Reinforcement Learning for Portfolio Management

Thesis Presentation

Angelos Filos

Supervisor: Professor Danilo Mandic

Electrical & Electronic Engineering, Imperial College London

June 26, 2018

Table of Contents

Background

Asset

An **asset** is an item of economic value, such as: cash (i.e., in bank or in hand), stocks and loans. Useful abbreviations (Kennedy, 2016) include AAPL for Apple and GE for General Electric.

Portfolio

A **portfolio** is a collection of multiple financial assets (i.e., master asset), characterized by its:

- **Constituents**: M assets of which it consists;
- **Portfolio vector**, $\boldsymbol{w}_t$: its i-th component represents the ratio of the total budget invested to the i-th asset, given by:

$$\boldsymbol{w}_t = \begin{bmatrix} w_{1,t}, & \dots, & w_{M,t} \end{bmatrix}^T \in \mathbb{R}^M$$
$$\text{such that: } \sum_{i=1}^{M} w_{i,t} = 1 \tag{1}$$

Problem Definition

The aim of this report is to investigate the effectiveness of context-independent Reinforcement Learning agents on sequential asset allocation*. A finite universe of financial instruments, assets, such as stocks, is selected and the role of an agent is to construct an internal representation (model) of the market, allowing it to determine how to *optimally allocate funds* of a finite budget to those assets.

The agent is trained on both synthetic and real market data. Then, its performance is compared with standard portfolio management algorithms on an out-of-sample dataset; data that the agent has not been trained on (i.e., test set).

*The terms *Asset Allocation* and *Portfolio Management* are used interchangeably throughout the presentation (Zhang and Wang, 2017).

Let $\rho_{i,t}$ the **log return**, or the difference of the log prices, $\log(p_{i,t})$, of asset i from time $(t-1)$ to time t, given by:

$$\rho_{i,t} \triangleq \log\left(\frac{p_{i,t}}{p_{i,t-1}}\right) \in \mathbb{R} \tag{2}$$

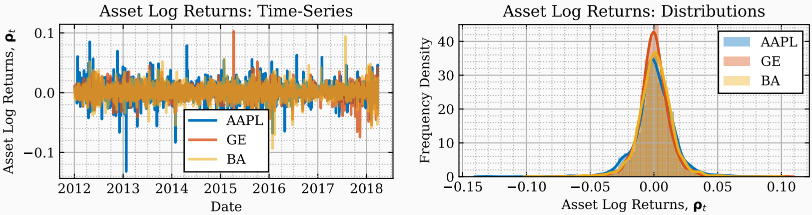


Figure 1: Log returns time-series (left) and distributions (right).

The log returns are recorded per asset at each time step, hence they form a multivariate time-series, $\vec{\boldsymbol{P}} \in \mathbb{R}^{M \times T}$.

Financial Risk and Performance Metrics

The **expected value** of log returns:

$$\mathbb{E}[\rho] = \mu_\rho \approx \frac{1}{T}\sum_{t=1}^{T}\rho_t \overset{\text{(portfolio)}}{=} \boldsymbol{w}_t^T \boldsymbol{\mu}_\rho \in \mathbb{R} \tag{3}$$

Greedy Criterion

For the same level of risk, choose the portfolio that maximizes the expected returns (Wilmott, 2007).

The **variance**, or **squared volatility** of the log returns:

$$\mathrm{Var}[\rho] = \sigma_\rho^2 \approx \frac{1}{T-1}\sum_{t=1}^{T}(\rho_t - \mu_\rho)^2 \overset{\text{(portfolio)}}{=} \boldsymbol{w}_t^T \boldsymbol{\Sigma}_\rho \boldsymbol{w}_t \in \mathbb{R} \tag{4}$$

Risk-Aversion Criterion

For the same expected returns, choose the portfolio that minimizes the volatility (Wilmott, 2007).

