Sentiment Analysis of News for Effective Cryptocurrency Price Prediction

Anh-Dung Vo, Quang-Phuoc Nguyen, and Cheol-Young Ock

***Abstract*—With the rapid development of e-commerce, financial industry and blockchain technology, cryptocurrencies have become a global phenomenon known to most people. The historical prices show that cryptocurrencies have experienced significant price fluctuations on both daily and long term valuations. Cryptocurrency market movement prediction systems have emerged to help people make an informed decision. Traditional supervised learning algorithms were used for predicting changes in cryptocurrency prices based on historical price data. Nowadays, the number of news available on the internet is increasing rapidly. Discerning the impact of news on price movement can provide a buying and selling advantage to an investor. In this paper, we describe a method for predicting cryptocurrency prices utilizing news and historical price data. Our paper analyses the ability of news data to predict price fluctuations for the second largest cryptocurrency in terms of market capitalization: Ethereum. The model is able to directly predict price direction by indicating whether to buy, sell, or hold. The final version of the model was able to correctly predict cryptocurrency price using historical data and sentimental information gained from news data. The important key in our model is the application of a set of natural language processing algorithms to identify the public moods for cryptocurrency fluctuations. We showed that sentiment analysis is an important perspective for cryptocurrency price prediction due to the interactive nature of financial activities.**

***Index Terms*—Cryptocurrency market prediction, Ethereum, long short-term memory, machine learning, natural language processing, sentiment analysis, text mining.**

I. Introduction

With the rapid development of e-commerce, financial industry, and blockchain technology, cryptocurrencies have become a global phenomenon known to most people. The cryptocurrency market has generated interest of investors through the Internet. A cryptocurrency is a digital currency or asset created to work as a medium of exchange that exploits strong cryptography to protect financial transactions, handle the creation of additional units, and certify the transfer of assets. Bitcoin (BTC) and Ethereum (ETH) are two largest cryptocurrencies in terms of market capitalization. Both BTC and ETH have experienced significant price fluctuations on both daily and long term valuations. Recently, cryptocurrency market movement prediction systems have emerged to help investors make an informed decision. Traditional supervised learning algorithms were used for predicting changes in cryptocurrency prices based on historical price data. However, the efficient market hypothesis (EMH) [1] states that the cryptocurrency market always follows a haphazard pattern and its prediction is always a challenging task.

As traditional financial markets, there is a relation between public moods and the cryptocurrency market. Discerning the impact of news on price movement can provide a buying and selling advantage to investors. Nowadays, more and more prevalence of related news articles and social media posts is progressively becoming available on the internet. However, it is overwhelming for people to read and figure out the most common public feelings and emotions. Therefore, automated sentiment discovery systems are needed to assist investors in making a better informed trading decision.

By taking this need, we aim to analyze the impact of news data to produce a prediction for cryptocurrency prices. Our model draws inspiration from our prior work on opinion-aspect relation analysis and distributed representation of phrases [2], [3]. In this paper, we investigate the influencers of news data to predict price fluctuations of ETH. As our main contribution, we present a system that includes two stages - sentiment analysis and price prediction. In the first stage, the system takes news data as input and produce sentiment score as output. We developed this stage using a set of natural language processing (NLP) tools: dependency parser (DP), co-reference (CR), and named entity recognition (NER). The sentiment detection process shares its probabilistic foundation with distributed representation of phrases and continuous bag of words in word2vec. We parametrize the model in a manner, that aims to capture both semantic and sentiment similarities among word. Thus, the proposed model achieved a sentiment score which may indicate price movement. In the second stage, the sentiment score from the first stage is fed back as additional input into an artificial recurrent neural network (RNN) architecture to predict cryptocurrency prices. We used long short-term memory (LSTM) network which is well-suited for making predictions based on time series data. The fact that most of the researchers use LSTM network to predict cryptocurrency prices, but this system improves the accuracy and usefulness of its news sentiment data by leveraging sentimental information. The important key in our model is the application of a set of natural language processing algorithms to identify the public moods for cryptocurrency fluctuations. We showed that sentiment analysis is an important perspective for cryptocurrency price prediction due to the interactive nature of financial activities.

