ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Technical Analysis of Stock Market Trends Using Trend Agents and Machine Learning

Abhijith P S, Atul B Kumar, Bharath M N, Nayak Ajay Ashok, V K Annapurna

Student, Department of Computer Science and Engineering, NIE, Mysuru, Karnataka,India.

Student, Department of Computer Science and Engineering, NIE, Mysuru, Karnataka, India.

Student, Department of Computer Science and Engineering, NIE, Mysuru, Karnataka, India.

Student, Department of Computer Science and Engineering, NIE, Mysuru, Karnataka, India.

Assosciate Professor, Department of Computer Science and Engineering, NIE, Mysuru, Karnataka,India.

ABSTRACT: Identifying the trend that the stock market will take, always support in taking better decisions in terms of buy, sell or hold the stock taken into consideration. Historical data is used to predict the trend of the stocks. The model developed uses a two layer approach that employs domain knowledge of technical analysis in first layer to guide the machine learning in the second layer. Based on number of tests using different price data, the prediction model successfully outperforms the market i.e. the National Stock Exchange (NSE).

KEYWORDS: technical analysis; trend agents; stochastic; stock; ADX; Moving Average Crossover; OHLC data.

I. INTRODUCTION

Forecasting the trend of future stock prices is a widely studied topic in many fields including trading, finance, statistics and computer science. The motivation for which is naturally to predict the direction of future prices such that stocks can be bought and sold at profitable positions. Professional traders typically use fundamental and/or technical analysis to analyse stocks and make investment decisions. Fundamental analysis is the traditional approach involving a study of company fundamentals such as revenues and expenses, market position, annual growth rates, and so on [3].Technical analysis, on the other hand, is solely based on the study of historical price fluctuations. Practitioners of technical analysis study price charts for price patterns and use price data in different calculations to forecast future price movements [4].The technical analysis paradigm is thus that there is an inherent correlation between price and company that can be used to determine when to enter and exit the market

In this paper, an attempt to identify and predict proper trend of stock market is being made using technical analysis and machine learning. For the purpose, a two layer approach is used. The first layer is implemented using set of technical indicators employing technical analysis on specific data subset. The generated values are then forwarding to machine learning algorithm to learn a classification model that aggregates and places the values derived from previous layer in context to each other.

Section III introduces the approach taken to generate profitable investment strategy. Section IV provides with the dataset used for the purpose, as well the methodology to use the dataset to test the profitability of the model developed. Section V introduces the technical indicators used in the model. Input to evolutionary tree as explained in section VI comes from the technical indicator outputs. Section VII provides with results when the model is run with the methodology explained in Section IV. Section VII is followed with concluding remarks over the model with few future enhancements that are possible with the model.

II. RELATED WORK

[1] Introduces to simplest and most popular trading rules which are used in stock trading, moving average and trading range break. It introduces to concept of buy, sell signal generation based on the above trading rules and compares the result with other models like random walk. It provides promising results for technical analysis over stock market. [2]

Introduces with different applications of technical stock market indicators and the factors that affect the technical indicators. [3] Provides with various technical analysis methods and it's applications for trading. The comparison of technical analysis and fundamental analysis as well as various trading methods that traders used based on results provided by technical analysis. [4] Provides with similar approach as of [3]. [5] Helps in understanding the behaviour of stock market prices and if such understanding of behaviour is of any help towards stock trading. [7] Provides with various technical indicators that are used by traders like moving average crossover (MAC) and average directional index (ADX) for identifying early entry and exit points in trading lifecycle. [8] is a comprehensive about decision tree learning and how evolutionary trees help in maintaining different phases of natural evolution like survival of the fittest, reproduction, mutation and recombination.

III. PROPOSED ALGORITHM

The approach taken towards generating signals (buy, sell or hold) for profitable trading consists of two distinct modules. The modules are developed such as to accommodate future modifications and additional functionalities if necessary. The first module is enough if human monitoring is involved as it provides with analysis on historical data which is the core of technical analysis.

