2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

PREDICTING STOCK PRICE FOR A LISTED COMPANY

N. SUKEERTHI[1], B.SAINATH[2], M.PRANATHI[3], MR.K. RANJITH REDDY[4], DR. B. KRISHNA[5]

Student, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India[1,2,3]

Professor, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India [4,5]

Abstract:

A stock market is considered as one of the highly complex systems, which consists of many components whose prices move up and down without having a clear pattern. The complex nature of a stock market challenges us on making a reliable prediction of its future movements. In this paper, we aim at building a new method to forecast the future movements of Standard & Poor's 500 Index (S&P 500) by constructing time-series complex networks of S&P 500 underlying companies by connecting them with links whose weights are given by the mutual information of 60-minute price movements of the pairs of the companies with the consecutive 5,340 minutes price records. We showed that the changes in the strength distributions of the networks provide an important information on the network's future movements. We built several metrics using the strength distributions and network measurements such as centrality, and we combined the best two predictors by performing a linear combination. We found that the combined predictor and the changes in S&P 500 show a quadratic relationship, and it allows us to predict the amplitude of the one step future change in S&P 500. The result showed significant fluctuations in S&P 500 Index when the combined predictor was high. In terms of making the actual index predictions, we built ARIMA models. We found that adding the network measurements into the ARIMA models improves the model accuracy.

Keywords: ARIMA models, standard and poor 500 index,predictor, accuracy

1.Introduction:

A stock market is considered as one of the highly complex systems, which consists of many components whose prices move up and down without having a clear pattern. The complex nature of a stock market challenges us on making a reliable prediction of its future movements. In this paper, we aim at building a new method to forecast the future movements of Standard & Poor's 500 Index (S&P 500) by constructing time-series complex networks of S&P 500 underlying companies by connecting them with links.

2.Literature Review:

Review of literature on speech recognition systems genuinely demands the very first attention towards the discovery of

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Alexander Graham Bell about the process of converting sound waves into electrical impulses and the first speech recognition system developed by Davis et for recognizing telephone quality digits spoken at normal speech rate. This effort for automatic recognition of speech was basically centered on the building up of an electronic circuit for recognizing ten digits of telephone quality. Spoken utterances were analyzed to get a 2-dimensional plot of formant 1 vs formant 2. For pattern matching, a circuit was designed for determining the highest relative correlation coefficient between a set of new incoming data and each of the reference digit patterns. It was also observed that circuit adjustment helps the recognition system to perform well for the speech of different speakers.An indication circuit was built to display the recognized spoken digit. The approaches to speechrecognition, evolved thereafter, had a major stress on finding speech sounds and providing appropriatelabels to these sounds. Various approaches and types of speech recognition systems came intoexistence in last five decades gradually. This evolution has lead to a remarkable impact on thedevelopment of speech recognition systems for various languages worldwide. Automatic speechrecognition has been viewed as successive transformations of acoustic micro structure of speech signalinto its implicit phonetic macro-structure. In other words, a speech recognition system is a speech-toytext conversion wherein the output of the system displays text corresponding to the recognized speech.Languages, on which so far automatic speech recognition systems have been developed, are just afraction of total around 7300 existing languages. Russian, Portuguese, Chinese, Vietnamese, Japan,Spanish,

Filipino, Arabic, English, Bengali, Tamil, Malayalam, Sinhala, Hindi are prominent among them. English is the language for which maximum work for recognition isdone.

Existing System:

There have been many studies and developments on predicting stock market movements using many different approaches including deep learning algorithms with neural networks. Having machines learn huge sets of data such as historical stock prices, trading volumes,accounting performances, fundamental features of the stocks, and even the weather, and produce the future values of stocks or index is one big branch of stock market forecasting methods. It utilizes many learning, regression, classification, neural networks algorithms such as support vector machine, random forest, logistic regression, naive Bayes, and reccurent neural networks, and tries to make accurate predictions by adjusting itself according to the market changes. Another popular method is to use natural language processing techniques that let machines extract and understand information written and spoken in human languages, and try to capture stock market sentiments for making investment decisions based on the mood or the sentiments of the stock market. Traditional finance and modern financial engineering also attempt to forecast the stock market using the fundamental and technical analysis. While the fundamental analysis is interested in evaluating the intrinsic values of the stocks based on companies' performances and the economic status, technical analysis focuses on the price and volume dynamics, and tries to capture the investing timing by developing technical indicators. Another study constructed a role-based trading network for each company characterizing the daily trading relationship among its investors with transaction data. Particularly, nodes are traders involved in the transactions of a stock, and for each transaction between two traders, there is a link from the seller to the buyer. By categorizing the nodes into three types (Hub, Periphery, and Connector) according to the node's connectedness, this study created 9 different link types, and found that the time-series of fraction of the link type P- H and C-H have a predictive power with the maximum accuracy of 69.2% [9]. Network science has been used and developed for many different fields. However, a few studies were conducted in terms of financial market time-series forecasting. Also, the previous studies did not show that whether the network analysis helps improve the performances of financial market time-series forecasting models.

