Journal of the Korea Convergence Society
Vol. 12. No. 7, pp. 31-36, 2021

ISSN 2233-4890 / eISSN 2713-6353
https://doi.org/10.15207/JKCS.2021.12.7.031

Research model on stock price prediction system through real-time Macroeconomics index and stock news mining analysis

Sunghyuck Hong
Associate Professor, Division of Smart IT, Fintech major, Baekseok University

실시간 거시지표 예측과 증시뉴스 마이닝을 통한 주가 예측시스템 모델연구

홍성혁
백석대학교 스마트IT공학부 핀테크 전공 부교수

Abstract As the global economy stagnated due to the Corona 19 virus from Wuhan, China, most countries, including the US Federal Reserve System, introduced policies to boost the economy by increasing the amount of money. Most of the stock investors tend to invest only by listening to the recommendations of famous YouTubers or acquaintances without analyzing the financial statements of the company, so there is a high possibility of the loss of stock investments. Therefore, in this research, I have used artificial intelligence deep learning techniques developed under the existing automatic trading conditions to analyze and predict macro-indicators that affect stock prices, giving weights on individual stock price predictions through correlations that affect stock prices. In addition, since stock prices react sensitively to real-time stock market news, a more accurate stock price prediction is made by reflecting the weight to the stock price predicted by artificial intelligence through stock market news text mining, providing stock investors with the basis for deciding to make a proper stock investment.

Key Words : Stock analysis, Big data, Text mining, RNN, Prediction, Convergence

요 약 중국 우한발 코로나 19 바이러스로 인하여 세계 경제가 침체하여, 미국연방준비제도를 비롯한 대부분 국가에서는 통화량을 늘려 경기를 부양하는 정책을 내놓았다. 주식 투자자들 대부분은 기업에 대한 재무제표 분석이 없이 유명 유튜버의 추천종목이나 지인의 말만 듣고 투자하는 경향이 있어서 주식투자의 손실 가능성이 크다. 따라서, 본 연구에서는 기존 자동매매 조건에서 발전된 인공지능 딥러닝 기법을 이용하여 주가에 영향을 미치는 거시지표를 분석하고 예측하여 주가에 미치는 상관관계를 통한 개별주가예측에 가중치를 부여하고 주가를 예측한다. 또한, 주가는 실시간 증시뉴스에 민감하게 반응하기 때문에 증시뉴스 텍스트 마이닝을 통하여 인공지능으로 예측된 주가에 가중치를 반영하여 더 정확한 주가 예측을 하여 주식 투자자에게 매매의 판단 근거를 제공하여 건전한 주식투자가 되도록 이바지하였다.

주제어 : 주가분석, 빅데이터, 텍스트마이닝, RNN, 예측 시스템, 융합

*This research was supported by 2021 Baekseok University Research Fund.
*Corresponding Author : Sunghyuck Hong(shong@bu.ac.kr)
Received April 17, 2021 Revised June 4, 2021
Accepted July 20, 2021 Published July 28, 2021

1. Introduction

1.1 Macro Indicators Affecting on Economy

The global economy stagnated due to the corona 19 infection in Wuhan, China, which began in December 2019, and most countries, including the Federal Reserve System, introduced policies to boost the economy by increasing money. Market capital overflowed and liquid capital flowed into the stock market, and domestic stock markets, including the US, rose sharply from the end of March 2020, leading to an overheating of the stock market. Young adults in their 20s and 30s in the United States and Korea opened securities accounts, increasing new accounts by more than 4.3 times from last year, and the KOSPI soared to 3,000 points. Most stock investors tend to invest only by listening to the recommendations of famous YouTubers or acquaintances without analyzing the company's financial statements, so there is a high possibility of stock investment loss. Therefore, this study analyzes and predicts macro indicators that affect stock prices using artificial intelligence deep learning techniques developed under existing automatic trading conditions, and gives weight to individual stock price predictions through correlations that affect stock prices. Predict. In addition, since stock prices are sensitive to real-time stock market news, weights are applied to the stock prices predicted by artificial intelligence through stock market news text mining to make more accurate stock price predictions, providing a basis for the predicted stock price, and making a proper stock investment to proceed with research.

