Optimization of Fuzzy Metagraph Based Stock Market DSS Using Genetic Algorithm

A. Thirunavukarasu, S.Uma Maheswari

Abstract—Investing in stock market is not an easy task. It needs careful decision making skills to maintain profit in the long run. Though many research works are available to guide the investors, they are not reliable and do not guarantee maximum profit possible. In this paper a fuzzy metagraph based stock market decision support system is proposed for short term investors. Technical indicators like RSI and William-%R are used to support the system. Membership functions are optimized using genetic algorithm. The proposed system reduces the risks involved in stock investing by significantly. The genetic algorithm optimization improves overall profit by 30% to 40%.

Keywords—DSS, Fuzzy Metagraph, Genetic Algorithm, RSI, Will %R, stock market.

I. INTRODUCTION

Stock market is one of the most popular investments due to its high returns. High return does not come without risk. If proper care was not taken while investment it may lead to adverse effects. There are numerous stock prediction models to help investors to predict a stock direction. Stock prediction models usually belong to two categories. They are fundamental analysis based model and technical analysis based model. Fundamental analysis involves global economic conditions, whether conditions, annual budget and company information. Technical analysis involves predicting stock price using technical indicators like RSI, MACD and William -%R. Most of the proposed prediction models are based on technical analysis. Classifiers like SVM, Fuzzy logic and neural networks are used to classify stock status.

Recently research work in stock market predictions are gaining momentum since stock investment is still a mystery and difficult to correctly predict. The predictions so far been based on portfolio management, fundamental analysis and technical analysis. Fuzzy metagraph is an emerging technique used in the design of many information processing systems like transaction processing systems, decision support systems, and workflow systems. In our work a fuzzy metagraph based fuzzy rule management and genetic algorithm based optimization are proposed.

The rest of the paper is organized as follows. Section 2 gives the related work. Section 3 deals with metagraph technique and also two technical indicators of the stock market. Proposed model DSS is presented in section 4. The membership function optimization using GA is presented in section 5. In section 6 the system results and discussion have been discussed. Section 7 concludes the paper.

II. RELATED WORK

Ahmed A. Gamil, and Raafat.S have proposed a multi agent and fuzzy logic model for stock market decision making based on technical analysis. They used short term to long term Moving Average indicators to develop a decision making system. The model is tuned and modified using genetic algorithms. The DSS has been integrated into an agent based framework to enhance the stock information retrieval process, and to be accessible through the Internet [2]. They developed a novel hybrid model based on a chaotic firefly algorithm and support vector regression (SVR) for stock market price forecasting. Integration of chaotic motion with a firefly algorithm as a simple and novel optimization method [3]. An-Sing Chen et al [7] developed a system to predict the direction of return on the Taiwan Stock Exchange Index. The probabilistic neural network is used to forecast the direction of index return. The performance of the PNN based system is compared with that of the generalized methods of moments (GMM) with Kalman filter and random walk. The results showed that PNN outperforms other models.

L. J. Cao et al [11] proposed a SVM based stock market prediction system for futures contracts collated from the Chicago Mercantile Market are used as the data sets. The performance of the system is investigated by varying kernel parameters. They have proposed a multiple-kernel support vector regression approach for stock market price forecasting. A two-stage multiple kernel learning algorithm is developed to optimally combine multiple-kernel matrices for support vector regression. The learning algorithm applies sequential minimal optimization and gradient projection iteratively to obtain Lagrange multipliers and optimal kernel weights [13]. Jacek Ma ndziuk et al[23] developed an ANN and GA based system for short term stock index prediction. Technical variables are
taken as neural network inputs. GA is used to find an optimal set of input variables. The system performs well still it is time consuming and complex.

Kasemsan.K, and Radeerom.M have proposed a decision-making model based on the application of Neuro fuzzy systems. The model has been applied in order to make a one-step forward decision, considering historical data of daily stock returns. They have used RSI, MACD and other indicators to support the trading system o formulate a trading strategy which achieves more stable results and higher profits when compared with Neural Networks and the Buy and Hold strategy [24]. Kyoung-jae Kim et al [25] proposed a support vector machine based stock prediction system. They used SVM to predict the direction of daily stock price change in the Korea composite stock price index. For this study they selected 12 technical indicators to make up the initial attributes.

