Stocks selected using SOM and Genetic Algorithm based Backpropagation Neural Network gives better returns.

Asif Ullah Khan
Asst. Prof. Dept. of Computer Sc. & Engg.
All Saints' College of Technology, Bhopal

T. K. Bandopadhyaya
Professor, Bansal Institute of Science and Technology, Bhopal

Sudhir Sharma
Professor, RGPV, Bhopal

Abstract

Investment in stock market is one of the most popular type of investment. There are many conventional techniques being used and these include technical and fundamental analysis. The main aim of every investor is to earn maximum possible return on investments. The main issue with any approach is the proper weighting of criteria to obtain a list of stocks that are suitable for investments. This paper proposes an improved method for stock picking using self-organizing maps and genetic algorithm based backpropagation neural networks. The stock selected using self-organizing maps and genetic algorithm based backpropagation neural networks outperformed the BSE-30 Index by about 30.17% *based on one and half month of stock data.*

Key Words: Neural Network, Stocks Classification**,** Technical Analysis, Fundamental Analysis, Self-Organizing Map (SOM), Genetic algorithm based backpropagation neural network(GA-BPN).

1. Introduction

Investments in stock market has long been an attraction in the minds of investors. The major forecasting method used in financial area is either technical or fundamental. Technical Analysis [1] provides a framework for studying investor behavior, and generally focuses only on price and volume data. Technical Analysis using this approach has short-term investment horizons, and access to only price and exchange data. Fundamental analysis involves analysis of a company's performance and profitability to determine its share price. By studying the overall economic conditions, the company's competition, and other factors, it is possible to determine expected returns and the intrinsic value of shares. This type of analysis assumes that a share's current (and future) price depends on its intrinsic value and anticipated return on investment. As new information is released pertaining to the company's status, the expected return on the company's shares will change, which affects the stock price. So the advantages of fundamental analysis are its ability to predict changes before they show up on the charts. Growth prospects are related to the current economic environment. Due to the fact that stock markets are affected by many highly interrelated economic, political and even psychological factors that interact with each other in a very complex fashion, it is very difficult to forecast the movement in the stock market. The purpose of this paper is to develop a method so that investors get maximum returns in short period of time. Stocks have been selected by us on the bases of fundamental analysis criteria. These criteria are evaluated for each stock and compared in order to obtain a list of stocks that are suitable for investments. Stocks are selected by applying common criteria on the stocks listed on Indian National Stock Exchange (NSE).After selection of stocks using fundamental analysis, classification of selected stocks is done in to fixed number of classes by Self-Organizing map. Each of the class is having its own properties; stocks having properties closer to a particular class get assigned to it. Among the classified stocks we then select stock for investments using genetic algorithm based backpropagation neural network.

2. Stocks Classification

Stocks are often classified based on the type of company it is, the company's value, or in some cases the level of return that is expected from the company. Some companies grow faster than others, while some have reached what they perceive as their peak and don't think they can handle more growth. In some cases, management just might be content with the level of business that they've achieved, thus stalling to make moves to gain further business. Before investing in a particular company, it is very important to get to know the company on a personal level and find out what the company's goals and objectives are for the short and long term. In order to prosper in the world of stock investing, a person must have a clear understanding of what they are doing, or they shouldn't be doing it at all. Stocks can be a very risky investment, depending on the level of knowledge held by the person(s)

making the investment decisions. Below is a list of classifications which are generally known to us- Growth Stocks, Value Stocks, Large Cap Stocks, Mid Cap Stocks, and Small Cap Stocks. Stocks are usually classified according to their characteristics. Some are classified according to their growth potential in the long run and the others as per their current valuations. Similarly, stocks can also be classified according to their market capitalization. The classifications are not rigid and no rules are laid down anywhere for their classification. We classified stocks by taking in account the Shareholding Pattern, P/E Ratio, Dividend Yield, Price/Book Value Ratio, Return on Net worth (RONW), Annual growth in Sales, Annual growth in Reported Profit After Tax, Return on Capital Employed (ROCE) and Adjusted Profit After Tax Margin (APATM) with Self-Organizing Map.

