

Bitcoin Stock Prediction Using Deep Learning and Sentiment Analysis

Balaji Mudaliyar | Neelam Babel

mudaliyar.b@husky.neu.edu

babel.n@husky.neu.edu

CSYE 7245 Big Data Sys & Intelligence Analytics, Spring 2018 Northeastern University

Abstract

Programmatically deriving sentiment has been the topic of many a thesis: it's application in analyzing 140-character sentences, to that of 400-word Hemingway sentences; the methods ranging from naive rule-based checks, to deeply layered neural networks. Unsurprisingly, sentiment analysis has been used to gain useful insight across industries, most notably in digital marketing and financial analysis.

An advancement seemingly more excitable to the mainstream, Bitcoin, has risen in number of Google searches by three-folds since the beginning of this year alone, not unlike it's exchange rate. The decentralized cryptocurrency, arguably, by design, a pure free market commodity – and as such, public perception bears the weight in Bitcoins monetary valuation.

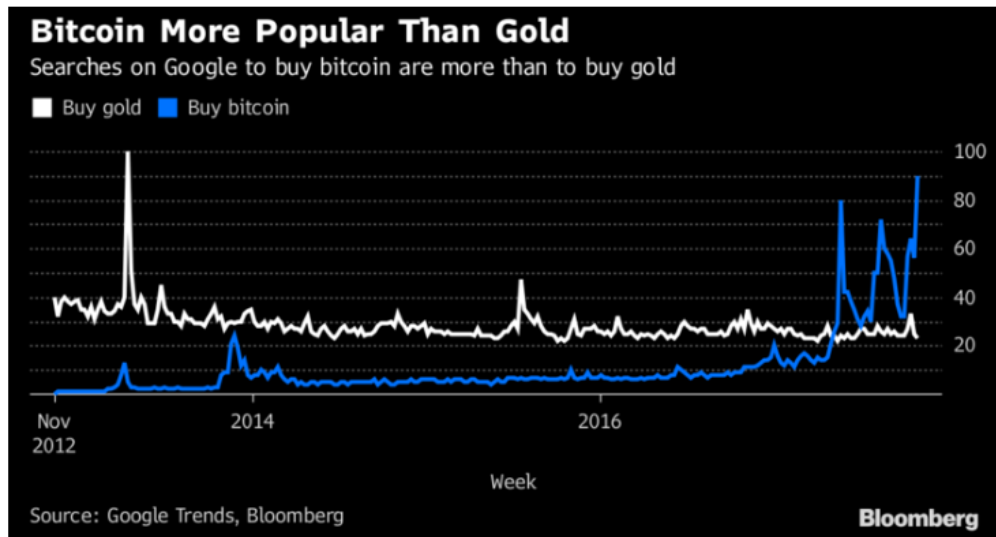
This thesis looks toward these public perceptions, by analyzing Bitcoin-related tweets for sentiment fluctuations that could indicate a price change in the near future.

Introduction

It's 2018, the people of the world generate 2.5 million terabytes of information a day. 500 million tweets, 1.8 billion pieces of shared information on Facebook, each and every day. These snippets of information regard anything under the sun; from what the user had for lunch, to their disgust over a referee in a football match. Twitter specifically has become known as a location where news is quickly disseminated in a concise format.

When regarding a financial commodity, the public confidence in a particular commodity is a core base of its value. Social media has served as platform to express opinions since their inception, and as such tapping into the open APIs provided of the likes of Facebook and Twitter, these arguably biased pieces of information become available with a sea of meta-data.

Bitcoin (BTC), the decentralized cryptographic currency, is similar to most commonly known currencies in the sense that it is affected by socially constructed opinions; whether those opinions have basis in facts, or not. Since the Bitcoin was revealed to the world, in 2009, it quickly gained interest as an alternative to regular currencies. As such, like most things, opinions and information about Bitcoin are prevalent throughout the Social Media sphere.



The motive behind this research is that there are a number of design flaws in Bitcoin, and people are trying to invent new coins to overcome these defects hoping their inventions will eventually replace Bitcoin. To June 2017, the total market capital of all cryptocurrencies is 102 billion in USD, 41 of which is of Bitcoin. Therefore, regardless of its design faults, Bitcoin is still the dominant cryptocurrency in markets. As a result, many altcoins cannot be bought with fiat currencies, but only be traded against Bitcoin. Hence, I chose Bitcoin as my commodity to make wiser future investments for my cryptocurrency portfolio.

Dataset

The datasets used in this project are from www.kaggle.com and <https://www.quandl.com/>.

Quandl.com:

Dataset set include the Adjusted Open, Adjusted High, Adjusted Low, Adjusted Close, Adjusted Volume for BTC, Adjusted Volume for Currency and Weighted Price for Bitcoin retrieved using [Quandl's free Bitcoin API](#) for dates ranging from January 7, 2014 to April 12, 2018.

Kaggle.com:

Dataset has been downloaded from the Kaggle website.

