Beat China's stock market by using Deep reinforcement learning

Article · April 2020

Huang Gang
Chongqing University

Beat China’s stock market by using Deep reinforcement learning

Gang Huang rockingapple@cqu.edu.cn

Xiaohua Zhou zhxiaoh@aliyun.com

Qingyang Song song211@126.com

Introduction

Pass

NOMENCLATURE

DRL	Deep reinforcement learning
RL	Reinforcement learning
MV	Mean-variance

Modern portfolio theory was established by Markowitz(1952), who creatively introduced mean-variance(MV) method to optimize the allocation of portfolio. Based on Markowitz’s theory. Sharpe(1964) specify that expected return of asset can be determined by risk factor. Whereas the mean-variance approach can’t get stable results, history price movement does not guarantee a similar price movement in future. Some other scholar develop Markowitz’s theory by analyzing capital return with multiple factors. Ross(1976) proposed the arbitrage theory of capital asset pricing model, in which expected return of asset can be represented by multiple factors. Fama and French(1993) created 3 factors model, which include book-to-market factor, earing-to-price factor and market risk factor, to explain the cross-section of average returns on stocks and bonds, and more than 20 years later, Fama and French(2015) add other 2 factors to extend the model that they proposed before, this five-factor asset pricing model achieves better interpretation than the three-factor asset pricing model.

Along with the development of asset pricing model, there has been a debate that nobody can defeat market. Fama(1970) proposed efficient market hypothesis, that’s to say, stock prices have fully reflected all information, so trying to beat market is futile in an efficient market. Then, a lot of market anomalies had been discovered on empirical study of efficient market. Based on Fama's hypothesis, most of market are not enough efficient, it means that investors can get excess return by trading, and market also can be defeated.

The desire of getting excess return drives investors to use econometric method to allocate their portfolio based on asset pricing model. Nonetheless, there have some drawbacks by using it. Harvey et al(2016) argues that it is a serious mistake to use the usual statistical significance cutoffs in asset pricing model, and they attempt to establish new benchmarks to guide empirical asset pricing tests. Laloux et al(2008) shows Markowitz's portfolio optimization scheme is not adequate, since its lowest eigenvalues are dominated by noise. In addition, a lot of evidence has shown that traditional econometric analysis is not proper approach (David, 2014; Chen & Pear, 2013; Prado & Lewis, 2018).

So, how can we defeat market to get excess return? A lot of scholar try to beat market by using machine learning method, especially reinforcement learning. Early work had been done by John Moody et al(2001), Dempster et al(2006), Yue Deng et al(2016) and Lu(2017), they have achieved good results of predicting single asset movement. Not until in 2015, A major technological breakthrough occurred, Volodymyr Mnih et al(2015) associate deep learning and reinforcement learning to play computer games and prove that human can using DRL to get Artificial Intelligence. This breakthrough inspire a lot of scholars to apply DRL in allocation of portfolio. Jiang et al(2017) first apply DRL in portfolio of digital currency, Motivated by Jiang et al, Liang et al(2018), Filos(2018) use DRL to analyze portfolio with respect to China, USA and Europe capital market. Nonetheless, the way created by Jiang can't short asset. Different from Jiang's approach, Wang et al(2019) extend MV portfolio allocation by using RL to analyze portfolio of S&P 500 stocks. Guo et al(2018) proposed Robust Log-Optimal Strategy with RL to analyze constituent stocks of the CSI300.

Our contributions

Our approach based on Jiang et al(2017)'s work. Since except for the stock index, most of stocks can't be short in China, to ours best knowledge, at present, no scholar has realized short selling in Chinas stock market using deep reinforcement learning.

Therefore, we improve the model adopted by Jiang to add short function. We take CSI300 index into our portfolio.

3 Methodology

In this section, we

State Space If there something happen, there must be a reason. We are inspired by Fama's work with respect to factor analysis, and Liu et al(2018)'s four-factor model which can explain most of return anomalies in China. We take some factors to represent states, that is PE(Earnings-to-Price), PB(Price-to-Book), PS(Price-to-Sales), PC(Price Cash Flow Ratio), and turn over factors. At a given time step, we take the past 53 observations of each feature to form a single state. A list of our features is below:

Price data: close price, open price, high price, low price. And all the price data should be normalized by the last close price of a given time window (53 observations).

