Selecting Portfolios Directly Using Recurrent Reinforcement Learning (Student Abstract)

Lin Li*
CDSB Graduate School of Economic and Social Sciences
University of Mannheim
lin.li@gess.uni-mannheim.de

Abstract

Portfolio selection has attracted increasing attention in machine learning and AI communities recently. Existing portfolio selection using recurrent reinforcement learning (RRL) heavily relies on single asset trading system to heuristically obtain the portfolio weights. In this paper, we propose a novel method, the direct portfolio selection using recurrent reinforcement learning (DPS-RRL), to select portfolios directly. Instead of trading single asset one by one to obtain portfolio weights, our method learns to quantify the asset allocation weight directly via optimizing the Sharpe ratio of financial portfolios. We empirically demonstrate the effectiveness of our method, which is able to outperform state-of-the-art portfolio selection methods.

Introduction

Portfolio selection usually aims to maximize the return over time and/or minimize the investment risk simultaneously. Investors typically gain profits by dynamically allocating their wealth among selected assets at the initial period and rebalancing their wealth afterwards. With the fast development of machine learning in recent years, financial portfolio selection is increasingly studied in the machine learning community. On-line portfolio selection problems have been investigated by many researchers (Borodin, El-Yaniv, and Gogan 2004; Li et al. 2015). In addition, Moody and his cooperators (Moody and Wu 1997; Moody et al. 1998; Moody and Saffell 2001) proposed to use reinforcement learning algorithms to optimize trading systems and portfolios. Specifically, they used the RRL method to optimize the differential Sharpe ratio to trade single financial security and obtained some results for portfolio selection from the single security trading system. Afterwards, other researchers followed their scheme with certain variations, either by mildly extending the standard RRL method (Maringer and Ramtohul 2012) or by optimizing different objective functions of the method (Almahdi and Yang 2017). However, these previous work about portfolio selection using RRL construct portfolios simply based on the single asset trading system. To be more specific, they trade each asset separately using RRL, from which they subsequently obtain the corresponding portfolio weights. For example, Maringer and Ramtohul (2012) selected portfolios based on signals of each stock trading system and simply assumed investors held equally weighted portfolio consisting of 12 stocks. This process is rigid and, more importantly, merely heuristic. None has systematically shown the effects of selecting the portfolio directly. This is what we aim to show in this paper.

Methods

In this paper, we assume that the trader takes only long or neutral positions and there is no income or consumptions. Also, the period profit is not re-invested.

Portfolio Construction Function. At the beginning of each period, the trader should rebalance the portfolio which is composed of several securities with different weights. Assuming there are m securities with price series $\{\{p_t^a\} : a = 1, ..., m\}$, the market rate of return r_t^a for price series p_t^a for the period ending at time t is defined as $r_t^a = \frac{p_t^a}{p_{t-1}^a} - 1$ and thus the return vector of m securities is defined as $\boldsymbol{r}_t = [r_t^1, r_t^2, ..., r_t^m]^\top$. Defining portfolio weight of the a^{th} security at period t as F_t^a, $\boldsymbol{F}_t = [F_t^1, F_t^2, ..., F_t^m]^\top$ and the vector $\mathbf{1} = [1, 1, ..., 1]^\top$, then the trader that takes only long or neutral positions should have portfolio weights that satisfy $\boldsymbol{F}_t = \frac{\exp[\boldsymbol{f}_t(\boldsymbol{Y}_t)]}{\mathbf{1}^\top \cdot \exp[\boldsymbol{f}_t(\boldsymbol{Y}_t)]}$, where $\boldsymbol{f}_t(\boldsymbol{Y}_t) = \tanh(\boldsymbol{Y}_t)$, $\boldsymbol{Y}_t = (\boldsymbol{X}_t \circ \boldsymbol{\Theta}_t) \cdot \mathbf{1}$, and $\boldsymbol{X}_t = [\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, ..., \boldsymbol{x}_t^m]^\top$, which is the input feature matrix to the portfolio selection system, while $\boldsymbol{\Theta}_t = [\boldsymbol{\theta}_t^1, \boldsymbol{\theta}_t^2, ..., \boldsymbol{\theta}_t^m]^\top$ is the system parameter matrix to be learned during the training process, and $\circ$ represents the Hadamard product between matrices. Note $\boldsymbol{x}_t^a = [1, r_t^a, ..., r_{t-M+1}^a, F_{t-1}^a]^\top$, the parameter $\boldsymbol{\theta}_t^a \in R^{M+2}$ where $a \in \{1, ..., m\}$, and M is the look-back window size of historical return series inputs to the trader. In this case, the trader will directly decide the portfolio weights F_t^a for each security at each period.

