KOSPI Fluctuation Prediction based on S&P500 Index using a Neural Network Model

Taeseung Kim, Soowon Lee

Department of Software Convergence, Soongsil University
Email : taeseungkim@soongsil.ac.kr, swlee@ssu.ac.kr

Abstract - Stock price forecasting is an interesting subject that is studied extensively in various fields such as economics, mathematics, statistics, and artificial intelligence. However, it is not easy to predict stock price because stock price movements contain a lot of noise and generally have nonlinear characteristics. In the past, there have been studies to quantify the degree of correlation between various variables and stock prices and to predict future volatility through time series analysis of past stock price patterns. There is also a method based on artificial intelligence that generates a predictive model by learning a machine learning model using data affecting stock price formation. In this paper, we propose a stock price prediction model based on LSTM that predicts the next day 's stock price fluctuation of Korea' s stock index, KOSPI, using S&P500, the US stock index. Comparing the prediction accuracy of the proposed method with the prediction accuracy of the comparison models, the prediction accuracy of the proposed method is about 2% higher than the other models.

1. Introduction

Stock is an essential presence that affects many aspects of society. Stock is a means of securing funds for companies and is also used as an indicator of the economic level of the country. Therefore, stock market analysis and stock price forecasting have been studied in various fields such as economics, business administration, and statistics. However, stocks have nonlinear movements and are affected by many factors, making it difficult to predict stock movements. As a typical research method, there is a method of predicting future stock movements by applying past stock price movements patterns to time series models such as ARMA(AutoRegressive Moving Average) and ARIMA (AutoRegressive Integrated Moving Average). there is also a study to quantify the degree of relationship between stock prices and factors affecting stock price formation using correlation analysis. In addition, there are artificial intelligence methods that predict the stock index by learning algorithms such as artificial neural network and SVM(Support Vector Machines) using various variables forming stock price. Selection of training data for prediction model generation is highly related to prediction accuracy. Therefore, analyzing variables affecting the stock market and using influential factors for model learning makes it possible to generate models with better performance. In this paper, we propose forecasting model based on deep learning using US stock index S&P500 (Standard and Poors) to predict the closing price of Korea's stock index KOSPI(KOrea composite Stock Price Index). In this paper, we use LSTM(Long Short Term Memory Network), which is known to have good performance for time series data processing, to generate predictive model.

2. Related works

2.1 Predicting stock price based on artificial intelligence

Woosik Lee [1] analyzed the directional prediction of KOSPI by using decision tree, SVM and deep learning, and compared each model. Yoojeong Song [2] by applying MLP(Multi Layer Perceptron) forecasted fluctuation of the closing price of Samsung Electronics of 5 days later. Dong-Ha Shin [3] predicted the closing price of five KOSPI stocks using RNN(Recurrent Neural Net) and LSTM, which are suitable for time series data, considering that stock prices are time series data. Chongda Liu [4] studied 4 models of Logistic Regression, GDA(Gaussian Discriminant Analysis), Naive Bayes, and SVM to predict the direction of S&P500 stock price and compares the accuracy of each models. Ping-Feng Pai [5] presented a method of predicting the closing prices of 10 individual stocks of the S&P500 using a hybrid model combining ARIMA and SVM. Sreelekshmy Selvin [6] predicted the stock prices of several minutes later by training RNN, LSTM, and CNN(Convolusion Neural Net) using price information of individual stocks listed on NSE(National Stock Exchange) respectively. Junyoung Heo [7] uses SVM based on financial information of company(EPS, BPS etc) to predict the fluctuation of stock price of a month later and compares proposed model with artificial neural network, decision tree, AdaBoost.