None-price data: volume, PE, PB, PS, PC, turn over. And all the none-price data should be normalized by its mean and variance.

Action Space In China's stock market, most of equities can't be short, therefore, it is difficult for ordinary investors to get excess returns. Lets imagine this situation, if some equities are overvalued, it's not wise to buy these equities at this time. The best way to get excess returns is to buy those undervalued stocks, and have a short position of stock index at the same time. Consequently, we select CSI300 stock index in our portfolio, and only the index can be short.

Motivated by (Jiang et al., 2017), we specify portfolio weights represent actions, that's to say, actions = $\boldsymbol{W}_t$, t represents trading period t. Therefore, $\boldsymbol{W}_t = (\mathbf{w}_{0,t}, \mathbf{w}_{1,t}, \mathbf{w}_{2,t}, \ldots, \mathbf{w}_{m,t})$. The first weight $\mathbf{w}_{0,t}$ is the weight of cash asset that investors have, the last weight $\mathbf{w}_{m,t}$ is the weight invested in stock index (CSI300).

Here we must know that making short in stock market just like margin trading in FX market, it will occupy your money or asset when shorting. We specify the maximum leverage for shorting is 1:1. Since only the stock index can be short in China's market, the range of $\mathbf{w}_{m,t}$ is (-1, 1), and the other weights' range are (0, 1). All the weights are subject to:

$$\sum_{i=0}^{n} |wi\,| = 1$$

i denote the ith asset in our portfolio, we initialize our portfolio with $\mathbf{W}_{0} = (1, 0, \ldots, 0)^{T}$, that means we don't invest money in any asset at the beginning of trading.

Reward Function we take the daily logarithmic rate of return as our reward function. Inspired by (Jiang et al., 2017), we define the price relative vector of t^{th} trading period as $\boldsymbol{Y}_t \triangleq \mathbf{P}_t \oslash \mathbf{P}_{t-1} = (1, p_{1,t}/p_{1,t-1}, \ldots, p_{n,t}/p_{n,t-1})^{T}$, where $\oslash$ represents the element-wise division, and $\mathbf{P}_t$ denote the close prices in the t period. We believe the growth rate of cash value during the trading period is zero, therefore, the first element of $\boldsymbol{Y}_t$ is always 1. If $\boldsymbol{V}_t$ denotes the portfolio value at period t, ignoring transaction cost, the portfolio value is then $\boldsymbol{V}_t = \boldsymbol{V}_{t-1}[(\ln\boldsymbol{Y}_t \cdot \boldsymbol{W}_{t-1}) + 1]$. The daily logarithmic rate of return is:

$$r_t = \ln(\boldsymbol{V}_t / \boldsymbol{V}_{t-1})$$

Transaction cost Inspired by (Jiang et al., 2017), the portfolio weights at the beginning of period t is $\boldsymbol{W}_{t-1}$. Due to price movements in the market, at the end of the same period, the weights evolve into $\boldsymbol{W}_t^{'} = (\boldsymbol{Y}_t \odot \boldsymbol{W}_{t-1}) / (\boldsymbol{Y}_t \cdot |\boldsymbol{W}_{t-1}|)$, where $\odot$ represents the element-wise multiplication. And then the objective function(reward function) will change weights from $\boldsymbol{W}_t^{'}$ to $\boldsymbol{W}_t$ in order to optimize portfolio return, therefore, the transaction cost $\boldsymbol{C}_t = \mu_t(\sum_{i=1}^{m} \left|\boldsymbol{w}_{i,t}^{'} - \boldsymbol{w}_{i,t}\right|)$, μ_t is the trading cost rate that include commission and fees when changing weights of portfolio. Taking transaction cost into consideration, the portfolio value is then $\boldsymbol{V}_t = \boldsymbol{V}_{t-1}(1 - \boldsymbol{C}_t)[(\ln\boldsymbol{Y}_t \cdot \boldsymbol{W}_{t-1}) + 1]$.

