

# Research on Financial System Safety Performance under the Basis of Data Mining Environment

Zhu Shanwen

Department of Finance  
Shanghai University  
Shanghai, China  
93905457@qq.com

Hu Yawei

Department of Finance  
Shanghai University  
Shanghai, China

Han Jun

Exchange Connectivity  
Morgan Stanley  
Shanghai, China  
447809822@qq.com

**Abstract**—The daily business operation of financial industry makes the abundant data collected by information system piling and expanding, which results in the existed relations and regulations in date can't be identified in time that makes many valuable information hiding inside but being ignored or omitted. At the same time, if these magnanimity data cannot be effectively managed, the risk hiding in data can't be avoided in time and the high security of finance can't be reached. Therefore, this essay analyzes and researches the application in financial industry and existing problem based on data mining technology, whose objective is highlighting the advantage effect of it in data processing of financial industry and avoiding financial risk. Financial data are collected by many organizations like banks, stock exchange authorities, taxation authorities, big accounting and auditor offices specialized data bases, etc. and in some cases are publicly available. The application of data mining techniques on financial data can contribute to the solution of classification and prediction problems and facilitate the decision making process. Typical examples of financial classification problems are corporate bankruptcy, credit risk estimation, financial distress and corporate performance prediction.

**Keywords**—data mining; data mining technology; financial system; financial data; security

## I. INTRODUCTION

The daily business of financial department will all produces abundant data, the functions of input, inquiry, statistics, etc. of data can be effectively realized by using existing data base system, but the existing relations and regulations in data cannot be discovered, and we cannot forecast the future developing trend according to existing data. We are lack of method of mining hiding knowledge behind data, which result in the phenomenon of data explosion but lack of knowledge. At the same time, financial risk must be existed in the operation of financial institution and risk management is the important work of every financial institution. Using data mining technology

can not only discover hiding laws from abundant data, but also well decrease the existing risk of financial institution. Learning and applying data mining technology has important significance to our country's financial institution.

In financial data transmission field, the country put effort in developing special network for financial data transmission and provides a unified platform for financial institution's data transmission, but this is far from satisfying the actual requirement of different financial institutions. As Chinese finance gradually stepping in network finance age and the constant change and development of financial business, now financial business requests that Chinese software industry can constantly develop software product which can bitterly satisfy business requirement.

## II. DATA MINING AND DATA MINING TECHNOLOGY

So called data mining means the process that we extract the information hiding in it, which people didn't know beforehand, but the information is potential and useful through setting certain learning algorithm from abundant not completely noise dim random data.

The basic process of data mining, first of all, need the definition of commercial questions, and then select the group to set up data mining library. Data Mining library information can be extracted from the data warehouse, if necessary also other information, from external data sources can be extracted directly. After Data Mining library set up, data analysis was required, to draw up the initial data model, including variable selection, select Record collection, transformation of variables or create new variables. After the initial model was set up, the model required a reasonable assessment, if the model and the actual system has a larger error, the model need to re-modify, and time until the model system closer to the model in the explanation, and as a decision support information to managers.

From the angle of technology, data mining means the process to extract hiding potential and useful information

and knowledge, which is not known by people beforehand from abundant, not complete, with noise, dim and random actual applicative data. It is a generalized interdisciplinary, which involves data base technology, information science, machine learning, neural network, statistics, mode identification, knowledge base system, knowledge acquisition, information inquiry, high performance computing, data visualization and other multi-disciplinary fields and it is constantly developing, as showed in Fig .1. Currently there are plenty challenging fields, such as text data mining, Web information mining, space data mining and so on.