2.2 Factor analysis affecting stock market

Hyungkyu Kam [8] selected macroeconomic variables (interest rate, CPI growth, industrial production index growth, economic growth index, exchange rate growth rate, etc.) that can systematically affect stock returns and then to see how they affect stock returns, he used general regression analysis and causal relationship analysis using VAR(Vector AutoRegression) model. Jung-il Choi [9] selected stock markets in five countries the US, China, Japan, Germany, etc., which are expected to have a great impact on the Korean stock market and for examining the synchronization phenomenon of each stock market in each country analyzed data for the past 151 months.

3. Methodology

The proposed model consists of stock price preprocessing module and stock price fluctuation prediction module. The stock price preprocessing module performs normalization to apply the input data to the LSTM. Input data consists of stock price(opening price, closing price, modified closing price, high price, low price, trading volume) and (closing price of the day - closing price of the previous day) in Korea's stock market and the United State stock market. The stock price fluctuation prediction module learns the preprocessed data and predicts the next day's closing price fluctuation compared to the day.

3.1 Stock price preprocessing

In the stock price preprocessing module, the z-score was applied to (closing price of the day - closing price of the previous day) variable, and min-max normalization was performed for other variables. After comparing the opening dates of the two markets, we removed data corresponding to the date of opening on one country only, in order to solve the problem that the US and the Korean stock market's closed day are not matched.

3.2 Stock price fluctuation prediction

The stock price fluctuation prediction module learns LSTM when data are given as an input, and anticipates the fluctuation of the closing price of the next day relative to the closing price of the day. This module is learned by minimizing a loss function by targeting KOSPI's closing price of the next day relative to the closing price of the day. The preprocessed input data is constructed in the form of a given time step to learn the sequence information. The time step determines how much historical data are used to predict the target. This module consists of input layer, output layer and 3 hidden layers. When data of t-1, t-2, ..t-n days are inputted through each of the layers, The closing price fluctuation of t day compared with t-1 day is predicted. Figure 1 shows a diagram of the stock price fluctuation prediction module.
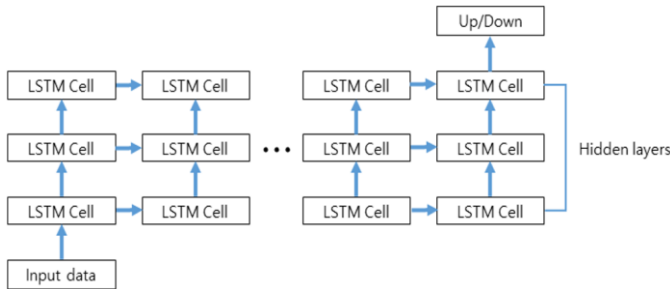

Figure 1. Stock price fluctuation prediction module

4. Experiment

4.1 Data set

In this study, we used price information of KOSPI and S&P500 collected from Yahoo Finance. To include degree of changing of the closing price of t day compared with t-1 day, we added (closing price of t day - the closing price of (t-1) day) item in the input vector. TABLE I is characteristic of input data. The input vector used for learning is shown in TABLE II.

TABLE I. Characteristic of input data

Item	Explanation
$Open_t$	Stock price, when market of day t open
$High_t$	The highest stock price of day t
Low_t	The lowest stock price of day t
$Close_t$	Stock price, when market of day t close
$AbjClose_t$	Modified closing price of day t reflecting business activity
$Volume_t$	Total trading volume of day t
$Close_t - Close_{t-1}$	Closing price of the day t - Closing price of the day (t -1)

TABLE II. An example of input vector

	KOSPI	S&P500
Date	2018.10.17	
Open	2,169.4	2,767.05
High	2,182.7	2,813.4
Low	2,159.3	2,766.9
Close	2,167.5	2,809.9
AdjClose	2,167.5	2,809.9
Volume	275,200	3,428,340,000
Close_t – Close_t-1	22.3	59.1

The data were divided into training set, validation set, and test set to evaluate the learning model. The "training set" is used to create the model, while the "validation set" is used to qualify performance. The data set used to evaluate the final model performance is called the "test set". TABLE III summarizes training set, validation set, and test set.