3. Stock Market Index

A stock market index is a method of measuring a stock market as a whole. Stock market indexes may be classed in many ways. A *broad-base* index represents the performance of a whole stock market — and by proxy, reflects investor sentiment on the state of the economy. The most regularly quoted market indexes are broad-base indexes comprised of the stocks of large companies listed on a nation's largest stock exchanges, such as the American Dow Jones Industrial Average and S&P 500 Index, the British FTSE 100, the French CAC 40, the German DAX, the Japanese Nikkei 225, the Indian Sensex and the Hong Kong Hang Seng Index. Movements of the index should represent the returns obtained by "typical" portfolios in the country. Ups and downs in the index reflect the changing expectations of the stock market about future dividends of country's corporate sector. When the index goes up, it is because the stock market thinks that the prospective dividends in the future will be better than previously thought. When prospects of dividends in the future become pessimistic, the index drops.

3.1 Composition of Stock Market Index

The most important type of market index is the broad-market index, consisting of the large, liquid stocks of the country. In most countries, a single major index dominates benchmarking, index funds, index derivatives and research applications. In addition, more specialised indices often find interesting applications. In India, we have seen situations where a dedicated industry fund uses an industry index as a benchmark. In India, where clear categories of ownership groups exist, it becomes interesting to examine the performance of classes of companies sorted by ownership group. We compared BSE-30 SENSEX with the stock selected using SOM and GA-BPN. We choose BSE-30 SENSEX for comparison because SENSEX is regarded to be the pulse of the Indian stock market. As the oldest index in the country, it provides the time series data over a fairly long period of time (From 1979 onwards). Small wonder, the SENSEX has over the years become one of the most prominent brands in the country. SENSEX is calculated using the "Free-float Market Capitalization" methodology. As per this methodology, the level of index at any point of time reflects the free-float market value of 30 component stocks relative to a base period. The market capitalization of a company is determined by multiplying the price of its stock by the number of shares issued by the company. This market capitalization is further multiplied by the free-float factor to determine the free-float market capitalization. The base period of SENSEX is 1978-79 and the base value is 100 index points. This is often indicated by the notation 1978-79=100. The calculation of SENSEX involves dividing the Free-float market capitalization of 30 companies in the Index by a number called the Index Divisor. The Divisor is the only link to the original base period value of the SENSEX. It keeps the Index comparable over time and is the adjustment point for all Index adjustments arising out of corporate actions, replacement of scrips etc. During market hours, prices of the index scrips, at which latest trades are executed, are used by the trading system to calculate SENSEX every 15 seconds and disseminated in real time.

BSE-30	A.C.C., JAIPRAKASH ASSOCIATS, BHARTI TELEVENTURES ,BHEL, CIPLA LTD, DLF LTD, GRASIM IND, GUJARAT AMBUJA CEMENT, HDFC, HDFC BANK, HINDALCO, HINDUSTAN LEVER, ICICI BANK, INFOSYS TECHNOLOGIES, ITC LTD., LARSEN & TOUBRO, MAHINDRA & MAHINDRA, MARUTI UDYOG, NATIONAL THERMAL POWER, ONGC, RANBAXY LAB., RELIANCE, RELIANCE COMMUNICATIONS, RELIANCE ENERGY, SATYAM COMPUTER, STATE BANK OF INDIA, TATA CONSULTANCY, TATA MOTORS, TATA STEEL, WIPRO LTD.

Table 1: List of companies of BSE-30 index

4. Application of Neural Networks in Stocks

4.1 Overview

The ability of neural networks to discover nonlinear relationships [3] in input data makes them ideal for modeling nonlinear dynamic systems such as the stock market. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A neural network method can enhance an investor's forecasting ability [4]. Neural networks are also gaining popularity in forecasting market variables [5]. A trained neural network can be thought of as an expert in the category of information it has been given to analyze. This expert can then be used to provide

projections given new situations of interest and answer "what if" questions. Traditionally forecasting research and practice had been dominated by statistical methods but results were insufficient in prediction accuracy [6]. Monica et al's work [7] supported the potential of NNs for forecasting and prediction. Asif Ullah Khan et al. [8] used the back propagation neural networks with different number of hidden layers to analyze the prediction of the buy/sell. Neural networks using back propagation algorithms having one hidden layer give more accurate results in comparison to two, three, four and five hidden layers.

