The Investment Strategy Decision Support System Applying Rule-based Neural Network and Trading Rules Extracted from Qualified Foreign Institutional Investors by Data Mining

Workshop on Artificial Intelligence

Deng-Yiv Chiu, Kuang-Ming Hsu

Department of Information Management of Chung Hua University

30 Tung Shiang, Hsin Chu City, Taiwan, 30067, R.O.C

TEL: (03) 5374281-6519

E-mail: **chiuden@mi.chu.edu.tw**, **m9010004@mi.chu.edu.tw**

Contact author: Kuang-Ming Hsu

Abstract

In the stock market, investors can invest in stocks to obtain profit. But there are many individual persons without professional investment knowledge about the stock market. It is difficult for them to choose stocks with high returns. After they have chosen the good stocks, it is also difficult for them to choose the time for selling or buying the stock. So, most of investors would lose money in investing stocks.

The purpose of this study is to propose a Decision Support System, DSS, which could support investors to invest in the stock market. The DSS could recommend investors which stocks may have high returns, and recommend the appropriate time to buy or sell the stocks. For the knowledge acquirement of DSS, we would apply data mining technique to relative electronics stocks information owned by Qualified Foreign Institutional Investors (QFIIs) to obtain the QFII's investment knowledge. The QFII's investment knowledge is composed of many trade rules in stock market investment. For the timing of buying or selling stocks, the different trade rules may provide different significance. So, in different time, each trade rule has different reference weight. We would use rule-based neural network to train each trade rule to obtain reference weight. In this way, we expect to support investors to invest in the stock market to obtain high returns or to decrease damage.

Key word: **QFII, DSS, Data Mining, Rule-based Neural Network**

1. Introduction

In the stock market, most of investors are individual persons who do not have professional investment knowledge. When someone says that some stocks are worthy of investing, many investors would buy the stocks. When someone says that some stocks will fall, they would just sell the stocks. Most of time, they buy unworthy stocks, and buy or sell stocks in wrong timing. Instead of obtaining profit, they lose money in investing stocks.

Recently, there are many people, called investment experts or teachers. They teach investors to invest in stocks on TV. They earn commission by teaching investors how to buy or sell stocks. But some of those teachers may not have enough professional investment knowledge. They may make many wrong decisions, and the investment market is full of garbage information. So, their members still lose money in investing stocks.

Although many investors have lost much money in stock market, some experts with professional investment knowledge still obtain much profit by investing in stocks. They analyze relative information and make right decisions to buy or sell stocks by professional skills. The purpose of this study is to propose a DSS that could support investors to invest in the stock market. The DSS could extract expert trade rules from those transaction data of the experts with professional investment knowledge. Hopefully the DSS has the capability of choosing stocks with high returns from the stocks market. The DSS also has the capability of recommending investors to buy or sell stocks in right time.

This study uses data mining technique to analyze experts' investment transaction data and

extract their investment knowledge. We would know what information they use, how they analyze these information to invest, and how they make decisions to buy or sell stocks at right time according to rules. The DSS would use these rules to recommend investors how to invest to obtain profit or how to avoid losing money.

Many factors would influence the stock market. The stock market changes any time. So, the rules extracted by data mining technique are not appropriate at any time. For example, there are two rules, A and B, extracted by data mining. Sometimes rule A may be more significant than rule B in the stock market. So rule A has much reference weight and rule B has little reference weight. At another time, rule B may be more significant than rule A. Different rules at different time have different appropriate weight in the stock market. So, in this study, we would use artificial neural network to get the appropriate weights for each expert rule. We would input the recent relative information and use neural network to train the expert rules. Then we would obtain the appropriate reference weight for each rule at that time in the stock market. By this way, the DSS model could use these rules to recommend investors the appropriate investment strategy, which stocks to buy or sell.

