**STOCK MARKET TREND PREDICTION SYSTEM**

**Abstract**

Forecasting stock return is an important financial subject that has attracted researchers’ attention for many years. It involves an assumption that fundamental information publicly available in the past has some predictive relationships to the future stock returns. This study tries to help the investors in the stock market to decide the better timing for buying or selling stocks based on the knowledge extracted from the historical prices of such stocks. The automated computer programs using data mining and predictive technologies do a fare amount of trades in the markets.

Data mining is well founded on the theory that the historic data holds the essential memory for predicting the future direction. This technology is designed to help investors discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. Stock Market is an untidy place for predicting since there are no significant rules to estimate or predict the price of share in the stock market. Many methods like technical analysis, fundamental analysis, time series analysis and statistical analysis etc are all used to attempt to predict the price in the share market but none of these methods are proved as a consistently acceptable prediction tool.

Artificial Neural Network (ANN), a field of Artificial Intelligence (AI), is a popular way to identify unknown and hidden patterns in data which is suitable for stock market prediction. Neural networks, as an intelligent data mining method, have been used in many different challenging pattern recognition problems such as stock market prediction.

Contents

1. Introduction

1.1 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

1.2 Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2. Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3. Problem Statement

3.1 Problem. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.2 Execution Of Problem Statement . . . . . . . . . . . . . . . . . . . 9

4. Project Description

4.1 Architecture. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

4.2 Data Flow Diagram. . . . . . . . . . . . . . . . . . . . . . . . . . . . . .11

4.3 Sequence Diagram. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .12

4.4 Activity Diagram. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .13

5. Implementation Details

5.1Methodology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .14

6. Conclusion and Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

1. Introduction

The stock market is essentially a non-linear, non- parametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. To achieve those objectives, and according to some research used the techniques of fundamental analysis, where trading rules are developed based on the information associated with macroeconomics, industry, and company. Analyzing the company’s operations and the market in which the company is operating can do this. Consequently, the stock price can be predicted reasonably well. Most people believe that fundamental analysis is a good method only on a long-term basis. However, for short- and medium- term speculations, fundamental analysis is generally not suitable.

Some other research used the techniques of technical analysis , in which trading rules were developed based on the historical data of stock trading price and volume. Technical analysis refers to the various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid.

Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area . Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis.

Following the assumption of technical analysis that patterns exist in price data, it is possible in principle to use data mining techniques to discover these patterns in an automated manner. Once these patterns have been discovered, future prices can be predicted.

Today, the grand challenge of using a database is to generate useful rules from raw data in a database for users to make decisions, and these rules may be hidden deeply in the raw data of the database. Traditionally, the method of turning data into knowledge relies on manual analysis; this is becoming impractical in many domains as data volumes grow exponentially. The problem with predicting stock prices is that the volume of data is too large and huge. In this project we use one of the data mining methods which is the classification approach on the historical data available to try to help the investors to build their decision on **whether to buy or sell that stock in order to achieve profit.**

The main objective is to analyze the historical data available on stocks using artificial neuron network technique as one of the classification methods of data mining in order to help investors to know when to buy new stocks or to sell their stocks.

Analyzing stock price data over several years may involve a few hundreds or thousands of records, but these must be selected from millions.

* 1. Motivation

From the day stock was born, the movement of prediction has been the focus of interset for years since it can yield significant profits. There are several motivations for trying to predict stock market prices. The most basic of these is financial gain. Any system that can consistently pick winners and losers in the dynamic market place would make the owner of the system very wealthy. Thus, many individuals including researchers, investment professionals, and average investors are continually looking for this superior system which will yield them high returns. There is a second motivation in the research and financial communities. It has been proposed in the efficient market hypothesis (EMH) that markets are efficient in that opportunities for profit are discovered so quickly that they cease to be opportunities . The EMH effectively states that no system can continually beat the market because if this system becomes public, everyone will use it, thus negating its potential gain.

* 1. Objectives

The objective of prediction research has been largely beyond the capability of traditional AI research which has mainly focused on developing intelligent systems that are supposed to emulate human intelligence. With the development of artificial neural networks investors are hoping that the market mysteries can be unrevealed because networks have great capability in pattern recognition and machine learning problems such as classification and prediction. These days’ artificial neural networks are considered as a common data mining method in different fields like economy, business, industry, and science .