2. Related Research

2.1 Correlation between macro indicators and stocks

The economic situation of a particular country can be easily grasped through macroscopic indicators. Representative macro indicators include exchange rate, interest rate, money volume, GDP growth rate, unemployment rate, inflation rate, current account balance, consumer index, consumer survey, production index, new housing supply, number of claims for unemployment benefits, long-term government bond interest rates, raw materials, etc. There are many studies on the relationship between macro indexes and stock price volatility, and although there are some differences, nominal economic variables such as money and inflation usually affect stock price fluctuations, but other real economic indicators do not have a significant effect [1-3]. There is a variety of stock stocks, and it is expected that more accurate predictions can be made if the macro-indices are predicted for reference when predicting individual stocks by analyzing the high correlation with the macro-indicators among individual stocks. Table 1 shows the changes in incremental values for comparison and correlation analysis because the units of macro indicators are different. Past data was downloaded and used on a monthly basis from the fred.stlouisfed.org site. In addition, the increment value was set to 0 based on the start date, and the difference from the next data was divided by the start date data and displayed. Table 2 shows the correlation between the macro index and stock price items (NASDAQ, DOW) as an index by analyzing the incremental values of the data for 10 years from 2011 to 2021. The closer the index is to 1, the stronger the correlation, the closer to 0 means there is no correlation, and the stronger the opposite correlation. Table 2 shows only those with a correlation index of 0.5 or more among the various macro indicators. In Table 1, since GDP is aggregated by quarter, there is no data after October 2020, and for the unification of monthly data, the incremental value by quarter was calculated as an equal ratio and expressed by circle.

Table 1. Macro indicator incremental value(2011~2021)

Date	Consumer index	Production index	New home sale	Unemployment rate	GDP	Number of claims for unemployment benefits	raw materials
2011-01-01	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...
2020-06-01	0.005	0.115	0.165	-0.17	0.0259	-0.126	0.003
2020-07-01	0.004	0.003	-0.002	-0.08	0.0051	-0.139	0.036
2020-08-01	0.002	-0.001	-0.012	-0.18	0.0051	-0.160	0.018
2020-09-01	0.001	0.000	0.000	-0.07	0.0050	-0.339	0.067
2020-10-01	0.002	0.016	-0.131	-0.12	-	-0.252	0.035
2020-11-01	0.002	-0.008	0.055	-0.03	-	-0.112	0.011
2020-12-01	0.003	-0.057	0.043	0.00	-	-0.083	0.013

Table 2. DOW/NASDAQ and macro-indicator correlation

Correlatoin	Consumer index	Production index	New home sale	Unemployment rate	GDP	Number of claims for unemployment benefits	raw materials
Dow vs. others	0.94	0.88	0.78	0.0567	0.35	0.19	0.93
NASDAQ vs. others	0.93	0.88	0.77	0.0978	0.25	0.24	0.92

The larger the number of claims for unemployment benefits, the more it affects the stock price decline, showing the opposite correlation. Therefore, after making predictions for individual stock prices through LSTM, making predictions on macro indicators that are highly correlated with or opposite to individual stock prices, the weight of the macro index values is adjusted and repeated to match the actual stock price after comparing it with the actual stock price. By learning, the accuracy of prediction can be improved compared to the method of analyzing only with historical stock price time series data. In addition, stock prices may fluctuate due to unpredictable stock market news or words of celebrities, so if stock market news is judged whether it is good or bad in the stock price through real-time text mining and reflected in the stock price, the stock price prediction system that has been used so far. It is expected that the accuracy can be improved.

2.2 Deep Learning Networks

Artificial intelligence is defined as a computer program that artificially embodies the capabilities of human perception, reasoning, and learning, or a computing system including the same [7]. Artificial intelligence is a general term.Machine learning is a field in the field of artificial intelligence that repeatedly learns by itself and learns by itself without human intervention, and deep learning is further developed in machine learning to make more accurate predictions with multiple hidden layers. This is the way to do it. Examples of deep learning networks include ANN (Artificial Neural Network), DNN (Deep Neural Network), RNN (Recurrent Neural Networks), and LSTM (Long Short Term Memory) [4-6]. It is used to predict based on past data by applying a deep learning network.

2.3 RNN and LSTM prediction method

RNN and LSTM are widely used in stock price Prediction because they are easy to analyze sequential and repetitive data [8]. It makes predictions by continuously iterating through weights that data affects the present in the past. Therefore, RNN and LSTM are widely used for stock price prediction [9]. In this study, not only using historical data of the stock price, but by analyzing each macro-indicator, predicting the stock price, comparing it with the actual stock price, adjusting the weight value as much as it differs, and performing iterative learning to best match the actual stock price to predict the stock price. In addition, the weight value is adjusted according to the correlation of each macro indicator to make an optimal prediction, and the weight value is also adjusted through repetitive learning.