The modules applied are:

A. *Indicator Generation:*

The module implements a knowledge intensive process that generates a set of discrete feature-values from the price data using domain knowledge from technical stock analysis. The generated features represent aspects of the data that are quintessential to stock price prediction. The process is essentially executing a coarse-grained analysis of the price data, filtering out the seemingly unimportant details and forwarding the seemingly important details in relation to stock price prediction. It serves as input to next prediction model however, can be used as individual decision support tool. The technical analysis is performed using various technical indicators based on Occam's razor application. The module consists of:

$$\text{Agent}((Ps), t) \rightarrow \{v1, v2 \ldots\}$$

Where Ps is the price history for some stock s and {v1, v2...} is a set of discrete feature values generated by the agent.

B. *Indicator Aggregation*:

The module employs machine learning to aggregate and place the generated features in context, creating a unified (and hopefully) profitable investment strategy.

$$\text{Invest}(A(Ps), t) \rightarrow \{\text{Buy, Sell, Hold}\}$$

In order to successfully apply the domain knowledge implemented in the feature generation module we need to aggregate and place the generated features in context of each other. That is, we need to create a classification model those classifies different combinations of the generated features according to the target function. .We consequently need a data structure that aggregates the different features and assigns classifications to different feature-value combinations. Decision tree classification is a framework that facilitates the preceding requirements.

IV. DATASET AND METHODOLOGY

The dataset consists of open, high, low and close (OHLC) as well as volume of trade done each day of each company on which the technical analysis is to be performed. The dataset which is used for testing the modules are taken from the National Stock Exchange (NSE) over wide range of dates. To test the model developed the dataset available is divided into two sets, training data and test data. The training data is used to train the model which then develops the decision tree for the next step. The next step consists of running the developed decision tree over test set which will generate signals as a decision support mechanism for the buy, sell or hold of stocks. For the purpose, the data used from NSE ranges about 1 year (245 days).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

V. TREND AGENTS

As mentioned in section IV, various trend agents are used for doing coarse grain technical analysis over the OHLC data. These indicators are part of first module and the dataset is directly fed to these indicators. Here P_s, $V_s$is the price history and volume history of stock *s* .*t* is time. The agents used for the generation of module are:

C. *Trend Agent:*

Formally, a trend in any given time series X=(x_1... x_n) is a prolonged period of time where xi ∈X rise or fall faster than their historical average (i.e., the prevailing direction of the data during some interval). A general approach used to identify the trending direction of a time series X is to first create a smoothed representation, G, that roughly describes X. Next, the first difference of G is calculated,

,

$$F(t) = G(t) – G(t−1),$$

And a trend is said to be established in intervals where the sign of the first difference is constant.
Trend Agent (Ps, t) → {Uptrend, Downtrend, No trend}

D. *Moving Average Crossover(MAC):*

Two simple moving average lines with different lengths, n and m where n < m, are plotted simultaneously and buy and sell signals are generated at points where the two moving averages intersect.

MAC n, m (t) = {
Buy if SMA n (t−1) < SMA m (t−1) and SMA n (t) > SMA m (t),
Sell if SMA n (t−1) > SMA m (t−1) and SMA n (t) < SMA m (t),
Hold otherwise.}

E. *Stochastic Agent:*

The Stochastic Agent generates buy and sell signals by a set of heuristics on values generated by a momentum indicator called the Stochastic Oscillator. Momentum indicators are a class of tools that are used measure trend momentum and to detect situations when a stock is overbought or oversold (Murphy, 1999)

$$\%K\, n(t) = 100\,(Ct – L\, n) / (H\, n – Ln)$$

Where Ct is the latest closing price in the period, Ln is the lowest low in the period, and H n is the highest high in the period.

$$\%D\ fast = SMA3\ (\%K)$$
$$\%D\ slow = SMA3\ (\%D\ fast)$$

The signal line is either a 3-day simple moving average of the %K line called the %D fast line, or another 3-day simple moving average of the %D fast line creating a smoother line called the %D slow

Stochastic (Ps, t) = {
Buy if %K (t−1) < %D (t−1) < 20 and %D (t) < %K (t) < 20,
Sell if %K (t−1) > %D (t−1) > 80 and %D (t) > %K (t) > 80, Hold otherwise. }

F. *Volume Agent:*

The Volume Agent employs a volume oscillator that use two simple moving averages with different lengths over the volume data. The volume agent is thus defined as:

Volume-Agent n, m (Vs, t) = {
Strong-Volume if SMA n (t) > SMA m (t),
Weak-Volume otherwise. }

G. *Average Directional Index (ADX) Agent:*

.The ADX measure the degree to which the price is trending, but provides no information on the direction of the trend. It should thus be very useful in relation to the Trend Agent.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

$$ADX\ n = 100 \times EMA\ (DX, n)$$

The ADX is calculated by smoothing another indicator, the directional movement index (DX), with an exponential moving average (EMA) over some time period n (typically 14 days). In general, low values (i.e., less than 25) of the ADX are interpreted as the price being in a weak trend and high values (i.e., above 25) are interpreted as the price being in a strong trend.

