Incremental Reinforcement Learning for Portfolio Optimisation

Refiloe S. Shabe, Andries P. Engelbrecht

Stellenbosch University

Introduction

Portfolio optimisation is a decision-making process of allocating a proportion of capital to a number of diverse financial assets to increase expected return on investment. Performance of traditional methods such as the Markowitz model relies on the accuracy of prior predictions of the market behaviour [3]. Traditional methods are found to be limited to adapting to characteristics of financial markets, such as the fact that market prices fluctuate in a noisy way over time. However, a basic portfolio optimisation model does not make any random walk or Brownian motion assumptions about the market. The absence of statistical presumptions makes the model simple and adaptable to form an online learning problem and to apply machine learning algorithms [1]. Dynamic portfolio selection and allocation is an essential research area in financial engineering. Machine learning techniques have been utilised in financial analysis to predict asset price movements successfully, but the challenge comes with automation of optimal portfolio allocation and asset trading. To resolve the challenge, researchers introduced applying reinforcement learning (RL) to the portfolio optimisation problem [4]. The goal of RL is to learn a proficient way to maximise accumulative future returns on investment, and minimise the associated risk, simultaneously. RL has been widely used in financial domains such as algorithmic trading and execution algorithms.

Aims and Objectives

- Investigate the effectiveness of RL agents in solving the portfolio optimisation problem as a dynamic optimisation problem.
- Build a framework for a portfolio optimisation that is robust to regime change, and can be an alternative investment advisor to portfolio managers.
- Automate the portfolio optimisation process.

Environment

In order to implement a trading strategy based on RL, a simulation environment that represents the financial market is created in OpenAI Gym [2]. The environment is comprised of historical price data for Dow Jones 30 constituents. The environment was initially cut down and incremented as the RL agent trained overtime. The test environment included the entire period when the COVID-19 pandemic broke out. The intention was to investigate the ability of RL agents to adapt to changes in the underlying distribution of the environment brought by the pandemic.

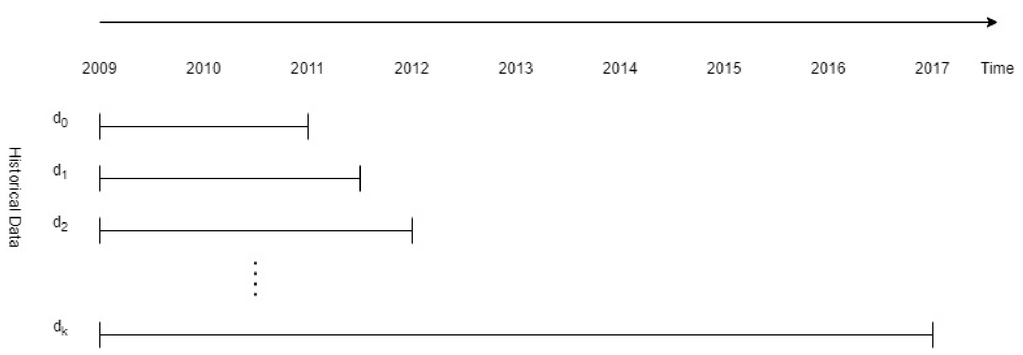

Fig. 1: A schematic illustration of environment expansion for incremental learning

Method

A recurrent proximal policy optimisation (PPO) algorithm was implemented. PPO is a family of policy gradient methods which alternates between sampling data through interaction with the environment, and optimising an objective function using stochastic gradient ascent [5].