Manuscript received May 22, 2019; revised October 10, 2019. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF- 2016S1A5B6913773).

The authors are with the Department of IT Convergence, University of Ulsan, Ulsan, 44610, Republic of Korea (e-mail: voanhdung@mail.ulsan.ac.kr, nqphuoc@mail.ulsan.ac.kr, okcy@ulsan.ac.kr).

doi: 10.18178/ijke.2019.5.2.116

II. Literature Review

A. Stock Market Prediction

Stock market always follows a haphazard pattern and its prediction always a complicated process. With the rise of technology in the world of finance, researchers have attempted a variety of analysis approaches such as technical analysis, quantitative analysis, and so on. In recent years, advances in machine learning allow many investors to do market research better, manage funds more efficiently, invest and expand more effectively. Regression and SVM are classic approaches that dominate prediction methods [4]. However, regression models are widely applied since they allow us to determine the impact of each variable included and evaluate the importance of factors by dropping them out. After close review, we can see in Kumar' survey [4] that overly complicated systems commonly perform poor precision and low accuracy. Meanwhile, linear models do not always exist in real-life cases because their foundation strongly relies on hypotheses. In spite of this challenging task, researchers seek to estimate stock price movement by applying neural network neural network such as self-organizing fuzzy neural network [5], neural tensor network [6], and so on. Recent research has shown greater predictability in stock data but these researches is often under introduced in the academic literature. Based on the researches we found, more progress has been developed in predicting near-term [7] and long-term price fluctuations [8] . A high accuracy of long-term prediction has shown when only considering limited number of stocks in a particular industry [9]. These models focus on targeting price change in near-term, short-term, and long-term by relying upon particular company information. However, these valuable datasets are hard to collect and do not work with short-term prediction. Most researches in short-term prediction have introduced classification models [10], [11]. Although the results might look satisfactory, the actual systems are too coarse for most current trading applications.

Our work, cryptocurrency price prediction, is related to but different from ordinary stock price prediction in several ways. First, we aim to predict price fluctuations in cryptocurrency market. People often consider the cryptocurrency market as another version of the stock market with no difference. But the fact is that cryptocurrency market is very much different from the ordinary stock market in tangible value, user base, price volatility, and so on. Second, in our study, the information gained from the news is combined with numerical data make a robust input data stream for prediction process. Stocks show ownership interest in companies; cryptocurrencies do not. Cryptocurrency is a means of exchange whose value is based solely on popularity. Therefore, our model achieved satisfactory experimental results by relying on particular information obtained from news data. Finally, our test models shows that naively passing in time series data of price performs no better than random guessing, so we implemented sentiment analysis system to analyze news data before running machine learning models for time series data. In our case, an ensemble system can be used to manipulate the combination either on a decision level or factor level.

B. Technical Analysis Indicator

As traditional financial markets, when predicting cryptocurrency price direction, investors typically use one of three approaches. The first is the fundamental method which analyses economic features that drive the price of cryptocurrency. The second method is to apply traditional technical analysis to foresee what others are thinking based on the price and volume of the cryptocurrency. Indicators are calculated from historical prices and past volumes and these are used to predict future changes in prices. Surveys shows that most of investors use at least some form of technical analysis [12], [13]. The last method is quantitative technical analysis that has more quantitative and statistical approach [14].

The common technical analysis indicators are rates of changes, moving averages, regression, moving variance rations, relative strength index, and so on. Studies [14]-[17] has shown technical analysis to be useful to predict price movement. With recent breakthroughs in technology, advanced algorithms allow us to outperform existing published methods of stock market direction. Our work builds upon those developments in machine learning for time series data.

C. Semantic Modelling

Most of the varied techniques which have been developed in NLP research can be classified into two categories: expressiveness and language rules. The early NLP approach used bag-of-words to represent textual data. In this model, the semantics of a piece of text is represented by a set of words and their frequencies. An obvious drawback of this model is that semantic information will not be obtained. This problem is sold by using Latent semantic analysis (LSA) that was designed to model word directly. In this vector space model (VSM), semantic word vectors are obtained using singular value decomposition. With the progress of machine learning techniques, it is possible to train word embedding models to use more complex NLP systems on much larger data sets. These models typically outperform the simple models of the past. Rich relational structure of the lexicon was captured by recognizing similarities between words and distance between word vectors in a high-dimensional space [18]. Early work focused on word vector representations using neural network models [19]. With the recent advances in word embedding, efficient models was introduced for learning high-quality distributed vector representations [20]-[22].

