Annual Spring Conference of KIPS

Anticipating Bitcoin Price Fluctuation Based on a Sentiment Analysis-Driven Indicator and Macro-Economic Data

주혜민[1], 양은혜[2], 최인수[1], 한승우[3], 김우창[1,*]

[1]한국과학기술원 산업및시스템공학과, [2]숭실대학교 산업정보시스템공학과, [3]서울대학교 조선해양공학과

Introduction

- In the midst of innovation surrounding blockchain technology, cryptocurrency has rapid price volatility, but it is settling in people seeking decentralization, a P2P network, and anonymity. Cryptocurrency is both new technology and a new currency, and the role of cryptocurrency in terms of global portfolio diversification can be positively moved forward.
- Bitcoin(Bitcoin) is based on a decentralized structure and provides trust functions such as safety and transparency. It also provides innovative utility, such as providing transaction anonymity through encryption through complex algorithms. In reality, transactions are carried out through bitcoin, and bitcoin itself is also being traded. For this reason, bitcoin can be incorporated into the portfolio, and high returns can be expected.
- In order to analyze the fluctuation of bitcoin, **this study identifies not only the sentiment analysis but also the relationship with other assets and conducts predictions through boosting models and LSTM**.

Methodology

- XGBoost (Chen, T., and Guestrin, 2016)
- LightGBM (Ke et al., 2017)
- CatBoost (Dorogush et al., 2018)
- Random Forest (Ho, 1995; Breiman, 1996)
- Long-Short Term Memory (Hochreiter, 1997)

Sentiment Analysis
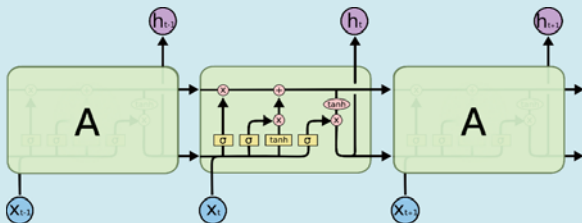
- **Sentiment analysis is a technique for grasping emotions about a specific topic by analyzing attitudes or tendencies to write or talk to text**. It is used to determine the sentiment state of texts occurring in online spaces such as articles, movie reviews, and reviews, or texts occurring in personal social network services such as Instagram, Facebook, Twitter, etc.
- In this study, we analyzed text represented by articles, editorials, and comments in the Reddit, and used a machine learning-based approach to perform sentiment classification using a trained model
- The methodology of machine learning is the most commonly used method in sentiment analysis and long-term memory (LSTM) as shown in the photo on the left, and the same text sensitivity score is calculated using the left formula by combining data and reference amount data

Data Features & Results

(1) Data Features
We collected the below data from 2014-01-01 to 2021-12-31. Bitcoin price, Bitcoin daily return rate, 10-year-bond yield rate. The change rate of 10-year-bond yield rate, GDP(PPP), and the Inflation rate of China, France, Germany, Kazakhstan, Korea, Turkey, UK, US, currency exchange rate for each country, and the sentimental score of Bitcoin, and whether Bitcoin price rises or not.

(2) NAN process
We filled out the blank cells by backfilling method and fulfilling method of 'fillna()' method.

(3) Result

	Bitcoin Only						With Macroeconomic Data and Sentiment Indicator				
Accuracy											
Period	XGB	LGBM	CAT	RandomForest	LSTM		Period	XGB	LGBM	CAT	Random Forest	LSTM
1	55.7%	54.8%	53.3%	53.3%	52.3%		1	53.6%	53.7%	54%	53.6%	54.5%
2	52.6%	53.0%	51.0%	56.9%	53.4%		2	57.2%	57.1%	57.3%	54.6%	53.9%
3	58.2%	58.2%	56.5%	58.7%	55.2%		3	61.6%	65.7%	60.2%	58.3%	57.6%
4	59.5%	58.0%	54.8%	61.4%	53.4%		4	71.6%	68.2%	61.2%	58.7%	69.3%
5	58.3%	62.1%	59.5%	65.2%	51.1%		5	74.0%	74.3%	64.5%	58.3%	51.2%

Period	XGB	LGBM	CAT	Random Forest	LSTM		Period	XGB	LGBM	CAT	Random Forest	LSTM
F1 Score												
1	68.7%	63.2%	69.1%	68.7%	68.3%		1	68.7%	61.7%	66.7%	68.7%	68.3%
2	69.7%	61.7%	70.0%	69.7%	69.0%		2	69.7%	67.8%	68.9%	67.7%	69.0%
3	70.7%	68.0%	71.3%	70.7%	70.9%		3	70.7%	74.0%	70.2%	70.7%	70.9%
4	70.7%	69.7%	71.5%	70.7%	69.3%		4	75.4%	75.6%	72.4%	70.7%	69.3%
5	70.0%	68.3%	70.0%	70.0%	67.1%		5	83.1%	81.6%	72.0%	70.0%	67.1%

Discussion & Conclusion

- This study researched the correlation between macro data, sentimental data, and bitcoin and predicted the fluctuation of bitcoin.
- Through this fluctuation prediction, it was confirmed that bitcoin is a somewhat predictable asset. It was also confirmed that macro data affect Bitcoin and that the boosting model brought the most remarkable accuracy, precision, and f1-score improvement.
- This research shows that predicting bitcoin moving is meaningful, using boosting, bagging, and deep learning model based on macroeconomic data, bitcoin price and volume data, and sentimental data.
- We analyzed the reason for **XGBoost's good performance, the regularization of the objective function** of XGBoost works well.
- This gives us the inspiration for cryptocurrency investment.

Reference

- Chen, S. S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. Journal of Banking & Finance, 33(2), 211-223.
- Gabralla, L. A., & Abraham, A. (2013). Computational modeling of crude oil price forecasting: A review of two decades of research. International Journal of Computer Information Systems and Industrial Management Applications, 5, 729-740.

ASK 2022 한국정보처리학회 KIPS Korea Information Processing Society