Predicting Ethereum Prices using Machine Learning and Block Chain Information

Emergent Research Forum (ERF)

Han-Min Kim  
Sungkyunkwan University, Korea  
hanmin8809@gmail.com

Gee-Woo Bock  
Sungkyunkwan University, Korea  
gwbock@gmail.com

Gunwoong Lee  
Sungkyunkwan University, Korea  
gwlee@skku.edu

Abstract

With the growing interest in cryptocurrency and its algorithm, studies on cryptocurrency price predictions have been extensively conducted in various academic disciplines. Since the cryptocurrency is generated and consumed by the Blockchain system, it has been considered that Blockchain-specific information would be the main components in predicting cryptocurrency prices. Specifically, this point of view has been largely employed in the studies of Bitcoin price predictions. However, this study recognizes that Ethereum, a popular and leading cryptocurrency in the market, has distinct Blockchain information as compared to that of Bitcoin. We attempt to investigate the relationships between inherent Ethereum Blockchain information and Ethereum prices. Furthermore, the research examines how Blockchain information of other coins in the market is associated with Ethereum prices. The results of data analysis show that Ethereum Blockchain information and Blockchain information of other coins have strong correlations with the final Ethereum prices.

Keywords
Ethereum, Prediction, Machine Learning, Blockchain, Blockchain Information

Introduction

With the growing interest in Blockchain, cryptocurrency, a component of the Blockchain system, has also received increasing attention from academic researchers. Cryptocurrency is a compensation made when a new block is created in the Blockchain system and is registered formally in the Blockchain system (Antonopoulos, 2014). Currently, cryptocurrency creates new economic values beyond its purpose of operating the Blockchain system. Cryptocurrency is used not only for transactions between coins, but also for purchasing products and can also be exchanged in legal currency (Mensi et al., 2018). As the cryptocurrency has attracted much attention, researchers have attempted to predict and analyze the price of cryptocurrency. Representative cryptocurrencies are Bitcoin and Ethereum, and the prices of these coins are oftentimes unpredictable by general economic indicators, unlike existing currencies or gold price (Ciaian et al., 2016). As price fluctuations of cryptocurrency change drastically, previous studies have attempted to predict the price of Bitcoin, a cryptocurrency of the first-generation Blockchain system. Previous studies on Bitcoin found that search information such as Google Trends and Wikipedia are related to Bitcoin prices (Kristoufek, 2013; Abraham et al., 2018). Ciaian et al. (2016) finds that market forces and Bitcoin attractiveness for investors and users affect Bitcoin prices.

On the other hand, recent literature on Bitcoin price prediction using Blockchain information has conducted considering that Bitcoin is generated and consumed in a Blockchain system. Prior studies revealed that Bitcoin’s Blockchain information and Bitcoin price are related by conducting machine learning technique (Saad & Mohaisen, 2018). Some studies also reported that Macroeconomic factors, global currency ratio, and Bitcoin Blockchain information are associated with Bitcoin price prediction (Jang & Lee, 2018; Mallqui & Fernandes, 2019). These studies have revealed that Blockchain information and
machine learning techniques can improve the overall performance of predicting Bitcoin prices (Jang & Lee, 2018; Saad & Mohaisen, 2018).

Recently several research works have attempted to predict Bitcoin prices, but predicting Ethereum prices has not been attracted much attention. Some studies have shown that social media information is associated with the bubble and fluctuation of Ethereum price (Kim et al., 2016; Abraham et al., 2018). Ethereum has four different characteristics in terms of Blockchain system compared to Bitcoin (Antonopoulos & Wood, 2018). First, unlike Bitcoin that contains only transaction information in a block, Ethereum has extensibility because it can contain various types of information in addition to transaction information in a block. Second, unlike the case where the Bitcoin does not pay the compensation coin for the uncle block, Ethereum provides some compensation coins for the uncle block. Third, Ethereum introduces the concept of gas to adjust the block size and speed up the system, unlike Bitcoin. Ethereum gas is the running cost of registering information in a block, it makes Ethereum consume in the Blockchain system. Also, the Ethereum block size cannot be increased like a Bitcoin block, because the limit of gas is set for each block. Finally, the average block generation time is about 15 seconds, because Ethereum is relatively smaller in size than Bitcoin and does not deal with much information. On the other hand, Bitcoin block is generated every 10 minutes on average.