Profit of DPS-RRL. We use multiplicative profits to measure the model's performance, as suggested by Moody and Wu (1997). Since F_t^a should be re-adjusted at each time step, a transaction cost rate δ should be applied to the

*I would like to thank Mr Jinsong Zheng and Dr Hao Ni for helpful discussions.

method. In this case, the wealth of the portfolio at time T is $W_T = W_0 \prod_{t=1}^{T}(1+R_t)$, where $R_t = (1+\boldsymbol{F}_{t-1}^{\top}\boldsymbol{r}_t)(1-\delta\cdot\mathbf{1}^{\top}|\boldsymbol{F}_t-\boldsymbol{F}_{t-1}|)-1$ is the period profit. Here for simplicity of computation, we assume the influence of price series movements on portfolio weights is negligible. We also set $W_0 = \$1$ for simplicity. Then the cumulative profit after the T periods is $P_T = W_T - W_0$.

Sharpe Ratio. In this paper, we aim to optimize the Sharpe ratio of the portfolio among other performance criteria for the reason that it well reflects the balance between profit and risk. With the estimate of the first and second moments of returns distributions over the horizon T, the Sharpe ratio formula of a portfolio is $S_T = \frac{E[R_t]}{\sqrt{E[R_t^2]-(E[R_t])^2}} = \frac{A}{\sqrt{B-A^2}}$, where $A = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $B = \frac{1}{T}\sum_{t=1}^{T}(R_t)^2$.

Gradient Ascent. By maximizing the Sharpe ratio, the DPS-RRL method is adapted to choose the optimal parameters for the portfolio selection system. Therefore, we need to evaluate the influence of Sharpe ratio on the portfolio trading system during the training stage. We attain this goal by computing the first order derivative of Sharpe ratio with respect to $\boldsymbol{\Theta}_t$. To be more specific, $\frac{dS_T}{d\boldsymbol{\Theta}_t} = \sum_{t=1}^{T}\left\{\frac{\partial S_T}{\partial A}\cdot\frac{\partial A}{\partial R_t} + \frac{\partial S_T}{\partial B}\cdot\frac{\partial B}{\partial R_t}\right\}\cdot\left\{\mathrm{diag}(\frac{\partial R_t}{\partial \boldsymbol{F}_t})\frac{\partial \boldsymbol{F}_t}{\partial \boldsymbol{\Theta}_t} + \mathrm{diag}(\frac{\partial R_t}{\partial \boldsymbol{F}_{t-1}})\frac{\partial \boldsymbol{F}_{t-1}}{\partial \boldsymbol{\Theta}_t}\right\}$, where $\mathrm{diag}(\frac{\partial R_t}{\partial \boldsymbol{F}_t})$ and $\mathrm{diag}(\frac{\partial R_t}{\partial \boldsymbol{F}_{t-1}})$ stand for square matrices whose main diagonal entries are from vectors $\frac{\partial R_t}{\partial \boldsymbol{F}_t}$ and $\frac{\partial R_t}{\partial \boldsymbol{F}_{t-1}}$, respectively, and all other entries are 0. Since we are trading several risky assets simultaneously, it is easy to understand that partial derivatives $\frac{\partial R_t}{\partial \boldsymbol{F}_t}$ and $\frac{\partial R_t}{\partial \boldsymbol{F}_{t-1}}$ are both vectors, while $\frac{\partial \boldsymbol{F}_t}{\partial \boldsymbol{\Theta}_t}$ and $\frac{\partial \boldsymbol{F}_{t-1}}{\partial \boldsymbol{\Theta}_t}$ are the Jacobian matrices. Once the term $\frac{dS_T}{d\boldsymbol{\Theta}_t}$ has been calculated, the system parameters $\boldsymbol{\Theta}_t$ is updated according to the gradient ascent rule with considering the ℓ_2 regularization to avoid overfitting the noise in the data. The process is repeated for N epochs, where N is chosen such that Sharpe ratio has converged during training.

