Reinforcement Learning Project

RAPHAEL MONGES

February 2024

1 Introduction

Project: Enhancing Algorithmic Trading Strategies with Reinforcement Learning

1. Motivation

The goal is to leverage reinforcement learning (RL) to enhance trading strategies, potentially augmented with human feedback. This approach is motivated by the hypothesis that RL can uncover complex market dynamics not easily captured by traditional models or human analysis.

2. Environment Description

State Space (S)

The state space represents the market and internal state at time t, defined as a vector of features $s_t \in S$. This includes:

- Price vectors: $P_t = [p_t^1, p_t^2, \ldots, p_t^n]$, where p_t^i is the price feature of the i^{th} asset at time t.
- Technical indicators: $I_t = [i_t^1, i_t^2, \ldots, i_t^m]$, where each i_t^j is a calculated indicator at time t.
- Account features: $A_t = [a_t^1, a_t^2, \ldots, a_t^k]$, including current portfolio balance and positions.

The state at time t could thus be represented as $s_t = [P_t, I_t, A_t]$.

Action Space (A)

The action space defines the possible actions $a_t \in A$ the agent can take at each timestep:

- Buy x shares of stock i: $a_t^i = x$, where $x > 0$.

- Sell x shares of stock i: $a_t^i = -x$, where $x < 0$.
- Hold: $a_t^i = 0$, indicating no action.

Reward Space (R)

The reward function $R(s_t, a_t)$ provides feedback based on the outcome of actions, guiding the agent's learning:

$$R(s_t, a_t) = \Delta PV_t - \lambda \cdot \text{Risk}$$

where ΔPV_t is the change in portfolio value resulting from action a_t at state s_t, and λ is a risk aversion coefficient.

3. Implemented Agent

Reinforcement Learning Model

The goal is to learn a policy π^* that maximizes expected returns. For Q-learning:

$$Q^*(s, a) = \mathbb{E}[R(s, a) + \gamma \max_{a'} Q^*(s', a')]$$

where γ is the discount factor. For DQN, $Q^*(s, a)$ is approximated using a neural network.

Learning from Human Feedback (Optional)

Integrate human feedback H into the learning process:

$$R'(s_t, a_t) = R(s_t, a_t) + \eta H(s_t, a_t)$$

where η scales the influence of human feedback.

4. Expected Outcomes and Challenges

The objective is to optimize π^* to maximize the expected cumulative reward, accounting for both market returns and risk. Challenges include modeling financial market stochasticity and ensuring model generalization.