TABLE III. Summary of data

Item	Term	Days
Training Set	2000.11.2 ~ 2014.11.3	3351
Validation Set	2014.11.4. ~ 2016.10.31	478
Test Set	2016.11.1 ~ 2018.11.8	477

4.2 Assessment method

In this study, Accuracy was used to evaluate the performance of the model. The accuracy formula is shown in Equation 1.

$$\text{Accracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 1. The accuracy formula

TP(True Positive) and TN(True Negative) are predicted values of the model as actual values, and FP(False Positive) and FN(False Negative) are prediction failures.

4.3 Experimental environment

In this study, we use the Scikit-learn library to generate the comparative model and the Tensorflow library to generate the stock price fluctuation prediction module. Scikit-learn is an open source library that provides various classifications, regression and clustering algorithms and is compatible with other libraries. Tensorflow is an open source library developed by Google for machine learning and deep learning. A summary of the experimental environment is shown in TABLE IV.

TABLE IV. Experimental environment

Environment	Tool
OS	Ubuntu 16.04
Programming Language	Python 3.5.2
IDE	Pycharm 2017.3 (professional edition)
Library	Scikit-Learn 0.19.1
	Tensorflow 1.4.1

5. Discussions/Results

5.1 Optimization of proposed model parameters

In order to find the optimal parameters of the proposed model, the accuracy was measured by changing the number of hidden layers and time step. We set the time steps to 3 when measuring the accuracy according to the change of the hidden layers and to set the hidden layers to 3 when measuring the accuracy according to the change of the time steps. Figure 2 and Figure 3 show the accuracy obtained by changing hidden layers and time steps, respectively.
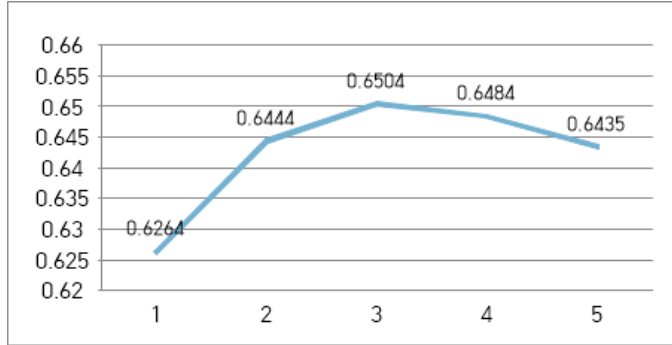

Figure 2. Accuracy with hidden layer changes

In Figure 2, the x-axis is the number of layers and the y-axis is the accuracy of the validation set. When the hidden layer is 3, it can be confirmed that model has the most accurate.
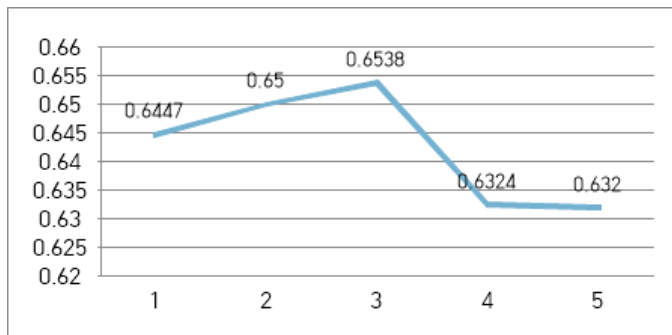

Figure 3. Accuracy with time step changes

Figure 3 shows the prediction accuracy result, according to the change of the time steps of the prediction model. When the time step is 3, it can be confirmed that the model has the highest accuracy. When the time step is 3, it can be assumed that the sequential characteristics of the data are captured better than other cases.

