

CS 230

Cryptocurrency Price Prediction Based on Twitter, News, Reddit and Trade Volume Data

Alex Fine

Introduction: Many papers have investigated the existence of relationships between social media data, news data, previous currency performance and bitcoin price fluctuation over discrete time periods. However, none have examined the exact relationship between these factors over continuous time periods. Additionally, papers either examine currency price prediction by classifying currencies as discrete entities or as a singular one. No papers have investigated the dynamic relations between them.

This paper aims to answer a few of the following types of questions:

1. How do different time periods of social media data and news data influence price change? For example, how might the past days news, and the past hours tweet data, influence the price change over the next six hours? This algorithm answers that question for 343 different time permutations.
2. How are the behaviors of different currency exchanges interconnected? How can we optimize Bitcoin data to predict the price change of Ethereum? This question is answered for 10,000 different currency permutations.
3. What data is most closely correlated to the positive price-change of bitcoin? This question will be answered for every currency.

The inputs for this algorithm are tweets, reddit forums, news articles and currency performance data over the past two years for all major currencies. First, this data will be aggregated for each major currency. Next, written text will be converted into usable data through a customized Word2Vec model. Data will then be analyzed using two distinct Neural Networks.

The first network is a deep Neural Net which optimizes for low variance. This network is designed to establish a general relationship between the aforementioned data and aggregate currency performance. The weights from this more general model are intended to initialize the weights of the second model.

The second model utilizes transfer learning from the first model and is customized price predictions per currency. The second algorithm is designed to optimize for high positive-prediction recall, as opposed to high accuracy. This double-network approach was chosen in order to keep variance low, yet fully optimize predictions per currency.

Milestone Report:

- Preliminary Shallow Model Created: I have a very preliminary shallow model created in tensorflow. This is intended to act as more of a structure when designing the larger model as opposed to a specific working model.
- Very detailed algorithm designed: I have created a *very* detailed plan for this algorithm and how every step will be executed. This plan makes the implementation much more trivial as the heavy lifting has been finished. I have put considerable thought into every facet of this algorithm. This outline is to the point, tight, and condensed. It is ~5 pages and attached in this zip.
- Data acquisition approach finalized: Though I have not finished my mass data collection, I have figured out exactly how I will do it, and have started the scripting algorithms. I have also prescribed the help of a friend to assist me in data collection. With my friend's help, I hope to have my finished dataset by next Friday.