Bitcoin Trading Strategy Using Alternative Data

ZIYUN CHEN

February 2020

Abstract

Quantitative investing strategies have been under-performing broader equities market and many other strategy types in the recent years. Investors have been actively looking to generate alpha signals with innovative methodologies and algorithms. Empirical researches suggest evidence of predictive power on financial market from alternative data sources such as web search data and twitter data. This project creates a framework to analyze and generate alpha signals from these alternative data sources.

1 Introduction

1.1 Background

The quantitative investing community has been trying to identify correlations and signals from alternative data since the beginning of the internet era. Tumarkin and Whitelaw (2001) examined relationships between online postings on a stock investing forum and the stock prices. However, factor investing based on modern portfolio theory dominated the mainstream. "Conventional" financial data including secondary market trading data as well as corporations' financial data were the focus of studies for most investors.

As more investors adopted similar factor based investing styles, market became crowded and alpha evaporated quickly. With a growing interest in "big data" and its application, investors started to look for new ways to identify signals. As the amount of secular data explodes in the form of text, speeches and images, many innovative quant funds have been working on the adoption of alternative datasets to generate trading signals to maximize returns. Popular approaches such as data mining web search data and text-based social media data are reviewed by researchers to predict financial market movements.

1.2 Empirical Research

Behavioral economics suggests emotions can affect individual behavior and decision-making. Bollen(2010) investigated that whether public mood can predict economic indicators. From the collection of 10mm tweets in 2008 posted by 2.7mm users, the researchers analyzed public sentiment and found that "calm" has the highest predictive power over DJIA index prices. Similarly Zhang and Yuan (2017) found that market related sentiment in Weibo data has strong correlation with stock return, turnover, and volume. Inclusion of Weibo sentiment improved the accuracy of a neural network model that the researches constructed. Alternatively, Zhong (2019) constructed an adaptive search-term rebalancing strategy that outperformed significantly market returns with google correlate service and trend data.

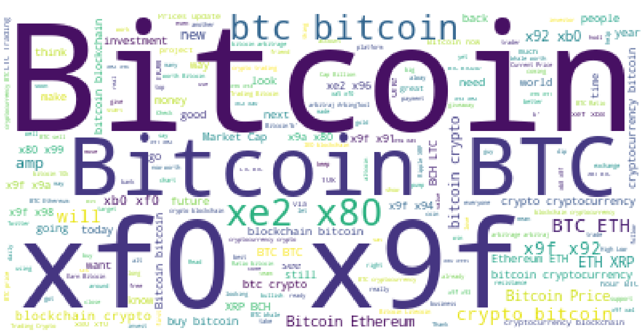
2 Project

The project builds a generic framework to gather alternative data and generate strategies based on the signals from these data samples. More specifically we model bitcoin price movements with alternative data along with popular technical indicators. The alternative dataset that is gathered and examined includes web search data (Google and Baidu) and text based social media data (Twitter and Weibo data). The paper proposes strategies of three different frequencies. The low frequency model examines daily price data based on which a daily trading strategy is proposed. The medium frequency model examines hourly data based on which the strategy will trade at every hour. The high frequency model on the other hand tries to create a strategy that trades every five minutes.

3 Data

This project creates models to predict bitcoin prices with both conventional and alternative data sets. For conventional data, the study will include historical pricing data as well as popular technical indicators including momentum features, simple moving averages, exponential moving averages, MACD, and bollinger bands. Pricing data will include the open, close, high and low for each time interval of investigation. To gather alternative data, server is setup in Google cloud to scrape historical google trend, Baidu trend and Tweets related to bitcoin. Due to the limitation of the need for consistent and unbiased samples of twitter data, we only include twitter sentiment in the high frequency model.

Table 1: Word Cloud of a Sample of Tweets



Optionally we will also include other financial market data including equity indices, bonds, and commodities in our model.

3.1 Low Frequency Model

The low frequency model examines daily historical bitcoin price data since March 2016. 1500 observations are available.

3.2 Medium Frequency Model

The medium frequency model examines hourly historical bitcoin price data since Jan 2019. 10000 observations are available.

3.3 High Frequency Model

The high frequency model examines historical bitcoin price data for every five minutes. For one week, about 2000 observations are available.

4 Plan

1. Milestone 1: Data Preparation
   (a) Web-search data: assemble Google and Baidu trend data of the three frequencies
   (b) Social media data: set up server in Google Cloud Platform to gather consistent Twitter data
   (c) Technical indicators: select and populate data for selected technical indicators
2. Milestone 2: Training and Testing
   (a) Feature selection and sentiment analysis on Twitter dataset
   (b) Design and implement deep learning models
   (c) Parameters fitting and model optimization
3. Milestone 3: Design Trading Strategies
   (a) Based on model result, build trading strategies of three frequencies
   (b) Back-test and Visualization
4. Final: Implement Strategies and Start Trading
   (a) Implement connection to bitcoin exchange and algorithm
   (b) Analyze trading result and presentation

5 References

1. Tumarkin, R., and R. Whitelaw. 2001. “News or Noise? Internet Posting and Stock Prices.” Financial Analysts Journal 57 (3): 41–51.

2. Bollen, Johan, Huina Mao, and Xiaojun Zeng. “Twitter Mood Predicts the Stock Market.” Journal of Computational Science 2.1 (2011): 1–8. Crossref. Web.

3. Zhang Xindong;Yuan Dongliang. “ Research on the Impact of Investor Sentiment on Stock Market Based on Micro-blog” Journal of Intelligence 2018-08

4. Zhong, X., Raghib, M. Revisiting the use of web search data for stock market movements. Sci Rep 9, 13511 (2019). https://doi.org/10.1038/s41598-019-50131-1