Research Article

Stock Market Volatility Prediction using Time Series Data

Vinayak Kusumkar and Prof. K. C. Waghmare

Department of Computer Engineering Pune Institute of Computer Technology

Abstract

Time series data is quite different from the traditional machine learning dataset as it presents order dependency between observations. The main aim of time series forecasting is to understand the behavior of observed series and predict future values of that series based on the previous pattern of the series. Stock market movement is one of the ideal and the most volatile examples of time series. Forecasting stock volatility can give rise to better trading strategies which can limit the risks and enhance the return. Apart from the historic data, information from different news, discussion boards, and social media can be used to predict the future movement or volatility of the stocks. Regression, classification, deep learning, etc. are some approaches that can be applied individually or ensemble of these techniques on the stock market data. The ensemble of techniques gives better performance as compared to techniques applied individually. The proposed approach uses historic data and news content to forecast stock market volatility using an ensemble of machine learning models.

Keywords: Time Series Forecasting, Machine Learning, Ensemble Methods, Stock Market Prediction

Introduction

Time series data can be simply considered as a collection of continuous observations. Continuous observations because they are collected sequentially with time. One more important aspect of time series data is that the current instance of data will be dependent on the previous instance. This is a very important series of data and forecasting of time series focuses on the determination of future values of series.

There are a variety of application areas where one can observe the time-dependent data and forecasting of such time series can be crucial. The application areas can be: (a) forecasting patient's seizure using EEG trace in seconds (b) forecasting birth rate at all hospitals each year (c) forecasting the closing price of stock each day (d) forecasting unemployment for a state each quarter.

Stock market value is the best example of time series data. It can be considered as the most volatile time series data as the price of every stock or indices fluctuates rapidly along with time. Stock market prediction is of great importance for financial analysis and it can help in determining trading strategies and earn better returns on investments. Efficient market hypothesis (EMH) [1] stated that the stock market does not get affected by the old prices of the stock which encouraged research in this area. Stock market prediction can be performed as either detecting the trend of price movement or determining the future price of the stock. Former focuses on whether the price will rise or fall, whereas later emphasizes on more accurate information about the risk associated with the share.

Initially, prediction of future values of the stock is carried out with the help of single-source data such as previously available numeric data of that stock. And in traditional finance, it is considered as stock movement cannot be dependent on the sentiments or beliefs of investors. But due to ubiquity of the Internet, insights about the firm's performance, interests of the traders, financial news sentiment has been proven to be an important aspect which is responsible for the fluctuations in stock markets. This volatility of the stock is entirely dependent on the release, dissemination and acceptance of such financial information. Considering both sources that are historical prices, as well as web media information in forecasting stock market volatility, can be beneficial in achieving the better performance.

Literature Survey

In recent years, several attempts were made in order to prove EMH wrong. This leads in exploring the performance of different machine learning approaches along with information from a variety of sources for predicting the stock market volatility.

Different approaches have been proposed for stock market prediction by numerous researchers in order to

achieve better performance. Based on information used to analyze the pattern in data and analysis models, different methodologies have been used in this area of research. Some of the researchers used a single form of information with only one analysis model, whereas others used multiple information along with an ensemble of the analysis model.

Bin Weng et al. [2] developed a financial expert system to predict the short-term stock prices based on the information from different sources. Their work enhances that, the features derived from online sources supplement other financial metrics instead of substituting them which leads into the improvement of prediction performance. The proposed system is comprised of data library and artificial intelligence model. Data library is the collection of different forms of data containing previous stock prices, sentiment scores, etc. The dataset contains Citi Group stock data from 01/01/2013 to 31/12/2016. This data is preprocessed and extracted features are forwarded to the AI platform. AI platform consists of an ensemble of support vector regression, random forest, neural network and boosted regression tree. This system showed mean absolute percentage error (MAPE) less than 1.50% for 1-day ahead stock price and MAPE less than 1.89% for 1-10 day ahead stock price.

Xiongwen Pang et al. [3] conducted the research to come up with better solution to stock market prediction based on neural network approach. The basis of their research was traditional neural network algorithms may fail to predict the stock market correctly as they can fall into local optimal resulting in incorrect predictions. Recurrent Neural Network (RNN) technique can be used effectively for this problem but RNN has the drawback of vanishing gradients. So, they have proposed a deep long shortterm memory neural network (LSTM) with an embedded layer and the long short-term memory neural network with an automatic encoder to predict the stock market. Data were obtained from the livestock market from 1 January 2006 to 19 October 2016. The proposed system gives the accuracy of 57.2% and 52.5% for two models for Shanghai A-shares composite index and accuracy for individual stock is 52.4% and 52.5% respectively.

Adam Atkins et al. [4] worked on determining the influence of quantitative information of past behavior and qualitative information coming via various forms on news feeds on the behavior of time series data from financial markets. They have proposed a machine learning model comprising of natural language processing techniques and contextual data from news feeds is represented using Latent Dirichlet Allocation (LDA) and direction of share price movement is predicted using simple naive Bayes classifiers. the data of two stock indices and two equities for the range 9 September 2011 to 7 September 2012 were collected. This model shows an accuracy of nearly 56%.

Dang Lien Minh et al. [5] proposed a novel framework to predict the directions of stock prices by using both financial news and sentiment dictionary. Daily news events and their speed of release can influence the fluctuations rapidly, hence an automated decision support system is essential. They have used a two-stream gated recurrent unit (TGRU) network and Stock2Vec sentiment word embedding method. Their dataset contains stock prices and daily financial news from October 2006 to November 2013. The overall accuracy given by this model was 66.32% which outperforms previous models such as GRU, LSTM.