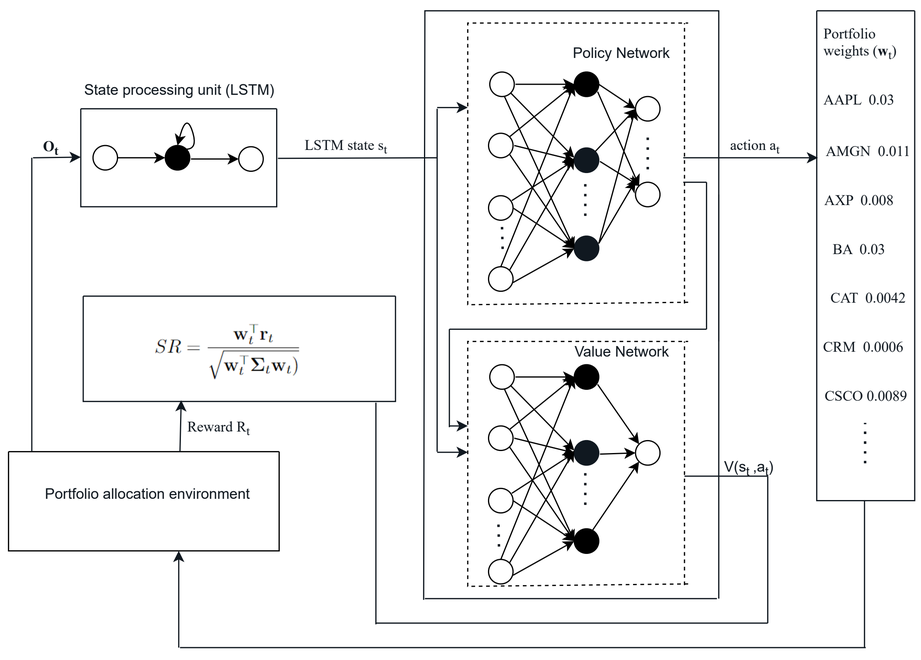

Fig. 2: A schematic illustration of a recurrent PPO model for portfolio optimisation.

Results

The experiments investigated the performance of an incremental PPO agent in comparison with the recurrent PPO, DJIA and minimum variance model. The incremental PPO is the recurrent PPO trained using incremental learning. Figure 3 illustrates the performance of the four implemented models.

Fig. 3: Big fancy graphic.

Table 1 shows the performance metrics for each strategy. Out of the four strategies, the incremental PPO agent scored the best CR of 52.46%, SR of 0.60 and CMR of 0.30, outperforming the other strategies. In terms of risk, the minimum variance model outperformed the incremental PPO. Incremental PPO portfolios were associated with high risk levels as compared to recurrent PPO portfolios.

Table 1: Evaluation measures considered to assess and compare the incremental PPO strategy with recurrent PPO and benchmark strategies

	CR (%)	AV (%)	MDD(%)	SR	CMR	VAR
Incremental PPO	52.46	21.61	−36.94	0.60	0.30	−0.027
Recurrent PPO	49.65	21.29	−36.75	0.58	0.29	−0.026
Min-Variance	31.29	17.83	−30.55	0.47	0.23	−0.022
DJIA	46.63	22.29	−37.09	0.54	0.27	−0.028

The recurrent PPO and incremental PPO agents began trading with equal stock weights on the first trading day as seen in Figure 4 (a) and (b). The portfolios showed that both strategies selected all the assets in the portfolio and assigned more weights to assets with a history of higher returns. The recurrent PPO strategy became obsolete over time while the incremental PPO strategy the incremental was more dynamic, adaptive to change and proficient.

date	AAPL	AMGN	AXP	BA	CVX	MMM	MSFT	TRV
2018/01/01	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483
2018/01/02	0.04184	0.031936	0.031936	0.035856	0.041121	0.031936	0.031936	0.031936
2018/01/03	0.046468	0.03026	0.03026	0.03775	0.043654	0.03026	0.03026	0.03026
2018/01/04	0.050054	0.028795	0.028795	0.048114	0.045118	0.028795	0.028795	0.03058
2018/01/05	0.065592	0.028245	0.028245	0.048785	0.037546	0.028245	0.028245	0.028353
2018/12/31	0.018853	0.018853	0.048842	0.051249	0.018853	0.051249	0.018853	0.051249
2019/12/31	0.040373	0.023024	0.062586	0.04305	0.023024	0.056544	0.023024	0.024071
2020/03/31	0.042162	0.021119	0.021119	0.057406	0.021119	0.057406	0.021119	0.032143
2020/07/31	0.020704	0.020704	0.020704	0.05628	0.020704	0.044293	0.020704	0.05628
2020/11/30	0.019923	0.019923	0.054156	0.033033	0.019923	0.054156	0.019923	0.045039
2020/12/31	0.06405	0.023563	0.023563	0.06405	0.053156	0.023563	0.023563	0.035129
2021/03/31	0.044006	0.025239	0.050876	0.032476	0.025239	0.031978	0.025239	0.031551
2021/06/30	0.049624	0.019826	0.019826	0.053893	0.019826	0.053893	0.019826	0.053893
2021/09/30	0.062922	0.023148	0.028206	0.062922	0.023148	0.062922	0.023148	0.023148
2021/12/29	0.066042	0.024296	0.056424	0.04016	0.024296	0.05461	0.024296	0.024296

