Data mining Techniques to Analyze the Risks in Stocks/Options Investment

Anupama Surendran

Dept of Computer Science

Cochin University of Science & Technology, CUSAT

Cochin-22, Kerala, India

anupama_surendran@rediffmail.com

anupama.deepak@gmail.com

Abstract—**Data mining is one of the most optimal methods to analyze the data .Nowadays most of the people will be reluctant to invest money in shares and options because of the global economic crisis. In this research paper, a proposal is made in order to analyze the stocks/sectors and options based on various financial parameters using data mining techniques. The neuro-fuzzy logic technique is proposed to use in this proposal, to develop this work and the mining technique is applied to classify the result obtained**

Keywords- Data mining, Options, EPS, DPS, Neuro-Fuzzy logic

I. INTRODUCTION (HEADING 1)

Data mining is the analysis of large quantities of data. In this proposed work, the various factors affecting the options and shares should be studied and these factors are classified using the neuro-fuzzy techniques in data mining, so that the people can decide whether to invest their money in share market and if so, in which company they have to invest [1].The main aim of combining neural network and fuzzy logic technique is to make the learning algorithms of neural networks simpler by combining the linguistic methods and rules of fuzzy logic

II. DATA MINING REVISITED

Data mining is the analysis of large quantities of data, so as to retrieve useful and meaningful patterns and rules. The volume of data is increasing day by day. In order to overcome the deficiencies of manual analysis, data mining techniques can be used, so that an accurate and optimal result is obtained. Data mining involves a series of steps. In classification, the incoming data is grouped by comparing their features to the predefined elements of a class. In estimation, a border limit is established and checked whether the data value is above or below that limit and the classification is done. Association rules helps to decide which combinations are best, so that the outcome is best. In clustering the grouping of data is done. There may be one or more similar groups. There is no predefined data in clustering. The clustering is done on the basis of the similarities in the data obtained. Either directed or undirected style of data mining can be used .In directed data mining predictions are made which can be applied to the future and in undirected data mining, data is analyzed to establish its properties so as to retrieve some useful information from data.

III. ANALYSIS OF STOCKS/SECTORS BASED ON FINANCIAL PARAMETERS.

In this section investor is provided with versatile information on various financial parameters. Investor is also provided comparison of price movement to financial parameters. Based on historical performances and future projection, expert analysis valuation is provided .Expert ranking of stock/sector is provided. Investor is provided with market summary. For doing market summary huge volume of data will have to be considered, so the neuro-fuzzzy logic can be used to cluster or group the data.

The various financial parameters which are to be considered are:

1. Earnings per share or EPS and it is the Profit after tax divided by the number of outstanding shares.

2. Dividend per share or DPS is the divided paid on common stock divided by the number of outstanding shares.

3. Price earnings or P/E is the Market price divided by EPS.

4. Book value is the Shareholder's equity divided by the number of outstanding shares.

A glossary of the above financial parameters is constructed and a data base of the sectors and stock is created .Data mining techniques is then used to classify data. The database constructed will have the company sector, list of companies, financial data for each company (the financial data includes EPS, DIVIDEND, P/E).The data is grouped according the value of EPS, DPS, DIVIDEND, P/E range. For example the EPS range can be set as 0, 1-5, 5-10, 10-20, and 20-30 and so on. Expert analysis is done and then the highest and lowest P/E is found. Maximum downside target is found by multiplying the highest P/E and projected EPS and minimum is found by multiplying the lowest P/E and projected EPS. The next step is to provide a weekly price data base for the last 10 years and stock weekly data base for last 10 years and a 52 week high/low data base for index and stocks. To do these things data mining will be the most optimal method because large volume of data has to be considered .There will be thousands of companies and each company's financial data for the last 10 years we have to consider and process the data to obtain the needed result. The percentage of cumulative return

IAMA 2009

for index is calculated as the difference between the present value and the base period value divided by base period value .The cumulative stock return is also calculated. If the cumulative stock return is greater than cumulative index return, then rank is assigned as one. If up seven years is seven out of last 10 years for stock, then rank is one. Similarly a Stock scanner and an Industry scanner is developed which will help to analyze the best/worst performers.