Simone Bova, and Pietro Codara have proposed to analyze Mamdani type fuzzy control systems in logical terms, with special emphasis on the fuzzy inference process Mamdani type implementation of the fuzzy system. They have used RSI and MACD to analyze stock market. They have developed an alternative inference system to Mamdani system [37]. Takashi Kimoto et al [43] proposed a index based prediction of Tokyo stock exchange prices indexes using ANN. They used both technical indexes and economic indexes to predict buy and sell timing for one month in the future. They developed a new high-speed learning method called supplementary learning to train neural networks. The buying and selling signal were accurate but were not tested for selected stocks.

Financial markets are highly volatile and generate huge amount of data on a day to day basis. The present study applied the popular data mining tool of k-NN for the task of prediction and classification of the stock index values of BSE-SENSEX and NSE-NIFTY. The results of k-NN classifier are compared with the Logistic regression model and it is observed that the k-NN classifier outperforms the traditional logistic regression method as it classifies the future movement of the BSE-SENSEX and NSE-NIFTY more accurately [42]. SVM is a promising type of tool for financial forecasting. They proposed a combining model by integrating SVM with other classification methods. The weakness of one method can be balanced by the strengths of another by achieving a systematic effect. The combining model performs best among all the forecasting methods [47].

Tiffany Hui-Kuang et al [44] used neural networks to forecast Taiwan stock index. They also developed a new fuzzy time series model to improve the forecasting. Vaidehi.V, and Monica.S have proposed a subtractive clustering based fuzzy system identification method to model a prediction system that can predict future movement of stock prices by taking samples of past events. When recent data are given to the trained system, it gives the possibility of a rise or a fall along with the next possible value of data. The prediction model is trained by daily market price data. It can also be used as a weekly or a monthly predictor [46]. Yakup Kara et al [49] proposed a neural network and support vector machine based stock prediction system to predict securities of Istanbul stock exchange. The models use both artificial neural networks (ANN) and support vector machines (SVM) for classification. Ten technical indicators including RSI, MACD simple MA and will %R were selected as inputs of the proposed models. The proposed model performs well when neural network is used as a classifier. Zheng-Hua has proposed a Fuzzy Metagraph based knowledge representation. The FM has been applied to fuzzy rule-based systems for knowledge representation and reasoning. In the format of algebraic representation and FM closure matrix [50]. In our work a fuzzy metagraph based fuzzy rule management and genetic algorithm based optimization are proposed.

III. METHODOLOGY AND TECHNICAL INDICATORS

A. Metagraph and its adjacency matrix

A metagraph S = <X, E> is a graphical representation consisting of two tuples X and E. Here X is its generating set and E is the set of edges defined on generating sets. The generating set X of the metagraph S i.e. the set of elements X = {x1, x2, x3, ......,xn} represents variables and occurs in the edges of the metagraph. An edge e in a metagraph is a pair e = (<Ve, We>) ∈ E (where E is the set of edges) consisting of an invertex V ∈ X and an outvertex W ∈ X. A simple path h(x, y) from an element x to an element y is a sequence of edges e1, e2, . . . ., en such that xe1 invertex(e1), ye2 outvertex(e2), and for all ei=1,2,...,n-1, outvertex(ei) ∩ invertex(ei) ≠ ∅. The coinput of x in the path (denoted coinput(x)) is the set of all other invertex elements in the path’s edges that are not also in the outvertex of any edges in the path, and the cooutput of y (denoted cooutput(y)) is the set of all outvertex elements other than y. The length of a simple path is the number of edges in the path [9].