Our work builds upon these researches in sentiment analysis of news for effective cryptocurrency price prediction. The sentiment detection process in our model shares its probabilistic foundation with distributed representation of phrases and continuous bag of words in word2vec. We parametrize the model in a manner, that aims to capture both semantic and sentiment similarities among word.

D. Sentiment Analysis

NLP has become increasingly influential due to data availability, various techniques, and powerful hardware developed in the past decade. Sentiment analysis refers to the use of many NLP sub-tasks, including subjectivity detection, aspect extraction, named entity recognition, sarcasm detection, personality recognition, user profiling to systematically extract, and subjective information analyzing [23]-[27]. Sentiment analysis based techniques are widely applied to reviews, survey, and online data. Recently, researchers introduced attempts to predict fluctuations in the price of cryptocurrencies. Coliannni *et al*. presented a supervised learning algorithm to predict price movement using an online text sentiment service [28]. Similarly, Connor *et al*. analyzed the ability of news and social media data to predict price fluctuations [29]. Abraham *et al*. presented a method for predicting changes in Bitcoin and Ethereum prices utilizing Twitter data and Google Trends data [30]. Meanwhile a hidden Markov model was introduced to predict such bubbles for a number of cryptocurrencies [31], [32].

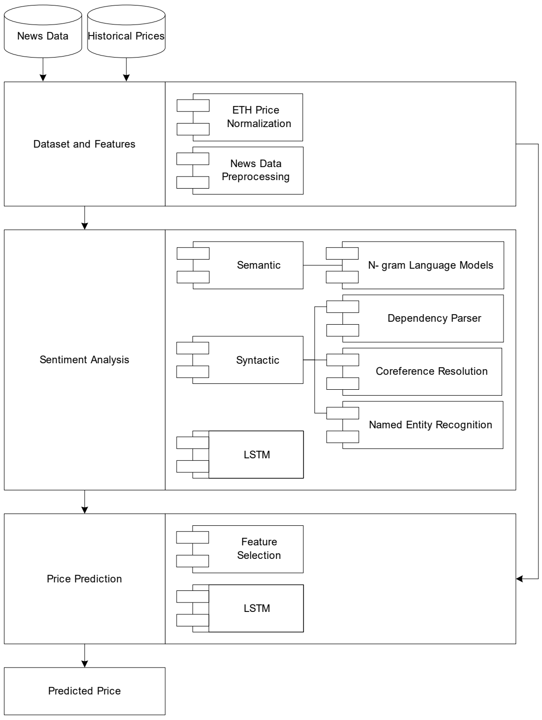

Fig. 1. The system proposed in this study.

Our work builds upon those developments in cryptocurrency price prediction. After extensive investigation, we did not find any works that combined advanced NLP with machine learning for time series data like we did. The sentiment analysis in our model is performed using a set of NLP tools: dependency parser, co-reference, and named entity recognition.

Because no regression system is perfect, we aim to improve the accuracy of prediction system by utilizing sentimental information gained from news data.

III. The Proposed System

Unlike traditional stock market, the cryptocurrency market is a more open, worldwide market where discussions and news are mostly public and easily accessible. In trading, psychology is a key concept which is often underestimated. It is critical to detect the feelings surrounding a particular market and get the global sentiment. Market psychology is certainly one of the most important factors to study. Since the discussions about cryptocurrencies are happening online, there is tremendous value in data mining from news and other social platforms to detect public mood and relevant insights.

Fig. 2. Time series data of ETH and labelled sentiment scores.

To stay up-to-date with the evolving cryptocurrency market, discerning the impact of news on price movement can provide a buying and selling advantage to investors. The analysis of web and social media conversations brings a lot of value to better understanding the specific expectations of the public and the emerging changes in market trends.

