Incorporating Social Media Sentiment Analysis to Enhance Cryptocurrency Forecasting

Modulus Global, Inc.- Scottsdale, Arizona

Introduction

As global financial markets have worked to grapple with uncertainty stemming from issues such as Covid19 and the U.S. Presidential election, investors are taking a more serious look at Cryptocurrencies. The recent resurgence in Bitcoin value has been impressive, setting records even higher than the 2017 spike, but not entirely surprising, considering the current state of the world and the decentralized nature of blockchain technologies.

Crypto traders use of variety of data sources to base their investment decisions on, but the significance of social media and message forum sites, such as Twitter and Reddit, is undeniable. Researchers around the world are more heavily investigating how to incorporate social media sentiment analysis into price forecasting models, with promising results (Mohapatra et al 2020; Wang & Chen 2020). At Modulus Global, we have a proven history in both sentiment analysis and cryptocurrency price forecasting using machine learning. Here we layout our plans to keep pace with bleeding edge research and update our forecasting systems accordingly.

System Description

(i.) Sentiment Analysis

Sentiment analysis is a branch of Natural Language Processing that focuses on understanding people's opinions. We can formalize this goal as follows:

$$f : M \rightarrow S$$

Where M is the message space we obtain data from (i.e. tweets) and S is an arbitrary sentiment space. There are numerous ways to quantify sentiment. More complex, emotion-based representations, $S = \mathbb{R}^n$, assign continuous real values to a set of n emotions (ie anger, sadness, etc.). Alternatively we can describe sentiment as a set of discrete values, $S = \{1, 2, 3, 4, 5\}$, or a range of continuous values, $S = [-1, 1]$, indicating negative, neutral, or positive sentiment.

(ii.) VADER

The VADER system, which stands for Valence Aware Dictionary for sEntiment Reasoning, is a rule-based sentiment model, developed by researchers at Georgia Institute of Technology (Hutto & Gilbert 2014). The weighted lexicon used in this model was created specifically for sentiment analysis in microblog or social media contexts. It functions by calculating positive, neutral, and negative scores for a body of text, in addition to a compound score ($S_{Compound}$). This compound score provides an effective and fast way to understand the overall sentiment of each message, $m \epsilon M$.

(iii.) Data Transformations

Our main objective is to incorporate sentiment from social media data into our existing forecasting model architecture. While sentiment scores are valuable in and of themselves, not all tweets are created equal. Therefore, we will weight each sentiment score according to Mohapatra et al (2020):

$$S_{Final} = (S_{Compound})(|U_{follow}| + 1)(|U_{like}| + 1)(|U_{share}| + 1)$$

Where U_{follow} represents the set of users who follow the original message creator, U_{like} represents the set of users who liked the message, and U_{share} represents the set of users that shared the message. We sum these final scores, $S_d = \Sigma S_{Final}$ where d represents a given day sampled, and normalize them $\overline{S} = S/\|S\|$ with respect to the whole dataset. To quantify the trend of sentiment surrounding cryptocurrency, we also include the linear momentum of sentiment:

$$p = \overline{S}v$$

where v is the rate of change in sentiment.

(iv.) Prediction Model

This final sentiment score and momentum are fed into our main price forecasting system:

$$f : (\overline{S}, p, T) \rightarrow \overline{P_{BTC}}$$

Where $T \epsilon \mathbb{R}^n$ represents the set of other technical indicators used in the dataset and $\overline{P_{BTC}}$ is the normalized price of Bitcoin. Through our experience using artificial neural networks for price forecasting, without social media data, we found that using relatively small training datasets led to more accurate short term predictions. However, we believe that the addition of sentiment information, especially our measure of momentum, will help our models identify indicators for sudden, large price fluctuations that we have struggled to capture until this point. Therefore, we will also test and compare results based on training data size.

References

Mohapatra, Shubhankar, Nauman Ahmed, and Paulo Alencar. "KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments." 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019.

Wang, Yu, and Runyu Chen. "Cryptocurrency price prediction based on multiple market sentiment." Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020.

Gilbert, C. H. E., and Erric Hutto. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf. Vol. 81. 2014.