International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 08 | Nov-2015 www.irjet.net p-ISSN: 2395-0072

Survey: Stock Market Prediction Using Statistical Computational Methodologies and Artificial Neural Networks

A. Rao[1], S. Hule[2], H. Shaikh[3], E. Nirwan[4], Prof. P. M. Daflapurkar[5]

[1,2,3,4] Department of Computer Engineering, MMIT, Pune, Maharashtra, India
[5] Head of Department, Department of Computer Engineering, MMIT, Pune, Maharashtra, India

---

Abstract - ***This paper aims at outlining the research work in the field of Stock Market Prediction focusing mainly on the technical approaches that have been proposed and/or implemented with varying levels of accuracy and success rates. It surveys mainly two approaches – the Statistical Computational Approach and the Artificial Neural Networks Approach. It also describes the attempts that have gone in combining the two approaches in order to achieve higher accuracy predictions.***

Key Words: *Artificial Neural Networks, Genetic Algorithm, Back Propagation, Statistical Computation, Autoregressive Integrated Moving Average*

1. INTRODUCTION

Stock prices are considered to be chaotic and unpredictable. The Efficient Market Hypothesis deems stock prices to follow the Random Walk Model. This very problem is challenged by Stock Market Prediction Models. Predicting the future stock prices of financial commodities or forecasting the upcoming stock market trends can enable the investors to garner profit from their trading by taking calculated risks based on reliable trading strategies. There are various number of approaches based on which a prediction model can be developed and implemented. They can broadly be classified as Fundamental Analysis, Technical Analysis and Machine Learning Analysis. Fundamental Analysis is governed by the school of thought that the market may initially miscalculate the price of a tradable financial asset but the right price will be reached eventually. Thereby, one can profit from this miscalculation by accumulating such assets and releasing them when the trends are favorable. Technical Analysis, on the other hand, deals with finding the probable stock movements of an asset by taking its price history into account. This is achieved using time-series analysis and other statistical computational techniques. It aims at identifying noticeable trends when the parameters are known and there exists a relationship between the elements that define the system. However, stock trends are also affected by various exterior elements whose effect on the stock movement may go unnoticed by conventional statistical means leading to prediction of inaccurate trends. That is where analysis using Machine Learning comes into play. Artificial Intelligence Classifiers can be used to train a prediction model to identify trends in a chaotic environment that is not governed by any quantifiable rules. The most popular among these three approaches, based on their success rates and prediction accuracy, are the approaches using Statistical Computational Methodologies and Artificial Neural Networks (ANN). It has been observed that these two approaches can be integrated to develop a hybrid prediction model in order to achieve higher prediction accuracy levels. This paper surveys the relevant research that has been done or is going on currently with respect to these two approaches and their integrated applications to predict stock prices.

2. LITERATURE SURVEY

2.1 Statistical Computational Approach

Debadrita Banerjee's Stock Prediction Model is based on the statistical computational approach that utilizes Autoregressive Integrated Moving Average (ARIMA) [1]. The author has analyzed month-wise data that has been obtained from the S&P Bombay Stock Exchange Sensitive Index. The dataset spans across a duration of six consecutive years from 2007 to 2012. Based on the analysis, the author has tried to predict the unobserved sensex indices. Durbin-Watson Test is applied on the accumulated dataset to determine whether Regression Analysis is required or Time-Series Analysis is required to be performed. Once the time-series characteristic of the dataset is ensured, Auto-Correlation and Partial Auto-Correlation are calculated to obtain a basic idea of the optimal stock prediction model. The most suitable ARIMA model is shortlisted by trying out different combinations of the number of auto-regressive terms, non-seasonal differences and lagged forecast errors in the prediction equation. Quantitative evaluation of this model is carried out for measuring the prediction accuracy.