IV. INVESTING IN OPTIONS

An option is a contract hat gives the holder a right, without any obligation, to buy or sell an underlying asset at a given strike price on or before a specified expiration period [4]. The underlying asset could be a share or any other asset. Call option is the right to buy an asset and put option is the sell an asset. A buyer of a call option on a share will exercise the deal right when the actual share price at expiration(S) is higher than the exercise price (E).Similarly the buyer of a put option will exercise the deal if the exercise price is higher than the share price, and will not exercise the option if the share price is equal to or greater than exercise price [5].

TABLE I. FACTORS AFFECTING OPTION PRICE

CALL OPTION	POSITION	PUT OPTION
Exercise if S>E	In the money	Exercise if E>S
Do not exercise if S<E	Out of the money	Do not exercise if E<S
Do not exercise if S=E	At the money	Do not exercise if S=E

The five factors that affect the value of share option are the share price, exercise price, volatility (standard deviation) of the share return, risk free rate of interest and the options time to expiration. A call option's value will increase with the increase in share price, rate of interest, volatility and time of expiration. It will decrease with increase in the exercise price. A put option's value will increase with increase in the share price and the rate of interest. The binomial tree approach is used to determine the value of option in simple situation and in complex situations the Black and Scholes model is used [4]. The result obtained from these methods can be used to train the neuro-fuzzy networks, so that it will be easy to classify large volume of data and then the outcome can be predicted.

V. FUSION METHODS OF FUZZY SYSTEMS AND NEURAL NETWORKS

This method stared in late 1980s, rapidly increased in 1990s and is applied to many application areas. The fuzzy system are applied to various application areas and does not require complex mathematical modeling and represent and manage uncertain and vague knowledge and easy to describing expert knowledge and experience [2]. Linguistic terms and IF-THEN rules are used and they resembles human decision making. Neural Networks is the computation model of the operation of human brain and the nodes are connected by links and the weights are used as long-term storage [3]. The learning capability is by updating the weights and is simple and easy to apply, but black-block model. The fusion of fuzzy and neuro systems will give modified fuzzy systems with neural networks making membership functions with neural networks.

VI. FUZZY NEURAL NETWORKS

It was proposed by Kwak and Lee in 1994
The Five layers were
1st: input nodes
2nd: linguistic terms of inputs
3rd: antecedent parts of rules
4th: consequent parts of rules
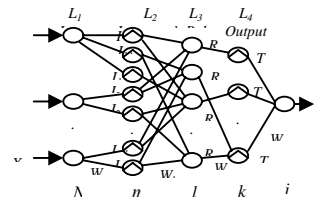5th: output node with defuzzification

Figure 1. Fuzzy neural systems

First layer (Inputs)

$$f_j^1(x) = x \tag{1}$$

Second layer (Input linguistic terms)

$$f_j^2(x) = \mu_{A_j}(x) \tag{2}$$

$$\mu_{A_j}(x) = \begin{cases} 1 - \dfrac{x_i - c_{A_j}}{sr_{A_j}}, & x_i \in (c_{A_j}, c_{A_j} + sr_{A_j}] \quad (3) \\[2mm] 1 + \dfrac{x_i - c_{A_j}}{sl_{A_j}}, & x_i \in [c_{A_j} - sl_{A_j}, c_{A_j}] \quad (4) \\[2mm] 0, & \text{otherwise} \qquad\qquad (5) \end{cases}$$

Third layer (Antecedent parts)

"IF X_1 is I11 and ••• and Xm is Im1 THEN"

$$f_j^3(x_1, x_2, \cdots, x_p) = \begin{cases} \min_{i=1}^p (x_i) & \text{if minimum used} \quad (6) \\[2mm] \prod_i^p (x_i) & \text{if product used} \quad (7) \end{cases}$$