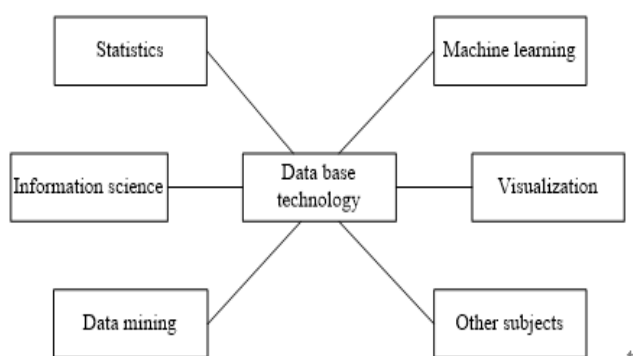


Figure 1. Data mining is an interdisciplinary subject

### III. ANALYSIS ON FINANCIAL INDUSTRY'S SECURITY AND CHARACTERISTIC OF FINANCIAL DATA

Analysis on financial industry involves contents of different aspects, which includes analysis and research on financer and financing market, such as analysis on the financial condition of company, credit status, and company's value, analysis on investor's trading activity, investing decision method and risk of investment, analysis on change of price, interest rate and exchange rate in market and so on. Uncertainty is the essence of financial market, and also the source of financial analysis. No matter financer, investor, supervisor or intermediary service institution usually use relevant quantitative analysis and method to "ensure" the uncertain factors, whose actual target is to forecast unknown variable and serving for final strategy and activity. The general process of financial analysis is establishing mathematics model on basis of various hypothesis, then using historical or current data to proceed with forecasting analysis, and getting the conclusion according to relevant financial theory at last, such as the variable forecast of price, exchange rate, interest rate, volatility, risk, earnings, etc. are all gained forecasting output by establishing mathematics model under certain hypothesis conditions, then using historical

or current data as input variable. As showed in Fig .2.

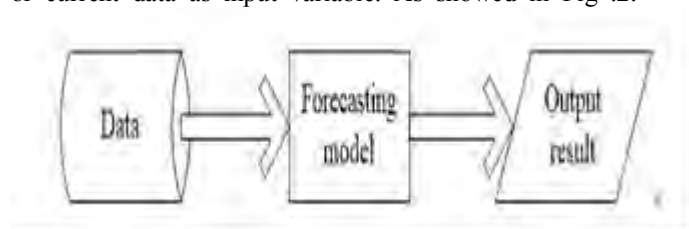


Figure 2. The general system model of financial analysis

### IV. THE APPLICATION OF DATA MINING TECHNOLOGY IN FINANCIAL INDUSTRY

#### A. Establishing the group analysis system of bank customers

Analysis on customer relation: i.e. the classification and clustering to targeted marketing's customers. For example: we can put the customers who has the same savings and loan paying back activity into the same group, and use effective clustering and collaborative filtering method to identify customer group to ensure the characteristic of customer so as to provide customers with targeted service and push targeted customer market.

Analysis on customer value: proceed with customer value analysis according to customer classification. We usually apply a series of data processing, transforming process, and other data mining technology to find key customers, and identify which are the customers the bank wants to keep and which are outflowing customers through detailed analysis to trading data, and have a targeted compensation to those customers who have similar feature before outflow.

Analysis on customer activity: we analyze their activity after ensuring key customers, and will discover the activity preference of customers and customize characteristic service. Analysis on customer activity is also divided into whole behavior analysis and group behavior analysis. Through the inter-mining analysis between different customer groups, we can discover the change rule between customer groups and automatically input the feedback of customers to market into data base to formulate relevant market strategy.

#### B. Proceed with multi-market trend forecasting mechanism

We can explain the data mining in applicative region and its using technical type in financial industry through table. The table reflects the analysis application of data mining technology in the price trend forecast of market, the change trend of customer's requirement, loan paying back forecast and credit policy of customers. At the same time, it can also be used in the control of credit risk, identifying financial cheat, and the detection of economic crimes like money laundering. The method is gathering the information of many data base and then applying data mining tool to search for abnormal mode and find the suspicious flow of huge cash between minority personnel in a short period to search for important detected clue.

### C. Connect business relation between industries

The bank has stored the trading information of abundant customers, which can proceed with mining and analysis to the income level, consuming custom, categories of purchasing and other indicators of customers to find out the potential requirements of customers and the connection hidden in data base through connection analysis. The bank can be the media between manufacturers and consumers to connect the decision or plan strategy between them through connection analysis.

