reason these rules to formulate appropriate marketing strategies and tactics; also can use this information to identify the customer concerns and consumption trend. In order to improve the market share of products and the competitive ability of the enterprise. Data mining can help the enterprise to find out that has important sense to the enterprise customers, including enterprises can bring huge profits to the gold customers and enterprise to further development of the potential customer (Zhang et al. 2005).
2. Risk identification and risk management. To build a classification model, bank loans to the safety or risk classification also the use of data mining technology in the control of credit risk. Credit risk management include: risk identification, risk measurement, risk management tool, effect evaluation. All process involves information includes the relevant commodity market situation, industry, industry development, enterprise management, personnel, financial condition and a series of content, information caused by manual evaluation, complex management difficulty to increase greatly. While the existing bank credit system is generally a business operation system, not for the decision analysis and the establishment of its application, data integration, integrity, accessibility, analysis are difficult to meet the demand of credit risk analysis. For this reason, it can establish a set of independent business system data warehouse, specifically addressing the credit analysis and risk and management issues (Xiong et al. 2004).

We can use a Bayesian network to establish the bank operational risk management system framework. Bayesian networks can be used to help identify risk factors, calculation of sensitivity and volatility, simplify the loss distribution based on scene and excess of loss event generation. The operational risk based on Bayesian network model first need to establish business model, then according to the business model of Bayesian network structure, using historical data or analog data to train the model, get the conditional probability distribution, using the model of production loss distribution or operational risk metric.

3. Market trend prediction. Data mining technology can forecast the trend of data, such as financial market price trend forecast, customer demand changing trends.

Identification of 4 financial fraud, money laundering crime of economy. Financial crime is one of the toughest issues, including malicious overdraft, Pirates of the card, credit card fraud, theft of the account password and money laundering, a serious threat to people's property and the normal functioning of the financial system. To detect money laundering and other financial crimes, it is important to multiple database information integration, and then using a variety of data mining tool to find the abnormal pattern. Found within a short time, a small number of staff between the huge cash flow, the suspicious trail (Kim 2003).

### 130.3 Conclusion

Data mining in the financial field has wide application, at present in China's financial industry has just started, need to be further research, development, promotion. We should focus on the study of data mining technology in the market, price trend, fraud detection, transaction pattern recognition, customer relationship management and other applications. With data mining software maturity and financial industry to accelerate the pace of information technology, the data mining technology in the financial field in our country will be able to play its great power, in order to improve the banking, securities and other enterprise management level and the competition strength contribution.

**Acknowledgments** This paper is supported by Department of science and technology of Hebei Province, China. (104572117)

### References

- Amihud Y, Mendelson H (1988) Liquidity and asset prices: financial management implications. *J Financial Management* 17(1):5–15
- Basel Committee (2000) Range of practice in bank's internal rating systems. *Data mining and knowledge discovery* 8(1):53–87
- Berson A, Smith S, Thearling K (2000) Building data mining application for CRM. McGraw-Hill Companies Inc
- Boyd JH, Hu J (2005) The stock market's reaction to unemployment news: why bad news is usually good for stocks. *J Finan* 60(2):649–672
- Brank J, Grobelnik M et al (2005) A survey of ontology evaluation techniques. Conference on data mining and data warehouses (SIKDD 2005)
- Han J-W, Kamber M (2001) *Data mining concepts and techniques*. China Mach Press 8
- Han J-W, Kamber M (2006) *Data mining: concepts and techniques*. Morgan Kaufmann, Leipzig, Germany
- Han J, Pei J, Yin Y, Mao R (2004) Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data mining and knowledge discovery* 8(1):53–87

- Kim K (2003) Financial time series forecasting using support vector machines. *Neurocomputing* 55:307–319
- Wang J (2009) Financial fraud detection platform design and anti money laundering system in a project application. China master dissertations full-text database
- Wang Y, Cao C (2002) KDD method in financial fraud detection application. *Comput Eng Des* 5:43–45
- Xiong T, Zhang D, Deng A (2004) Financial prediction using fuzzy logic model. In: *Proceedings of ISCIIA*
- Yang C (2002) International financial fraud case set. University of International Business and Economics press. Beijing, China
- Zhang D, Jiang Q, Li X (2005) A hybrid mining model based on neural network and kernel smoothing technique. *Lecture notes in computer science*, p 3516
- Zurell N (2002) Operational and analytical technologies for optimizing your CRM system. *DM Review* 82