RL Algorithm

Deep Deterministic Policy Gradient (DDPG) A DDPG is a off-policy algorithm, which learns a Q function and a policy for environments with continuous action spaces. The optimal action-value function $\boldsymbol{Q*}$ is given by:

$$\boldsymbol{Q}^{*}(\boldsymbol{s},\boldsymbol{a}) = \underset{\boldsymbol{s'}\sim\boldsymbol{P}}{\mathbf{E}}\left[\boldsymbol{r}(\boldsymbol{s},\boldsymbol{a}) + \boldsymbol{\gamma}\max_{\boldsymbol{a'}}\mathbf{Q}^{*}(\boldsymbol{s'},\boldsymbol{a'})\right]$$

where $s \sim P$ is abbreviation for saying that the next state, s , is sampled by the environment from a distribution $P(\cdot \mid s, a)$.

we set mean-squared Bellman error (MSBE) loss function to tell us how closely Q_{ϕ} comes to satisfying the Bellman equation:

$$L(\phi, D) = \mathop{\mathbf{E}}_{(s,a,r,s',d)\sim D}\left[\left(Q_{\phi}(s,a) - \left(r + \gamma(1-d)\max_{a'} Q_{\phi}(s',a')\right)\right)\right]$$

DDPG algorithm is an improvement of DQN, more details can be found in Silver et al(2014) and Lillicrap et al(2016).

4 Experiments

Deep learning network design

In order to make short, the weight of stock index should be negative, therefore, we replace softmax with tanh at the last layer of deep learning network.

Backtest

Since deep learning network can simulate almost any function, that's to say, we can find a opportunity in a not efficient market to get profit.

Here we randomly select 2 groups of portfolios, each portfolio contains 7 stocks and 1 index. Based on the analysis above, only the index can be short. Because of the shell-value contamination(Liu et al, 2018), our portfolio exclude stocks that rank in the last 30% of their market capitalization