5.2 Predictive accuracy by model

Figure 4 shows the comparison between the prediction accuracy of the proposed model and the prediction accuracy of the other models. Each model was optimized using a validation set. In the proposed model, LSTM, the prediction accuracy is slightly higher than other comparative models. The reason seems to be the advantage of the LSTM model that it can contain sequential information of time series data. TABLE V is optimal parameters of LSTM.
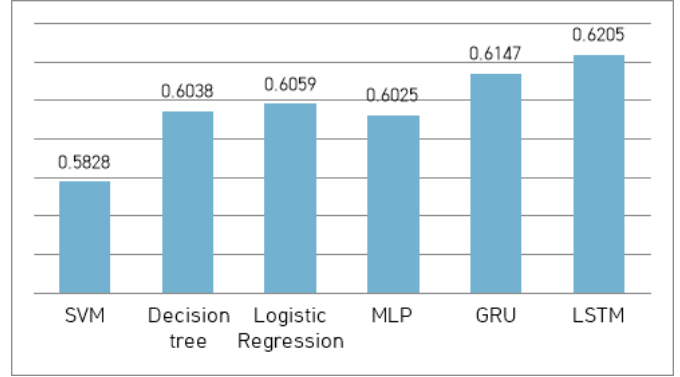

Figure 4. Accuracy of the proposed model and comparison models

TABLE V. Optimal parameters of LSTM

Parameter	Value
Learning Rate	0.001
Epoch	15
Batch Size	4
Hidden Layer	3
Time Step	3
Hidden Unit	16

6. Conclusion

In this study, for predicting KOSPI fluctuation we proposed a forecasting model based on a neural network. The proposed model preprocessed the S&P500 index and KOSPI data and predicted the closing price of the next day relative to the closing price of the day using LSTM. Experiments using the same data showed that the proposed model had a prediction accuracy of about 2% higher than the comparison models and showed a slightly higher accuracy than GRUs with similar structures. This paper does not include the degree of fluctuation, but simply classifies it as rising or falling. Therefore, it is necessary to study numerical prediction of rates of change in the future plan. Also, considering technical analysis indicators like RSI and MACD and the basic analysis, such as financial information of the company, will be helpful to improve the accuracy of the model.

References

[1] Woosik Lee. "A deep learning analysis of the KOSPI`s directions." *Journal of the Korean data & information science society, vol. 28, no. 2,* 2017, pp. 287-295.

[2] Yoojeong Song, Jongwoo Lee. "A Design and Implementation of Deep Learning Model for Stock Prediction

using TensorFlow." *Proceedings of Korea Computer Congress on Korean Institute of Information Scientists and Engineers(KIISE), vol. 2017, no. 06,* 2017, pp. 799-801.

[3] Dong-Ha Shin, Kwang-Ho Choi, Chang-Bok Kim. "Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM." *Journal of the Korean Institute of Information Technology, vol. 15, no. 10,* 2017, pp. 9-16.

[4] Chongda Liu, Jihua Wang, Di Xiao, Qi Liang. "Forecasting S&P 500 Stock Index Using Statistical Learning Models." *Open Journal of Statistics, vol. 6, no. 6,* 2016, pp. 1067-1075.

[5] Ping-Feng Pai, Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting." *Omega, vol. 33, no. 6,* 2005, pp. 497-505.

[6] Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P. "Stock price prediction using LSTM, RNN and CNN-sliding window model." *Proceedings of International Conference on Advances in Computing, Communications and Informatics(ICACCI),* 2017, pp. 1643-1647.

[7] Junyoung Heo, Jin Yong Yang. "SVM based Stock Price Forecasting Using Financial Statements." *Journal of KIISE Transactions on Computing Practices(KTCP), vol. 21, no. 3,* 2015, pp. 167-172.

[8] Hyungkyu Kam, Yongjae Shin. "The Impact of Macroeconomic Variables on Stock Returns in Korea." *Journal of Daehan Association of Business Administration Korea, vol. 30, no. 1,* 2017, pp. 33-52.

[9] Jung-il Choi. "Synchronization Phenomenon and Correlation Analysis of Global Stock Market." *Journal of the Korea Contents Association, vol. 16, no. 1,* 2016, pp. 699-707.