The graphical structure can be represented by the adjacency matrix of a metagraph. The adjacency matrix A of a metagraph is a square matrix with one row and one column for each element in the generating set X. The ijth element of A, denoted aij, is a set of triples, one for each edge e connecting xj to xi. Each triple is of the form <Ci, COi, Ei>, in which Ci is the coinput of xj in e and COi is the cooutput of xi in e. Table 1 represents adjacency matrix A of given metagraph. For example in the figure 1, aij = a11 = ∅, since there are no edges connecting x1 to itself or connecting x1 to x1. On the other hand, a13 contains one tripie, since there is one edge connecting x1 to x3. The first component of the triple is the coinput of x1 for this edge, the second component is the cooutput of x3, and the third component is the edge. Since x1 has coinput x2, while x3 has no cooutput, and the edge is e1, we have a13 = {< {x2}, ∅, < e1 >}. Similarly, since e1 connects x2 to x4 with no coinputs or cooutputs and no other edges connect x2 to x4 we have a24 = {< ∅, ∅, < e2 >>}. The metagraph construct is unable to tackle the issue of uncertainty and imprecision. Originally, metagraph focused on structural aspects, i.e., the connectivity relationships among components of systems. As an outcome, this approach cannot support uncertain knowledge representation and approximate reasoning.
Graph
Fuzzy set theory is primarily concerned with quantifying the vagueness in human thought and perception, where linguistic terms can be properly represented by the approximate reasoning of fuzzy sets. Linguistic variables are used in fuzzy logic. In the fuzzy theory, fuzzy set \( A \) of universe \( X \) is defined by function \( \mu_A \) called the membership function of set \( A \). Membership Functions are used to convert crisp inputs into fuzzy values. \( \mu_A : X \rightarrow [0, 1] \), where \( \mu_A(x) = 1 \) if \( x \) is totally in \( A \), \( \mu_A(x) = 0 \) if \( x \) is not in \( A \), \( 0 < \mu_A(x) < 1 \) if \( x \) is partly in \( A \). A graph is a symmetric binary relation on a nonempty set \( V \). Similarly, a fuzzy graph is a symmetric binary fuzzy relation on a fuzzy subset. Let \( V \) be a non empty set. A fuzzy graph is a pair of functions \( G = (V, \sigma, \mu) \) Where \( \sigma \) is a fuzzy subset of \( V \) and \( \mu \) is a symmetric fuzzy relation on \( \sigma \). i.e. \( \sigma : V \rightarrow [0,1] \) and \( \mu : V \times V \rightarrow [0,1] \), a fuzzy edge set of \( G \) such that for all \( x, y \in V \), \( \mu(x, y) \leq \sigma(x) \land \sigma(y) \).

C. Technical Indicators and Its Calculation
Technical indicators are numerical values calculated using mathematical formulas on historical stock information. There are numerous technical indicators calculated by well known mathematicians. Each indicator can be used independently or in combination with other indicators. Among them relative strength index (RSI) and William %R are well known and widely used indicators. In our work Relative strength index (RSI) and William-%R are used.

Relative strength index: It is a momentum oscillator that measures the speed and direction at which prices are moving. RSI oscillate within the band of zero to 100. The formula used to calculate RSI is

\[
\text{RSI} = 100 - \left( \frac{100}{1 + \text{RS}} \right)
\]

Where, \( \text{RS} = \text{Average gain/Average loss} \)

Simple 14 period averages are used as initial value for average gain and average loss. For subsequent values the following formula is used

\[
\text{Average Gain} = \left[ \text{(previous Average Gain)} \times 13 + \text{current Gain} \right] / 14.
\]

\[
\text{Average Loss} = \left[ \text{(previous Average Loss)} \times 13 + \text{current Loss} \right] / 14.
\]

\( \text{A stock is considered overbought when the value of RSI is above 70} \)

\( \text{and oversold when below 30. Signals can also be generated by looking for divergences, failure swings and centerline crossovers. RSI can also be used to identify the general trend. Buy signal is generated when RSI is near 30 and sell signal is generated when RSI is near 70. The value between 35 and 65 is considered as normal \[45\]. The RSI indicator for TCS, BSE and India is shown in figure 2.} \)

William-%R: Developed by Larry Williams, William-%R is a momentum indicator which reflects the level of the close relative to the highest high and lowest low for the look-back period.
William-%R oscillates from 0 to -100. A stock is considered overbought when William-%R Read from 0 to -20. Readings from -80 to -100 are considered oversold. The default setting for William-%R is 14 periods, which can be days, weeks, months or an intraday timeframe. Buy signal is generated when William-%R is near -80. Sell signal is generated when indicator is near -20. The will %R indicator for TCS, BSE and India is shown in figure 3.