Qing Li et al. [6] performed an extensive literature survey to analyze the impact of data from web media along with big data approaches on the future direction of stock market price. They have reviewed 229 research articles to point out the interplay between web media and stock markets and various automated approaches that can help in minimizing the risk involved in the stock market.

With the advent of the Internet, social media became one of the fastest media of information dissemination. J. Bollen et al. [7] explained the effect of tweets on the predictability of stock fluctuation. X. Luo et al. [8] used microblogs, websites to extract sentiment features and proposed sentiment-oriented trading options.

Emotional impulses of the traders can be responsible for the fluctuation of the stock market prices. Sometimes, it is possible that traders may get affected by the opinions of others over social media, websites and news platforms. B. Pang et al. [9] opted word-based and sentiment-based sentiment analysis techniques to analyze the opinions of experts.

Proposed Methodology

In the proposed system, a different type of data related to the focused stock will be collected from different sources for prediction of the volatility of that stock. The historic data will contain the quantitative information comprised of open, high, low, close (OHLC), traded volume, etc. News content will be the qualitative information about the stock. The architecture and steps involved in the proposed system are given below,
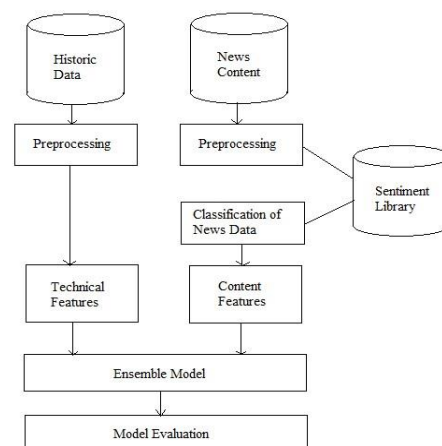

Fig. 1: System Overview

• Historic data and news data collected from different sources comprising of quantitative and qualitative information will be the input to the proposed system.
• In the preprocessing stage, data will be cleaned and transformed in order to derive the necessary features from it.
• Qualitative information will be classified according to related stock mentioned in it using natural language processing techniques (NLP).
• The content classified into different stocks will be further divided into a positive and negative category using sentiment libraries intended for the financial domain.
• Technical and content features derived after preprocessing and feature extraction stage will be passed to the ensemble of the neural network model for training. The adaboost ensemble approach will be used to enhance the performance of the system.
• The trained model will be tested with the new information and results will be evaluated using root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

Dataset

Historic prices of intended stocks for last the 3 months will be collected using Quandl API which provides quantitative information about every stock. Media data will comprise the articles collected from different financial websites, news sites, etc.

Result and Discussions

Predicting the volatility of the stock market with respect to previous data and contextual data is the main goal of the proposed system. The time-series behavior of stock values will be captured using quantitative data and qualitative information captures the movement of the stock based on sentiment of the news content.

The results of the proposed system will be evaluated using following model evaluation measures and based on these measures, model will be considered as good if the results show minimum error.

$$RMSE = \sqrt{\frac{\sum^{n}_{i=1} (X_{obs,i} - X_{model,i})^2}{n}}$$

$$MAE = \frac{\sum^{n}_{i=1} (X_{obs,i} - X_{model,i})}{n}$$

$$MAPE = \left(\frac{\sum^{n}_{i=1} \frac{X_{obs,i} - X_{model,i}}{X_{obs,i}}}{n}\right) * 100$$

Conclusions

The proposed system will predict the direction of stock market volatility using historic data and contextual data. This will be beneficial in deriving better trading strategies in order to mitigate the risks and improve profits.

References

[1]. E.F. Fama, "The behavior of stock-market prices", J. Bus., vol. 38, no. 1, pp. 34-105, 1965.
[2]. Bin Weng, Lin Lu, Fadel M. Megahed, Waldyn Martinez, "Predicting Short-Term Stock Prices using Ensemble Methods and Online Data Sources", Expert Systems with Applications, pp.258-273, 2018.
[3]. Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, Victor Chang, "An innovative neural network approach for stock market prediction", Journal of Supercomputing, 2018. https://doi.org/10.1007/s11227-017-2228-y\
[4]. Adam Atkins, Mahesan Niranjan, Enrico Gerding, "Financial news predicts stock market volatility better than close price", The Journal of
[5]. Finance and Data Science 4, Elsevier, pp.120-137, 2018.
[6]. https://doi.org/10.1016/j.jfds.2018.02.002
[7]. D. Minh, H. Huy, K. Min, H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network", IEEE Access, vol.6, pp.55392-55404, 2018.
[8]. Q.Li, Y. Chen, J. Wang, Yuanzhu Chen, H. Chen, "Web Media and Stock Markets : A Survey and Future Directions from a Big Data Perspective", IEEE Transactions on Knowledge and Data Engineering ,2017.
[9]. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market", J.Comput. Sci., vol. 2, no.1, pp. 1-8, 2011.
[10]. X. Luo, J. Zhang, and W. Duan, "Social media and firm equity value", Inf. Syst. Res., vol. 24, no, 1, pp. 146-163, 2013.
[11]. B. Pang and L. Lee, "Opinion mining and sentiment analysis", Found. Trends Inf. Retr., vol. 2, no. 1-2, pp1-135, Jan. 2008.