Table I. Application example enumerating

Applicative field	Applicative technology	Applications
Forecasting breakup of bank	Artificial neural network method and rule induction	forecast
Forecasting bad loans	Rule induction	forecast
Forecasting price of stock	Artificial neural network method	forecast
Management of investing combination	Rule induction	forecast
Forecasting fluctuate of interest rate	Artificial neural network method and case-based reasoning	forecast
Forecasting futures index	Artificial neural network method and case-based reasoning	forecast

### D. Bankruptcy Prediction

Bankruptcy prediction seems to be most popular topic of the application of data mining techniques on financial data. Corporate bankruptcy causes economic damages for management, investors, creditors and employees together along with social cost. For these reasons bankruptcy prediction is an important issue in finance. Bankruptcy prediction by using financial statements data attracts its origin from the work of Altman in 1968. Altman argues that corporate failure is a long period process and that the financial statement data should include warning signals for the imminent bankruptcy. By applying Multiple Discriminant Analysis techniques he developed a model for bankruptcy prediction. Since the work of Altman many researchers developed alternative models by using statistical techniques.

Tung et al. (2004) employed a hybrid model integrating NNs and fuzzy systems. The model called "Generic Self-organizing Fuzzy Neural Network" was a rule-base consisting of IF-THEN fuzzy rules that can self-adjust the parameters of the fuzzy rules using learning algorithms derived from the NN paradigm. The main advantage of the fuzzy NN was mentioned to be its ability to model a problem by using easily comprehensible high level linguistic model instead of complex mathematical expressions.

The model was applied to predict bank failures. Input variables are 9 financial variables, which have been found to be significant in previous studies. The sample contained data about 2555 non failed and 548 failed banks. 20% of the data were used as training set and 80% as testing set. To reduce Type 1 errors the sample was balanced to include equal number of failed and not failed banks.

### E. Credit Risk Estimation

The task of credit risk analysis becomes more demanding due to the increased number of bankruptcies and the competitive offers of creditors. Data mining techniques have been applied to facilitate the estimation of credit risk.

Huang et al. (2003) performed credit rating analysis by using Support Vector Machines (SVMs), a machine learning technique. Two data sets were used; one containing 74 Korean firms and the other containing 265 US firms. For both data sets 5 rating categories were defined. Two models for Korean data set and two models for US data set, each one having a different input vector were built. SVMs and a backpropagation NNs were used to predict credit rating. SVMs performed better in the three of the four models. Another consideration of the study was to interpret the NN. The Garson method was used to measure the relative importance of the input values.

Mues et al. (2004) used decision diagrams to visualize credit risk evaluation rules. Decision diagrams have the theoretical advantage over decision trees that they avoid the repetition of isomorphic subtrees. Two data sets, one containing German data and two containing Benelux data were used. A NN was employed to perform the classification. The rule extraction methods Neurorule and Trepan were applied to extract rules from the network. Additionally C4.5, C4.5 rules and Entropy-based Oblivious Decision Graphs methods were used to produce decision trees and rules. The performance of Neurorule and Trepan was comparable with the performance of the NNs and superior to the performance of the other methods. Finally the rules were visualized in the form of decision diagrams.

### F. Corporate performance prediaction

Lam (2003) developed a model to predict the return rate on common shareholders equity. She used backpropagation NNs and inferred rules from the weights of the connections by applying the GLARE algorithm. The input vector included 15 financial statement ratios and 1 technical analysis variable. In an additional experiment 11 macroeconomic variables were also included. The data sample contained 364 firms .

Back et al. (2001) developed two models to cluster companies according to their performance. Both models used SOMs. The first model operated over financial data of 160 companies. By employing text mining techniques, the second model analyzed the CEOs' annual report of the companies. The authors concluded that there are differences between the clustering results of the two methods.

Kloptchenko et al. 2004 built on the previously mentioned research effort. Two models were developed, one analyzing financial ratios and the other analyzing the CEOs' reports. In this study a different method, the Prototype-Matching Text Clustering, was used for

analyzing the reports. By comparing the results of the qualitative and the quantitative methods the authors concluded that the text reports tend to foresee changes in the financial state before these changes explicitly influence the financial ratios.

