

Title	【課題研究報告書】ソーシャルメディア感情推定と機械学習を用いた仮想通貨の価格変動予測
Author(s)	Zhou, You
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18363
Rights	
Description	Supervisor : 岡田 将吾, 先端科学技術研究科, 修士(情報科学)

Abstract

Recently virtual currency market is drawing more and more attention. Comparing to stock market, price change in the virtual currency market is more drastic. Meanwhile, it is difficult to predict price in short term in the virtual currency market. Generally, only basing on market indices, which is a common practice in stock and foreign exchange (FX) markets, in many cases short-term price prediction takes a combination of multiple indices into account. However, beside markets, there are other types of factors usually influence price changes. Social media is one of them.

According to previous researches, price change of Bitcoin (BTC), a representative of virtual currencies, is known to be highly related to text contents which are posted on social media like Twitter, Reddit, and so on. In a previous research, a price prediction methodology with a machine learning algorithm was proposed. The proposal took results of sentiment analysis score from relevant tweets and virtual currency market data as features for the machine learning algorithm. The experiment results indicated that comparing to using market information only, by taking Emotion Induction results into account, some of models yielded to higher accuracies. However, it only extracted features from data of that single day. It didn't consider market direction changes in time sequence. Thus, the prediction accuracy is not good enough. Moreover, all tweets are processed with the same importance, thus it didn't reflect real sentiment.

My proposed research takes tweets from Jan. 1, 2020 to Dec. 31, 2020 and market information as research targets for virtual currencies BTC, ETH, XRP and LTC. Emotion Induction scores of each tweet are calculated with VADER. This research proposes to introduce weights of each tweet for sentiment score, Moving Average (MA) curves and their slopes for the predictions. Weights of tweets are calculated from number of followers, number of replies, number of retweets, number of Likes, overall number of tweets on that single day, etc.

Number of followers, number of replies, number of retweets and number of Likes are used to improve accuracy of the sentiment score. This proposed research utilizes 7 types of weights. sentiment score, MA curves of market information and their slopes reflect features in time sequence. In this proposed research, Simple Moving Average (SMA) curves and Moving Average Convergence/Divergence (MACD) curves with 3-day, 5-day, 7-day time spans respectively, as well as slopes of individual MA curves, are used.

A combination (weights of sentiment score and the MA type) with the best results is extracted with machine learning models MLP, SVM and RF. An analysis is done on the prediction results with a model using this combination to show case effectiveness of this proposed research.

Finally, prediction results of this proposed research are compared to previous researches. Since number of data and its duration used in this research differ to those in previous researches, the comparison can not be a quantitative one, but is one to compare prediction accuracies before and after Sentiment Induction in each research. This proposed research shows effectiveness comparing to previous researches in more scenarios.