2.2 Artificial Neural Networks Approach

O. Abdalla, A. Elfaki, and Y. Murtadha have introduced a new methodology for the optimization of ANN parameters [2]. Their method introduces a process of training the ANN which is effective and also less human-dependent. Based on their survey of the related work in this field, they have addressed ANN designing problems with the help of a Genetic Algorithm. They have gathered datasets from two sources. The first is the PETRONAS Penapisan (Melaka) Sdn Bhd from January, 2007 to February, 2007 while the other dataset is the XOR standard problem dataset. Data collection is followed by the pretreatment and analysis of the dataset during which outlier errors and invalid observations are removed using MATLAB functions. Then the preparation and scaling of the dataset is carried out. Their ANN Development Model utilizes the combination of ANN with Genetic Algorithm from which the latter is used to obtain the optimal ANN design and training parameters. The parameters represented using Genetic Algorithm Encoding are namely, network architecture, activation function, training algorithm, learning rate, momentum rate and number of epochs. The population size, mutation rate, and the crossover rate of these parameters are set according to predefined specifications initially to obtain population of elements based on which the model carries out its prediction function. Further populations are generated using repopulation and recombination operators. The fitness of each population is evaluated. Evaluation continues until a termination criterion is satisfied.

B. Chauhan, U. Bidave, A. Gangathade, and S. Kale have dealt with the working of Back Propagation algorithm to train an ANN by outlining its control-flow intricacies with the help of a mathematical model [3]. They have also described the relevance of ANN in the field of Data Mining and Machine Learning by emphasizing its pattern-recognition capabilities in seemingly unpredictable environments with no perceivable order.

D. Kumar and S. Murugan discuss the integration of time-series data with ANN [4]. They first describe their literature survey in the field of ANN and stock market prediction. They proceed by detailing the one-step-ahead forecasting strategies. Their prediction model is designed for single step-ahead short term predictions. Their prediction model is based on a feed-forward ANN with back propagation. The system's learning function utilizes Gradient Descent algorithm. The Error Back Propagation learning algorithm, which is a gradient descent method, is explained step-wise. The hidden layer and the output layer are deployed with sigmoid function as an activation function. The datasets used are explained by the two types of indices, NIFTY and SENSEX, which are described in relation to their corresponding stock exchanges, National Stock Exchange and Bombay Stock Exchange, respectively. Performance analysis is done on the basis of two complex time-series, BSE100 and NIFTY MIDCAP50. There are five experiments each carried out for the two time-series datasets and the performance is charted out with the help of Regression Analysis. Performance measurement by statistical approach is derived by analyzing the prediction performance in terms of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Percentage Mean Absolute Deviation (PMAD), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Performance evaluation and graphical representation is achieved using MATLAB. They conclude by giving a comparative study of the results.

K. Abhishek, A. Khairwa, T. Pratap, and S. Prakash have made use of an ANN that utilizes feed-forward architecture to predict share market prices [5]. Their prediction model was trained using a dataset of Microsoft Corporation from January 1, 2011 to December 31, 2011. Their method of implementation involves a two-step procedure. In the first step, the ANN undergoes a training process where a genetic algorithm teaches the model to identify the trends. The genetic algorithm used here is the Levenberg-Marquardt algorithm that is an extension of the quasi-Newton method of Standard Back Propagation. The advantage of this approach is that the comparatively complex computation of Hessian Matrix is approximated owing to the feed-forward architecture. The calculation of the gradient depends on the Jacobian Matrix whose computation is carried out using back propagation. The second step involves obtaining results experimentally by applying the trained ANN prediction model to the dataset.

Z. Khan, T. Alin, and Md. A. Hussain have developed a Forecasting Model for the Bangladesh Stock Exchange [6]. They have highlighted the challenges encountered in the prediction of stock price and have described the various types of inputs that can affect the stock. Their system architecture is based on a multilayer feed-forward network. They have surveyed the different approaches that are typically used for modeling a stock market prediction system. They have explained how a back propagation algorithm can be used in the training phase of an ANN to effectively train the hidden units by the backward propagation of errors from the previous layers. They have targeted five inputs namely, General Index (GI), P/E ratio, Net Asset Value (NAV), Earnings per Share (EPS) and volume. These are the inputs that are used to train the prediction model in the training phase and also to derive the predictions in the test phase. In the training phase, the inputs are first normalized then fed to the network to derive weights based on acceptable errors. These weights help the network to determine the relevancy of the inputs. The trained network is then applied to the test dataset to predict the future stock

values. Using this architectural setup, they have predicted the stock values of ACI, a pharmaceutical company, for 8 days of November, 2010 based on the company's historical data.