Experiments

Artificial Dataset. Inspired by Moody and Saffell (2001), the dataset is generated using random walks with autoregressive trend processes. Specifically, we generate 3 price series with 2000 sample points each, representing 3 different securities' prices. Moreover, we divide the price series into 2 equal parts for training and test, respectively. To evaluate the performance of our model, we compare the results obtained by our model with other benchmark strategies, i.e. equally weighted buy and hold (EW-B&H) strategy, maximum Sharpe ratio buy and hold (Max-SR-B&H) strategy, ANTICOR (Borodin, El-Yaniv, and Gogan 2004) and OLMAR (Li et al. 2015). For the EW-B&H strategy, the investment weight is set equally among assets for the whole horizon. The Max-SR-B&H strategy aims to select portfolios from the Markowitz mean-variance efficient frontier (Almahdi and Yang 2017). The ANTICOR and OLMAR are competitive on-line portfolio selection strategies, which transfers wealth from outperforming stocks to underperforming ones via cross-correlation and auto-correlation and uses the moving average reversion pattern of stock price relatives, respectively. We firstly train our model on the training set with initial given hyper-parameters and then, the model is directly tested on the test set. Here, we empirically set the hyper-parameters of our model, especially we set $\delta = 0.1\%$ for ANTICOR, OLMAR and our strategy, while EW-B&H and Max-SR-B&H are under no transaction cost.

Table 1: Out-of-sample results on artificial dataset.

Methods	Sharpe ratio	Cumulative profit ($)
EW-B&H	0.0430	0.0735
Max-SR-B&H	0.0039	0.0067
ANTICOR	-0.3447	-0.5615
OLMAR	-0.3838	-0.6619
DPS-RRL	**0.1723**	**0.3760**

Results. Table 1 clearly shows that the DPS-RRL strategy behaves the best in terms of both cumulative profits and Sharpe ratio. Therefore, the results indicate that our proposed strategy is more efficient than other benchmarks even if it is subject to 0.1% transaction costs.

Conclusion and Future Work

We introduce the DPS-RRL method systematically, which aims to select portfolios directly using RRL to optimize the Sharpe ratio of a portfolio. The dependence of portfolio selection on single asset trading system using RRL is deleted. We obtained encouraging results. In the future, we will test our method on broader data sets (e.g. real market data sets).

References

Almahdi, S., and Yang, S. Y. 2017. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications* 87:267–279.

Borodin, A.; El-Yaniv, R.; and Gogan, V. 2004. Can we learn to beat the best stock. In *Advances in Neural Information Processing Systems*, 345–352.

Li, B.; Hoi, S. C.; Sahoo, D.; and Liu, Z.-Y. 2015. Moving average reversion strategy for on-line portfolio selection. *Artificial Intelligence* 222:104–123.

Maringer, D., and Ramtohul, T. 2012. Regime-switching recurrent reinforcement learning for investment decision making. *Computational Management Science* 9(1):89–107.

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

Moody, J., and Wu, L. 1997. Optimization of trading systems and portfolios. In *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)*, 300–307. IEEE.

Moody, J.; Wu, L.; Liao, Y.; and Saffell, M. 1998. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting* 17(5-6):441–470.