The application of artificial neural networks in prediction problems is very promising due to some of their special characteristics.

1. First, artificial neural networks can find the relationship between the input and output of the system even if this relationship might be very complicated because they are general function approximations. Consequently, artificial neural networks are well applied to the problems in which extracting the relationships among data is really difficult but on the other hand there exists a large enough training data sets.
2. Second, artificial neural networks have generalization ability meaning that after training they can recognize the new patterns even if they haven’t been in training set. Since in most of the pattern recognition problems predicting future events (unseen data) is based on previous data (training set), the application of artificial neural networks would be very beneficial.
3. Third, artificial neural networks have been claimed to be general function approximations. It is proved that an neural network can approximate any complex continuous function that enables us to learn any complicated relationship between the input and the output of the system.

2. Literature Review

In recent years, there have been a growing number of studies looking at the direction of movements of various kinds of neural network computing to traditional statistical methods of analysis. Both academic researchers and practitioners have made tremendous efforts to predict the future movements of stock market prices index or its return and devise ﬁnancial trading strategies to translate the forecasts into proﬁts .

Over the past two decades many important changes have taken place in the environment of financial markets. The development of powerful communication and trading facilities has enlarged the scope of selection for investors. Forecasting stock return is an important financial subject that has attracted researchers’ attention for many years. It involves an assumption that fundamental information publicly available in the past has some predictive relationships to the future stock returns . In order to be able to extract such relationships from the available data, data mining techniques are new techniques that can be used to extract the knowledge from this data.

For that reason, several researchers have focused on technical analysis and using advanced math and science. Extensive attention has been dedicated to the field of artificial intelligence and data mining techniques . A empirical study on building a stock buying/selling alert system using back propagation neural networks (BPNN) was made. The system was trained and tested with past price data from Hong Kong and Shanghai Banking Corporation Holdings over the period from January 2004 to December 2005. The empirical results showed that the implemented system was able to predict short-term price movement directions with accuracy about 74%.

El-Baky et al., proposed a new approach for fast forecasting of stock market prices. The proposed approach uses new high speed time delay neural networks (HSTDNNs). The authors used the MATLAB tool to simulate results to confirm the theoretical computations of the approach.

The effectiveness of the neural network models was examined and used for level estimation and classification. The results showed that the trading strategies guided by the neural network classification models generate higher profits under the same risk exposure than those suggested by other strategies.

3. Problem Statement

3.1 Problem

The purpose is to develop an artificial neuron network system for predicting stock market trends.

3.2 Execution Of Problem Statement

•Here firstly we collect the raw data i.e. historical stock data of various companies from respective company’s stock exchange web sites.

•Now, from the collected data we decide input and output factors which are use for future stock price prediction.

•The selected parameters are process using some technical indicators . This process data is needed for giving the input to network.

•This processed data is feed to model i.e. ANN which will provide us a prediction results.

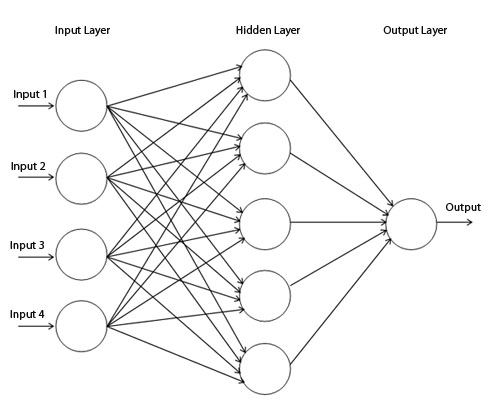
•The system will generates the suitable numerical output indicating the type of decision that will be suitable for the investor.

Along with that simultaneously we will validate the data using some statistical method and

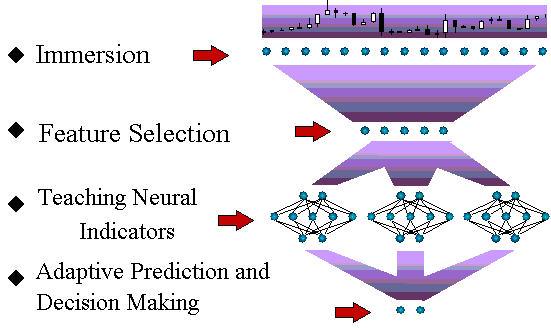
validate the data sets to get more appropriate results.