Proposed System:

This study aims at building a reliable stock market prediction model based on the stock market networks analysis, and prove that the network measurements indeed improve the ARIMA models.We computed eigenvector and betweenness centralities of the nodes in our S&P 500 networks, and formed metrics using their mean, median and maximum values. In the same sense, we also formed metrics using network modularity. Modularity is a clustering measure that is to find the community structure of a network. A high modularity means that there are more links within a specific group in a network than when links are randomly distributed among groups in the network. For each metric we built, we tested their prediction performance against the actual changes, squares of the changes and absolute values of the changes in S&P 500 index. The second and third-order of polynomial regression and a simple linear regressions are used for performing the fitting tests.

3.System Requirements:

The hardware requirements used in developing this project are a system which contains 15 processor 2.23GHz , hard disk with 300GB which is used to store the data and Ram with 16GB which is a type of computer memory that can be accessed randomly.

The software requirements used in developing this project are an operating system which can be either Windows/Ubuntu, Language used during implementation is python,Front end is python and the back end is PostgreSql which is a part of a computer or application that is not directly acccessed by the used, typically responsible for storing and manuplating the data.

International Journal For Recent Developments in Science & Technology

2581-4575

Crossref

A Peer Reviewed Research Journal

4.Architecture:

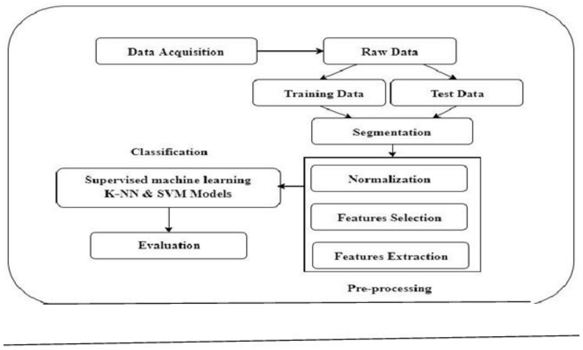

Figure 4.1

Firstly,the data is acquired as a dataset and in the first step we usually take this raw data and then go through the data to remove the inconsistencies in the data and make it error free. Then we need to seperate the data for training and testing the model. Usually we use 90% data for training and the remaining 10% is used for testing. This may subject to change based on the acceptance criteria for the accuracy of the model. Then dataset is segmented so as to make small portions which can be fed to the model for testing. The next phase is Preprocessing phase which consists of three other subphases.They are Normalization,Feature selection and Feature extraction. In Normalization we remove the repetitive data from the dataset which as it increases the accuracy of the model. Feature selection is one of the important aspects in datamining. Its necessity is felt due to very high dimensionality of data sets and growing computational methodologies of the target problems.Data mining aids in storing huge data and these data is full of noise i.e. redundant and irrelevant features. Feature selection is the pre -processing step where the noise is filtered, resulting in reducing the dimensionality of the data set and aids in creating computationally effective models with less time and cost.feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non- redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.When the input data to an algorithm is too large to be processed and it is suspected to be redundant then it can be transformed into a reduced set of features. Determining a subset of the initial features is called feature selection.Then we use the supervised machine learning algorithms like KNN and SVM models for classification using machine learning which is then used for evaluation of realtime data.

5.Working:

First, we acquire the raw data set of stock price records.

Second, we pre-process the data set with an appropriate imputation method to deal with missing data.

Third, we compute the pairwise mutual information of the stocks in S&P 500. Fourth, we construct networks using the mutual information as link weights. Fifth, we compute the node strength distribution of each network.

Sixth, we build the several predictors using the strength distribution, network centrality and modularity.

Seventh, we build linearly combined predictors with the two best performing metrics that maximizes the correlation between the metric and the changes in S&P 500.

Eighth, we built ARIMA models to predict the actual S&P 500 index. Lastly, we show whether or not the network measurements help improve the ARIMA models.

6.Conclusion:

In this study, we demonstrated a new approach to forecast future S&P 500 changes using networks science, and showed that the predictors we built were strongly correlated to the amplitude of the S&P 500 changes. This result was because that we could be able to capture the market dynamics by analyzing the S&P 500 networks. The networks showing high connectedness among all the companies(nodes) means the stocks are more highly correlated. Stocks are highly correlated when the stocks are bought or sold together. As a whole market point of view, when most of the stocks are moving together, the index is likely to move in the same direction as the majority of stocks is moving. The results can be used as a new indicator that might advise financial policy makers in dealing with huge sudden market fluctuations that definitely bring the market serious problems. Also, the result can be used for the quantitative investors to improve their existing ARIMA models. In this paper, we tested the usefulness of network measurements with ARIMA model only.

7.Acknowledgement:

The authors would like to acknowledge the support of Chairman, Director and Head of the Department. Department od Computer Science and Engineering CMR Technical Campus, Medchal, Hyderabad, Telangana for their encouragement to the authors.

8.References:

[1] J, M.A., Yueshen, B.Z.: The Flash Crash: A Cautionary Tale about Highly Fragmented Markets (2016). doi:10.2139/ssrn.2243520. https://ssrn.com/abstract=2243520

[2] U.S. Commodity Futures Trading Commission, U.S.S..E.C.: Findings regarding the market events of may 6, 2010 (2010)

[3] KIRILENKO, A., KYLE, A.S., SAMADI, M., TUZUN, T.: The flash crash: High-frequency trading in an electronic market. The Journal of Finance, (2017). doi:10.1111/jofi.12498

[4] Frank, Z.: High-Frequency Trading, Stock Volatility, and Price Discovery (2010). doi:10.2139/ssrn.1691679. https://ssrn.com/abstract=1691679