Stock Market Prediction System
W.N.N De Silva¹, H.M Samaranayaka², T.R Singhara³, D.C.H Wijewardana⁴.
Sri Lanka Institute of Information Technology, Malabe, Sri Lanka,
¹nathashanirmani55, ²malmisamaranayaka, ³thamalisinghara@gmail.com; ⁴chamodiw@hotmail.com

Abstract-This paper presents two approaches for forecasting the stock price. They are graphical approach and a mathematical approach. Mathematical model is executed utilizing late strategies, for example, Markov model and Artificial Neural Network, where recorded information is been broke down for to foresee a worth. The graphical technique is chiefly utilized for template distinction. Pattern matching is the technique which will be utilized to perceive the example of the chart and the other piece of the graphical strategy is discovering the future prices according to the matching pattern. At last both graphical and scientific methodologies will be consolidated utilizing combination method as a part of request to get a more accurate result.

Keywords-Stock Market Prediction System, Hidden Markov Model, Artificial Neural Networks, Pattern Matching, Back Propagation Algorithm

I. INTRODUCTION

Stock exchange expectation is the demonstration of attempting to focus the future estimation of an organization stock exchanged on a money related trade. To foresee the eventual fate of stock costs is a typical issue for everybody as no one realizes that which stock is prone to give rates of profitability in future. Anyway a financial specialist can make some clever and instructed speculations about what is going to happen and what steps could be taken so as to minimize the danger to the least level. There were different strategies utilized within the past so as to anticipate the fate of stock costs [1].

This system consists of two main parts which are graphical approach and mathematical approach. The graphical approach includes analysis of graphs where system tries to find out certain patterns that are followed by stocks. The stock chart patterns are meaningful patterns formed by trend lines in the chart. These chart patterns are identified through years of trading experience by analysts. It is meaningful because different patterns indicate different future price movement trend. So the system can predict the chart pattern and can potentially predict the future price movements which enable to make a lot of money.

Mathematical module will be used to find an exact value for the prices. Artificial Neural Networks and Markov Model will be used to define algorithms. Finally graphical and mathematical module will be integrated using the Fusion technique to get the most efficient value.

The main objective of this research is to answer the question of how to develop a model that would accurately predict the future price of a particular stock. Stock market prediction system provides a clear idea about the company stock price movements. Proposed system would be less expensive compared to other existing systems and it is an affordable software system. Other than that it will provide a proper idea about stock price movements to the new investors and it will give a clear idea about a company's future value. Developing system with accuracy, efficiency, understandability, flexibility and satisfying other non-functional requirements is highly appreciated. The opportunity to expand knowledge in finance and investing, as development team had only little prior exposure to these fields may be another goal of this research.

II. METHODOLOGY

The research project stock market prediction system is based on Technical analysis and Machine learning approach. Two major components of this are Graphical and Mathematical components. Graphical component includes pattern matching and ad-hoc method. Mathematical component includes two algorithms implemented using Artificial Neural Network and Markov Model.