[]

date	AAPL	CSCO	CVX	JPM	MMM	MSFT	PG	TRV
2018/01/01	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483	0.034483
2018/01/02	0.034631	0.035444	0.034814	0.034578	0.034751	0.034286	0.034811	0.034758
2018/01/03	0.034486	0.036601	0.034973	0.034603	0.034967	0.034112	0.034973	0.035623
2018/01/04	0.034659	0.037373	0.035026	0.035033	0.034777	0.033968	0.034717	0.036421
2018/01/05	0.034199	0.039192	0.035376	0.034867	0.034993	0.033737	0.034949	0.037584
2018/12/31	0.027393	0.074461	0.049284	0.027393	0.051314	0.027393	0.042144	0.074461
2019/12/31	0.028671	0.073895	0.042926	0.028671	0.038781	0.028671	0.035973	0.077936
2020/03/31	0.028168	0.071704	0.045971	0.028168	0.04365	0.028168	0.037577	0.076569
2020/07/31	0.027888	0.075806	0.04644	0.027888	0.046061	0.027888	0.039613	0.075806
2020/11/30	0.035439	0.054986	0.035952	0.036637	0.031876	0.031552	0.032946	0.05254
2020/12/31	0.035338	0.039718	0.035376	0.035901	0.034036	0.033563	0.034249	0.038757
2021/03/31	0.032734	0.058266	0.037238	0.035126	0.03388	0.031211	0.034211	0.05475
2021/06/30	0.028487	0.074988	0.043915	0.028487	0.040136	0.028487	0.03672	0.077436
2021/09/30	0.0279	0.07584	0.04675	0.0279	0.045648	0.0279	0.038928	0.07584
2021/12/29	0.029874	0.071057	0.041784	0.028968	0.036223	0.028968	0.034774	0.078744

[]

Fig. 4: Figure (a) shows a sample of portfolios obtained by recurrent PPO. Figure (b) shows a sample of portfolios obtained by incremental PPO.

Conclusion and Future Work

The study concludes that,

- The simulation environment plays an important role in training a robust and adaptive RL-based portfolio optimisation strategies.
- The ability of RL agents to perform tasks successfully and to improve over time is dependent on the reward received.

For future work,

- Solving a constrained optimisation problem using RL may grant construction of a well structured portfolio.
- High frequency data may be considered to build a more detailed understanding of financial markets and to construct strategies that are able to scale to larger markets.
- In stead of evaluating RL agents in one market, transfer learning can be explored to significantly improve sample efficiency of RL agents in other markets and to investigate generalisation

Acknowledgements

Major thanks to my supervisor and my mentors from DeepMind for their guidance and support.

References

[1] Amit Agarwal et al. “Algorithms for portfolio management based on the newton method”. In: *Proceedings of the International Conference on Machine learning*. 2006, pp. 9–16.

[2] Greg Brockman et al. “Openai gym”. In: (2016).

[3] Harry Markowitz. “Portfolio Selection”. In: *The Journal of Finance* 7 (1952), pp. 77–91.

[4] John Moody et al. “Performance functions and reinforcement learning for trading systems and portfolios”. In: *Journal of Forecasting* 17.5-6 (1998), pp. 441–470.

[5] John Schulman et al. “Proximal Policy Optimization Algorithms”. In: *ArXiv* abs/1707.06347 (2017).