URL: <https://www.kaggle.com/jessevent/all-crypto-currencies>

Description: All historic open, high, low, close, trading volume and market cap info for all Cryptocurrencies.

Model Architecture Design

My first framework is a Recurrent Neural Network trained on 3 popular stock market indicators and past prices as key data points to find an optimal technique for cryptocurrency stock market prediction.

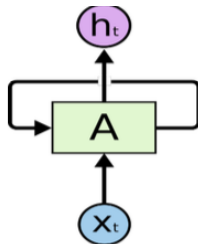
My second framework is a Recurrent Neural Network trained on sentimental analysis score calculated from twitter comments and past prices as key data points to find an optimal technique for cryptocurrency stock market prediction.

Comparison will be made based on their performance. Both techniques have some advantages and disadvantages. My research will analyze advantages and limitations of these techniques to find which technique is comparatively better for specifically Bitcoin stock market prediction.

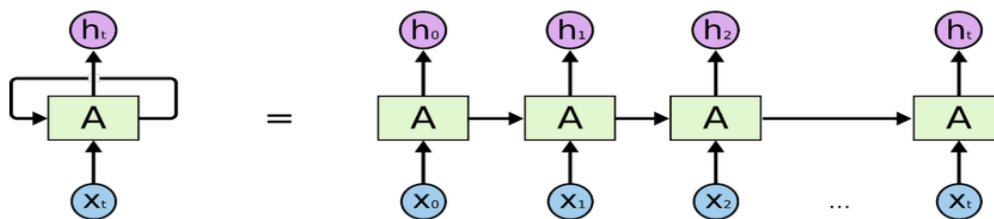
Also, prediction will be done using Random forest algorithm and it will be a forecast for the next 1 month.

Recurrent Neural Networks

Recurrent neural networks are networks with loops in them, allowing information to persist.



A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.



RNN's might be able to connect with previous information to the present task, i.e. using previous video frames might inform the understanding of the present frame.

Problem with RNN

In cases, where the gap between the relevant information and the place that's needed is small, RNN's can learn to use the past information. Unfortunately, as that gap grows, RNN's become unable to learn to connect the information.

Sentiment Analysis Approach to Crypto currency Speculation

We have analyzed user sentiments related to crypto currencies on social media, e.g., Twitter, and converted the feedback into scores. We then study correlation between the scores and fluctuations in price and trade volume to determine any relation.

Articles on crypto currencies, such as Bit coin, are rife with speculation these days, with hundreds of self-proclaimed experts advocating for the trends that they expect to emerge. So, we felt that analysis of the current trending tweets for the term Bit-coin to predict its closing price for the next day seemed like the most unbiased approach to resolving the biased opinions strewn around the web.

This approach also resonates with our personal approach to track the closing prices of the crypto currencies we have invested in. We always find ourselves skimming through the trending tweets on the matter.

Sentimental Analysis

Sentiment Analysis is the process of ‘computationally’ determining whether a piece of writing is positive, negative or neutral. It’s also known as **opinion mining**, deriving the opinion or attitude of a speaker.

Why sentiment analysis?

- **Business:** In marketing field companies use it to develop their strategies, to understand customers’ feelings towards products or brand, how people respond to their campaigns or product launches and why consumers don’t buy some products.
- **Politics:** In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!
- **Public Actions:** Sentiment analysis also is used to monitor and analyse social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

3 major steps are done in program:

- Authorize twitter API client.
- Make a GET request to Twitter API to fetch tweets for a particular query.
- Parse the tweets. Classify each tweet as positive, negative or neutral.