In Graphical approach, make predictions using two main methods which are pattern matching and an Ad-hoc method. These two methods are different from each other though they help to determine the future trend of a particular stock. In pattern matching, according to a given sets of predefined patterns understand the matching pattern of future. In Ad-hoc method, find the future trend using some given set of values for each pattern. Pattern matching is basically a chart pattern analysis [2] and a graphical representation of historical prices which helps to determine the future trend of a stock. For a better identification of short term trend in the stock market, the candle sticks are used to plot the graphs.
There are some special patterns for stock time series pattern matching shown in Fig. 1. Before starts the pattern matching process, first rank the predefined stock patterns according to the above Fig.1. However, there are some special requirements for the stock data pattern matching due to their specialized shape for frequently occurred patterns as shown in Fig.1. Therefore, it is required to consider the separate shift for each point in the patterns found in stock data for differentiating the specific patterns of stock when doing pattern similarity comparison. The matching process starts from the beginning of the predefined patterns and in the stock time series graph, matching process starts at the end. The ratio between first two points (the end points in the graph) and next two points in the graph are matched with ratio between same points in each and every pattern. Then according to that system matches the matching pattern and one or more patterns can be matched along with that. By using an accuracy level system predicts the most accurate matching pattern. Designing of ad-hoc method is basically based on historical data. Historical data is collected from various stocks in various companies by considering long time duration. Data sets are collected from the history and these data sets are stored in separate file in pattern wise. So each and every stock market chart pattern there is a separate file which contains data sets from historical data. Each data set contains seven data points. In order to get a coHMMon structure for each pattern, content of each file is analysed. Average values of each data point are considered as the coHMMon structure for a particular pattern. System has detected a coHMMon structure for each and every stock chart pattern using the proposed ad-hoc method. Average values of each data point are further analysed in order to find the minimum ratio between these seven data points. Stock chart patterns are separately identified by the values of corresponding minimum ratio as shown in Fig.3. There after stock chart patterns can be graphically represent by using the same ratio values. This is how template of each stock chart patterns are recognized by using the above mentioned ad-hoc method. After designing the template, system has to detect the matching part of the drawn graph against each template. According to the algorithm, drawn graph is matching from the end while templates are starting to match from the beginning. Prices of drawn graph and values of template points are compared with use of ratio value. Initially find the ratios of price changes between two adjacent prices. Then find the ratios of value changes between two adjacent points in the template. Finally compare the drawn graph’s ratio values against each and every template’s ratio values. If parts of the drawn graph’s ratio values are similar to a particular template, system will be continued the rest of the future graph according to the matching template. If drawn graph matches with more than one pattern, system will be continued the rest of the future graph according to the matching template separately. After identifying the future pattern of the graph, forthcoming prices are calculated according to a matching pattern.

There are some special patterns for stock time series pattern matching shown in Fig.1. Before starts the pattern matching process, first rank the predefined stock patterns according to the above Fig.1. However, there are some special requirements for the stock data pattern matching due to their specialized shape for frequently occurred patterns as shown in Fig.1. Therefore, it is required to consider the separate shift for each point in the patterns found in stock data for differentiating the specific patterns of stock when doing pattern similarity comparison. The matching process starts from the beginning of the predefined patterns and in the stock time series graph, matching process starts at the end. The ratio between first two points (the end points in the graph) and next two points in the graph are matched with ratio between same points in each and every pattern. Then according to that system matches the matching pattern and one or more patterns can be matched along with that. By using an accuracy level system predicts the most accurate matching pattern. Designing of ad-hoc method is basically based on historical data. Historical data is collected from various stocks in various companies by considering long time duration. Data sets are collected from the history and these data sets are stored in separate file in pattern wise. So each and every stock market chart pattern there is a separate file which contains data sets from historical data. Each data set contains seven data points. In order to get a coHMMon structure for each pattern, content of each file is analysed. Average values of each data point are considered as the coHMMon structure for a particular pattern. System has detected a coHMMon structure for each and every stock chart pattern using the proposed ad-hoc method. Average values of each data point are further analysed in order to find the minimum ratio between these seven data points. Stock chart patterns are separately identified by the values of corresponding minimum ratio as shown in Fig.3. There after stock chart patterns can be graphically represent by using the same ratio values. This is how template of each stock chart patterns are recognized by using the above mentioned ad-hoc method. After designing the template, system has to detect the matching part of the drawn graph against each template. According to the algorithm, drawn graph is matching from the end while templates are starting to match from the beginning. Prices of drawn graph and values of template points are compared with use of ratio value. Initially find the ratios of price changes between two adjacent prices. Then find the ratios of value changes between two adjacent points in the template. Finally compare the drawn graph’s ratio values against each and every template’s ratio values. If parts of the drawn graph’s ratio values are similar to a particular template, system will be continued the rest of the future graph according to the matching template. If drawn graph matches with more than one pattern, system will be continued the rest of the future graph according to the matching template separately. After identifying the future pattern of the graph, forthcoming prices are calculated according to a matching pattern.