4.2 Kohonen self-organizing map

Self-organizing maps (SOM) belong to a general class of neural network methods, which are nonlinear regression techniques that can be applied to find relationships between inputs and outputs or organize data so as to disclose so far unknown patterns or structures. It is an excellent tool in exploratory phase of data mining [9]. It is widely used in application to the analysis of financial information [10]. The results of the study indicate that self-organizing maps can be feasible tools for classification of large amounts of financial data [11]. The Self-Organizing Map, SOM, has established its position as a widely applied tool in data-analysis and visualization of high-dimensional data. Within other statistical methods the SOM has no close counterpart, and thus it provides a complementary view to the data. The SOM is, however, the most widely used method in this category, because it provides some notable advantages over the alternatives. These include, ease of use, especially for inexperienced users, and very intuitive display of the data projected on to a regular two-dimensional slab, as on a sheet of a paper. The main potential of the SOM is in exploratory data analysis, which differs from standard statistical data analysis in that there are no presumed set of hypotheses that are validated in the analysis. Instead, the hypotheses are generated from the data in the data-driven exploratory phase and validated in the confirmatory phase. There are some problems where the exploratory phase may be sufficient alone, such as visualization of data without more quantitative statistical inference upon it. In practical data analysis problems the most common task is to search for dependencies between variables. In such a problem, SOM can be used for getting insight to the data and for the initial search of potential dependencies. In general the findings need to be validated with more classical methods, in order to assess the confidence of the conclusions and to reject those that are not statistically significant. In this contribution we discuss the use of the SOM in searching for dependencies in the data. First we normalize the selected parameters and then we initialize the SOM network. We then train SOM to give the maximum likelihood estimate, so that we can associate a particular stock with a particular node in the classification layer. The self-organizing networks assume a topological structure among the cluster units [2]. There are m cluster units, arranged in a one or two dimensional array: the input signals are n-dimensional. Fig. 1 shows architecture of self-organizing network (SOM), which consists of input layer, and Kohonen or clustering layer.

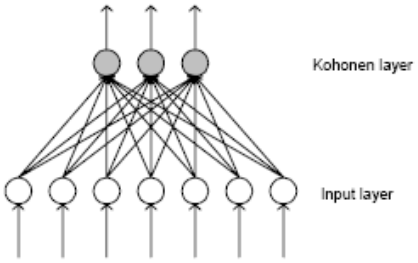

Fig.1: Architecture of Kohonen self-organizing map

The shadowed units in the Fig. 1 are processing units. SOM network may cluster the data into N number of classes. When a self-organizing network is used, an input vector is presented at each step. These vectors constitute the "environment" of the network. Each new input produces an adaptation of the parameters. If such modifications are correctly controlled, the network can build a kind of internal representation of the environment.

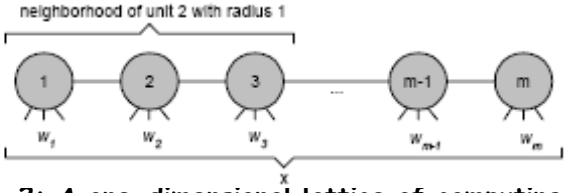

Fig. 2: A one-dimensional lattice of computing units.

The n-dimensional weight vectors $w_1, w_2, \ldots, w_m$ are used for the computation. The objective of the clustering for each unit is to learn the specialized pattern present on different regions of input space as shown in Fig. 2. When an input from such a region is fed into the network, the corresponding unit should compute the maximum excitation. SOM may distinctly reduce misclassification errors [12]. Kohonen's learning algorithm is used to guarantee that this effect is achieved. A Kohonen unit computes the Euclidian distance between an input x and its weight vector w. The complete description of Kohonen learning algorithm can be found in [2] and [3].

5. Genetics Algorithm

5.1 An overview

A genetic algorithm is an iterative procedure maintaining a population of structures that are candidate solutions to specific domain challenges [13]. During each temporal increment (called a generation), the structures in the current population are rated for their effectiveness as domain

solutions, and on the basis of these evaluations, a new population of candidate solutions is formed using specific genetic operators such as reproduction, crossover, and mutation. Genetic Algorithms (GAs) are search algorithms based on the mechanics of the natural selection process (biological evolution). The most basic concept is that the strong tend to adapt and survive while the weak tend to die out. That is, optimization is based on evolution, and the "Survival of the fittest" concept. GAs has the ability to create an initial population of feasible solutions, and then recombine them in a way to guide their search to only the most promising areas of the state space. Each feasible solution is encoded as a chromosome (string) also called a genotype, and each chromosome is given a measure of fitness via a fitness (evaluation or objective) function. The fitness of a chromosome determines its ability to survive and produce offspring. A finite population of chromosomes is maintained.
GAs use probabilistic rules to evolve a population from one generation to the next. The generations of the new solutions are developed by genetic recombination operators

Biased Reproduction: selecting the fittest to reproduce
Crossover: combining parent chromosomes to produce children chromosomes
Mutation: altering some genes in a chromosome. Crossover combines the "fittest" chromosomes and passes superior genes to the next generation. Mutation ensures the entire state-space will be searched, (given enough time) and can lead the population out of a local minima.
Determining the size of the population is a crucial factor. Choosing a population size too small increases the risk of converging prematurely to a local minimum, since the population does not have enough genetic material to sufficiently cover the problem space. A larger population has a greater chance of finding the global optimum at the expense of more CPU time. The population size remains constant from generation to generation.
Fitness Function drives the Population toward better solutions and is the most important part of the algorithm.