VI. EVOLUTIONARY DECISION TREES

In a decision tree the problem attributes are organized in a tree structure where each internal node is assigned to an attribute with outgoing arcs corresponding to the possible values of the attribute. Possible output values for the target function. In our case, we have a set of discrete feature-value pairs represented by the agents as mentioned in section V. In the general decision tree classification framework, a classification is reached by performing a sequence of tests on the problem attributes. Thus, the output classification depend on the tree structure and the arrangement of classification nodes. An Agent Decision Tree Learning (ADTL) algorithm is an evolutionary algorithm that evolves an agent decision tree for a particular stock through a process of artificial evolution.

A.Evolutionary Cycle:

The artificial evolution can be expressed as biological evolution using concepts of inheritance, natural selection, mutation etc. In evolutionary biology the population goes through a cycle of development (i.e., growing up), natural selection by surviving and being able to reproduce in some environment, and death. Reproduction also results in heredity as each parent transmits some of their genetic material (i.e., the blueprint of the organism, such as DNA in biological organisms) to the offspring. The genetic material of an organism is known as the genotype while its manifestation as an organism is known as the phenotype. Natural selection then ensures that only the strongest individuals are able to reproduce by favoring certain individual traits over others (i.e., survival of the fittest). Thus, inheritance works solely on the genotype, while natural selection works solely on the phenotype. As this process of recombination, heredity and natural selection continues over successive generations, the population gradually evolves and adapts to its environment. This same process, including concepts of a genotype, phenotype, population, diversity, heredity and selection is modeled in most applications of evolutionary algorithms. Thus, individuals in evolutionary algorithms are typically represented by a genotype and a phenotype that can be derived from the genotype.

B.Representation:

As our goal with the ADTL algorithm is to find an agent decision tree that maximizes potential profits, the phenotypes (i.e., candidate solutions) are decision trees. The genotype representation is typically implemented as a simple linear representation of the phenotype that facilitates the reproduction operators. However, as reproduction operators are easily specified on tree structures by separating and recombining sub-trees, the ADTL algorithm employs a direct mapping between genotype and phenotype. That is, the ADTL algorithm operates directly on a population of decision trees.

C.Fitness Testing and Portfolio Management:

The fitness criteria is important in terms of giving objective to the phenotypes. The fitness of the decision trees are evaluated based on the fitness function and only the strongest ones remain and take part in the decision making procedure. Once the tree is generated, the result evaluation is just tracing the tree based on the parameters generated by the individual indicators if their involvement is there in the final tree. All indicators may not be present in the final tree based on the evolution of tree and pruning performed on the tree during the evolution phase. Once the decision tree is ready, it is traced over test data with initial investment amount. As a simple strategy based on the decision tree, stocks are bought with the entire amount when a buyoccurs. Similarly all available stocks are sold when sell occurs. The end result is evaluated based on the factor of outperforming the market. Outperforming the market compares the change in the market value of the concerned stock to change in the investment made on the stock. When the change in investment value is better than the market value change, it concludes that the model outperforms the market. In terms of a bearish market, the loss incurred should be less as compared to the market value change. In terms of bullish market the profit achieved should be comparable to the market value change. .

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

D.Selection Strategy and Reproduction:

In the process of evolution, the trees generated are selected and removed based on different criteria. The purpose of such criteria is to select only the fittest trees and use those for reproduction of the next generation trees. This next generation of trees similar to natural phenomenon are well adapted to the changes caused in the environment, in this case the changes in stock market trend. The trees are first sorted based on the fitness value, in this case the trees yielding maximum profit are selected. The selected trees undergo through mutation and pruning. This ensures maintaining favourable traits and presence of each indicator only once along any path. The trees are evolved with initial count of 100 (random creation with pruning). These trees undergo selection resulting in 50 trees. These trees are then used to creation of next generation. The fittest tree based on fitness function among this generation is then used as the IDTL. The model serves well, however comes with a disadvantage that the parameters of the number of trees in each step needs to be set manually