<table>
<thead>
<tr>
<th>Pattern Name</th>
<th>Head &amp; Shoulder</th>
<th>Double Top</th>
<th>Triple Top</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern Figure</td>
<td><img src="image1" alt="Head &amp; Shoulder" /></td>
<td><img src="image2" alt="Double Top" /></td>
<td><img src="image3" alt="Triple Top" /></td>
</tr>
<tr>
<td>Position in the descending order</td>
<td>[6.24;15.37]</td>
<td>[6.41;25.7]</td>
<td>[6.15;25.37]</td>
</tr>
<tr>
<td>Rank</td>
<td>653.25451</td>
<td>452.5652</td>
<td>653.1542</td>
</tr>
</tbody>
</table>

Fig. 1. Eight Pattern Models and Ranks

<table>
<thead>
<tr>
<th>Pattern Name</th>
<th>Head &amp; Shoulder Reversed</th>
<th>Double Top Reversed</th>
<th>Triple Top Reversed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern Figure</td>
<td><img src="image1" alt="Head &amp; Shoulder Reversed" /></td>
<td><img src="image2" alt="Double Top Reversed" /></td>
<td><img src="image3" alt="Triple Top Reversed" /></td>
</tr>
<tr>
<td>Position in the descending order</td>
<td>[1.53;74.62]</td>
<td>[1.56;74.1]</td>
<td>[1.53;64.72]</td>
</tr>
<tr>
<td>Rank</td>
<td>153.5357</td>
<td>355.5552</td>
<td>155.5356</td>
</tr>
</tbody>
</table>

Fig. 2. Ranking Values in Ad-hoc Method

The Mathematical model is consist of two algorithms implemented based on ANN and HMM. In the implemented system ANN is trained to perform a specific task by adjusting these weights. The weights are continually adjusted by comparing the output of the network with the target until the output of the network ‘matches’ the target in the sense that the error function measuring the difference between the target and the output is minimised. Many pairs of input and output are used to train the network and this mode of adjustment is called, ‘supervised’ learning. There are several approaches for minimisation of the error function including standard optimisation techniques like conjugate gradient and Newton-type methods. However, the most popular algorithm in this context is the back propagation algorithm of Rumelhart et al [3].
This process will include two phases known as the training phase (learning) and the testing phase (prediction).

Training of the network categorize into two sections; incremental and batch. In case of incremental training, individual pairs are fed one at a time to the network. Output is compared with the target for each input and adjustments of the weights are made using a training algorithm, most often the back propagation algorithm. The next pair is then fed to the network after making the adjustment for the previous pair and the same process is repeated. Incremental training is sometimes referred to as ‘on line’ or ‘adaptive’ training. In case of batch training, all the pairs of input and output are fed to the network at the initial stage and the weights are adjusted.

Neural networks will be used to predict stock market prices because they are able to learn nonlinear mappings between inputs and outputs.

- The inputs that will be used are opening price, low and high during the day, volume and the previous closing price.
- The output was taken to be the closing price of the day [9].

ANNs generally have at least three layers. It contains the artificial neurons: input, hidden (or middle), and output. Artificial neurons are arranged in these layers. The input layer will take the inputs and passes to the middle layer. Even though it depends on the implementation, generally there occurs no data processing at the input layer. The middle (hidden) layer is where all the complexity resides and the computation is done [10].

In artificial neural system the information or data is circulated through the system and put away as weighted interconnections. The interconnections between simulated neurons are called weights. Artificial neurons live in layers. The data variables are mapped to the yield variables by squashing or changing by an exceptional capacity known as activation function. Back propagation algorithm will be used to train the network.

Back propagation is the process of back propagating errors through the system from the output layer towards the input layer during training. Back propagation is necessary because hidden units have no training target value that can be used, so they must be trained based on errors from previous layers. The output layer is the only layer which has a target value for which to compare. As the errors are back propagated through the nodes, the connection weights are changed. Training occurs until the errors in the weights are sufficiently small to be accepted [9].

