The Application of Data Mining in Securities Industry

Li Ma[1] and Mingfeng Jiang[2]

***Abstract*—**The rapid development of electronic business promotes the popularization and application of securities business on the Internet in recent years. Traditional business is transferred to network platform with lower cost, more efficiency and convenience. In securities business, data is an important source of information, and data mining plays an important role in getting information. Combined with the definition, basic techniques and principles, this paper confirms the necessity and feasibility of data mining in securities business, discusses its widespread use in stock prediction, customer management and anomaly detection with cases, summarizes applications of data mining in securities business and looks into the future.

***Keywords*—**data mining, securities industry, stock prediction, customer management, anomaly detection.

I. Introduction

DATA mining is widely applied in e-commerce, financial industry, healthcare, government, etc. It is defined from Wikipedia as the analysis step of the "Knowledge Discovery in Databases" process, [1]a computational process of discovering patterns in huge data sets ("big data") involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems, [2]and an interdisciplinary subfield of computer science. [3-4]The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Fig. 1 illustrates the conception of data mining, and TABLE I is a collection of typical applications in industries, in which the analysis of price volatility and customer management are often used for securities business.

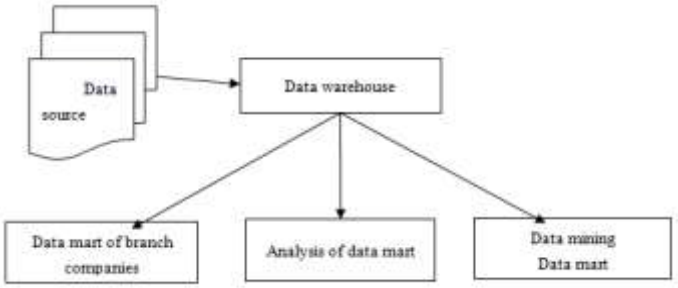

Fig.1: The conception of data mining [5]

TABLE I

THE FUNCTION AND APPLICATION OF DATA MINING

Categories	Functions	Algorithms	Typical applications
Prediction model	Classification	Decision tree, neural net, variance analysis, Logistic regression, probit regression	Targeted marketing, quality control, risk analysis, medical diagnosis
	Numerical prediction	Linear regression, non-linear regression, radial basis function	Analyzing profitability
Segment	Cluster analysis	K means, demographic, neural net	Market segmentation, customer segmentation
Link analysis	Association discovery	Statistics, set theory	Market basket analysis, cross-selling
	Sequential pattern discovery	Statistics, set theory	Sequential market basket analysis
	Similar time sequence discovery	Statistics, set theory	Sales to cash flow ratio, price volatility
Prediction	Time series prediction	ARIMA, Box-Jenkins, neural net, etc.	Sales forecasting, rate forecasting, inventory control, loss forecasting, etc.

II. The Necessity and Feasibility of Data Mining in Securities Industry

Securities market is characterized as a social structure represented by the networks of actors who traded on the floor

Li Ma, School of Business, Dalian University of Technology, Panjin, P. R. China

Mingfeng Jiang, Institute of Science Studies and S&T Management, Dalian University of Technology, Dalian, P. R. China

or on the net of a major securities exchange. Trading among different actors exhibits distinct social structural patterns that dramatically affect the direction and magnitude of price volatility. [6]Securities market encompasses equity markets, bond markets and derivatives markets. Securities market is a barometer of national economy with digitization as one of the most important features, in which brokers require correctness, timeliness and safety of data for high. Therefore data mining plays a particular role in securities business.

A. The Necessity

Advanced technology has guaranteed the securities trading system with amount of data accumulated for a long time. Those data, divided into two major categories: stock quotes data and trading data, could be obtained through data mining.

- Stock quotes data mining provides the basis for trade. As the most important activities, decisions of securities trading have higher demands for dependence and sensitivity of data with faster development of securities business. Data mining collects and analyzes the accumulated data and transfers them to graphic information which could be understood and accepted easily. And then the graphic information can be passed to brokers and customers for making decisions of trade through trading software (e.g. real-time market analysis system).
- Dealing data mining is essential to improving customer services. Analysis, learning and guidance of customer demand have become important issues with the "customer-oriented theory" enjoying popular support in industries, so did the Securities business. Dealing data emerging from different stock exchanges are scattered and private relatively. Data mining could exploit these data and use them to reflect capital positions, holdings and transactions, which could be of a high analysis value in Securities Company and benefit customer orientation. Customer services could be more efficient with data mining.