2. Backgrounds and Literature Review

2.1 KDD and Data Mining

This study expects to extract usable investment knowledge from huge QFII's transaction data and relative information. We would use the extracted knowledge to form the DSS knowledge base. The traditional method of KDD (Knowledge Discovery in Databases) is analyzing, explaining and extracting huge data by manual ways. But today, the amount of information gets more and more. And the real world changes very fast. The traditional method is not

efficient and cannot satisfy our demands. Now, many people use data mining to extract knowledge. Data mining is a step in the KDD process that consist of applying data analysis and discovery algorithms that produce a particular enumeration of patterns or models over data [1]. The KDD process proposed by Usama Fayyad is shown in Figure 1.
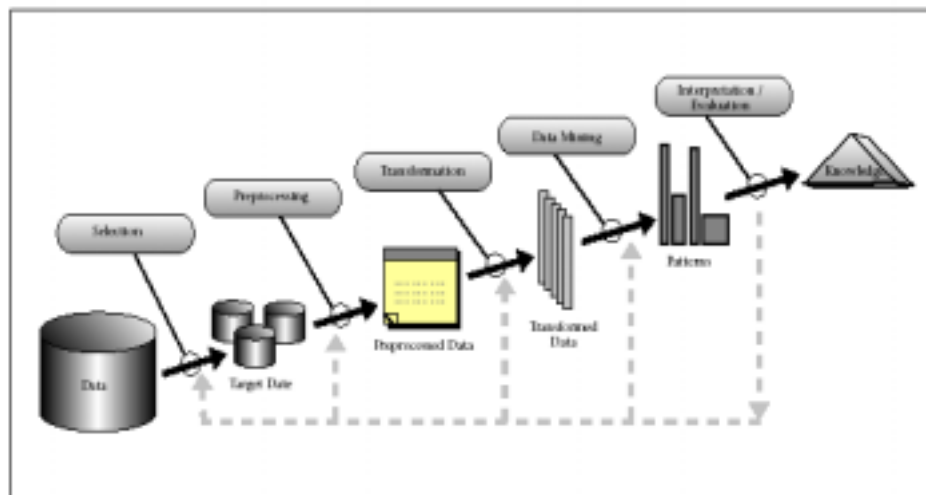

Figure 1. An Overview of the Steps That Compose the KDD Process [1].

2.1.1 Data Mining Application

Data mining is using automatic or semi-automatic method to explore and analyze huge data, and to find unknown and meaningful patterns, associations and rules. Data mining builds effective model and rules for decision-making. The data mining application is divided into classification, estimation, prediction, clustering, association and description [2].

The application and researches of data mining are very wide, and they are usually used in retail business, communication, medical treatment and finance. Tim Chen used data mining

5

and event study to explore the relationship between financial statement and investing returns, and then built a model to estimate cumulated abnormal return (CAR) [3].

2.1.2 Data Mining Step

Different data mining model is used for different purpose and different data type. For example, some models are used for dealing with continuous data type, and others are used for dealing with categorical data type [4]. Groth proposed five data mining steps: data clean, study definition, model construction, model understanding and prediction [5]. Jiang considered that different algorithms are used for different data type, and data transformation needs to cooperate with selected algorithm [4]. Pyle considered that all processes were called data preparation before doing data mining step. And the data preparation step is eighty percent of the whole data mining processes [6]. This study would use data mining techniques according to KDD steps proposed by Usama Fayyad [1] to extract the expert rules by analyzing the transaction data of QFII.

2.2 Qualified Foreign Institutional Investors (QFII)

Recently, many researches showed that the QFII's investment trends had great influence on Taiwan stock market. The QFII's overbuy and oversell affect the Taiwan stock market. The QFII's overbuy makes the stock price rising and the QFII's oversell makes stock price falling down. Stock price is influenced by many factors, for example, policy, economy, national state, etc. But those factors may make the fluctuation of stock price in the short time, and it would come back to its proper price of the stock. The QFII owns many experts, and they often perform long-term investment. So we expect to apply the professional skill derived from QFII to help us to invest in the stock market.

The ROI (Return of Investment) of QFII is significantly higher than that of TAIEX [7]. It indicates that QFII has good performance in investing stock market. Their outstanding return has made their investment strategy perceived in the market. A lot of researches have proved that returns better than the average can be obtained if financial-ratio-based performance indicates are used as investment criteria. The correlation between QFII holdings and financial performance is significantly positive. This result supports the argument that the increase/decrease of QFII shareholding on a particular company is a good proxy to the positive/negative change of the company's financial performance [8].