3. Proposed Model

3.1 Price Prediction Algorithm

Stock price Prediction largely uses basic Prediction and technical analysis. Basic Prediction is to analyze and predict the value of a company through the company's financial statements, industry trends, and CEO capabilities, and in this study, it means stock price prediction through technical analysis. Candle Chart, Moving Average, RSI, (Relative Strength Index), CCI (Commodity Channel Index), MACD (Moving Average Convergence and Divergence), along with basic stock price data such as high, low, close, and trading volume.), Disparity, SIGNAL, and stochastic1 [10-12]. In this study, it is expected that accurate stock price prediction can be made by learning technical Prediction through LSTM, adjusting weight values by analyzing various macroindices and correlations, and predicting the final stock price after text mining real-time stock price news [13-15].

3.2 Python Stock Prediction Code

Fig. 1 to ① used pandas for using csv files, numpy for matrix operation, matplotlib.pyplot for data visualization, and Keras for creating deep learning models. ③ The window size is analyzed in units of 20 and the meaning of 20 means that one data is predicted after analyzing 20 data each. ④ Using a function called Sequential(), out of 20 window sizes, the training set is set to 90%, and the remaining 10% is set as the test set.

In Fig. 1, the NASDAQ stock index prediction was analyzed by dividing it into a training set and a test set. The same result as in Fig. 2 was obtained. Red is the actual stock price data, and blue is the predicted value.

```
①
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense, LSTM, Dropout
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
data = pd.read_csv("./NASDAQ.csv")
...
②
def MinMaxScaler(data):
    numerator = data-np.min(data,0)
    denominator = np.max(data,0) - np.min(data,0)
    return numerator / (denominator + 1e-7)
...
③
window_size = 50
data_NASDAQ = []
data_NASDAQ_next = []

for i in range(len(NASDAQ) - window_size):
    data_NASDAQ.append(NASDAQ[i : i + window_size])
    data_NASDAQ_next.append(NASDAQ[i + window_size])

train_size = int(len(data_NASDAQ) * 0.9)
train_NASDAQ = np.array(data_NASDAQ[0 : train_size])
train_NASDAQ_next = np.array(data_NASDAQ_next[0 : train_size])

print(train_NASDAQ.shape)
print(train_NASDAQ_next.shape)

test_size = len(data_NASDAQ_next) - train_size
test_NASDAQ = np.array(data_NASDAQ[train_size :
len(data_NASDAQ)])
test_NASDAQ_next = np.array(data_NASDAQ_next[train_size :
len(data_NASDAQ_next)])

print(test_NASDAQ.shape)
print(test_NASDAQ_next.shape)
...
④
model = Sequential()
model.add(LSTM(units = 10, activation = 'relu',
return_sequences=True, input_shape=(window_size,1)))
model.add(Dropout(0.1))
model.add(LSTM(units=10, activation='relu'))
model.add(Dropout(0.1))
model.add(Dense(units=1))
model.summary()

model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(train_ebix,train_NASDAQ_next, epochs=30, batch_size =
20)
pred_NASDAQ = model.predict(test_NASDAQ)

plt.figure()
plt.plot(test_NASDAQ_next, color = 'red', label='real price')
plt.plot(pred_NASDAQ, color='blue',label = 'predicted price')
plt.title('price prediction')
plt.legend()
plt.show()
...
```

Fig. 1. Python Stock Price Prediction Partial Code

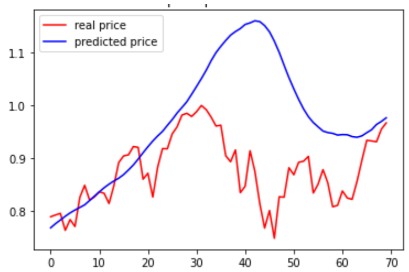

Fig. 2. Python Stock Price Predict Result

Table 3. Incremental Predicted Value

Type	Consumer index	Production index	New home sale	Unemployment rate	GDP	Number of claims for unemployment benefits	raw materials
2021-01-01 (Prediction)	**0.001**	**0.007**	**0.013**	**0.050**	**0.002**	**0.109**	**0.008**
2021-01-01 (Real value)	0.0026	-0.0574	0.0429	-0.0597	0.00511	-0.0830	0.0125
(Prediction-Real)*100 ①	-0.160	6.440	-2.990	10.970	-0.311	19.200	-0.450
Correlation ②	0.93	0.88	0.77	0.0978	0.25	0.24	0.92
①x②=W	-0.1488	5.6672	-2.3023	1.072866	-0.07775	4.608	-0.414
Weight	$N_d \times W_i / \sum_{1}^{n} W_i$						

*N_d=NASDAQ real value -NASDAQ Prediction value

In Table 3, if the predicted NADAQ value is 13,309 and the actual value is 13,048, the weight value is calculated to adjust the part where the difference between the predicted value and the actual value is 261. The calculation method is to add and subtract the W_i value obtained by multiplying each index's correlation multiplied by each index predicted value-actual value by 100(① in Table 3), and then adjust each weight value (W_i) to match the actual value by adding and subtracting it by the ratio of each index, and reflecting the weight value on the stock price (② in Table 3). When a final prediction price comes, then match the real price. If there is a differential between the final prediction and the real one. We assume that the differential based on the stock news. That is why the gap between prediction and real one from news effect. Therefore, we can determine that how to affect this stock news on the stock price. After accumulating news data, then we can determine whether it is good news or bad news through news text mining, and then compared it with the actual value and adjusted it.