IV. PROPOSED MODEL OF THE DSS

The structure of the proposed model is shown in figure 4. The proposed stock market decision support system consists of several stages. Initially historical stock information is collected from stock market. This historical information of selected stocks is collected. Technical indicators are calculated using historical stock information. In our work RSI and William-%R are used.

A fuzzy metagraph is a triple $\tilde{S} = \{X, \tilde{E}, \tilde{V}\}$ in which $X$ is a fuzzy set on $X$ and $\tilde{E}$ is a fuzzy relation on $X\times X$. A fuzzy set $X$ on $X$ is completely characterized by its membership function $\mu: X \rightarrow [0, 1]$ for each $x \in X$, $\mu(x)$ is the truth value of the statement of “$x$ belongs to $X$”. $\tilde{E}$ is a fuzzy edge set $\{\tilde{e}_m, m=1, 2, 3, \ldots, m\}$. Each component $\tilde{e}$ in $\tilde{E}$ is characterized by an ordered pair $<\tilde{v}_n, \tilde{w}_m>$. In the pair $\tilde{v}_n$ subset of $\tilde{X}$ is the in-vertex of $\tilde{e}_m$ and $\tilde{w}_m$ subset of $\tilde{W}$ is the out-vertex. The membership value of an edge is also called certainty factor of the edge [50].

Fuzzy Inference System and Rulebase

A Fuzzy Inference System is used to map an input space to an output space using fuzzy logic. Fuzzy inference system uses a collection of membership function and rules to drive output from a crisp input. A fuzzy inference system implements a nonlinear mapping from its input space to output space through a number of fuzzy if-then rules. The three most popular inference systems used in fuzzy logic are the: Mamdani fuzzy model, Tagaki-Sugeno fuzzy model, and Tsukamoto model. In this work Mamdani Fuzzy model is used.

Rule Base for Fuzzy Metagraph Based System

A rule base is an ordered pair $T = \{P, R\}$ in which $P$ is a set of propositions and $R$ is a set of rules. Given a generating set $X$ and a metagraph $S=\langle X, E \rangle$ on $X$ with $E$ and $A$ rule base $T=\langle P, R\rangle$, $S$ corresponds to $T$ if there are bijective mappings between $X$ and $P$ and between $E$ and $R$. If $S$ corresponds to $T$, then $X$ corresponds to $P$ and $E$ corresponds to $R$. The metagraph notation to refer the rule base system that is, $x_i$ will denote the $i^{th}$ proposition and $e_k$ will denote the $k^{th}$ proposition.
rule. For example, in the metagraph of Figure 1, \( e_1 \) represents the rule \( x_1 \land x_2 \rightarrow x_3 \), where “\( \land \)” denotes conjunction and “\( \rightarrow \)” denotes implication. \( e_2: x_3 \rightarrow x_4 \), \( e_3: x_4 \rightarrow x_5 \). Since \( M_2 \) is a metapath from \{x1, x2\} to \{x5\}, we can conclude that \( x_5 \) can be inferred from \( x_1 \) and \( x_2 \)-that is \( x_1 \land x_2 \rightarrow x_5 \). Fuzzy rules for fuzzy metagraphs are formed based on fuzzy metagraph shown in figure 5. Technical indicators RSI, WILLIAM-%R are used as input vector to the fuzzy Metagraph system. Decision vector (BUY, HOLD, and SELL) is output of the system. Where, RSI = \{x1, x2, x3\}, WILLIAM-%R = \{x4, x5, x6\} and DECISION = \{x10, x11, x12\}. Elements Proposition For generating set from figure 6. X1: RSI <30,
X2: 30<=RSI<=70, X3: RSI>70,
X4: WILLIAM - %R <-80
X5: -80<=WILLIAM - %R<=-20
X6: WILLIAM - %R> -80
X7: X1 \land X4 \rightarrow X7,
X8: X2 \land X5 \rightarrow X8,
X9: X3 \land X6 \rightarrow X9,
X10: BUY, X11: HOLD, X12: SELL