Risk-Aversion with Transaction Costs

Construct a portfolio out of single assets so that to: balance profitability-risk trade-off and eliminate transaction cost effects:

$$\begin{aligned} \underset{\boldsymbol{w}_t}{\text{maximize}} \quad & \overbrace{\boldsymbol{w}_t^T \boldsymbol{\mu}_\rho}^{\mu_\rho} - \overbrace{\boldsymbol{1}^T \beta \| \boldsymbol{w}_{t-1} - \boldsymbol{w}_t \|_1}^{\text{transaction consts}} - \alpha \overbrace{\boldsymbol{w}_t^T \boldsymbol{\Sigma}_\rho \boldsymbol{w}_t}^{\sigma_\rho^2} \\ \text{and} \quad & \boldsymbol{1}^T \boldsymbol{w}_t = 1 \\ \text{and} \quad & \boldsymbol{w}_t \succeq 0 \end{aligned} \tag{5}$$

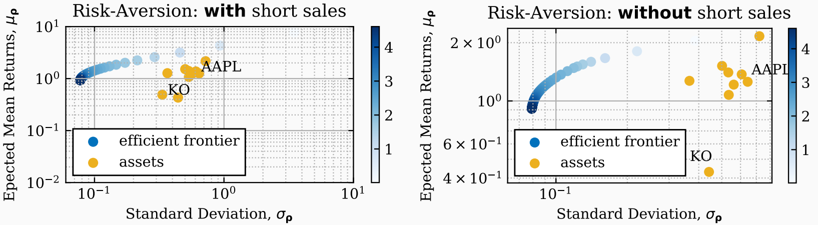


Figure 2: Risk-aversion with transaction costs efficient frontier for various α values and fixed coefficient $\beta = 0.0025$.

Innovation

Financial Market as Discrete-Time Stochastic Dynamical System

- **Observation**, $\boldsymbol{o}_t \in \mathbb{O}$, asset prices at t;
- **Action**, $\boldsymbol{a}_t \in \mathbb{A}$, portfolio vector at $(t+1)$;
- **Reward signal**, $r_t \in \mathbb{R}$, reward signal at t;
- **Environment state**, $\boldsymbol{s}_t \in \mathbb{S} \Rightarrow$ unobservable;
- **Reward generating function**, $\mathcal{R} \Rightarrow$ framework hyperparameter;
- **Environment dynamics** $\mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss}'} \Rightarrow$ intractable.

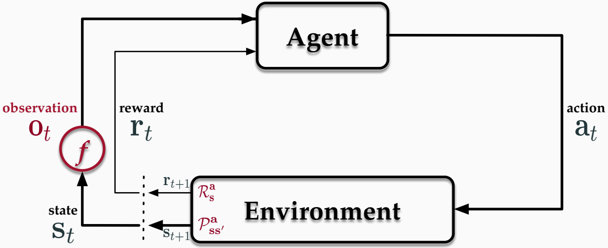


Figure 3: Partially Observable Markov Decision Process.

Trading Agent

Algorithm 1: General setup for trading agents.

inputs : trading universe of M-assets; initial portfolio vector $\boldsymbol{w}_1 = \boldsymbol{a}_0$; initial asset prices $\boldsymbol{p}_0 = \boldsymbol{o}_0$; objective function $\mathcal{J}$.

output : optimal agent parameters $\boldsymbol{\theta}_*, \boldsymbol{\varphi}_*$

repeat
    for $t = 1, 2, \dots T$ **do**
        observe 2-tuple $\langle \boldsymbol{o}_t, r_t \rangle$
        calculate gradients $\nabla_{\boldsymbol{\theta}} \mathcal{J}(r_t)$ and $\nabla_{\boldsymbol{\varphi}} \mathcal{J}(r_t)$ `// BPTT`
        update agent parameters $\boldsymbol{\theta}, \boldsymbol{\varphi}$
            using adaptive gradient optimizers `// ADAM`
        get estimate of agent state: $\boldsymbol{s}_t \approx f(\cdot\ ,\ o_t)$
        sample and take action: $\boldsymbol{a}_t \sim \pi(\cdot | \boldsymbol{s}_t; \boldsymbol{\varphi})$ `// portfolio rebalance`
    end
until convergence
set $\boldsymbol{\theta}_*, \boldsymbol{\varphi}_* \leftarrow \boldsymbol{\theta}, \boldsymbol{\varphi}$