Ramnik Arora has developed a feasible trading strategy using ANN that determines long term pricing relationships between stocks, thereby, profiting from their divergence from these relationships [7]. He has surveyed the research works that have been carried out in the domains of ANN and time-series econometrics. His work is based on the hypothesis that the seemingly complex patterns of stock movement can be learnt within a short period of time using ANN. Also, it is assumed that all the non-specific macro-economic news can be discounted and that, in their absence, relations can be derived between the different pricing elements of the market. He has taken into account the Indian Construction sector and the U.S. Technology sector. The end of day dataset of Unitech, which is a group of Indian Construction and Contracting companies, was sourced from Yahoo Finance. Intra-day data of Yahoo, Amazon and other U.S. Technology companies was also targeted. These datasets were adjusted and intersected to create 640 data vectors as inputs for the purpose of training and testing the ANN. Special care was taken in choosing the input variables so as not to include any lagged variables. His prediction model network incorporates 3 hidden layers to tackle the price predictions for the intra-day data. Transfer functions, with 7 neurons at the first hidden layer and 5 neurons each for the remaining hidden layers, are deployed. On the other hand, the price prediction of the end-of-day data is handled using a 2-hidden-layers network. Transfer functions are deployed at the hidden layers with 7 neurons and 5 neurons respectively. Judging from the results of his trading strategy, which mirrors statistical arbitrage, it can be concluded that the use of intra-day data is beneficial over the use of end-of-day data, although, further research is advised by the author on a dataset spread over the entire market rather than concentrating on particular sectors.

2.3 Hybrid Approach

B. Qian and K. Rasheed have studied the Dow Jones Industrial Average (DJIA) index [9]. They have used the Hurst Exponent to identify a date period with a capacity for high predictability. They have employed Rescaled Range Analysis to estimate the Hurst Exponent for a date period of 1024 trading days. By doing so, they have obtained a Dow Jones daily return time-series using which they have further generated training patterns for their stock prediction model. They have attempted to reconstruct the dynamical system using time-delay embedding vectors by heuristics from chaos theory to calculate the embedding dimension and the time-delay. They have used a multi-classifier approach for prediction. The first classifier used is a feed-forward one-hidden-layer network. The learning algorithm used to train the model is the Levenberg-Marquardt algorithm. The sigmoid transfer function is deployed at the hidden layer whereas the output layer is deployed with a linear transfer function. The instance-based learning algorithm, k-nearest neighbor algorithm, is used as the second classifier to assign unknown cases to the category of k most-similar cases. As the third classifier, CART algorithm is utilized to create decision trees. The nodes of these decision trees are split using Gini diversity index while cross-validation is used to prune the decision trees. The results from the above-mentioned experimental setup validate the need for the combination of multiple classifiers to achieve accurate predictions.

3. CONCLUSIONS

We have comprehensively examined the two significant stock market prediction approaches, the Statistical Computational Approach and the ANN Approach, that are used to tackle the dynamic complexity of stock market prediction arising due to a non-linear relationship between its market parameters and closing price. Our survey shows the efficacy of ANN-based stock market prediction models. It also shows the advantage of combining ANN with statistical computational methodologies to acquire noticeable improvements in the prediction accuracy.

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our Project Guide, Prof. P. M. Daflapurkar, whose constant support and expertise in the domain aided us in completing this work.

REFERENCES

[1] D. Banerjee, "Forecasting of Indian Stock Market Using Time-Series ARIMA Model", In: Proceedings of the 2nd International Conference on Business and Information Management, IEEE, 2014, pp.131–135.

[2] O. Abdalla, A. Elfaki, and Y. Murtadha, "Optimizing the Multilayer Feed-Forward Artificial Neural Networks Architecture and Training Parameters using Genetic Algorithm", International Journal of Computer Applications, Vol.96, Issue-10, 2014, pp.42–48.

[3] B. Chauhan, U. Bidave, A. Gangathade, and S. Kale, "Stock Market Prediction Using Artificial Neural Networks", International Journal of Computer Science and Information Technologies, 2014, pp.204–207.

[4] D. Kumar, and S. Murugan, "Performance Analysis of Indian Stock Market Index Using Neural Network

Time-Series Model", In: Proceedings of the International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE, 2013, pp.72–78.