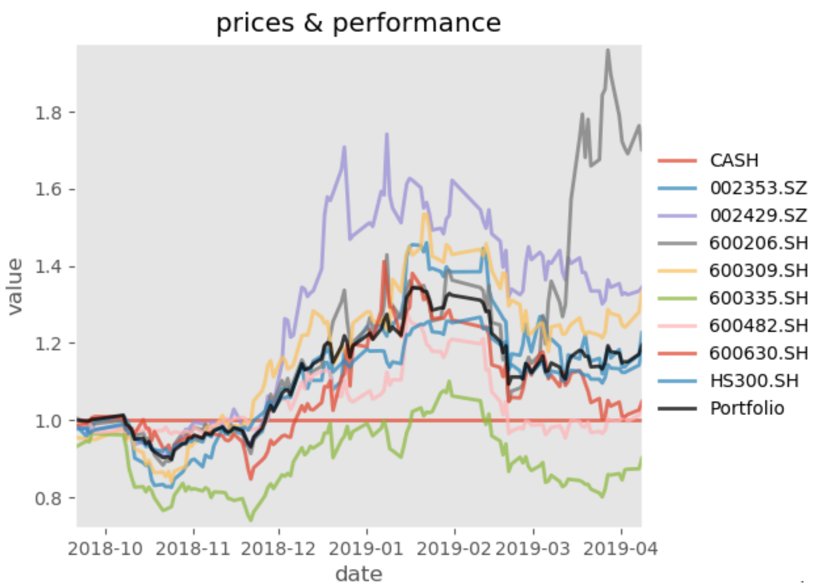

Figure 1 Max drawdown is 13.23%, sharpe ratio is 1.2806

Result

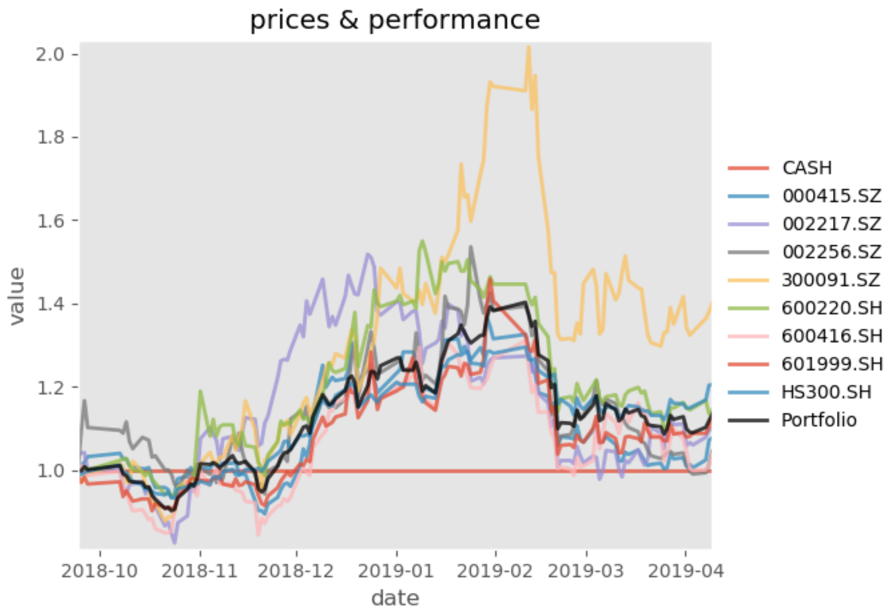

Figure 2 Max drawdown is 12.91%, sharpe ratio is 0.9011

Our empirical research shows that we can beat China's stock market by using deep reinforcement learning.

Conclusion and future work

pass

投资组合中的股票是随机选择的，而且神经网络也没经过优化。实证结果显

示，我们可以应用深度强化学习战胜市场。

通过数据驱动来获利，而不是寻找一个规则来获利。

论文还在完善中

Reference

Markowitz H. Portfolio Selection[J]. The journal of finance, 1952.

Sharpe W F. Capital asset prices: A theory of market equilibrium under conditions of risk[J]. The journal of finance, 1964, 19(3): 425-442.

Ross S A . The Arbitrage Theory of Capital Asset Pricing[J]. Journal of Economic Theory, 1976, 13(3):341-360.

Fama E F, French K R. A five-factor asset pricing model[J]. Journal of financial economics, 2015, 116(1): 1-22.

Malkiel B G, Fama E F. Efficient capital markets: A review of theory and empirical work[J]. The journal of Finance, 1970, 25(2): 383-417.

Harvey C R, Liu Y, Zhu H. … and the cross-section of expected returns[J]. The Review of Financial Studies, 2016, 29(1): 5-68.

Laloux L, Cizeau P, Bouchaud J P, et al. Noise dressing of financial correlation matrices[J]. Physical review letters, 1999, 83(7): 1467.

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values[J]. Royal Society open science, 2014, 1(3): 140216.

Chen B, Pearl J. Regression and causation: a critical examination of six econometrics textbooks[J]. Real-World Economics Review, Issue, 2013 (65): 2-20.

Lopez de Prado M, Lewis M J. Detection of false investment strategies using unsupervised learning methods[J]. Quantitative Finance, 2019, 19(9): 1555-1565.

Moody J, Saffell M. Learning to trade via direct reinforcement[J]. IEEE transactions on neural Networks, 2001, 12(4): 875-889.

Dempster M A H, Leemans V. An automated FX trading system using adaptive reinforcement learning[J]. Expert Systems with Applications, 2006, 30(3): 543-552.

Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529.

Filos A. Reinforcement Learning for Portfolio Management[J]. arXiv preprint arXiv:1909.09571, 2019.

Lu D W. Agent inspired trading using recurrent reinforcement learning and lstm neural networks[J]. arXiv preprint arXiv:1707.07338, 2017.

Wang H, Zhou X Y. Large scale continuous-time mean-variance portfolio allocation via reinforcement learning[J]. Available at SSRN 3428125, 2019.

Guo Y, Fu X, Shi Y, et al. Robust Log-Optimal Strategy with Reinforcement Learning[J]. arXiv preprint arXiv:1805.00205, 2018.

Liu J , Stambaugh R F , Yuan Y . Size and Value in China[J]. Social Science Electronic Publishing.

Silver D, Lever G, Heess N, et al. Deterministic policy gradient algorithms[C]. 2014.