To do this kind of market psychology analysis, we analyze the ability of news data to predict price fluctuations for the largest cryptocurrency in terms of market capitalization: Ethereum. The model is able to directly predict price direction by indicating whether to buy, sell, or hold. The final version of the model was able to correctly predict cryptocurrency price using historical data and sentimental information gained from news data. The important key in this model is the application of a set of natural language processing algorithms to identify the public moods for cryptocurrency fluctuations. We showed that sentiment analysis is an important perspective for cryptocurrency price prediction due to the interactive nature of financial activities.

As our main contribution, we present a system that includes two stages: sentiment analysis and price prediction. The overall experimental procedure is illustrated in Fig. 1.

A. Dataset and Features

1) New datasets

We aim to analyze the impact of news to generate a prediction for cryptocurrency prices. Although a large number of sources is needed for a sentiment analysis based system, we test a small experimental subset to demonstrate how our method works. The dataset is obtained from NewsNow (https://newsnow.co.uk). This dataset contained crypto-related news, which is labelled with respect to cryptocurrencies such as BTC, ETH, and so on. Preprocessing tasks were performed on these datasets to fit in our system input requirements.

2) Cryptocurrency datasets

We applied machine learning approach to predict prices and identify scams before they occur, based on historical data. We use Poloniex-API (https://docs.poloniex.com) to access the time series data of BTC and ETH traded at Poloniex, which is a crypto exchange platform that also provides API for data mining. The Poloniex-API allows to access intraday time series, daily time series, weekly time series and monthly time series data. Since the domain of our task is short-term prediction, we access daily time series data, which includes daily open price, daily high price, daily low price, daily close price and the daily volume. There were 1164 historical prices in the cryptocurrency datasets. A normalization process is applied to project the price chart to 10. In our case, the normalization function is defined as:

$$\text{scaled price} = log_2\left(1 + 1023 \times \frac{\text{input price}}{\text{maxPrice}}\right)$$

where $price$ is input price, $maxPrice$ is maximum price or price ceiling. Since logarithm of 1024 with base 2 gives output 10, prices are spread between 0 and 10.

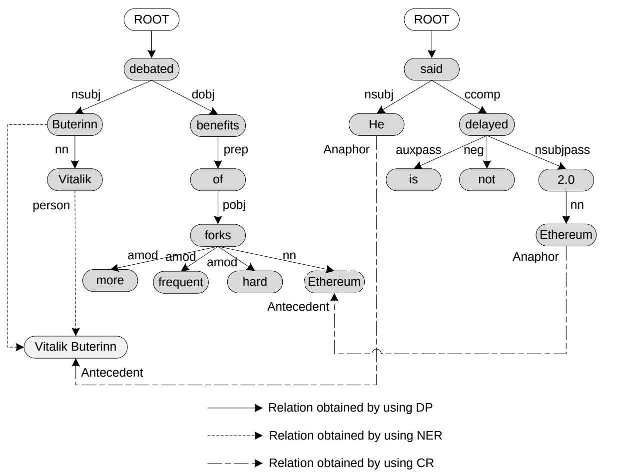

Fig. 3. Graphical dependency representation of the sentences "Vitalik Buterinn debated benefits of more frequent Ethereum hard forks. He said Ethereum 2.0 is not delayed." using dependency parser, co-reference, and named entity recognition.

3) Dataset labeling

The obtained news datasets are manually labelled with their sentiment scores relying on certain rises and drops. Fig. 2 illustrates time series data of ETH from 18-Nov-2017 to 9–Sep-2018. In this example, the real normalized ETH prices are plotted in black trace. The sentiment scores are represented by the red trace where 1 indicate positive sentiment, 0 means neutral sentiment, and -1 represents negative sentiment. The sentiment score is assigned a positive sentiment before a rise. Similarity, sentiment score is assigned a negative sentiment before a drop. Otherwise, sentiment score is assigned a neutral sentiment.

B. Sentiment Analysis

In the first state, our system takes a set of news as input and produce score as output. We developed this stage using a set of natural language processing tools: dependency parser, co-reference, and named entity recognition. The sentiment detection process shares its probabilistic foundation with distributed representation of phrases and continuous bag of words in word2vec. We parametrize the model in a manner, that aims to capture both semantic and sentiment similarities among word.