- System Identification such that:

$$\underbrace{\hat{\mathcal{P}}^{\boldsymbol{a}}_{\boldsymbol{ss'}}}_{\text{model}} \approx \underbrace{\mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss'}}}_{\text{environment}} \tag{6}$$

- **Predictive model & planning**:

$$\boldsymbol{s}_t \xrightarrow{\mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss'}}} \boldsymbol{s}_{t+1} \xrightarrow{\mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss'}}} \cdots \xrightarrow{\mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss'}}} \boldsymbol{s}_{t+L}, \quad \boldsymbol{a}_{t+1} \equiv \max_{\boldsymbol{a} \in \mathbb{A}} \mathcal{J}\left(\boldsymbol{a} \middle| \boldsymbol{a}_t, \boldsymbol{s}_t, \ldots, \boldsymbol{s}_{t+L}\right) \tag{7}$$

- Implementations: vector autoregressive processes (VAR), recurrent neural network (RNN) and Gaussian processes (GP).

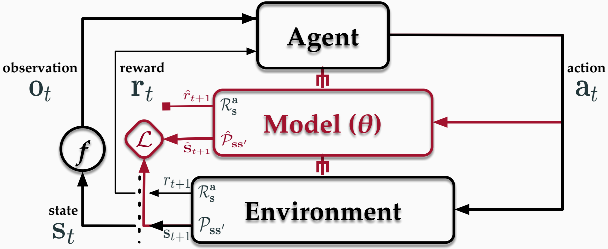


Figure 4: General setup for model-based reinforcement learning.

- Multi-stage decision making:

$$\boldsymbol{a}_{\geq t} = \underset{\boldsymbol{a} \in \mathbb{A}}{\operatorname{argmax}} \sum_{\tau=1}^{\infty} \gamma^{\tau} r_{t+\tau}, \quad \gamma \in [0, 1] \tag{8}$$

- Direct parametrization of policy π or/and action value function q:

$$\pi(\cdot|\boldsymbol{s};\ \boldsymbol{\theta}) : \quad \mathbb{A} \times \mathbb{S} \to [0, 1] \tag{9}$$

$$q(\cdot|\boldsymbol{s};\ \boldsymbol{\theta}) : \quad \mathbb{A} \times \mathbb{S} \to \mathbb{R} \tag{10}$$

- Given objective $\mathcal{J}$, follow gradient:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha \nabla_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) \tag{11}$$

Weaknesses

Trading Agents Comparison Matrix: 12-assets of S&P 500					
	Model-Based		Model-Free		
	VAR	RNN	DSRQN	REINFORCE	MSM
Non-Stationarity	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Long-Term Memory	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Non-Linear Model	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
End-to-End	$\times$	$\times$	$\times$	$\checkmark$	$\checkmark$
Linear Scaling	$\times$	$\times$	$\times$	$\times$	$\checkmark$
Universality	$\times$	$\times$	$\times$	$\times$	$\checkmark$
Low Variance	$\times$	$\times$	$\times$	$\times$	$\times$
Short Sales	$\checkmark$	$\checkmark$	$\times$	$\times$	$\times$

Table 1: Comprehensive comparison of trading agent weaknesses.