Defuzzification and Decision
Defuzzification is the process of finding one single crisp value that summarizes the fuzzy set that enters it from the inference block. The output crisp value is used to make decisions. Here any value between 0 and 35 can be used as a BUY signal. Value between 35 and 65 is used as STAY signal. Any value between 65 and 100 is used as SELL signal.

V. MEMBERSHIP FUNCTION OPTIMIZATION USING GENETIC ALGORITHM

The genetic algorithm is a model of machine learning algorithm based on the evolutionary ideas of natural selection and genetic. The steps of genetic algorithm includes
1. Randomly generate an initial population
2. Compute and save the fitness value for each individual in the current population
3. Define selection probabilities for each individual
4. Generate a new population set by probabilistically selecting individuals from random population to produce offspring via genetic operators
5. Repeat step 2 until satisfying solution is obtained.

The decision making ability or profit depends on the membership function of the fuzzy system. If the membership function was properly chosen the profit earned can be increased. One of the ways to find out the suitable membership function parameters is optimization of the membership function. Genetic algorithm is used to optimize the membership functions. The results of optimized membership functions are compared with traditional membership functions. Optimization increased the total profit by 30% -40%. The flow diagram for genetic algorithm based membership function optimization is shown in figure 7. In this work both trapezoidal and triangular membership functions are used to define RSI and William-%R. The parameters of input membership function are taken as gene of chromosomes. The parameters that are fixed are not taken. Initially five individuals are randomly generated with some constraint. Total profit for given time period is taken as fitness value of the membership function parameters.

The profit is calculated by assigning these values to the membership functions. Two individuals are chosen randomly and applied crossover and mutation. 30th bit from the initial point is taken as crossover point and mutation is done by complementing a randomly chosen bit. Profit for these two individual are calculated. The individual with maximum profit is taken as next generation. The randomly generated individual with minimum profit is replaced by the next generation individual. This process is repeated until maximum profit is achieved. The membership function of RSI and William-%R
VI. RESULTS AND DISCUSSIONS

To test the performance of the proposed system two stocks of Bombay stock exchange, India are considered. They are Tata consultancy service (TCS) and Reliance Industry Limited (RIL). Both companies are popular and well performing companies in India. The year 2011-2012 had been very challenging year for Indian share market. In this study stock information for that period is taken to analyze the performance of the system at hard times. Table 3 represents testing results of TCS for one year. Table 4 represents testing results of RIL for one year. The system is tested with last one year historical trading data of the respective companies. In order to test the effectiveness of the system the results obtained from genetic algorithm optimized system is compared with that of non optimized system and BUY and HOLD method. The findings are listed in the table 5 given below. From the experience the proposed system outperforms both fuzzy logic based stock market decision support system and traditional BUY and HOLD method. With our system one can expect 30% to 40% more profit than any other methods.

VII. CONCLUSION

Fuzzy metagraph based decision support system for stock market investing is proposed. Technical indicators are used to aid the decision making. Genetic algorithm is used to improve the overall profit of the system. The proposed system can be used in short term investing. The system performance is tested using stocks listed in Bombay stock exchange. The results are satisfactory. This method reduces the risk factor involved in stock investing. Future works may concentrate on optimization techniques applied for tuning the input parameters to further enhance the performance of the system.