Fourth layer (Consequent parts)

$$f_j^4(x) = \max_{i=1}^{q}\{w_{ji}x_i\} \tag{8}$$

Fifth layer (Defuzzification)

$$f_j^5(x_1, x_2, \cdots, x_t) = \frac{\sum_i^t Centroid(B_i, x_i) Area(B_i, x_i)}{\sum_i^t Area(B_i, x_i)} \tag{9}$$

$$(11) \quad \begin{aligned} Area(B_i, x_i) &= \sum^n \min(\mu_{B_i}(y_i), x_i) \\ Centroid(B_i, x_i) &= \frac{\sum_i y_i \cdot \min(\mu_{B_i}(y_i), x_i)}{\sum_i^n \min(\mu_{B_i}(y_i), x_i)} \end{aligned}$$

Learning algorithm is based on error back propagation and weights of links between 3rd and 4th layer and parameters in the nodes of 2nd and 4th layer.

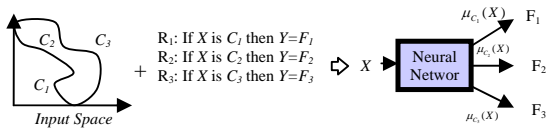## VII. MAKING MEMBERSHIP FUNCTIONS WITH NEURAL NETWORKS

Figure2. Steps in making membership function with neural systems

Building Neural networks on fuzzy partitioned input space builds a fuzzy classifier which clusters the given input-output data into several classes and builds a neural network per a class and trains the neural networks with the input-output data in the corresponding class and reduces the complexity of NNs by using Fuzzy rules dividing input space with fuzzy rules and constructing NNs for each fuzzy partitions [6] [7].
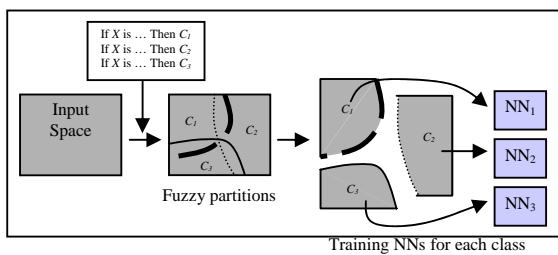

Figure3.Building fuzzy –neural networks

VIII. CONCLUSION

The Stock/Sector investment analysis is done with neuro-fuzzy logic methods and the data is classified. The values of various financial parameters are taken and the neuro-fuzzy networks are trained according to the result obtained from the various parameters. The neural networks are trained based on the linguistic rules obtained from fuzzy networks. Thus instead of calculating all the values and dividing the data into various categories, the data mining approach using neuro-fuzzy methods will train data and the huge volume of data can be categorized .This will make the investment decision easier for customers.

REFERENCES

[1] H. Takagi, "R&D in Intelligent Technologies: Fusion of NN, FS, GA, Chaos, and Human," Half-Day Tutorial/Workshop, IEEE International Conference on Systems, Man, and Cybernetics, Orlando, FL, Oct. 11, 1997

[2] Zimmermann, H.J. (1978), Fuzzy programming and linear programming with several objective functions, Fuzzy Sets and Systems.

[3] Ya –Junzhang and Zhi-Quiang Liu "self-splitting competitive learning: A new On-Line clustering paradigm" IEEE Transactions on Neural Networks, vol 13 NO.2 March 2002 PP 369-380.

[4] F.Black and M.Scholes,"The Pricing of Options and Corporate Liabilities,Journel of Political Economy,81 pp 637-654(May-June 1973)

[5] J.Cox and M.Rubistein,Option Markets,Prentice Hall,1985.

[6] Duch W., Hayashi Y., "Computational Intelligence Methods in Data Understanding", Springer Studies in Fuzziness and Soft Computing, 2000, Vol 54.

[7] S. J. Ovaska, H. F. VanLandingham, and A. Kamiya, "Fusion of Soft Computing and Hard Computing in Industrial Applications: An Overview," IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews 32, 72–79 (2002).