Pre-Training i

Algorithm 2: Pre-training supervised dataset generation.

inputs : number of pairs to generate N

number of assets in portfolio M

look back window size T

transaction costs coefficient β

output : dataset $\{(\boldsymbol{X}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$

for $i = 1, 2, \dots N$ **do**

- sample valid random initial portfolio vector $\boldsymbol{w}_{t_i}$
- sample random lower triangular matrix $\boldsymbol{L} \in \mathbb{R}^{M \times M}$ `// Cholesky decomposition`
- sample randomly distributed log returns: $\vec{\boldsymbol{\rho}}_{t_i - T \to t_i} \sim \mathcal{N}(\mathbf{1}, \boldsymbol{L}\boldsymbol{L}^T)$
- calculate empirical mean vector of log returns: $\boldsymbol{\mu} = \mathbb{E}[\vec{\boldsymbol{\rho}}_{t_i - T \to t_i}]$
- calculate empirical covariance matrix of log returns: $\boldsymbol{\Sigma} = \mathrm{Cov}[\vec{\boldsymbol{\rho}}_{t_i - T \to t_i}]$
- determine $\boldsymbol{a}_{t_i+1}$ by solving quadratic program (**??**)
- set $\boldsymbol{X}_i = [\vec{\boldsymbol{\rho}}_{t_i - T \to t_i},\ \boldsymbol{a}_{t_i}]$ and $\boldsymbol{y}_i = \boldsymbol{a}_{t_i+1}$

end

Pre-Training ii

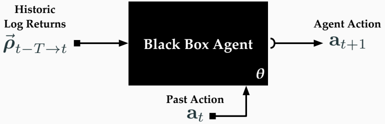


Figure 5: Interfacing with policy gradient agents as black boxes, with inputs (1) historic log returns $\vec{\rho}_{t-T\to t}$ and (2) past action (i.e., current portfolio vector) $\boldsymbol{a}_t$ and output the next agent actions $\boldsymbol{a}_{t+1}$.

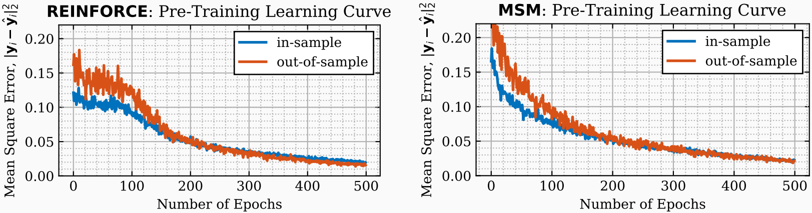
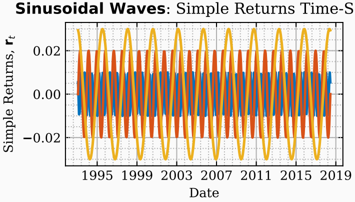


Figure 6: Pre-training mean square error (MSE) for policy gradient agents.

Experiments

Deterministic Synthetic Data



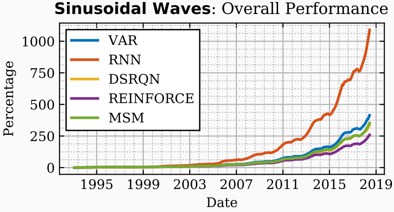
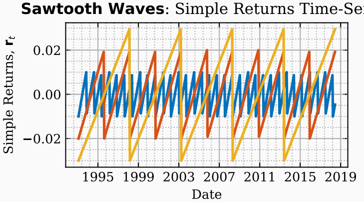


Figure 7: Synthetic universe of deterministic sinusoidal waves. (*Left*) Example series from universe. (*Right*) Cumulative returns of reinforcement learning trading agents.



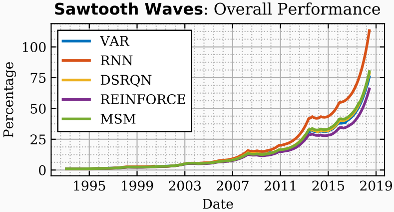


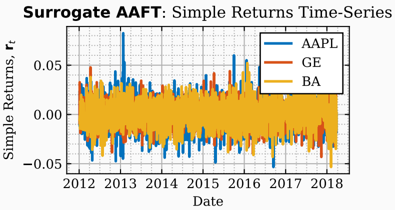
Figure 8: Synthetic universe of deterministic sawtooth waves. (*Left*) Example series from universe. (*Right*) Cumulative returns of reinforcement learning trading agents.