According to literature mentioned above, the QFII chose long-term and stable investment strategy. They use professional investment knowledge to analyze the financial data and related information to make decisions and to invest. If we could extract their trade rules and use these trade rules to invest, we would obtain much profit. So this study is designed to extract investment knowledge from QFII's transaction data.

2.3 Artificial Neural Network

Many researchers use static methods in financial study, such as regression and experimental design. Now there are many researchers using information technology to analyze and discuss financial data, such as Data Mining, Fuzzy, Artificial Neural Network, GA and so on.

Kiviluoto K. proposes a novel tool based on a hierarchy of two self-organizing maps (SOMs) for analyzing financial statements. The first-level SOM determines the company's

position each year. The second-level SOM turns out to give a more accurate description of the state of the company than the first-level SOM. It is easy to interpret, as each point on the second-level SOM corresponds to a trajectory on the first-level SOM. With this method, several different patterns of corporate behavior can be recognized [9].

Most researches only consider the quantitative factors, such as stock indicators and stock index. But qualitative factors are also important, such as policy and industry news. These qualitative factors deeply affect the stock market. KUO R. J. proposed an intelligent stock market forecasting system that combined ANN and Fuzzy Delphi. The system considered both quantitative and qualitative factors to predict the stock market [10].

Neural network has the capabilities of machine learning and solving the non-line problem. So, it is applied in many domains. Neural network suffers from two major drawbacks. First, it is difficulty in embedding existing knowledge into the system. The other is lack of explanation capability [11]. When the DSS recommends investors to invest, if the DSS output is without any explanation action, the decision could not convince investors of the DSS output. So in this study, we would use rule-based neural network to solve the two problems. Rule-based Neural Network is a network combining both connectionistic and symbolic reasoning [12]. We use inference mechanism in neural network, and then the investors could know how the DSS make the decision through the expert rule. Most researches are proposed to predict the market index. But for the individual persons, it is important for the stocks that they buy to raise or to fall. So the purpose of this study is to predict a single stock's rising or falling in the future and to recommend investor to buy or to sell the stock.

3. DSS Framework

This study proposes a DSS framework, and the DSS framework is composed of two parts. The first part is KDD model. We would use it to extract the expert rules from the QFII's transaction data. The second part is rule-based neural network. We would use it to train these trade rules to get appropriate reference weight. The DSS framework that we proposed in this research is shown in figure 2.
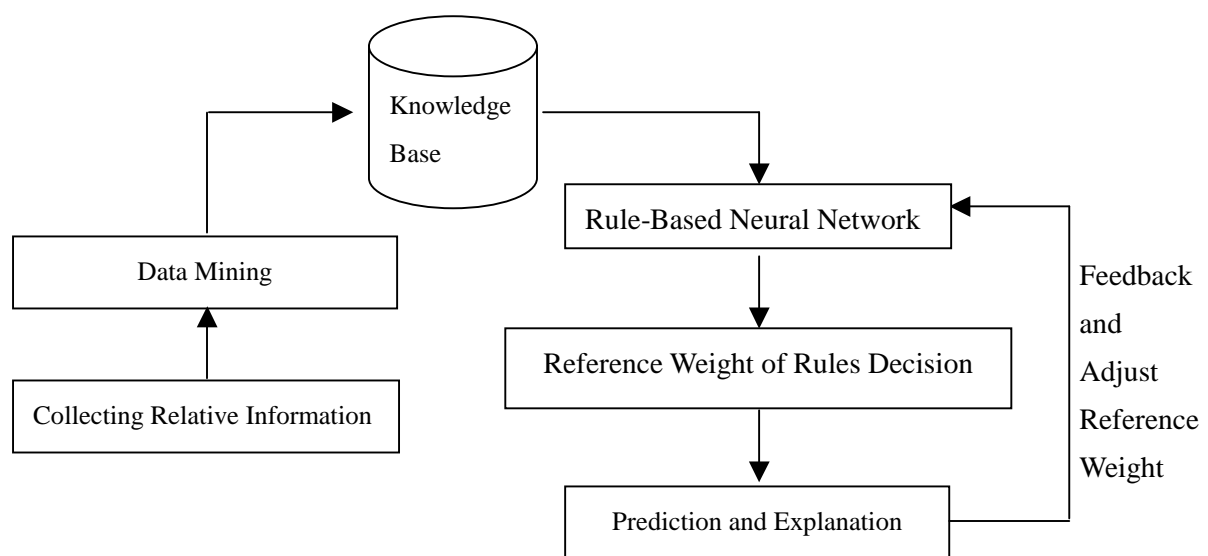