VII. RESULTS

The methodology explained in section IV was tested over the dataset available from NSE. The dataset when passed over the first module of indicator generation results in coarse grain analysis over historical data. This module can be used as self-sustained tool for decision support over the stocks. However, this paper focuses on classifying the results obtained from the analysis and use this data for creation of a decision tree as explained in section VI. The model was tested over different data sets from NSE, providing with promising results. Most of the results were resulted in decrease in the initial investment, however this cannot be directly associated with the model design as the condition of the Indian Stock Market was unstable over the last few years. Thus, the factor of outperforming the market comes into picture. On an average the market values of stock was shifting in a downtrend losing about 14%. The model when tested over the test data had loss involved however with change of 3%. The results specify that the losses incurred are quite favourable due to the unstable market state. The results are however restricted to Indian Stock Market and availability of current data.

VIII. CONCLUSION

The results documented in the previous section shows that the developed prediction model using domain knowledge and machine learning keeps a track of the market trend and minimizes the loss incurred in a market in bearish state. The model however consists of pure technical analysis and is favourable for short term trading as long term trading requires data like company turnover, products, sales etc. which is unavailable to the public. The two layer reasoning is deemed successful based on section VI. For one, the indicator-oriented design of the first reasoning layer allows for easy integration with new analysis techniques and adaptability by simply adding and/or removing indicators from the population, facilitated by strictly defined indicators and layer interfaces. As each layer may serve as independent prediction models, we get the added ability to evaluate each layer without interference from other parts of the system. The second layer of reasoning is also highly adaptable as the fitness function employed by the IDTL algorithm can be easy extended with additional constraints or other measures of success. The omission of transaction costs in the portfolio simulation procedure is perhaps the most apparent flaw in the results documented. The most apparent problem with the prediction model is the inherent stochasticity in the model. The determination of training data plays a vital role in developing the model thus needs to be chosen properly. However there is no well-defined criteria in doing so. A few future enhancements are specified that may be used to mitigate the flaws present in the model and increase confidence in the prediction model.

One approach to mitigate the apparent risk and create more stable result is to extend the Feature Generation module with additional domain knowledge. As the current system uses historical prices as its sole basis for prediction, it seems natural to extend the system with agents that performs fundamental analysis. An agent that monitors the stock exchange for financial statements and classifies those as good or bad may thus be very useful. Stock prices are also influenced by expectations caused by news reports. A web-mining agent that tries to analyse if a stock has received positive or negative news coverage may thus provide the model with an important second source of information. Moreover, the global economy and stock markets influences each other in many ways, which might motivate an agent that monitors stock markets in other parts of the world. In order to increase confidence in the generated predictions, an explanation module for providing reasons for the predictions generated would be a highly regarded addition to the model.

For the risk-averse trader the model may seem too unstable in its present state. However, the results are satisfactory, more so than what was expected when the work was initiated, both in terms of the model architecture and the documented results.

REFERENCES

[1] Brock, J. Lakonishok, and B. LeBaron. "Simple technical trading rules and the stochastic properties of stock returns". The Journal Of Finance.
[2] Aseel Hmood, Mostafa Erfani, Iman Keivanloo, Juergen Rilling, "Applying technical stock market indicators to analyze and predict the evolvability of open source projects". IEEE Conference of Software Mainatinence.
[3] J. Murphy,"Technical analysis of the financial markets: a comprehensive guide to trading methods and applications".
[4] Turner, "A Beginner's Guide to Day Trading Online".
[5] Fama," The behavior of stock-market prices. The Journal of Business".
[6] Browne and W. Whitt," Portfolio choice and the bayesian kelly criterion. Advances in Applied Probability".
[7] J. I. Larsen. "An indicaor based approach towards predicting stock market trends."
[8] Thomas Grubinger [aut, cre], Achim Zeileis [aut], Karl-Peter Pfeiffer [aut]. " Evolutionary Learning Of Globally Optimal Trees."

BIOGRAPHY

Abhijith P.S , Atul B. Kumar, Bharath M. N., Nayak Ajay Ashokare students in The National Institute Of Engineering, Mysore,India. They are pursuing Batchelor Of Enginnering(At the time of publishing) in Computer Science Engineering.(Of batch 2016**). V. K. Annapurna** works as an Assosciate Professor in the same institution.