### Table .3 Testing results of TCS for one year

<table>
<thead>
<tr>
<th>Date(D D-MM-YY)</th>
<th>Price</th>
<th>Signal</th>
<th>Buy price</th>
<th>Profit</th>
<th>Total profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-02-2011</td>
<td>1098</td>
<td>BUY</td>
<td>1098</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>31-03-2011</td>
<td>1182</td>
<td>SELL</td>
<td>1182</td>
<td>84</td>
<td>-</td>
</tr>
<tr>
<td>20-06-2011</td>
<td>1069</td>
<td>BUY</td>
<td>1069</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14-10-2011</td>
<td>1134</td>
<td>SELL</td>
<td>1134</td>
<td>64</td>
<td>148</td>
</tr>
</tbody>
</table>

### Table .4 Testing results of RIL for one year

<table>
<thead>
<tr>
<th>Date(DD-MM-YY)</th>
<th>Price</th>
<th>Signal</th>
<th>Buy price</th>
<th>Profit</th>
<th>Total profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>27-01-2011</td>
<td>943</td>
<td>BUY</td>
<td>943</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>01-03-2011</td>
<td>988</td>
<td>SELL</td>
<td>988</td>
<td>45</td>
<td>-</td>
</tr>
<tr>
<td>20-06-2011</td>
<td>834</td>
<td>BUY</td>
<td>834</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25-07-2011</td>
<td>882</td>
<td>SELL</td>
<td>882</td>
<td>48</td>
<td>-</td>
</tr>
<tr>
<td>04-08-2011</td>
<td>812</td>
<td>BUY</td>
<td>812</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>08-09-2011</td>
<td>853</td>
<td>SELL</td>
<td>853</td>
<td>48</td>
<td>-</td>
</tr>
<tr>
<td>16-11-2011</td>
<td>786</td>
<td>BUY</td>
<td>786</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>07-12-2011</td>
<td>809</td>
<td>SELL</td>
<td>809</td>
<td>23</td>
<td>164</td>
</tr>
</tbody>
</table>

### Table .5 Comparative analyses of the various results

<table>
<thead>
<tr>
<th>Name</th>
<th>Price at the beginning of the investment period in Rs</th>
<th>Price at the end of the investment period in Rs</th>
<th>Profit per share without applying GA in Rs</th>
<th>Profit per share with GA in Rs</th>
<th>Profit using BUY and HOLD method in Rs</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCS</td>
<td>1178</td>
<td>1170</td>
<td>108</td>
<td>148</td>
<td>-8</td>
</tr>
<tr>
<td>RIL</td>
<td>1079</td>
<td>698</td>
<td>93</td>
<td>164</td>
<td>-381</td>
</tr>
</tbody>
</table>
REFERENCES


About the Authors

Mr. A. Thirunavukarasu completed his B.E Degree in Computer Science and Engineering from Coimbatore Institute of Technology, Coimbatore in the year 2006 and M.E. Degree in Computer Science and Engineering from Anna University of Technology, Coimbatore in the year 2009. Currently he is pursuing PhD degree from Anna University, Chennai. He is working as a Teaching Fellow, Department of Computer Science and Engineering in Anna University, University College of Engineering, Ramanathapuram Campus. He is having more than 5 years of teaching and research experience. He has published technical papers in National / International conferences/Journals. His areas of specialization include Data Structures and Algorithms, Compilers, Theory of computation, Data mining, Database Security and Metagraph.

Dr. S. Uma Maheswari received her B.E Degree in Electronics and Communication Engineering from Government College of Technology, Coimbatore in the year 1985 and M.E (Applied Electronics) from Bharathiar University in 1991. She received her Ph.D degree in the area of Biometrics from Bharathiar University, Coimbatore in the year 2009. She is Associate Professor of Electronics and Communication Engineering department in Coimbatore Institute of Technology. She is having more than 27 years of teaching experience. She has published technical papers in National /International Conferences/Journals. Her special fields of interest are Digital Image Processing and Digital Signal Processing. She is a Member of IE (India), Life Member in Indian Society for Technical Education (India), Life Member in Systems Society of India, and Life Member in Council of Engineers (India).