Figure 2. The DSS Framework

3.1 KDD Model

In this phase, we would use the KDD process proposed by Usama Fayyad [1] to extract the expert investment rules. The KDD steps are selection, preprocessing, transformation, data mining and interpretation / evaluation in turn. Because the criteria of each industry on financial ratios and technical indicators are different, we would focus on the electronic industry. We would use data mining techniques to mine relative information about electronic industry.

The time series is the record composing a series of real number, and each real number represents the corresponding value at a time point [13]. The tendency diagram formed by a series of numbers could represent stock price, sales volume and exchange rate, etc. In the financial domain, the information about time series is important. For example, the tendency of stock price or technical indicators changes over a period of time. Those information of time series is important for making investment strategy. So we would add the time series factors in the knowledge extracting mechanism to find the relationship between the long-term change of each referenced indicator and QFII's investment strategy. By this way, the extracted rules would describe the QFII's investment strategy accurately. In data representation, when we collect relative information, we also record its time stamp. Then we could analyze those long-term change tendencies of relative indicators by using time stamp.

We would use the relative indicators as the dimensions in data mining technology to analyze the QFII's transaction data and relative information. On this phase, the input data sets of data mining are as below.

1. The transaction data of the QFII overbuying and overselling electronic category stocks within top 10 transaction amount every week.
2. The financial ratios, technical indicators and stock price of selected electronic companies.
3. The relative information of the stock market.

The output of the data-mining model is association rules about QFII's investment strategy. The form of the extracted rules is shown as following.

IF (the change ratio of first season, second season and third season is positive) THEN QFII

would increase 5% of holding stock rate.

In data mining, we would use the Apriori algorithm proposed by Agrawal [14] to extract the input data sets and use support, confidence and improvement to determine whether the extracted rules are good enough. An efficient rule must satisfy that the support, confidence and improvement are equal to or greater than the threshold decided by us. The formulas are shown in following where X is condition of the expert rule, and Y is result.

$$\text{Support} = \frac{\text{number of transactions which contain X and Y}}{\text{number of transactions in the database}}$$

$$\text{Confidence} = \frac{\text{number of transactions which contain X and Y}}{\text{number of transactions which contain X}}$$

$$\text{Improvement} = \frac{\text{number of transactions which contain X and Y}}{P(X) \times P(Y)}$$

After knowledge extracting phase above, we would obtain the association rules about QFII's investment strategy. We would use these rules in DSS knowledge base to provide inference mechanism.

3.2 Reference Weight of Rules

By KDD process discussed above, we could find the expert rules about QFII's investment strategy in electronic category stocks. Some rules are significant. But some rules are unimportant. Different rules have different reference weight. We need to decide the reference

11

weight of each rule for prediction result. If we let the reference weights of rules decided by investment experts, it may be subjective. And the stock market changes very fast, we cannot depend on experts each time. It expends large cost. In this study, we would choose objective methods to obtain reference. We would use rule-based neural network to train the experts and obtain the appropriate reference weight.

3.2.1 Rule-Based Neural Network architecture

The input of rule-based neural network is a set of expert rules. The form of these expert rules is shown as following, proposed by Chou S. T. [11].

If　(A1 (cf1) and A2 (cf2) … and An (cfn))

Then B (CF)

Where Ai is condition, B is conclusion, cfi is the certainty factor of the ith condition, and CF is the rule's certainty factor. The rule-based neural network framework is divided into three layers. The input layer represents condition of the rule, the hidden layer represents the rule itself, and the output layer represents the prediction result. The interconnection weights between the input layer and the hidden layer come from the certainty factors of corresponding conditions in rule. The rules' certainty factors become the interconnection weights from the hidden layer to the output layer.

For instance, there are five rules extracted by data mining. The symbol A, B, C and D represent technical indicators.

