Comp 576 Final Project Proposal

Cryptocurrency Predictive Analysis

Minjun Park, Erin Song, Hao Xiang, Ying Xiong

November 2019

Background/ Motivation

Cryptocurrency payment systems such as bitcoin are gaining a lot of attention and adoption since their first emergence around the year 2009. Bitcoin has a current market capitalization of 9 billion USD and has over 250,000 transactions taking place per day. The decentralization of cryptocurrencies has greatly reduced the level of central control over them, impacting international relations and trade. Cryptocurrency predictive analysis is an interesting time series prediction problem as the market is still in it transient stage. Further, wide fluctuations in cryptocurrency price indicate an urgent need for an accurate way to forecast the price.

State of the art methods/ Tension/ Detailed related work

● Overview of the existing methods

Research on cryptocurrency price forecasts is at an early stage. Various methods are adopted to analyze the price behaviors of cryptocurrency; method to predict cryptocurrency price by considering various factors such as market cap, volume, circulating supply, and maximum supply based on deep learning techniques such as the recurrent neural network (RNN) and the long short-term memory (LSTM),which are effective learning models for training data, with the LSTM being better at recognizing longer-term associations.

One example for the price prediction is Shah and Zhang's application of Bayesian regression to Bitcoin price prediction which achieved high accuracy [6]. However, there remain other features, such as mining speed, that should be considered in the analysis.

● Challenges and Limitations

High Volatility: Due to high variance, validating machine learning predictions onto Bitcoin data remains a difficult task. In particular, price behavior of cryptocurrency is oftentimes counter-intuitive, making the price prediction challenging.
Lack of seasonality: classical approaches such as Holt-Winters exponential smoothing for time series prediction problems are dependent on linear assumptions, and these approaches may not be useful for predicting cryptocurrency price since there is no seasonal effect in cryptocurrencies, which are highly volatile in nature.

Broader scope and Goal

1. Use deep learning techniques, such as LSTM and RNN, and an ensemble learning method, random forest, to predict the nonlinear nature of cryptocurrency price
2. Identify useful parameters and features that lead to higher prediction accuracy
3. Based on the best model, design a decision strategy for buying or selling cryptocurrency

Proposed solution and major contributions

● Issue unaddressed and solution
Based on the paper [1] , Sliding Window validation approach is an optional technique which we will further explore more in our project.
In most studies, they predict future price movements and consider them as trading signal, converting directly the trend into an action for the investor. However, price predictions are not market actions. Moreover, the models do not predict the amount we should buy or sell for each cryptocurrency. So to build a real strategy, we need to consider another type of machine learning, Reinforcement learning, that simulates a real artificial intelligence.

● Major contributions
Random forest and generalized linear model is used by the paper [2]. In the paper, they mentioned extremely small and rapid changes in bitcoin is often due to the erratic nature of bitcoin fluctuations. Random forest outperformed GLM due to the RF's use non-parametric of decision trees, and thus outliers and linear separability of the data are not concerns to this model.
The Long short Memory network was slightly more accurate than the multi-layer Perceptron(MLP) network in this paper [4].

Proposed experiments

1) Dataset

Our dataset contains the historical price information (from January 2012 to March 2018) of top cryptocurrencies such as Bitcoin, Ethereum and etc (15 in total). The dataset is obtained from

Kaggle, and originated from coinmarketcap, Blockchain info and Etherscan [5]. For each type of cryptocurrency, the data file is stored in a csv format, with information such as the number of transactions per day, daily price of the coin, total value of the currency in circulation and many others (~20 per currency). In addition, we also have the following information for a given day:

1. The opening price
2. The highest price
3. The lowest price
4. The closing price
5. The number of transactions
6. Market capitalization in USD

2) Model and method

In general, modeling construction may go from basic neural network, to RNN with LSTM (or GRU) layer, to RNN with GRU layers with various ways of dropout. For the model architecture, we will be using Long Short-Term Memory(LSTM), Recurrent Neural Network(RNN), Ensemble Model. For the methods, we will implement variable selection method - Boruta (built around the random forest classification method - and Random Decision Forest.

Project execution plan

For the execution part of the project, we have decided to organize our tasks into weekly plans. The first week would mostly be investigation - data preprocessing (Hao & Ying) and literature review (Erin & Minjun) for finding novel techniques. We then plan to perform variable selection and implement different models for comparison (everyone). Finally, our last step would be to organize the findings and results into presentable formats such as report, and poster (everyone).

<u>Tasks</u>:
Report, Poster, Presentation, Code Compile (Git Repo)

<u>Roles</u>:
Everyone will participate in coding and problem solving.
Erin - Literature Review, Report
Ying - Data Processing , Poster
Hao - Data Processing , Presentation
MinJun - Literature Review , Code Compile (Git Repo)

Plan:
Week 1: Data Processing / Literature Review
Week 2: Variable Selection / Model
Week 3: Compiling
Presentation & Report

Feasibility

Based on the paper [1], which compared the performance time of training models on GPU and CPU, we assume GPU to outperform CPU. The paper used Intel Core i7 2.6GHz CPU and the utilized GPU was NVIDIA GeForce 940M 2GB. The challenge lies in our team's machine not being equipped with a GPU. After experimenting training with CPU, if we conclude that it is not feasible, we may attempt to set up a virtual machine with GPU.

Potential Impact

We will find out the best algorithmic mechanisms which help anticipate the short-term evolution of the cryptocurrency market. And we will further design a simple trading strategy assisted by state-of-the-art machine learning algorithms which could outperform standard benchmarks.

Reference

1. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8374483
2. http://cs229.stanford.edu/proj2014/Isaac%20Madan,%20Shaurya%20Saluja,%20Aojia%20Zhao,Automated%20Bitcoin%20Trading%20via%20Machine%20Learning%20Algorithms.pdf
3. https://pdfs.semanticscholar.org/ff94/fe775276b8557daf638a5aefed15bad9af85.pdf
4. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8780358
5. https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory/data
6. https://arxiv.org/abs/1410.1231