1) Semantic vector

Assume that we are at a time t and aim to compute the sentiment for time $t + 1$. We generate a vector that represents content of news in a window of size m until time t. Our model's semantic component shares its probabilistic foundation with word2vector model. To generate semantic vector, a coarse semantic vector is created by combining all words that occur in the above data. Subsequently, semantic vector is extended by adding bigram and trigram features. For each occurred word, the system looks at its neighbors in a windows size of $m = 2$ to capture the semantic information.

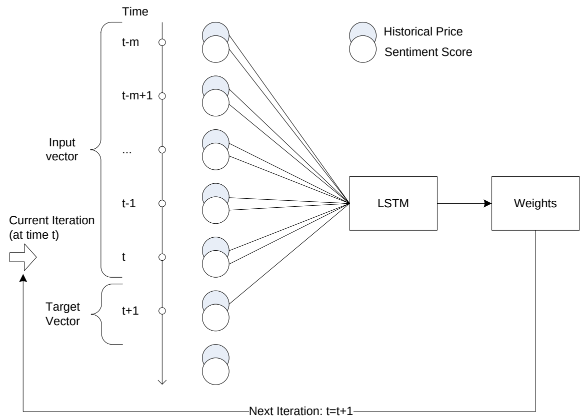

Fig. 4. Training process of LSTM to predict ETH price.

2) Syntactic vector

The syntactic vector is built on our prior work on opinion-aspect relationship analyses [2], [3]. To capture the sentiment semantic similarities among the words, the system consider not only the nearby word, but also looks for global contextual information. For this reason, cross-sentence context can provide a sufficient information source for enriching vector over the whole news data. The important step in our model is the application of a set of NLP algorithms to generate the syntactic. The phrase structure and dependency representation are the two main varieties of syntactic annotation. In general, the phrase structure representation contains clear constituent structures and is suitable for language representation. In contrast, dependency representations can be better for languages largely independent of word order. Fig. 3 illustrates the graphical dependency representation for an example. In this work, we used dependency representation to capture the semantic information. First, a DP is applied to detect relationships between words within sentences to find related words. Second, CR is used to precisely identify the participating term or entity. Third, NER is used to detect the types of terms. These processes increase the coverage of the semantic information extraction when information is sparse. Fig. 3 also shows how CR, NER are applied to the dependency representations.

The semantic vector and syntactic vector are then combined to generate an input vector. Simultaneously, the obtained labelled datasets is used to assign a target vector, which is a scalar which can be either 1, -1 and 0. Finally, a LSTM network is used to predict the sentiment score.

C. Price Prediction

In this stage, the sentiment score from the first stage is fed back as additional input into a RNN network to predict cryptocurrency prices. The prediction is built on a LSTM network, which is well-suited for making predictions based on time series data. The LSTM takes historical cryptocurrency price datasets and obtained sentiment scores as input and produces a predicted price as output.

Assume that we are at a time t and we aim to predict the price of cryptocurrency at time $t + 1$. First, the input vector is generated by collecting historical price data in a window of size n until time t. Subsequently, the input vector is enriched by adding obtained sentiment score. Simultaneously, the price at time $t+1$ is assigned as the target vector. The training process of LSTM-neural network is shown as in Fig. 4. Training a network requires a huge amount of data. We used 80% of the historical prices in the cryptocurrency datasets for training and the remaining 20% for testing.

Fig. 4. Comparison of different approaches to predict ETH price.

IV. Experiment and Discussion

We evaluated our model using daily time series data of ETH from 30-Jul-2017 to 5-Oct-2018. For sentiment analysis, we gathered the news of the past 7 days ($m = 7$). Similarity, we took into account the prices of past 7 days ($n = 7$) for price prediction. The ETH price prediction results are shown as in Fig. 5. The real normalized ETH prices are plotted in black trace. The red trace indicates the predicted prices using support vector regression. The yellow trace represents predicted prices using LSTM with historical prices. The blue trace means predicted prices using final version of the model. The experimental results show that there is a good accordance between the true and the estimated price. The worst performance of proposed model occurs when there is a large rise or drop in ETH price.