Surrogate Data

Algorithm 3: Amplitude Adjusted Fourier Transform (AAFT).

inputs : M-variate original time-series $\vec{\boldsymbol{X}}$
output : M-variate synthetic time-series $\hat{\vec{\boldsymbol{X}}}$

for $i = 1, 2, \dots M$ **do**
    calculate Fourier Transform of univariate series $\mathfrak{F}[\vec{\boldsymbol{X}}_{:i}]$
    randomize phase component
    calculate Inverse Fourier Transform of unchanged amplitude and randomized phase $\hat{\vec{\boldsymbol{X}}}_{:i}$
end



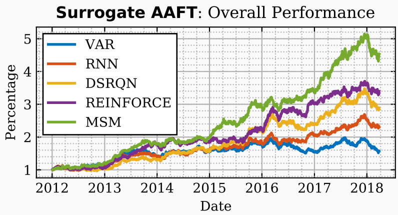


Figure 9: Prices time-series (left) and distributions (right).

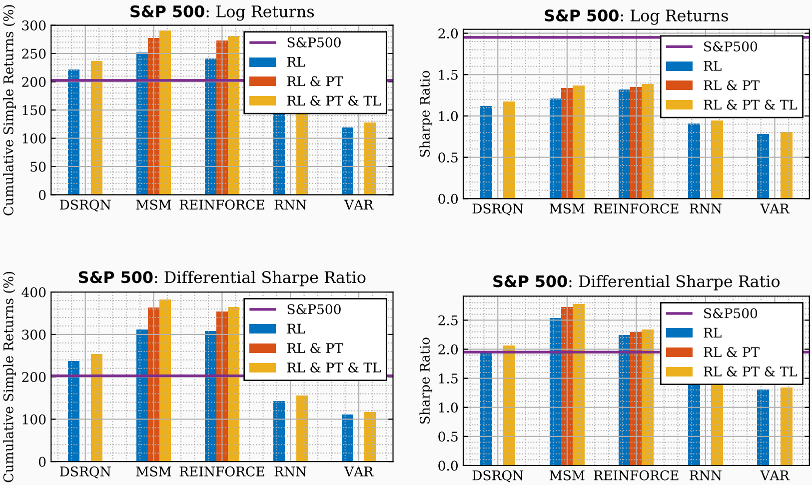


Figure 10: Comparison of reinforcement learning trading agents on S&P 500.

Standard & Poor's 500 ii

Trading Agents Comparison Matrix: S&P 500				
Reward Generating Function	**Differential Sharpe Ratio**		**Log Returns**	
	Cumulative Returns (%)	Sharpe Ratio	Cumulative Returns (%)	Sharpe Ratio
SPY	202.4	1.95	202.4	1.95
RNN	142.3	1.49	146.2	0.91
DSRQN	237.1	1.96	221.5	1.12
MSM	310.8	2.53	251.9	1.21
REINFORCE & PT	353.6	2.29	272.7	1.33
MSM & PT & TL	**381.7**	**2.77**	291.0	1.36

Table 2: Comprehensive comparison of evaluation metrics of reinforcement learning trading algorithms and their variants, namely pre-training (PL) and transfer learning (TL). Improvement 9.2% in annualized cumulative returns and 13.4% in annualized Sharpe Ratio, compared to the most recent models in (Jiang *et al.*, 2017).