4. Conclusion

Although many studies are being conducted to predict stock prices using deep learning algorithms of artificial intelligence, most of the existing stock price prediction algorithms analyze past time-series data. Therefore, there is a limit because it does not reflect an unexpected situation in real-time. However, if we learn repetitive patterns using machine learning techniques of artificial intelligence and correct prediction errors through real-time news mining, we expect to be able to buy and sell stock prices on their own by making judgments at the level of fund managers. This study does not predict the stock price by simply analyzing the stock price index only. It is a technique that predicts the stock price by finding new changing patterns through real-time iterative learning through macro-indicators and news mining. I hope that a healthy investment culture will be established.

REFERENCES

[1] S. H. Shin. (2013). Analysis on the Relation between Foreign Investors in Korean Stock Markets and Macroeconomic Variables. *INTERNATIONAL BUSINESS REVIEW, 17(2),* 89-107

[2] M. F. Elhusseiny. (2017). Industries Stock Return Reactions To Risk Factors: An Empirical Investigation On The G-7 Countries. *Journal of Financial and Monetary Economics, 4(1),* 196-204.

[3] K. Guru-Gharan, K. M. Rahman & S. Parayitam. (2009). Influences of selected macroeconomic variables on US stock market returns and their predictability over varying time horizons. *Academy of Accounting and Financial Studies Journal, 13(1),* 13.

[4] Y. Qian, Y. Fan, W. Hu & F. K. Soong. (2014, May). On the training aspects of deep neural network (DNN) for parametric TTS synthesis. *In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* (pp. 3829-3833). IEEE.

[5] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang & A. Yuille. (2014). *Deep captioning with multimodal recurrent neural networks (m-rnn).* arXiv preprint arXiv:1412.6632.

[6] R. Pascanu, C. Gulcehre, K. Cho & Y. Bengio. (2013). *How to construct deep recurrent neural networks.* arXiv preprint arXiv:1312.6026.

[7] I. Goodfellow, Y. Bengio, A. Courville & Y. Bengio.

(2016). *Deep learning (Vol. 1, No. 2)*. Cambridge : MIT press.

[8] L. R. Medsker & L. C. Jain. (2001). Recurrent neural networks. *Design and Applications, 5*.

[9] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon & K. P. Soman. (2017, September). Stock price prediction using LSTM, RNN and CNN-sliding window model. *In 2017 international conference on advances in computing, communications and informatics (icacci)*. (pp. 1643-1647). IEEE.

[10] D. H. Shin, K. H. Choi & C. B. Kim. (2017). Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM. *Korean Institute of Information Technology, 15(10)*, 9-16.

[11] T. J. Hsieh, H. F. Hsiao & W. C. Yeh. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing, 11(2)*, 2510-2525. doi:10.1016/j.asoc.2010.09.007

[12] M. S. Mahmud & P. Meesad. (2014). Time series stock price prediction using recurrent error based neuro-fuzzy system with momentum. *2014 International Electrical Engineering Congress (iEECON)*. (pp. 1-4) doi:10.1109/ieecon.2014.6925866

[13] G. P. C. Fung, J. X. Yu & W. Lam. (2003, March). Stock prediction: Integrating text mining approach using real-time news. *In 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings*. (pp. 395-402). IEEE.

[14] S. Hong. (2020). Research on Stock price prediction system based on BLSTM. *Journal of the Korea Convergence Society, 11(10)*, 19-24.

[15] S. Hong. (2020). A study on stock price prediction system based on text mining method using LSTM and stock market news. *Journal of Digital Convergence, 18(7)*, 223-228.

홍 성 혁(Sunghyuck Hong) [종신회원]

· 2007년 8월 : Texas Tech University, Computer Science (공학박사)
· 2012년 3월 ~ 현재 : 백석대학교 스마트IT공학부 핀테크 전공 주임
· 관심분야 : 핀테크, 딥러닝, DID, 사물인터넷 보안, 경량보안프로토콜
· E-Mail : shong@bu.ac.kr