Next HMM is a state machine for a system adherent to a Markov process with unobserved states. Each state is associated with observations. The stock market can also be seen in a similar manner. The underlying states are invisible to the investor. The company transitions between these underlying states based on company policy, decisions and economic conditions etc. The visible effect which reflects these is the value of the stock. Clearly, the HMM conforms well to this real life scenario.

Rabiner [6] tutorial explains the basics of HMM. It is characterized by the following:

1) Number of states in the model
2) Number of observation symbols
3) State transition probabilities
4) Observation emission probability distribution characterizes each state.
5) Initial state distribution

For the rest of this paper the following notations will be used regarding HMM.

\[ \lambda = (A, B, \pi) \] the overall HMM model.

To work with HMM, the following three fundamental questions should be resolved.

1. Given the model \( \lambda = \lambda \) how to compute \( P(O | \lambda) \), the probability of occurrence of the observation sequence \( O = O_1, O_2, \ldots, O_T \).

2. Given the observation sequence \( O \) and a model \( \lambda \), how to choose a state sequence \( q_1, q_2, \ldots, q_T \) that best explains the observations.

3. Given the observation sequence \( O \) and a space of models found by varying the model parameters \( A, B \), and \( \pi \), how to find the model that best explains the observed data.

There are established algorithms to solve the above questions. In this task forward-backward algorithm was used to compute the \( P(O | \lambda) \).

In stock market analysis the observations are not from a discrete space. In each state a series of observations could be received.
The observation would be consisting of following three of fractional values:

\[ O = \{ \text{open-low, high-low, close-low} \} \]

Hence HMM represents a Gaussian distribution. A continuous HMM model is trained on a particular stock, where emission probabilities are modelled as Gaussian Mixture Models conditioned on a particular state. It has 2 states (High, low) and there is a mixture of multi-Gaussianian distribution for each state [4][7].

There are a number of sources of stock and financial data on web including Google Finance and Yahoo Finance. In the specific case of historical prices, it seems that Yahoo Finance is performed well. Stock data from 2012 was loaded. In the application developed for this project, an interface is included to use when needed to load the data beforehand using the algorithms.

A likelihood value is calculated for each and every historical day through the evaluation method of Marcov chains. Above format of observation sequence (O) is taken as the parameter. If the difference between a particular likelihood day value and yesterday’s likelihood value is less than 0.5 those values are collected. The sum of those likelihood values is taken. Then it is divided by the number of days. Finally resulted average value is taken as future closing price.

III. RESULTS AND DISCUSSIONS

The SMPS is implemented to get accurate results of the stock market. Mathematical and graphical methods are planned to be combined by a fusion technique to get the results. To design the graphical approach, need two main methods which are pattern matching and ad-hoc method. Artificial Neural Networks and Markov Model are used to create algorithms. As per research findings mostly two techniques are not been combined. Most of the research uses only one algorithm. In graphical method stock price is analysed using the pre-defined patterns like double-top-bottom, triangles etc. Finally graphical and mathematical module will be integrated using the Fusion technique to get the most efficient value.

The system was tested on a number of stocks from different sectors of the economy for e.g. Automobiles, FMCG, Oil and Gas. The data used for development and testing of system was obtained from the website where “yahoo.com”. It provided the stock prices prevailing at any stock in any company. The data contains the stock prices on current day/week, the highest, open, close, lowest values and the volume of the stocks traded on that day/week. Data for the period April 2010 to April 2014 (one data point for each day), was used for testing the system. To test the results from the system, a Simulink model was developed to take inputs from available parameter values of various company stocks. The system then generates the suitable numerical output indicating the type of decision that will be suitable for the investor.

