

STAT 497 - Final Project

Concordia University

Bakr Abbas, Matthew Liu, Frederic Siino

Day Month Year

0 Introduction

One of the main challenges of managing a portfolio or a fund, is the decision process of how to allocate investment resources to each assets. The goal is to maximize returns but also to diversify, putting all the fund in the most profitable asset could be beneficial in the short run, but the possibility of that one asset crashing in value is larger than the possibility of multiple assets crashing in value. So diversifying is important to hedge against that outcome, and portfolio management in the modern day involves more automation than ever. Reinforcement & Deep learning methods are being used to optimize this decision making process.

Cryptocurrencies are decentralized digital assets that behave as an alternative to government issued money. They are being introduced into most investment portfolios, this paper address the challenge of managing five different cryptocurrencies with Bitcoin(BTC) being cash. Meaning that the four other currencies are quantified as the price relative to BTC. Different methods will be used to manage the weights assigned to each asset in the portfolio. The first three methods will be based on the contextual bandits model. The final method will be based on a policy gradient model. A more realistic model would also take transaction costs into account, but this will be overlooked for simplicity.

A previous research paper (Jiang, Xu, Liang, 2017) has been published on this exact topic based on a framework consisting of Ensemble of Identical Independent Evaluators (EIE), a Portfolio-Vector Memory (PVM), an Online Stochastic Batch Learning (OSBL) scheme and a fully exploiting strategy. The actions were determined using a policy gradient method and the tested evaluators were a Convolutional Neural Network (CNN), a Recurrent Neural Network (RNN) and a Long Short-Term Memory Network (LSTM). Experiments were run with a trading period of 30 minutes and the RL framework was observed to behave far better than other optimization frameworks in the test period of 2017 but inferior to some frameworks in the test period of 2018.

In this project, we opt for a much simpler approach, relying instead on an action preference function and a policy gradient when framing the task as a contextual bandits problem. We do however borrow some aspects of Jiang, Xu and Liang such as definitions for states and returns.

1 Problem description

In this project, the environment is solely dictated by the data of the historic prices of all cryptocurrencies relative to BTC. In the experiments, a trading period of $T = 120$ minutes is used, and we only consider the closing prices of each period (which are equivalent to the opening prices of the next period). In addition, the experiments rely on two key assumptions (Jiang, Xu, Liang, 2017):

- Instant liquidity: The liquidity of the market allows each trade to be carried out immediately at the price listed when the order is placed.
- No market impact: The capital invested by the agent is so insignificant as to have no impact on the market.

This means that the sequence of states and transition probabilities are already predefined based on the data at hand. In particular, we define a price relative vector y_t of the t th trading period like Jiang, Xu and Liang, namely:

$$y_t := v_t \oslash v_{t-1} = \left(1, \frac{v_{2,t}}{v_{2,t-1}}, \dots, \frac{v_{m-1,t}}{v_{m-1,t-1}}, \frac{v_{m,t}}{v_{m,t-1}}\right)^T$$

where $v_{t,i}$ denotes the closing price of the i th asset in the t th trading period. This price relative vector can then be used to calculate the reward, defined as

$$r_t = y_t \cdot w_{t-1}$$

where w_{t-1} is the portfolio weight vector at time $t - 1$. This reward is effectively the change in portfolio value for period t . The action a_t at time t is the portfolio weight vector, where the i th element is the proportion of asset i in the portfolio. The goal of the agent is then to maximize the growth rate (or cumulative return) r_t of the portfolio over the experimental time frame.

In addition, we also define a state vector $S_T := (s_{1,T}, \dots, s_{m,T})$ where $s_{i,T}$ denotes the state of the previous price changes of asset i at time T . Namely,

$$s_{i,T} = \sum_{t=T-n}^{t=T} \gamma^t y_{i,t}$$

where $y_{i,t}$ is the i th element of the previously defined price relative vector y_t , γ is a discount factor and n is the size of the history we consider. As such the state is a vector of n -step backward relative price changes for each asset discounted by a rate gamma.

2 Analyses

Analysis goes here

2.1 Method 1

Analysis of method 1

2.2 Method 2

Analysis of method 2

3 Conclusion

Conclusion goes here

4 References

References go here