B. The Feasibility

Applications in securities business of data mining have attracted universal attention of the whole society. Scholars examined the feasibility from the perspective of theory and got some fruits. There are some typical theories: Y. Song & Y. Wu (2013) summarized the key points of data mining on the basis of systematic examination, analyzed technical issues and limitation by virtue of individual cases, and provided technical and theoretical supports to credit rating. [7]S. Liao, P. Chu & Y. You (2011) proposed several possible portfolio alternatives in the Taiwan financial capital market including foreign exchange currencies and stock investment under different circumstances. [8]Karpio, et al. (2013) believed that in order to describe and to understand market's behavior, data mining techniques would be more flexible in use than for instance pricing models based on a finance theory. [9]Those researches provided theoretical basis for applications of data mining.

In reality, brokers help clients avoid risks as much as possible and obtain a higher return by providing them with "valuable information" for their investment and stock-picking. And that "valuable information", hidden in the data of financial statements published regularly by listed companies and daily K-line, could be obtained by data mining. For example, we often check the K-line as a reference when we want to trade.

Theoretical researches and reality applications have demonstrated and affirmed the necessary and feasibility of data mining in securities business.

III. Typical Applications of Data Mining in Securities Industry

The rapid development of network technology has brought a wide range of application of "big data", and then data mining would play greater effectiveness. There are two key factors in ensuring the effectiveness of data mining: addressing problems exactly and the correct data. The former is the premise of well reward, and the latter guarantees the effectiveness of information obtained. Applications of data mining aimed at different problems would be more effective by mining eligible data from a huge database and getting categories of information (see Fig. 2.). There are three essential applications in securities business illustrated with cases as follows: stock prediction, customer management and anomaly detection.

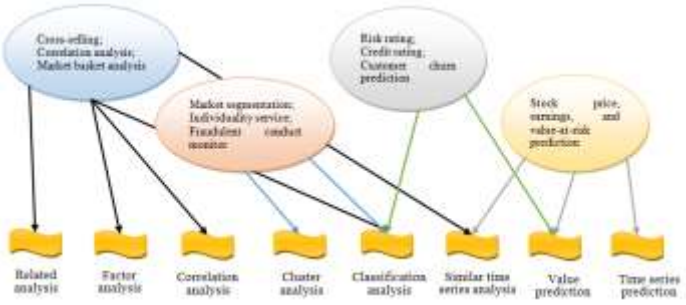

Fig. 2: Typical applications of data mining in securities business

A. Stock Prediction

People often use data mining to deduce the change of stock. Some scholars analysed and compared the principle and models to affirm the function of data mining. Y. Peng (2012) fitted dynamic process of change by Poisson distribution through historical data, determined the best fitting parameter values of Poisson distribution, and built probability-distribution function to predict the change of stock. Tested by Shanghai composite index, new Shanghai composite index and new Shenzhen composite index, the prediction showed a better result and more effective on Chinese securities market in predicting the change to some extent. [10]Shankar &

Purusothaman (2013) proposed a new analysis which deploys the Utility Based Data mining to generate utility emphasized trading rules which based analysis was composed of a pre-analysis and a core analysis. They also believed that unlike the conventional utility based data mining, the obtained trading rules would be optimal and utility efficient. [11]Tsai & Hsiao (2010) combined multiple feature selection methods to identify more representative variables for better prediction. In particular, three well-known feature selection methods, which are Principal Component Analysis (PCA), Genetic Algorithms (GA) and decision trees (CART), were used. The combination methods to filter out unrepresentative variables were based on union, intersection, and multi-intersection strategies. For the prediction model, the back-propagation neural network was developed. They resulted in 14 and 17 important features respectively which could be the important factors for stock prediction and could be used for future investment decisions. [12]

In addition, the rising of big data provides a new idea for data mining and stock prediction. With the help of big data mining, we could predict the change of stock from multiple perspectives. Scholars have tried about this. A report published by California of University, Riverside, in early 2012, predicted stock prices through the analysis of Twitter message. Similarly, Dr. Lai from Nankai University research team also used big data mining to analyze and explore the relationship between the public mood reflected in "microblog" and stock market volatility in 2014. [13]All these researches told us that we could estimate the tendency of stock market and conclude a better return on investment of listed companies for long-term purchase by using data mining.