(1) If　A>0 (0.6) and B<3 (0.5)　Then　QFII would overbuy (0.7)

(2) If   A>3 (0.5) and B<1 (0.4) and C>5 (0.7)   Then   QFII would overbuy (0.9)

(3) If   B<0 (0.3) and C<3 (1.0)   Then   QFII would oversell (0.6)

(4) If   C<0 (0.5) and D<0 (0.7)   Then   QFII would oversell (0.5)

(5) If   A>0 (0.7) and D<0 (0.5)   Then   QFII would oversell (0.8)

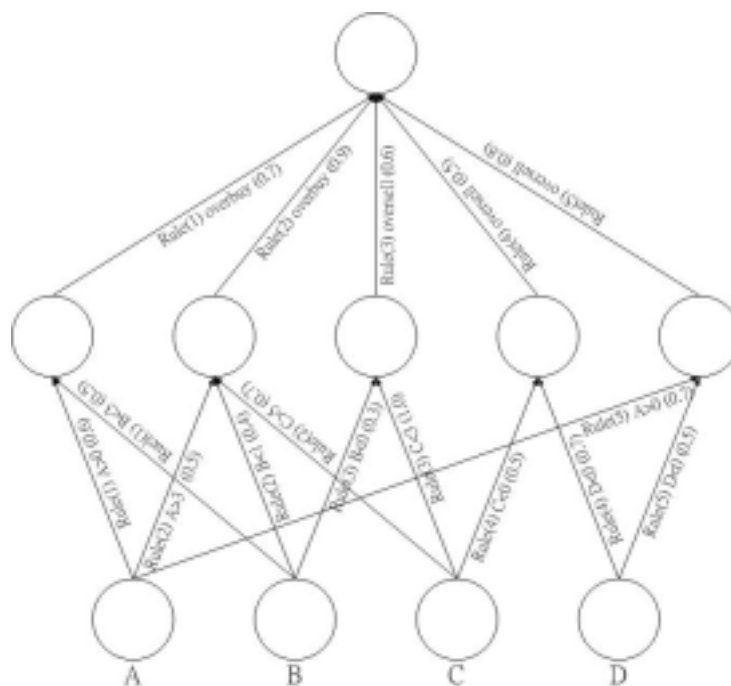The initial rule-based neural network architecture is shown in figure 3.

Figure 3. The Initial Rule-Based Neural Network Architecture

3.2.2 Rule Refinement

After deciding the initial reference weight, it is not constant forever. Over the time changing,

the reference weight of each rule becomes different. Sometimes, rule A has more significance.

But, at another time, rule B has more significance, and rule A becomes unimportant. In this

study, the rule-based neural network has the mechanism used to change the reference weight

of each rule dynamically and incrementally. After each prediction, the DSS would feedback the prediction result and the real result automatically and using rule-based neural network to adjust the expert rules to obtain appropriate reference weight.

On this phase, the input data sets of the rule-based neural network are the dimensions of the conditions of each rule, such as the technical indicators of a week. The output data sets of the rule-based neural network are the investment strategy of the DSS suggestions and the explanation.

IF output = +1 THEN buy the stock

IF output = 0 THEN hold the stock

IF output = -1 THEN sell the stock

When the prediction result and the real state are different, they would feedback to the rule-based neural network and adjust the weight of the rules to obtain appropriate reference weight. By the mechanism of refining the reference weight of rules, each rule has appropriate reference weight. The DSS would use these expert rules to inference and to recommend investors how to invest.

4. Experiment Design

We would divide our experiment into three parts, including model-training phase, testing phase and evaluating phase. The source data is also divided into training data set, testing data set and evaluating data set. We would use the training data set to extract the expert rules and to construct the rule-based neural network first. We expect to make the DSS model powerful,

and not restricted by the training data set. So we use the testing data set to adjust the DSS model. And then, we use the evaluation data set to estimate the DSS model performance in dealing with unknown data.

We would collect relative data for two years. The relative data includes companies' stock price, financial ratios and technical indicators. We would also collect the stock market information, such as market price, etc. We would use the data of the former 80 weeks to be the training data set. And the duration of the testing data set and the duration of evaluating data set are 12 weeks individually. We use those data to train, test and evaluate the DSS model.

After constructing the DSS model, we would evaluate the performance of DSS model. We would use the DSS to invest in the stock market, and explain how the DSS makes the decision. If the performance of DSS does not meet our demand, we would improve it on three phases.