Conclusions

1. Model-based Reinforcement Learning is appropriate when predictive models are accurate, while model-free Reinforcement Learning when environment dynamics are chaotic, but true objective is tractable;
2. Unified, versatile framework for training agents and investment strategies, based on a mathematical formulation of financial markets as discrete-time stochastic dynamical systems;
3. A comprehensive account of reinforcement agents has been developed, including traditional, baseline agents, a universal agent by virtue of principles of parameter sharing (Bengio *et al.*, 2003) and transfer learning (Pan and Yang, 2010);
4. Overcame the barrier of training deep architectures due to luck of large datasets, via data augmentation (a generative approach) and a choice of pre-training strategies, both of which are validated against current state-of-the-art models.

Future Work

1. **Intepretability**: Penalize model flexibility for understanding strategies (i.e., "gray-box" (Rico-Martinez *et al.*, 1994));
2. **Feature Engineering**: Extraction of financial signals and indicators to improve convergence properties;
3. **Reward Generating Functions**: Consider alternative risk metrics (e.g., VaR, CVaR);
4. **Probabilistic networks**: Replace point estimators with Bayesian Neural Networks (Nasrabadi, 2007);
5. **Exact Policy Gradient**: Fourier Policy Gradient (Fellows *et al.*, 2018).

Questions?

Email: angelos.filos14@imperial.ac.uk;
af814@ic.ac.uk;
filos.angel@gmail.com
Github: github.com/filangel/qtrader;
FYP.Presentation.pdf;
FYP.Final-Report.pdf

References

- Rico-Martinez, R, JS Anderson, and IG Kevrekidis (1994). "Continuous-time nonlinear signal processing: a neural network based approach for gray box identification". *Neural Networks for Signal Processing [1994] IV. Proceedings of the 1994 IEEE Workshop*. IEEE, pp. 596–605.
- Bengio, Yoshua *et al.* (2003). "A neural probabilistic language model". *Journal of Machine Learning Research* 3.Feb, pp. 1137–1155.
- Tsay, Ruey S (2005). *Analysis of financial time series*. Vol. 543. John Wiley & Sons.
- Nasrabadi, Nasser M (2007). "Pattern recognition and machine learning". *Journal of Electronic Imaging* 16.4, p. 049901.
- Wilmott, Paul (2007). *Paul Wilmott introduces quantitative finance*. John Wiley & Sons.

Pan, Sinno Jialin and Qiang Yang (2010). "A survey on transfer learning". *IEEE Transactions on Knowledge and Data Engineering* 22.10, pp. 1345–1359.

Goodfellow, Ian *et al.* (2016). *Deep learning*. Vol. 1. MIT press Cambridge.

Kennedy, Douglas (2016). *Stochastic financial models*. Chapman and Hall/CRC.

Jiang, Zhengyao, Dixing Xu, and Jinjun Liang (2017). "A deep reinforcement learning framework for the financial portfolio management problem". *arXiv preprint arXiv:1706.10059*.

Zhang, Xiao-Ping Steven and Fang Wang (2017). "Signal processing for finance, economics, and marketing: Concepts, framework, and big data applications". *IEEE Signal Processing Magazine* 34.3, pp. 14–35.

Fellows, Matthew, Kamil Ciosek, and Shimon Whiteson (2018). "Fourier Policy Gradients". *CoRR*.

Disclaimer

This presentation is for informational purposes only and does not constitute an offer to sell, a solicitation to buy, or a recommendation for any security; nor does it constitute an offer to provide investment advisory or other services. Nothing contained herein constitutes investment advice or offers any opinion with respect to the suitability of any security, and any views expressed herein should not be taken as advice to buy, sell, or hold any security or as an endorsement of any security or company. In preparing the information contained herein, we have not taken into account the investment needs, objectives, and financial circumstances of any particular investor. Any views expressed and data illustrated herein were prepared based upon information, believed to be reliable, available to us at the time of publication. We make no guarantees as to their accuracy or completeness. All information is subject to change and may quickly become unreliable for various reasons, including changes in market conditions or economic circumstances.