There is a case of "rising stock model" to illustrate the prediction process. The "rising stock model" could be used to predict the rising of stock price. For example, if we want to examine the change of stock market in the USA over a given period (e.g. 2009-2014), firstly, we should collect original data source of stock from 2009 to 2014 to build a database. Secondly, we should extract important features of the related data through correlation analysis, cluster analysis and related inspection about the change of stock, seasonal or cyclical earnings, the floating of transaction data, etc. Thirdly, we could build a feature library of rising stock and factors leading them rise, model by time window moving and filtering algorithm (TWMA), and then we could predict the change on the basis of the feature library and last sales.

B. Customer Management

Data mining could help us cluster and manage customers. Firstly, customers could be classified according to customer metrics (such as natural property, income contribution, volume of trade, value, etc.) by digging and analysing the data accumulated, then types of customer behaviour mode could be identified and corresponding measures would be taken to promote maximizing the profit of enterprises. This application has also been confirmed from theoretical viewpoints. X. Zhang & C. Qi (2008) examined data mining and generated customer analysis system of securities. They analyzed and pre-processed the data, modeled through K-means and C5.0 algorithm of SPSS CLEMENTINE8.0, used and verified this model to predict the most potential customers. [14]F. Xie (2011) described how to build a subscriber churn analysis system of securities in detail, with which securities companies could forecast subscriber churn and get early warning about potential subscriber churn. She also verified this system with real data from a securities company and achieved ideal results. [15]

Today big data is prevailing, and data collection is more convenient and extensive. If the data of one Securities Company were processed with big data mining, all the locations and consumption patterns would be demonstrated. It helps brokers transport information of securities targeted, estimate the degree of customer concentration and the most valuable customer, cement registered customer and explore the potential one in the area. Besides, the emergence of big data mining could connect the customer with the change of stock. For example, customer satisfaction has been approved to have influence on future performance of stock market. [16]

C. Anomaly Detection

The expansion of capital market depends on network increasingly as the application of E-business. The safety of account and system is essential requirements on securities business because of larger and more complex information system. Abnormal activities of securities trading system threat not only the safety of individual account but also the normal trading order and the development of securities market.

The main performances of anomaly are unusual behaviours in trading stocks, bonds and transferring funds. *A rule for real-time monitoring of abnormal trading securities of Shanghai stock exchange, lunched in 2013, defines* and classifies the anomaly as several trading categories: buying high and selling low for interest transfer, accessing fund frequently, transacting abnormally and frequently of long time idle accounts, a mass of trust or query ask, false declaration, Short-term manipulation, continuous concentrated trade, etc. At this point, data mining could help us realize the anomaly in time. There was a composite figure (see Fig.3) for illustrating the principle of anomaly detection.

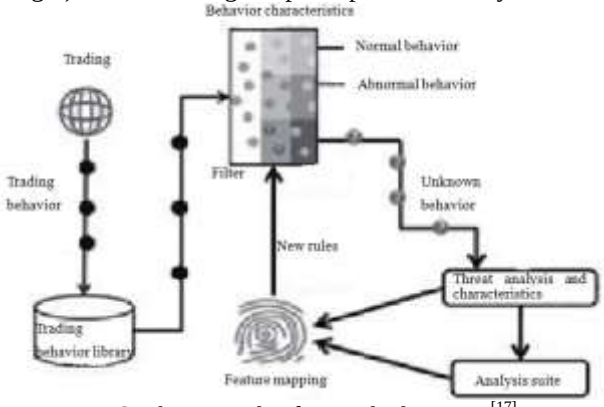

Fig. 3: The principle of anomaly detection [17]

Fig. 3 shows us the principle of data mining used in anomaly detection in securities trading. In this model, we

should collect lots of data of trading to build a huge database, adopt appropriate algorithm (such as GSP, PreFixSpan, and outlier mining) dealing with the data, classify features and abstract the abnormal one.

Anomaly detection is a powerful application of data mining in securities business, which could be implemented mainly because of the development of big data. For example, a founder of SAC capital, the magnate of hedge fund of Wall Street, was suspected of insider trading in 2013. In this incident big data mining was expected to do investigation because the field investigation couldn't be done. Similarly in China, a fund manager named L. Ma was suspected of trading on the information which is not properly disclosed to the public in August 2013 and he was under arrest in the end. In this case, with the help of technology of big data mining, abnormal activities were detected by Shenzhen stock exchange, reported to China Securities Regulatory Commission (CSRC) and solved smoothly.