4. Project Description

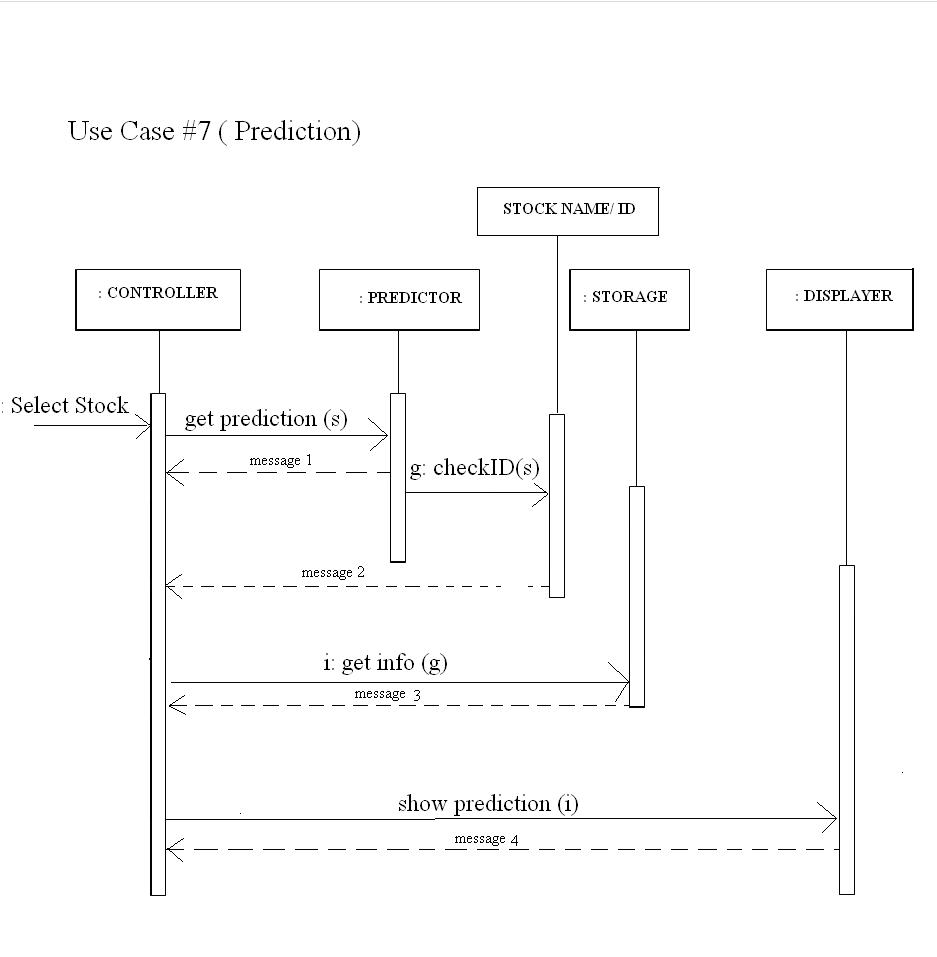
4.1 Architecture:



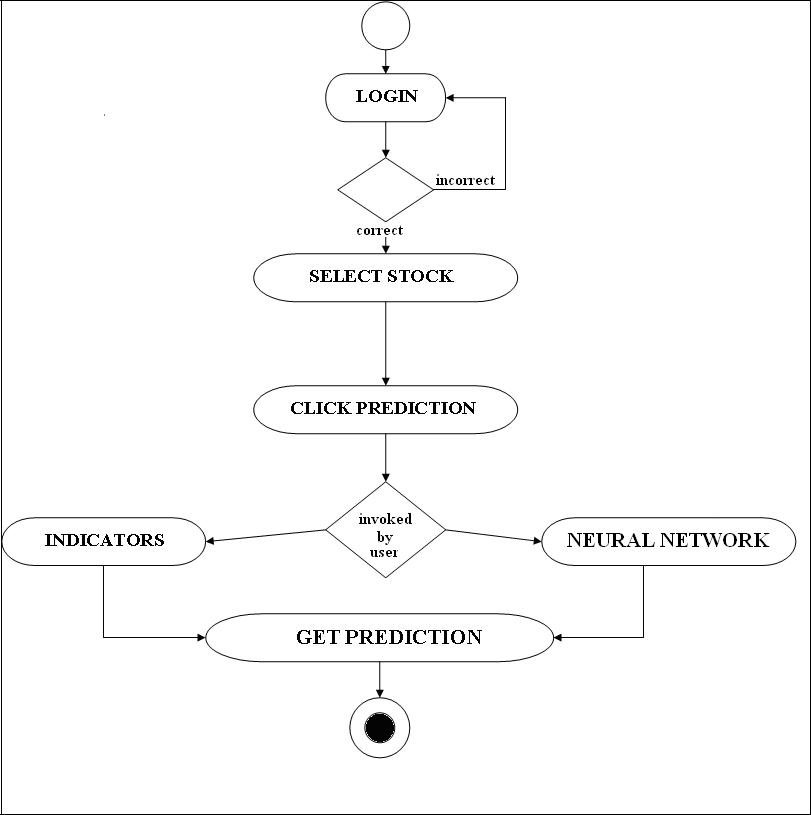
4.2 Data Flow Diagram:



4.3 Sequence Diagram:



4.4 Activity Diagram:



5. Implementation Details

5.1 Methodology

We have used Multilayer Feed forward Neural Network and such types of networks consist of input layer , one or more hidden layers and an output layer.

The model was generated in five steps :

a) Data Collection

b) Data pre-processing

c) Neural Network Creation and Training

d) Network Validation

1. Data Collection :

In order to train, validate and test the neural network, data is required and we collected historical data of various companies .

1. Data pre-processing :

The data must be prepared such that it covers the range of inputs for which the network is going to be used. Since the performance and reliability of the output from the neural network mainly depends on the quality of the data, therefore, the data must be pre-processed before it is fed to a neural network. First of all, we applied attribute relevance analysis on data so as to remove unwanted attributes from data and then the data is normalized . Since the input is in the normalized form, the output we get is also in the normalized form and hence, it must be de-normalized so as to have actual value.

In order to train the network, we divided the data into three subsets :

1. Training Data Set : This data set was used to train the network. The gradient was computed and biases and the weights of the connections between the neurons were adjusted accordingly. 2.Validation Data Set: This data was used to save the weights and biases at the minimum error and to avoid network over fitting data.

3. Testing Data Set : This data set was used to test the performance of the network.

1. Neural Network Creation and Training :

In this step neural network is created with two layers one hidden layer and one output layer .Of course, input layer is essential.

Artificial Neural Networks depend on the following parameters :

 Number of layers

 Number of neurons in input layer

 Number of neurons in hidden layer

 Learning rate

 Number of training iterations that are required to obtain the best result

 Training algorithm used

 Learning function used.

The network is created with some initial values of above mentioned network parameters. The network was trained using backpropagation algorithm with the aim to improve the network performance i.e. to reduce mean square error (mse). In this algorithm , the network is trained by repeatedly processing the training data set and comparing the network output with the actual output and reducing the error to the minimum possible. If the error between network output and the actual falls below the threshold value, then the training stops otherwise weights of the connections between various neurons are modified so as to reduce “mse”. The modifications are done in the opposite direction i.e. from output layer through each hidden layer down to the first hidden layer. Since the modifications in the weights of the connections are done in the backward direction so the name given is back propagation.

1. Network Validation :

After training the network output was validated using validation data so as to improve the network performance.

6. Conclusion and Summary

The project has reached its preliminary stage in which the Stock Market Dataset extraction, UI designing, and Programming Language selection has been completed . The next step of the project will be addition of more dataset, which will be done by end of December. After this project building will move to the business end of the project which will the classiﬁcation of data using improvised machine learning algorithm which will take around 3 months. Finally, project enhancement will be done by trying to include context-wise classiﬁcation, which is a research topic.