Although Ethereum is a second-generation Blockchain featured with structural factors comparable to Bitcoin, there is lack of studies investigating relationships between Ethereum Blockchain information and Ethereum prices. Previous studies have not considered Ethereum's distinct Blockchain information (gas limit, gas usage, gas price, uncle block) and other coin’s information in predicting Ethereum price. This study recognizes the necessity of using Ethereum Blockchain information in predicting Ethereum prices because Ethereum has a structure that is issued and consumed in a Blockchain system. In addition, we consider that there is a possibility that other coin’s block chain information can be related to the price of Ethereum, considering that other coins with Ethereum are actively trading in the market (Songmuang, 2018).

Therefore, this study answers the following salient research questions. First, is Ethereum Blockchain information related to Ethereum prices? Second, is Ethereum's special Blockchain information such as compensation for uncle block, gas limit, gas price, gas consumption associated with Ethereum prices? Third, is the Blockchain information of other coins related to Ethereum prices? This study bears significant implications to both practitioners and researchers in the related domains by employing Ethereum’s special Blockchain information and Blockchain information of other coins for predicting Ethereum prices.

**Blockchain and Ethereum**

The Blockchain is featured with a system that allows all participants to easily share the transaction ledger and verify the reliability of the ledger (Antonopoulos, 2014; Antonopoulos & Wood, 2018). Blockchain can provide transaction trust and transparency to all participants of Blockchain, because it is very difficult to change after you enter information. The block in a Blockchain has a hash value that can be called a unique identity. The block has a header and a body portion. The header includes a nonce value to be found for generating a new block, difficulty of block creation, hash value of the previous block, time stamp, length number of the block, and merkle root. A hash value of the block is created by hashing the header information. The block body has transaction information. Blockchain has a structure in which blocks are connected to each other in a chain form, because block contains unique hash value of previous block when creating new block. The hash value of a block provides a completely different hash value if there is a slight change in the input value. Therefore, it is almost impossible for the attacker to manipulate the information registered in the Blockchain.

Generally, the coins in the Blockchain are paid as compensation to the miners who formally create new blocks and register them in the Blockchain network. This process is called Proof-of-Work (PoW). Coin can be paid for adding transaction information in a Blockchain network and can be used to exchange for other coins. As above-mentioned, Ethereum is a second-generation Blockchain system designed to contain various types of information as well as transaction information (Antonopoulos & Wood, 2018). Ethereum is making a significant contribution to the practical application of Blockchain technology because of the scalability.
Method

Data description and Experimental design

This study uses Ethereum datasets gathered from July 30, 2015 (the date of Ethereum issuance) to November 28, 2018. Price and Blockchain information of Ethereum (from https://etherscan.io/charts) were recorded on a daily basis. Macro-economy development and Global currency ratio data were obtained from the university database. The Blockchain information of other coins was selected by two criteria. First, we employ the coins’ Blockchain information with high transaction weights in the market. Second, this study uses Blockchain information of other coins that match time with Blockchain information of Ethereum.

This study selects independent variables related to Bitcoin price drawing on previous studies (Jang & Lee, 2018; Poyser, 2018). We also consider the gas and uncle block variables theoretically related to the issuance and consumption of Ethereum (Antonopoulos & Wood, 2018). Ethereum price will be the dependent (outcome) variable. Blockchain information of Ethereum and other coins, macro-economic development index, and global currency ratio are used as the independent variables. The research variables used in the study are summarized in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Data (Daily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>Price of Ethereum (in USD)</td>
</tr>
<tr>
<td>Blockchain Information of</td>
<td>Transaction volume (in USD), Transaction count,</td>
</tr>
<tr>
<td>Ethereum</td>
<td>Generated coins, Transaction fee, Active address,</td>
</tr>
<tr>
<td></td>
<td>Block size, Block count, Difficulty, Uncle block,</td>
</tr>
<tr>
<td></td>
<td>Gas limit, Gas price, Gas used</td>
</tr>
<tr>
<td>Blockchain Information of</td>
<td>Transaction volume (in USD), Transaction count,</td>
</tr>
<tr>
<td>other coins (Bitcoin, Dash,</td>
<td>Generated coins, Transaction fee, Active address,</td>
</tr>
<tr>
<td>Litecoin)</td>
<td>Block size, Block count, Difficulty</td>
</tr>
<tr>
<td>Macro-economic Development</td>
<td>S&amp;P 500, DOW30, Eurostoxx, NASDAQ, Crude oil, SSE,</td>
</tr>
<tr>
<td>Index</td>
<td>Gold, VIX, Nikkei225, FTSE100</td>
</tr>
<tr>
<td>Global Currency Ratio</td>
<td>GBP/USD, JPY/USD, CHE/USD, CNY/USD, EUR/USD</td>
</tr>
</tbody>
</table>

Table 1. The summary of data

The correlations between the price of cryptocurrency (Ethereum, Bitcoin, Dash, Litecoin) and the corresponding Blockchain information are presented in Table 2. From the correlation analysis, we confirm that the Ethereum price is related to Ethereum Blockchain information and other coins’ Blockchain information.