[5] K. Abhishek, A. Khairwa, T. Pratap, and S. Prakash, "A Stock Market Prediction Model Using Artificial Neural Network", In: Proceedings of the ICCCNT, IEEE-20180, 2012.

[6] Z. Khan, T. Alin, and Md. A. Hussain, "Price Prediction of Share Market Using Artificial Neural Network (ANN)", International Journal of Computer Applications, Vol.22, Issue-2, 2011, pp.42–47.

[7] R. Arora, "Artificial Neural Networks for Forecasting Stock Price", Indian Institute of Technology, Kanpur (Study), 2008, pp.1–15.

[8] Z. Xiaotian, Hong Wang, Li Xu, and L. Huaizu, "Predicting Stock Index Increments by Neural Networks: The Role of Trading Volume under Different Horizons", Expert Systems with Applications, Vol.34, Issue-4, 2008, pp.3043-3054.

[9] B. Qian, and K. Rasheed, "Stock Market Prediction with Multiple Classifiers", Springer Science + Business Media, LLC, 2006, pp.25–33.

[10] Yevgeniy Bodyanskiy, and Sergiy Popov, "Neural Network Approach to Forecasting of Quasi-Periodic Financial Time-Series", European Journal of Operational Research, 2006, pp.1357–1366.

[11] K. K. Lai, L. Yu, S. Y. Wang, and C. X. Zhou, "Neural-Network-Based Meta-Modeling for Financial Time-Series Forecasting", In: Proceedings of the 9th Joint Conference on Information Sciences, Atlantis Press, Paris, 2006, pp.172–175.

[12] B. Qian, and K. Rasheed, "Hurst Exponent and Financial Market Predictability", In: Proceedings of the 2nd International Conference on Financial Engineering and Applications, Cambridge, MA, USA, 2004, pp.203–209.

[13] K. Okyay, "Artificial Neural Networks and Neural Information", Springer, ISBN 3540404082, 2003.

[14] J. V. Hansen, J. B. McDonald, and R. D. Nelson, "Time-Series Prediction with Genetic Algorithm Designed Neural Networks: An Empirical Comparison with Modern Statistical Models", Computational Intelligence, 2002, pp.171–184.

[15] A. Kanas, "Neural Network Linear Forecasts for Stock Returns", International Journal of Finance and Economics, 2001, pp.245–254.

[16] J. S. Zirilli, "Financial Prediction Using Neural Networks", International Thomson Computer Press, UK, 1997.

[17] D. J. E. Baestaens, W. M. van den Bergh, and H. Vaudrey, "Market Inefficiencies, Technical Trading and Neural Networks", In: Dunis C (ed) Forecasting Financial Markets, Financial Economics and Quantitative Analysis, John Wiley & Sons, Chichester, England, 1996, pp.254–260.

[18] H. S. Stern, "Neural Networks in Applied Statistics", Technometrics, Vol.38, Issue-3, 1996, pp.205–14.

[19] E. Gately, "Neural Networks for Financial Forecasting", Wiley, New York, 1996.

[20] A. Refenes, "Neural Networks in the Capital Markets", Wiley, New York, 1995.

[21] Cheng, B. and D.M. Titterington (1994), "Neural Networks: A Review from a Statistical Perspective", Statistical Science, Vol.9, Issue-1, pp.2–30.

BIOGRAPHIES

A. Rao is pursuing Bachelor of Engineering (Computer Engg.) from MMIT, Pune, affiliated to Savitribai Phule Pune University (formerly University of Pune).

S. Hule is pursuing Bachelor of Engineering (Computer Engg.) from MMIT, Pune, affiliated to Savitribai Phule Pune University (formerly University of Pune).

H. Shaikh is pursuing Bachelor of Engineering (Computer Engg.) from MMIT, Pune, affiliated to Savitribai Phule Pune University (formerly University of Pune).

E. Nirwan is pursuing Bachelor of Engineering (Computer Engg.) from MMIT, Pune, affiliated to Savitribai Phule Pune University (formerly University of Pune).

Prof. P. M. Daflapurkar
M.E. (CSE), Ph. D. (Registered)
She is the Head of Department for Computer Engineering in MMIT, Pune with a total experience of 19 years. Her areas of interest include Artificial Intelligence and Machine Learning.