5.2. Genetic Algorithm Based BPN Network Training

Step1: Randomly generate an initial population of, say, P strings of length d: $S(0)=\{\mathbf{s}_1, ..., \mathbf{s}_P\} \subset \Omega$.
Step 2: Compute the fitness score $\mathbf{f}(\mathbf{s}_k)$ of each individual string $\mathbf{s}_k$ of the current population S(t).
Step 3: Generate an intermediate population [termed mating pool] by applying the selection operator.
Step 4: Generate S(t+1) by applying recombination operators (crossover and mutation) to the intermediate population.
Step 5: t:=t+1 and continue with Step 2 until some stopping criterion applies [in this case designate the best-so-far individual as the result of the GA]. The first step generates an initial population S(0), i.e. $S(0)=\{\mathbf{s}_1, ..., \mathbf{s}_P\} \subset \Omega$.In GA each member of S(0) is a string of length d that corresponds to the problem coding. S(0) is usually generated randomly, because it is not known a priori where the globally optimal strings in Ω are likely to be found. From this initial population, subsequent populations S(1), ..., S(t), ... will be computed by employing the three genetic operators of selection (reproduction), crossover and mutation. After calculating the relative fitness for all the strings in the current population S(t) (Step 2), selection is carried out and then strings in the current population are copied (i.e. duplicated) and placed in the intermediate population proportional to their fitness relative to other individuals in the population. After selection has been carried out the construction of the intermediate population is completed. Then crossover and mutation are applied to the intermediate population to create the next population S(t+1) (Step 4). Crossover and mutation provide a means of generating new sample points in Ω. While partially preserving distribution of strings across hyper planes which are observed in the intermediate population. Crossover is a recombination mechanism to explore new regions in Ω. The two new strings, called offspring, are formed by the juxtaposition of the first part of one parent and the last part of the other parent. Continue with Step 2 until some stopping criterion applies to find final population. Then final weights are determined from it for backpropagation algorithms.

5.3. Genetic Algorithm Based Backpropagation Neural Network Organizations

Hybrid approach offer strong advantage over either rule based or unaided neural network approaches [14]. Genetic algorithm based back propagation neural network offer good generalization abilities although it may be difficult to determine the optimal network configuration and network parameters. Genetic algorithms based backpropagation neural networks gives higher prediction accuracy in comparison of backpropagation neural networks [15].
The architecture of GA based neural network used is as follows:
1). Input layer with 2 nodes
2). One hidden layer with 2 nodes
3). Output layer with one node.

6. Experimental Results

The system has been developed and tested on Windows XP operating system .We have used Visual Basic and Microsoft Access as front end and back end tool. Simulation data was sourced from Indian National Stock Exchange (NSE).From the 2007 compendium of Top 500 Companies in India we selected 100 companies as per their Shareholding Pattern, P/E Ratio, Dividend Yield, Price/Book Value Ratio, Return on Net Worth (RONW), Annual growth in Sales, Annual growth

in Reported Profit After Tax, Return on Capital Employed (ROCE) and Adjusted Profit After Tax Margin (APATM), with these inputs SOM divides them into different classes. As the SOM are more relevant to the problem where stocks of different companies are to be compared on some common parameters and arranges in the form of different classes. Out of these classes we compared stocks belonging to the best class with genetic algorithm based backpropagation neural networks. Normalisation is a key part of data pre-processing for neural networks and should enable more accurate predictable rates. Normalised data is used for training genetic algorithm based backpropagation neural network. We normalize inputs so that input values lies between 0 and 1. Input attributes should be carefully selected to keep the dimensionality of input vectors relatively small [16]. As we know close rate and volume are primary quantitative factors for individual equities and from quantitative factors the key qualitative factor of the market sentiment can be derived. So we used close rate and volume of stocks as our input in genetic algorithm based backpropagation neural network and next stock rate as our target for training networks. GA-BPN is trained on data set of classified stocks for the years of 1-Aug- 2005 to 30-Jul- 2007 after training, testing is done on data set of 31-Jul- 2007 to 31-Oct- 2007.Classified stocks are compared using GA-BPN. Selected stock using GA-BPN is then compared with BSE-30 index for the period 12/11/2007 to 01/01/2008 i.e near about one and half month data.. We have found that our selected stock gives 30.17% more returns in comparison to BSE-30 Index as shown in fig. 3.