1. Knowledge Extraction Target: The performance of DSS model could not meet our demand, because the knowledge extraction target may not appropriate. So, we could consider other investment experts and use their transaction data to extract expert knowledge. We also could use other existing expert investment rules in the DSS knowledge base. We expect to obtain the appropriate expert investment rules.

2. Knowledge Extraction Method: There are many methods in data mining to extract knowledge, such as associated rules, decision tree, etc. There are also many algorithms for each method. For example, the algorithms of decision tree may be C4.5, CART, CHAID, etc. We expect to find appropriate data mining method and algorithm, and to extract appropriate investment rules.

3. Rule-Based Neural Network: There are many choices on its algorithm. The different algorithms would affect the learning mechanism and the prediction output. So, we need to explore the appropriate structure of rule-based neural network.

After completing experiments above, we would find appropriate methods to extract expert rules. And then, we would use appropriate neural network to train the rules and find appropriate reference weight. We would obtain satisfactory performance from DSS, and then the DSS could recommend investors with appropriate investment strategies.

5. Conclusion

The purpose of this study would construct a DSS, which could support investors to invest in the stock market. By using the expert knowledge, the DSS could select the worthy investing and may have high returns electronic category stocks from the stock market. The DSS also could recommend investors when to buy or sell the stocks at right time. By above DSS and experiments, we expect to obtain a DSS with well performance and it can support investors to invest in the stock market to obtain high returns or to decrease damage.

Because the criteria of each industry on financial ratios and technical indicators are different, this study focuses on the electronic industry. In the future, we may discuss other industries. This study focuses only on single stock predicting and recommends investors investing at right timing. For deep discussion, we may make a study of optimal investment portfolio or other capabilities, and expect to develop a well performance DSS to recommend investors the investment strategy.

6. Reference

[1] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P., "From Data Mining to Knowledge Discovery in Databases," Proceeding American Association for Artificial Intelligence, 1996, PP.37-54.

[2] Berry, M.; Linoff, G., "Data Mining Techniques: for Marketing, Sales, and Customer Support," John Wiley& Sons, 1997

[3] Chen, T.; Liu, J. J., "A Study on Stock Price Abnormal Return of Software Industry in American Market -- Using On-line Public Information Mining the Relevant Financial Ratios," Thesis, 2000.

[4] Jiang, M. F.; Tseng, S. S.; Liao, S. Y., "Data Types Generalization for Data Mining Algorithms," IEEE Tran. Knowledge and Data Engineering, 1999.

[5] Groth, Robert, "Data Mining: Building Competitive Advantage," Prentice Hall PTR, 1999

[6] Pyle, Dorian, "Data Preparation for Data Mining," Morgan Kaufmann Publishers, 1999

[7] Johnhu; Liu W. M., "The Impact of Study of the QFII'S Overbuy and Oversell Information," Thesis, 2000.

[8] Hu, J. W.; Huang, S. L., "The Relationship Between Financial Performance of Publicly Listed Companies and Foreign Investment in Taiwan," Thesis, 2000.

[9] Kiviluoto, K.; Bergius, P., "Two-Level Self-Organizing Maps for Analysis of Financial Statements," Neural Networks Proceedings, 1998, IEEE World Congress on Computational Intelligence, The 1998 IEEE International Joint Conference, Volume: 1, PP. 189 -192.

[10] Kuo, R.J.; Lee, L.C.; Lee, C.F., "Systems, Man, and Cybernetics," 1996, IEEE International Conference on, Volume: 2, 1996, PP. 1073 -1078

[11] Chou, S. T.; Yang, C. C.; Chan, C. H.; Lai, F. P., "A Rule-Based Neural Stock Trading Decision Support System," Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on , 1996, PP. 148 –154

[12] Fu, L. M., "Knowledge-Based Neural Network," Neural Network in computer Intelligent, McGraw-Hill, 1994, PP. 115-143

[13] Wang, H. J.; Anthony; Lee, J. T., "Similarity Query and Data Mining in Time-series Data: Taking TAIWAN Stock Market for Example," Thesis, 1999.

[14] Agrawal, R.; Srikant, R., "Fast Algorithms for Mining Association Rules in Large Database," Proceedings of the 20th Internal Conference on Very Large Data Bases, September 1994.