To quantify the performance of proposed method, we computed the errors of the predicted prices compared to the correct prices. There exist multitudes of performance evaluations for predictive analysis in the current literature [32]. Normalized error is a statistical evaluation used to compare proficiency testing. However, we used a modified form of mean absolute percentage error (MAPE) and mean absolute normalized error (MANE), which is defined as:

$$mANE = median(E)$$

where E is absolute normalized errors measured at all samples. Each absolute normalized error E_n in E is computed as:

$$E_n = \left| \frac{\partial_{Pre}(n) - \partial_{Cor}(n)}{\partial_{Cor}(n)} \right|$$

where $\partial_{\mathrm{Pre}}(\mathrm{n})$ and $\partial_{\mathrm{Cor}}(\mathrm{n})$ are predicted price and correct price of test sample n, respectively. The $mANE$ was used because the MANE is too sensitive for outliers. For comparison, tested versions of our model using LSTM (historical prices only), the full model LSTM (historical price and sentiment analysis features), and an implemented support vector regression with a Gaussian kernel. As shown in Table I, we achieved satisfactory experiment results. Even without the use of sentiment information, the LSTM outperforms SVM because LSTM networks are well-suited to making predictions based on time series data. Our model performs best when combined with sentiment analysis features.

TABLE I: Prediction Results for mANE

Method	Features	mANE (%)
SVM	Historical prices only	3.52
Our LSTM	Historical prices only	2.47
Our Full LSTM	Historical price + Sentiment analysis	1.36

V. Conclusion

Our study addressed cryptocurrency price prediction using a natural language processing approach. We present a model that is able to directly predict price direction by indicating whether to buy, sell, or hold. The final version of the model was able to correctly predict cryptocurrency price using historical data and sentimental information gained from news data. The important key in our model is the application of a set of natural language processing algorithms to identify the public moods for cryptocurrency fluctuations. We showed that sentiment analysis is an important perspective for cryptocurrency price prediction due to the interactive nature of financial activities. This strategy offers a new approach to addressing cryptocurrency price prediction and can be applied to challenging tasks such as social media tracking, popularity, trending, and sentiment analysis based event detection.

References

[1] E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, vol. 38, no. 1, pp. 34-105, 1965.

[2] A. D Vo, Q. P. Nguyen, and C. Y. Ock, "Opinion–Aspect relations in cognizing customer feelings via reviews," *IEEE Access*, vol. 6, p. 6, 2018.

[3] A. D Vo, Q. P. Nguyen, and C. Y. Ock, "Automatic knowledge extraction for aspect-based sentiment analysis of customer reviews," in *Proc. the 10th International Conference on Computer Modeling and Simulation*, 2018, pp. 110-113, ACM: New York, NY, USA.

[4] B. S. Kumar and V. Ravi, "A survey of the applications of text mining in financial domain," *Know.-Based Syst.*, vol. 114, no. C, pp. 128-147, 2016.

[5] J. Bollen, H. Mao, and X. J. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, 2010.

[6] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. the 24th International Conference on Artificial Intelligence*. AAAI Press: Buenos Aires, Argentina, 2015, pp. 2327-2333.

[7] D. R. Michael, "Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction," University of Iowa, 2014.

[8] N. Milosevic, "Equity forecast: Predicting long term stock price movement using machine learning," *CoRR*, 2016.

[9] S. Shen, H. Jiang, and T. Zhang, "Stock market forecasting using machine learning algorithms," Department of Electrical Engineering, Stanford University, 2012.

[10] C. K. Leung, R. K. MacKinnon, and Y. Wang, "A machine learning approach for stock price prediction," in *Proc. the 18th International Database Engineering & Applications Symposium*, ACM: Porto, Portugal, 2014, pp. 274-277.

[11] Y. Dai and Y. Zhang, "Machine learning in stock price trend forecasting," Stanford University: Stanford University, 2013.

[12] L. Menkhoff, "The use of technical analysis by fund managers: International evidence," *Journal of Banking & Finance*, vol. 34, 2010.

[13] L. Menkhoff and M. P. Taylor, "The obstinate passion of foreign exchange professionals: Technical analysis," *Journal of Economic Literature*, vol. 45, 2007.