## V. THE FACED PROBLEM AND SOLUTIONS WHEN FINANCIAL INDUSTRY APPLIES DATA MINING TECHNOLOGY

### A. Establish data classification to produce rules

Establish data mining model needs full and accurate data support, i.e. guarantees the correctness and accordance of data values. For example: the assignment of male pregnancy is null in region, but because the shortage of this kind of data as input, this record will be deleted, when we integrate data from several different sources, the fault result will be got. Therefore, it is quite important to produce rules while establishing data classification, the target is making servicer able to complete business and indicator system, planning unified information standard, solving problems of intermediate business, and do the concentrated initial plan well to seek for appropriate measuring relation between concentrating and scattering.

### B. Establishing reasonable distribution decision-making tree

After general data concentration, the phenomenon of "data explosion but lack of knowledge" usually exists, then application for mining the knowledge and method hiding behind data shall be constantly perfected, establishing reasonable distribution decision-making tree to make these decision groups produce rules through the classification to this data groups and indicates hided, unknown identifying rules and further to make it modelling, which is used for gaining relevant business decision plan.

### C. Establishing rules for maintaining individual information security

Legislative regulations: according to the regulation of national privacy right, the over gathering of data, hyper-normal use of data, renting out or reselling data from customers shall all be stopped by law, no matter their starting point is well-meaning or vicious. Formulating enterprise regulations: the regulations will make the interior of enterprise and the industries cooperating with them to strictly observe relevant regulations of privacy right, and they will strictly punish the industry or staffs that reveal the privacy of customers.

### D. Observe the agreement and appointment with customer

Ensure which contents shall be private required by customers, which are the information that customers are willing to provide to enterprise so as to ensure the security of customers' information, at the same time, avoid existing privacy right dispute.

## VI. CONCLUSION

Data mining has widely use in financial field. Currently the financial industry of our country is still at developing stage, we shall pay attention to research the application of data mining technology in aspects of market forecast, price trend, cheat detection, customer relation management, etc. As the mature of data mining technology and the quicken step of informatization in financial industry, data mining will surely play huge effect in financial industry of our country.

## ACKNOWLEDGMENT

I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to Professor Li Xiang, my supervisor, for his constant encouragement and guidance. He has walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

Last my thanks would go to my beloved family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my fellow classmates who gave me their help and time in listening to me and helping me work out my problems during the difficult course of the thesis.

## REFERENCES

- [1] Jiawei Han Michline Kamber. The concept and technology of data mining. Translated by Fan Ming, etc. Beijing: China Machine Press, 2011.
- [2] Zhang Yuchun, The application of data mining in financial analysis, Southern financial computer, 2010.
- [3] Zhang Xian, Data mining technology and its application in financial field, Education and research on finance, 2009.
- [4] Mohan Atreya, Benjamin Hammond. Digital Signatures[M]. McGraw-Hill Education, 2012
- [5] J. He, Y. Geng and K. Pahlavan, Modeling Indoor TOA Ranging Error for Body Mounted Sensors, 2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Sydney, Australia Sep. 2012 (page 682-686)
- [6] Y. Geng, J. Chen, K. Pahlavan, Motion detection using RF signals for the first responder in emergency operations: A PHASER project[C], 2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), London, Britain Sep. 2013
- [7] S. Li, Y. Geng, J. He, K. Pahlavan, Analysis of Three-dimensional Maximum Likelihood Algorithm for Capsule Endoscopy Localization, 2012 5th International Conference on Biomedical Engineering and Informatics (BMEI), Chongqing, China Oct. 2012 (page 721-725)
- [8] Y. Geng, J. He, H. Deng and K. Pahlavan, Modeling the Effect of Human Body on TOA Ranging for Indoor Human Tracking with Wrist Mounted Sensor, 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, Jun. 2013.
- [9] Y. Geng, J. He, K. Pahlavan, Modeling the Effect of Human Body on TOA Based Indoor Human Tracking[J], International Journal of Wireless Information Networks 20(4), 306-317
- [10] Ding Qiulin, Li Shiqi, Customer relation management (first edition), Beijing: Tsinghua humanology press, 2012