CS221 Project Proposal
Group Members: Ellen Sebastian, Abaho Katabarwa, Guoxing Li
Project Title: How to Succeed with Bitcoin Without Really Trying

Introduction

Bitcoin is the world's first decentralized cryptocurrency. Since it was created by Satoshi Nakamoto in 2008, Bitcoin has become a popular international digital payment system because of its ability to bypass national financial regulatory bodies and stay resistant to the fluctuations of traditional currencies. Furthermore, the existence and volatility of bitcoin has created a vibrant and profitable electronic trading market. We would like to take advantage of various concepts in artificial intelligence and the vast amount of transactional data surrounding Bitcoin to create a Bitcoin trading bot whose goal will be to maximize profit by predicting future Bitcoin prices.

Evaluation Metric

Given one has an *a priori* knowledge of the future price of Bitcoin, there is a theoretical maximum profit that one can attain. Our evaluation metric shall be determined by grading the profit gained by our bot based on this theoretical upper bound.

Input And Output Behavior:

The input will include information about the market over time (since 2011) and recently (the last *n* transactions) and a sentiment index about bitcoin. The user can also provide a time limit (how long will they wait to buy/sell) and a cap on the amount of money they want to spend. We also need the user's current inventory of bitcoins in order to make sure we don't recommend they sell more than they have. We also need the current time in context of the example, which must be no later than (actual current time - time limit) in order to assess the quality of our prediction.

> *Note: There is a large amount of data related to bitcoin (ex. bitcoin transactions, various exchange data, bitcoin mining data). Therefore the nature of our inputs will change as we progress through the project.*

Input example:

> *current_time = Jan 1 2014; Time limit = 2 months; current_inventory = 100 bitcoins; spending_limit = $2000; bitcoin price between 2011 and 2014; most recent transactions; bitcoin sentiment index for the last 10 days*

Above is an example of the input data used to predict the bitcoin price and therefore the recommendation on whether to buy or sell.

Our output will consist of recommendations about how many bitcoins to buy or sell and at what time. This is based on our prediction of what the price will be at any point in time as well as our confidence in predictions. The more confident we are, the more bitcoins we will recommend the user buys/sells.

> *example: You should sell all 100 bitcoins on Jan 6, 2014. This produces an expected profit of $88,166.*

Data:

To test our baseline and orable, we gathered data from https://blockchain.info, which includes daily data on market price, number of transactions, trade volume, difficulty of mining a bitcoin, etc.

Note: The oracle and the baseline were geared toward predicting the future stock price of bitcoin. This is a subproblem of the general project goal which is to make recommendations to buy and sell. The idea is if we can predict the price, there is a simple extension of the problem that accounts for the recommendation to buy/sell.

Baseline:

The baseline is created by creating a simple moving average over the market price data of bitcoin for a window of size n days (n = 5 in our case). This moving average is defined to be the predicted price of bitcoin for the next day. We tested on 100 testing examples, and computed the mean squared error of the difference between predicted and actual prices. The MSE for the baseline was 459.7.

Oracle:

For the Oracle, we chose to give the model extra information. For day *d*, we created a feature vector composed of all metrics found at https://blockchain.info/charts for *d-100* to *d+100*. For the bitcoin price metric, we only included values for *d-100* to *d-1*. We used these feature vectors to train a Stochastic Gradient Descent model to predict Bitcoin's price for *d+1*. We trained and tested this model on 215 days between February 28, 2013 and July 13, 2014 which had complete data available. We acheived 0 training MSE and 34.124 leave-one-out-cross-validation MSE, much lower than the Baseline's error. Because of the success of the oracle and the availability of bitcoin metrics, we believe that bitcoin price prediction is feasible.

Challenges:
- **State space is very large**. If we chose to make a state for every possible decision, there would be an enormous magnitude of search space that we do not have the resources nor knowledge to traverse.
- **The market is not deterministic**. We need to use domain-specific information and trust our intuition about what we think impacts Bitcoin prices. We cannot predict some important political events, such as if bitcoin becomes illegal in the United States in the future.
- **input metrics are not always correlated with Bitcoin price.** By looking at https://blockchain.info/charts, you can see that most metrics do not match up with the Bitcoin market price (USD). This makes it more difficult to predict prices based on these metrics.

Possible approaches:
- **Machine learning** - We could potentially use machine learning on past data points to take advantage of some kind of pattern.
- **Games:** We could model the bitcoin market as a single player and ourselves as an agent. The market makes decisions that change the state of gameplay and we act to maximize our final score (profit).
- **MDP:** We could define an MDP where the start state is the current state of the market and our bitcoin inventory; transitions are buying and selling at a certain time; and rewards are the amount of money we gain.
- **Bayesian regression -** A recent paper used this methodology with some success.

Previous work:
- **http://www.btcpredictions.com/** uses a neural network to predict bitcoin prices up to the next 20 days. article: cryptocoinsnews.com/bitcoin-price-prediction-tool-amazingly-accurate/
- https://www.cryptocoinsnews.com/take-bitcoin-price-predictions-grain-salt/ "Bitcoin price predictions fail because humans have limited mental capacities and can never comprehend the intricacies of the Bitcoin economy" Our algorithm will have a higher limit on its mental capacities and will be better able to comprehend the intracacies of the Bitcoin economy.
- http://arxiv.org/abs/1410.1231 - uses Bayesian regression to trade bitcoin