<table>
<thead>
<tr>
<th>Category (Ethereum)</th>
<th>Log (price)</th>
<th>Category (Bitcoin)</th>
<th>Log (price)</th>
<th>Category (Dash)</th>
<th>Log (price)</th>
<th>Category (Litecoin)</th>
<th>Log (price)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction volume</td>
<td>0.556**</td>
<td>Transaction volume</td>
<td>0.583**</td>
<td>Transaction volume</td>
<td>0.393**</td>
<td>Transaction volume</td>
<td>0.426**</td>
</tr>
<tr>
<td>Transaction count</td>
<td>0.872**</td>
<td>Transaction count</td>
<td>0.363**</td>
<td>Transaction count</td>
<td>0.071*</td>
<td>Transaction count</td>
<td>0.638**</td>
</tr>
<tr>
<td>Generated coins</td>
<td>-0.703**</td>
<td>Generated coins</td>
<td>-0.691**</td>
<td>Generated coins</td>
<td>-0.246**</td>
<td>Generated coins</td>
<td>-0.159**</td>
</tr>
</tbody>
</table>
Conclusion and Research Plan

Theoretical Implications

This study will bear theoretical implications to the related literature. First, this research identifies whether Blockchain information of Ethereum is valid/relevant in predicting Ethereum price. Previous studies have found that Bitcoin blockchain information is related to predicting Bitcoin price (Jang and Lee, 2018). However, this point of view was rarely taken into account in the Ethereum price prediction study. We will find the relationship between Ethereum prices and Blockchain information.

Second, this study will find additional variables of the Ethereum Blockchain related to the price prediction besides the Blockchain variables related to the Bitcoin price prediction. As mentioned earlier, Ethereum has a different Blockchain structure compared to Bitcoin such as gas limit, gas usage, compensation for uncle blocks, and so on. Despite these differences, specific Blockchain information of Ethereum is hardly applied to Ethereum price prediction. Therefore, the research will find additional Blockchain variables related to Ethereum prices.

Third, in addition to the Ethereum Blockchain information, we will find the relationship between Blockchain information of other coins and Ethereum price. Ethereum is actively trading with other coins in the coin market. Considering that the Blockchain information and the coin price are related, there is a possibility that the Blockchain information of the other coins is related to the Ethereum prices. Although there is an opportunity to discover additional relevant variables for Ethereum price prediction, the research on this perspective have not been conducted yet. In addition to the Ethereum Blockchain variables, we may find the relationship between the Blockchain variable of other coins and Ethereum prices.

Practical Implications

The practical implications expected from this study are as follows. First, we will be able to identify and utilize a set of key variables required to predict Ethereum prices. Practitioners could carry out proper investment and sales of Ethereum using the Blockchain information found in this research.

Second, the findings of this study will suggest that it is necessary to consider the Blockchain information when adjusting the Ethereum prices and issuing other coins. The Blockchain information related to the price may help to design the structure of the Blockchain.
Research Plan

This study will take advantage of time-series analysis and advanced machine learning techniques to predict the Ethereum prices more precisely. Previous studies have used non-linear methods such as Artificial neural network (ANN) (Zhou et al., 2017), Bayesian neural networks (BNN) (Jang & Lee, 2018), and Support vector regression (SVR) (Xiong et al., 2014; Wang, 2011), and Support vector machine (SVM) (Kaytez et al., 2015) for time series analyses. They found that these methods can provide improved performance in various time series data (Zhou et al., 2017; Jang & Lee, 2018). This study will employ ANN, BNN, SVR, and SVM analyses to predict Ethereum price drawing on previous studies. We are currently collecting additional datasets for recent periods and applying a 10-fold cross validation to ensure the reliability of analysis results. This study will employ the criteria of root mean square error (RMSE) and mean absolute percentage error (MAPE) to evaluate the results of performances.

REFERENCES