[14] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *Journal of Finance*, vol. 55, 2000.

[15] C. H. Park and S. H. Irwin, "What do we know about profitability of technical analysis," *Journal of Economic Surveys*, 2007.

[16] S. Schulmeister, "Profitability of technical stock trading: Has it moved from daily to intraday data?" *Review of Financial Economics*, vol. 18, 2009.

[17] F. Westerhoff, "Technical analysis based on price-volume signals and the power of trading breaks," *International Journal of Theoretical and Applied Finance*, vol. 09, 2006.

[18] P. D. Turney and P. Pantel, "From frequency to meaning: vector space models of semantics," *J. Artif. Int. Res.*, vol. 37, no. 1, pp. 141-188, 2010.

[19] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *J. Mach. Learn. Res.*, vol. 3, pp. 1137-1155, 2003.

[20] T. Mikolov, I. Sutskever, and K. chen, "Distributed representations of words and phrases and their compositionality," in *Proc. the 26th International Conference on Neural Information Processing Systems*, Curran Associates Inc.: Lake Tahoe, Nevada, 2013, pp. 3111-3119, vol. 2.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, 2013.

[22] I. Chaturvedi, Y. S. Ong, I. W. Tsang, R. E. Welsch, and E. Cambria, "Learning word dependencies in text by means of a deep recurrent belief network," *Knowl.-Based Syst.*, vol. 108, pp. 144-154, 2016.

[23] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, pp. 102-107, 2016.

[24] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, pp. 74-80, 2017.

[25] S. Poriaa, E. Cambriab, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42-49, 2016.

[26] N. Majumde, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, pp. 74-79, 2017.

[27] A. D Vo. and C. Y. Ock, "Sentiment classification: A combination of PMI, sentiwordnet and fuzzy function," in *Proc. the 4th International Conf. on Computational Collective Intelligence: Technologies and Applications*, Ho Chi Minh City, Vietnam: Springer-Verlag, 2012, pp. 373-382, vol. 2.

[28] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic trading of cryptocurrency based on twitter sentiment analysis," Stanford University: CS229 Project, 2015.

[29] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency price prediction using news and social media sentiment," in *Proc. Conference Proceedings*, 2017.

[30] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," in *Proc. Conference Proceedings*, 2018.

[31] R. C. Phillips and D. Gorse, "Predicting cryptocurrency price bubbles using social media data and epidemic modelling," in *Proc. 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1-7.

[32] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, p. 8, 1992.

Cheol-Young Ock is a professor of the School of IT Convergence, University of Ulsan, Rep. of Korea. He received his BS (1982), MS (1984), and PhD (1993) degrees in computer engineering from the National University of Seoul, Rep. of Korea. His research interests include natural language processing, machine learning, text mining, word sense disambiguation, and sentiment analysis.

He has been a visiting professor at the Russia Tomsk Institute, Russia (1994), and Glasgow University, UK (1996). He was also a chairman of sigHCLT (2007 to 2008) in KIISE, Rep. of Korea. He has been a visiting researcher at the National Institute of Korean Language, Rep. of Korea (2008). He received an honorary doctorate from the School of IT, National University of Mongolia, (2007), and earned a medal for Korean development from the Korean government (2016). He has been constructing a Korean WordNet, namely Ulsan Word Map (UWordMap) since 2002.

Anh-Dung Vo received his B.Sc. degree in computer science from the Hanoi University of Science and Technology, Vietnam, in 2010 and M.Sc. degree in computer science from the University of Ulsan, Rep. of Korea, in 2013. Currently, he is a Ph.D. candidate at the University of Ulsan, Rep. of Korea. His current research interests include natural language processing, machine learning, text mining, web mining, and sentiment analysis.

Quang-Phuoc Nguyen is a Ph.D candidate of University of Ulsan, Rep. of Korean. He received his BS degree from Vietnam National University, Ho Chi Minh City - University of Natural Sciences, Vietnam in 2005 and MS degree from Konkuk University, Rep. of Korean in 2010 both in information technology. His research areas are natural language processing, machine learning, word sense disambiguation, and machine translation.