Several similar incidents caused heat debates in domestic. Nowadays data mining and big data have been familiar with all the fields. It showed that data mining plays an increasingly important role.

IV. Summary and Outlook

Applications of stock prediction, customer management and anomaly detection show the importance of data mining. With the development of "big data", data mining could help us improve the success rate of stock changes with new prediction models, the management efficiency with more precise positioning and the efficiency and sensitivity of anomaly with big data. Technology has injected new vitality for data mining and provided a new idea for securities business. However, the improvement of data mining's technology also brings about privacy problems inevitably. Customers' privacy has taken place indeed in reality and the security of trading has also been questioned. In China, because of the disadvantage of later development of related technique, the legislation is not perfect and security protection is inadequate. In the future, we should not only consolidate and develop new technology, but also pay more attention to the security of data and the perfection of relevant laws. Securities business would gain healthy and vigorous development only if the technology and laws give it high level support.

References

[1] G.Piatetsky-Shapiro. F. Usama, and S. Padhraic. "From data mining to knowledge discovery in databases". Ai Magazine, vol. 17, No. 3, pp. 37-54, 1996.

[2] ACMSIGKDDCurriculumCommittee. "Data mining curriculum". Available:http://www.kdd.org/curriculum/index.html, April 30, 2006.

[3] Christopher Clifton. "Knowledge Discovery in Databases". Available: http://www.britannica.com/technology/data-mining, 2010.

[4] T.Hastie, R.Tibshirani, J.Friedman. et al. "The elements of statistical learning: data mining, inference, and prediction". General Information, vol. 27, No. 2, pp. 567-567, 2001. http://dx.doi.org/10.1007/978-0-387-21606-5

[5] K. Zhang, G. Li. "Application of data mining on security bargain analysis system". Geological Technoeconomic Management, vol. 24, No. 6, pp. 31-34, 2002.

[6] W. E. Baker. "The social structure of a national securities market". American Journal of Sociology, vol. 89, No. 4, pp. 775-811, 1984. http://dx.doi.org/10.1086/227944

[7] Y. Song, Y. Wu. "An applied research on data mining technology in credit scoring". Credit Reference, vol. 177, pp. 24-28, Oct 2013.

[8] S. Liao, P. Chu, Y. You. "Mining the co-movement between foreign exchange rates and category stock indexes in the Taiwan financial capital market". Expert Systems with Applications, Vol. 38, No.4, pp.4608-4617, Apr 2011. http://dx.doi.org/10.1016/j.eswa.2010.09.134

[9] K. Karpio, P.ŁUKASIEWICZ, A.ORŁOWSKI, T. ZĄBKOWSKI. Mining Associations on the Warsaw Stock Exchange[J]. Acta Physica Polonica, A.., Vol. 123, No. 4, 3, pp.553-559, Mar 2013.

[10] Y. Peng. "The probability of stock changes based on data mining". Statistics and Decision, vol. 16, pp. 159-161, 2012.

[11] S.Shankar, T.Purusothaman. "A new utility-emphasized analysis for stock trading rules". Intelligent Data Analysis, Vol. 17, No. 2, pp.271-294, 2013.

[12] C. Tsai, Y. Hsiao. "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches". Decision Support Systems. Vol. 50, No. 1, pp. 258-269. Dec 2010. http://dx.doi.org/10.1016/j.dss.2010.08.028

[13] Nankai group. "Microblog is China's social mood 'pulse'". The Chinese news, Available: http://www.chinadaily.com.cn/dfpd/tj/bwzg/2014-03/29/content_17389504.htm, April, 28th 2014.

[14] X. Zhang, C. Qi. "Securities clients analysis system based on data mining". Computer Applications, vol. 28, pp. 369-371,375, Jun 2008.

[15] F. Xie. "Application of data mining in customer churns management of Securities Company". Science and Technology Management Research, vol. 10, pp. 180-183, 2011.

[16] R. Jacobson, N. Mizik." The financial markets and customer satisfaction: reexamining possible financial market mispricing of customer satisfaction". Marketing Science, Vol. 28, No. 5, pp. 810-819, Sep/Oct2009. http://dx.doi.org/10.1287/mksc.1090.0495

[17] H. Zhu, X. Chen. "An analysis based on data mining of abnormal behavior in online securities trading". Network & Computer Security, vol. 8, pp. 20-24, 2011.