goo.gl/zaQwzv

Reinforcement Learning for Portfolio Management

Final Year Project

Angelos Filos

December 4, 2017

Department of Electrical Engineering, Imperial College London

Abstract

We investigate the effectiveness of an agent to allocate funds within a finite universe of instruments Ω. Deep Reinforcement Learning algorithms and architectures are employed in the development of the agent, while its training is accomplished on a simulated universe $\tilde{\Omega}$. $\tilde{\Omega}$ is generated in an unsupervised manner such that it captures the statistical properties of Ω.

Keywords

Deep Reinforcement Learning (DRL), Deep Q-Network (DQN), Double Deep Q-Network (DDQN), Portfolio Management (PM), Recurrent Neural Network (RNN), Feedforward Neural Network (FNN), Markov Decision Process (MDP), Softmax Layer, Sharpe Ratio (SR).

Assumptions

Observability

Markov Decision Process (`MDP`)

A tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \gamma \rangle$, where:

- $\mathcal{S}$: a finite set of states
- $\mathcal{A}$: a finite set of actions
- $\mathcal{R}$: a reward function
  - $\mathcal{R}_s^a = \mathbb{E}\left[\mathcal{R}_{t+1} | \mathcal{S}_t = s, \mathcal{A}_t = a\right]$
- γ: discount factor, $\gamma \in \left[0, 1\right]$

Partially Observable Markov Decision Process (`POMDP`)

$$\mathcal{O}_t = \mathcal{S}_t^e \neq \mathcal{S}_t^a$$

Market

Zero Slippage
The liquidity of all market assets is high enough that, each trade can be carried out immediately at the last price when a order is placed.

Zero Market Impact
The capital invested by the trading agent is so insignificant that is has no impact on the market.

Stationarity

Definition

A stochastic process $\mathcal{G}$ is called **stationary** iff its statistical properties are time-invariant.

Parameters

Relax stationarity assumptions to **Wide-Sense-Stationarity**, requiring time-invariance for:

- $\mu_{\mathcal{G}}$: mean value of $\mathcal{G}$
- $\sigma_{\mathcal{G}}$: standard deviation of $\mathcal{G}$

Goal

Aiming to **drop any conditions on stationarity**.

Architecture

Reward

Sharpe Ratio

The reward at time t, $\mathcal{R}_t$, is the contribution to the sharpe ratio of the action at time $t-1$, $\mathcal{A}_t$:

$$\mathcal{R}_t = \frac{\mu_t^*}{\sigma_t} \tag{1}$$

where:

- μ_t^*: adjusted mean returns of portfolio at time t
- σ_t: standard deviation of portfolio at time t

Transaction Cost

Returns at time t, ρ_t, are adjusted to accommodate transaction costs[2].

State i

The state $\mathcal{S}_t$ at time t, is given by:

$$\mathcal{S}_t = \langle \mathbf{w}_t, \mathcal{C}_t^T \rangle \tag{2}$$

where:

- $\mathbf{w}_t$: the portfolio vector at time t
- $\mathcal{C}_t^{(T)}$: the cumulative adjusted returns at time t for a window of length T.

State ii

Portfolio Vector

For a finite universe Ω of instruments:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_M\}, \quad |\Omega| = M < \infty \tag{3}$$

$\mathbf{w}_t$ is the **portfolio vector**, such that:

- $\mathbf{w}_t \in R^M$
- $\sum_{i=1}^{M} |w_t^{(i)}| = 1$

State iii

Cumulative Returns

Index	$\mathcal{C}_t^3(\omega_1)$	...	$\mathcal{C}_t^3(\omega_M)$
t-3	$\rho^*_{t-3}(\omega_1)$	...	$\rho^*_{t-3}(\omega_M)$
t-2	$\rho^*_{t-3}(\omega_1)\rho^*_{t-2}(\omega_1)$	...	$\rho^*_{t-3}(\omega_M)\rho^*_{t-2}(\omega_M)$
t-1	$\rho^*_{t-3}(\omega_1)\rho^*_{t-2}(\omega_1)\rho^*_{t-1}(\omega_1)$	...	$\rho^*_{t-3}(\omega_M)\rho^*_{t-2}(\omega_M)\rho^*_{t-1}(\omega_M)$

Table 1: Cumulative Adjusted Returns at time t for a window of length $T = 3$.

where:

- $\mathcal{C}_t^T(\omega_m)$: the cumulative adjusted returns of instrument m for a window of length T.
- $\rho^*_j(\omega_m)$: the adjusted returns of instrument m at time j.

Actions i

Single Instrument Case

Portfolio Management objective can be interpreted as a Binary Classification problem:

- $\mathcal{A}_t \in \left\{\texttt{LONG}, \texttt{SHORT}\right\}$
- **confidence** (probability) of each position

Softmax Layer

Softmax Function is a generalization of the logistic function, normalising a M-dimensional vector $\mathbf{a}$ such that $\sum_{i=1}^{M} \tilde{a}_i = 1$, where:

$$\tilde{a}_i = \frac{e^{a_i}}{\sum_{j=1}^{M} e^{a_j}} \tag{4}$$

Actions ii

Generalisation

The single instrument case can be generalised to $\mathcal{M}$ instruments, provided that:

1. $\mathcal{M}$ Multi-Class Classifiers are trained for $\mathcal{A}_t \in \left\{\texttt{LONG}, \texttt{SHORT}, \texttt{HOLD}\right\}$
2. the associated probabilities are passed through a **softmax layer**
3. the signs of the weights are respected
    - $\texttt{LONG} \rightarrow +$
    - $\texttt{SHORT} \rightarrow -$
4. the portfolio vector $\mathbf{w}_t$ is updated accordingly

Roadmap

Winter Term (Oct-Dec 2017)

~~Problem Definition~~

Specify the goals for the project.

~~Mathematical Formulation~~

Translate Portfolio Management objective to `POMDP` problem.

~~Concept Proof~~

Build and test architecture on a simulated (toy) dataset.

`OpenAI Gym` Integration

Build a trading environment, compatible with `OpenAI Gym`.

Spring Term (Jan-Apr 2018)

Recurrent Network

Replace Feedforward Neural Network with a Recurrent Architecture.

Market Simulation

Capture statistical properties of universe Ω by data generation.

Backtest on True Data

Build and test architecture on true data.

Double DQN

Improve stability by decoupling $\mathcal{Q}$ from $\max_a \mathcal{Q}$ in Bellman Equation.

Questions?

Market Simulation

Generative Adversarial Networks

+ Statistical Properties
+ Conditional Distributions
– Data
– Time
– Stability

Frequency Domain Phase Randomisation

+ Data
+ Time
+ Stability
– Statistical Properties
– Conditional Distributions

D. Silver.
Markov Decision Processes, 2015.
[Online]. Available: http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching_files/MDP.pdf. [Accessed: 27- Nov- 2017].

Quantopian.
Commission Models, 2017.
[Online]. Available: https://www.quantopian.com/help. [Accessed: 03- Dec- 2017].

References ii

Wikipedia.
Imperial College London, 2017.
[Online]. Available: https://upload.wikimedia.org/wikipedia/commons/thumb/a/ad/Imperial_College_London_crest.svg/1200px-Imperial_College_London_crest.svg.png. [Accessed: 07- Nov- 2017].

Z. Jiang, D. Xu, J. Liang.
A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem, 2017.
[Online]. Available: https://arxiv.org/pdf/1706.10059.pdf. [Accessed: 17- Nov- 2017].