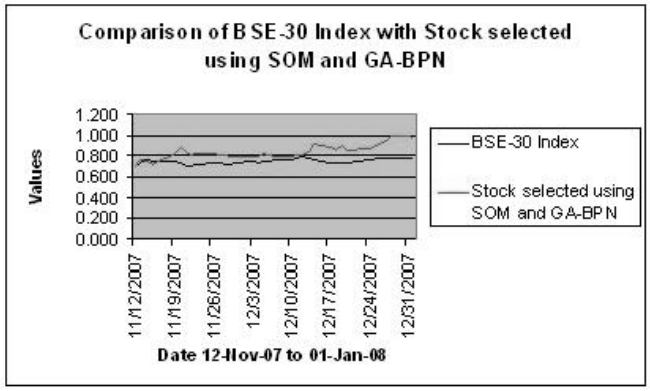

Fig. 3: Comparison chart between BSE-30 Index and stock selected using SOM and GA-BPN.

7. Conclusion

This paper compares the performances of the stock selected using self-organizing maps and genetic algorithm based backpropagation neural network with BSE-30 Index. The result shows that the performance of stock belonging to the best class among the classes generated by self-organizing maps and best prediction accuracy on test data using genetic algorithm based backpropagation neural network gives maximum return on investment. Stock selected using SOM and GA-BPN gives 30.17% more returns in comparison to BSE-30 Index

References

[1] Mizuno, H., Kosaka, M., Yajima, H. and Komoda N., "Application of Neural Network to Technical Analysis of Stock Market Prediction", *Studies in Informatic and Control*, 1998, Vol.7, No.3, pp.111-120.

[2] Haykin, Simon, "Neural Networks: A Comprehensive Foundation", *Macmillian College Publishing Company*, New York, 1994.

[3] Phillip D. Wasserman, Van Nostrand, "Neural Computing: Theory and Practice", *Van Nostrand Reinhold*, New York, 1989.

[4] Youngohc yoon and George swales, "Predicting stock price performance: a neural network approach", *IEEE publishing*, 1991.

[5] Shaikh A. Hamid, "Primer on using neural networks for forecasting market variables", in *proceedings of the conference at school of business*, Southern New Hampshire university, 2004.

[6] Ramon Lawrence, "Using Neural Networks to Forecast Stock Market Prices", *Course Project*, University of Manitoba Dec. 12, 1997.

[7] Monica Adya and Fred Collopy, "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation", *Journal of Forecasting*, 1998.

[8] Asif Ullah Khan et al., "Stock Rate Prediction Using Back Propagation Algorithm: Analyzing the prediction accuracy with different number of hidden layers", *Glow gift*, Bhopal, 2005.

[9] Juha Vesanto and Esa Alhoniemi, "Clustering of the Self-Organizing Map", *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, May 2000.

[10] Serrano, C., "Self Organizing Neural Networks for Financial Diagnosis", *Decision Support Systems Elsevier Science*, 1996, Vol 17, July, pp. 227-238.

[11] Tomas Eklund, "Assesing the feasibility of self organizing maps for data mining financial information", *ECIS*, June 6–8, 2002, Gdansk, Poland.

[12] Egidijus Merkevicius, Gintautas Garsva, "Forecasting of credit classes with the self organizing maps", *Informacines Technologies (ISSN 1392 – 124X)* Ir Valdymas, 2004, Nr.4(33).

[13] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning." *New York: Addison-Wesley*, 1989.

[14] K. Bergerson and D. Wunsch, "A commodity trading model based on a neural network- expert system hybrid", *IJCNN-91-Seattle International Joint Conference*, Volume I, Issue 8-14 Jul 1991, Page(s): 289 – 293.

[15] Asif Ullah Khan et al., " Comparisons of Stock Rates Prediction Accuracy using Different Technical Indicators with Backpropagation Neural Network and Genetic Algorithm Based Backpropagation Neural Network", pp. 575-580, 978-0-7695-3267-7/08 $25.00 © *2008 IEEE DOI 10.1109/ICETET.2008.59.*

[16] H. White, "Economic prediction using neural networks: The case of IBM daily stock returns", in *Neural Networks in Finance and Investing*, chapter18, pages 315–328, 1993.