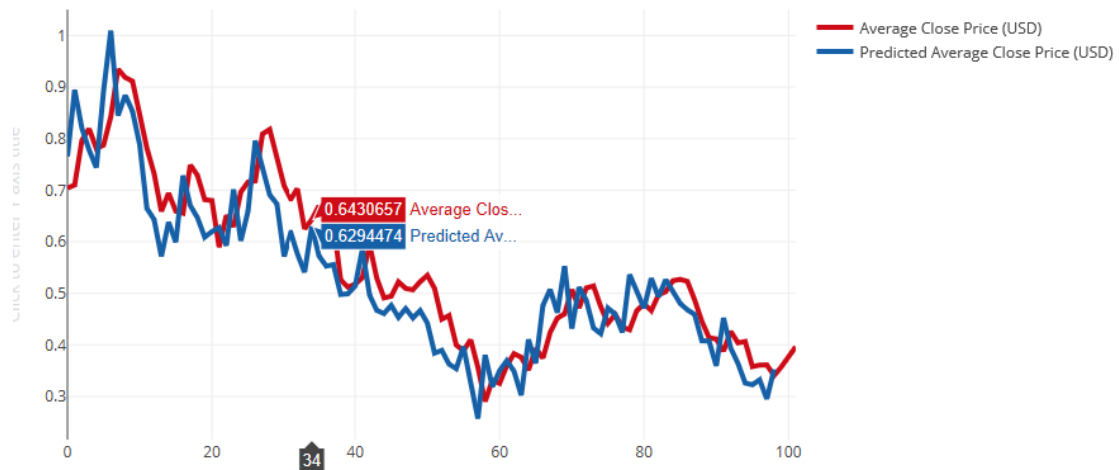
Code with Documentation

GitHub Link: <https://github.com/balaji-mudaliyar/Bitcoin-Prediction>

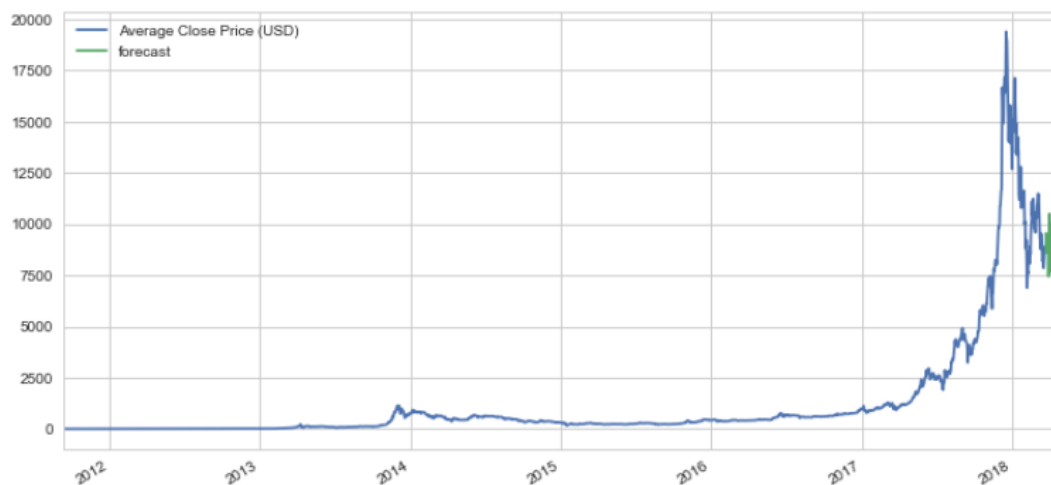
RESULT

1. Prediction using Recurrent Neural Network trained on Initial Features (Historical Bitcoin Prices) and Secondary Features (Fiat Currency Stock Market Technical Indicators)

Average Close Price (USD) vs Predicted Average Close Price (USD)



- Prediction using Random Forest algorithm. We predicted the bitcoin prices for the next one month.



Acknowledgment

We would like to show our gratitude to professor Nik Bear Brown for guiding us during this project.

Discussion

Despite their numerous strengths, neural networks suffer from a number of weaknesses that researchers should keep in mind.

Sample size: Neural networks need larger samples in order to be estimated properly. This is due to the large number of parameters introduced in such models that link the

inputs to the hidden neurons, which are then linked to the output variable. As the data available for the training and testing of the model increase, one would then expect the marginal gains in accuracy to increase. There is no rule of thumb rule for the “optimal” sample size for which one can expect neural nets to improve noticeably. For Sentiment Analysis, we have used Tweepy API to extract the data from twitter, limiting it to a count of 200 for testing our approach. Increasing the count should lead to more specific results.

References

- [1] Reid F, Harrigan M. An analysis of anonymity in the bitcoin system: Springer; 2013.
- [2] Böhme R, Christin N, Edelman B, Moore T. Bitcoin: Economics, technology, and governance. *The Journal of Economic Perspectives*. 2015;29(2):213–38.
- [3] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [4] Kondor D, Pósfai M, Csabai I, Vattay G. Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PloS one*. 2014;9(2):e86197 doi: [10.1371/journal.pone.0086197](https://doi.org/10.1371/journal.pone.0086197) [\[PMC free article\]](#) [\[PubMed\]](#)
- [5] Ron D, Shamir A. Quantitative analysis of the full bitcoin transaction graph *Financial Cryptography and Data Security*: Springer; 2013. p. 6–24.
- [6] Garcia D, Tessone CJ, Mavrodiev P, Perony N. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of the Royal Society Interface*. 2014;11(99):20140623. [\[PMC free article\]](#) [\[PubMed\]](#)
- [7] Kondor D, Csabai I, Szüle J, Pósfai M, Vattay G. Inferring the interplay between network structure and market effects in Bitcoin. *New Journal of Physics*. 2014;16(12):125003.
- [8] Kristoufek L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports*. 2013;3. [\[PMC free article\]](#) [\[PubMed\]](#)
- [9] Kristoufek L. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one*. 2015;10(4):e0123923 doi: [10.1371/journal.pone.0123923](https://doi.org/10.1371/journal.pone.0123923) [\[PMC free article\]](#) [\[PubMed\]](#)
- [10] Yelowitz A, Wilson M. Characteristics of Bitcoin users: an analysis of Google search data. *Applied Economics Letters*. 2015;22(13):1030–6.