Fig.3 shows what are the matching patterns according to the template matching for a given stock and forecast the future values according to the pattern. According to the Fig.1 the matching pattern is given as Triple Top Reversed pattern and the forecast values are shown in the bottom of the interface. The values are forecasted according to the five points in the graph. As same as Fig.4 shows what are the matching patterns according to the ad-hoc method for a given stock and forecast the future values according to the pattern. According to the Figure2 the matching pattern is given as Triple Top pattern and the forecast values are shown in the bottom of the interface. The values are forecasted according to the five points in the graph.

In Mathematical model which is implemented using Artificial Neural Networks and Markov Model have huge number of days which is having the difference is less than 0.5. Hence it takes time to process the predicted price. When the amount of days which is used as historical data is increased the accuracy of the predicted price increases. Most of accurate predicted prices are retrieved when the historical days are limited to about 90-100 days. Following Table 1 shows the resulted predicted prices and actual prices.
<table>
<thead>
<tr>
<th>Date</th>
<th>Predicted price</th>
<th>Actual Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/20/2014</td>
<td>33.79</td>
<td>35.12</td>
</tr>
<tr>
<td>8/19/2014</td>
<td>33.77</td>
<td>35.48</td>
</tr>
<tr>
<td>8/18/2014</td>
<td>33.71</td>
<td>35.07</td>
</tr>
<tr>
<td>8/14/2014</td>
<td>33.72</td>
<td>35.59</td>
</tr>
<tr>
<td>8/13/2014</td>
<td>33.70</td>
<td>35.30</td>
</tr>
<tr>
<td>8/12/2014</td>
<td>33.68</td>
<td>35.12</td>
</tr>
<tr>
<td>8/11/2014</td>
<td>33.66</td>
<td>35.20</td>
</tr>
<tr>
<td>8/8/2014</td>
<td>33.65</td>
<td>35.17</td>
</tr>
<tr>
<td>8/7/2014</td>
<td>33.62</td>
<td>34.82</td>
</tr>
<tr>
<td>8/6/2014</td>
<td>33.60</td>
<td>35.04</td>
</tr>
<tr>
<td>8/5/2014</td>
<td>33.58</td>
<td>35.07</td>
</tr>
<tr>
<td>8/4/2014</td>
<td>33.57</td>
<td>35.33</td>
</tr>
<tr>
<td>8/1/2014</td>
<td>33.53</td>
<td>35.19</td>
</tr>
</tbody>
</table>

By developing Stock Market Prediction System it will greatly ease the people who are knee to find the future prices of different companies. The implemented system can get user inputs and produce valuable outputs satisfying the users need. The system can produce different types of charts which show the patterns of past activities and predicted future trends which can help evaluate future decisions.

Stock Market Prediction System is a web-based system, which will have the capability of facilitating the client until the system is temporary halted. Its unique option of using optimized algorithms and data fetching through the yahoo API gives effective contribution to the systems performance by greatly decreasing the computational time. However, the systems tolerance and performance will be definitely challenged according to the changes in the stock market.

**IV. CONCLUSION AND FUTURE WORK**

The SMPS has achieved the most of the goals according to the current situation, where research objectives been reached and the problems is been resolved.

The main intention of the Stock Market Prediction System was to provide stock market clients to predict the future prices of any given company in order to ease their work and to encourage for buy the stocks which is a major need for the growth of economy in the world as well as the country.

According to the research finding the existing systems are very expensive, so the goal of creating a less expensive system was reached according to the current situation. Stocks are exchanged in the worldwide but most of the people have less knowledge about the price movements other than the people who are highly involved in the industry. SMPS helps the company owners as well as the investors to have an idea about the future values and the lesser’s to identify the price movements. The research objectives of gathering data is been cleared by using yahoo historical API. Forecasting period is been defined as daily. More accurate techniques such as ANN, HMM, pattern matching etc. is been used to resolve the research problems.

As the future work of the pattern matching, system is going to designed to get the accuracy level of each matching pattern and select the matching pattern as the pattern which is having the most accuracy and further we planned to create two more algorithms using some other strong techniques to increase the accuracy of the system. Finally, integrate both mathematical and graphical methods to produce accurate results.

**REFERENCES**


