Reinforcement Learning for Quantitative Investing

Ishmam Kabir

Outline

1. Introduction
2. Proposed Method
3. Experiments

Background

Portfolio selection (PS) aims to maximize the long-term returns of wealth by dynamically allocating the wealth among a set of assets.

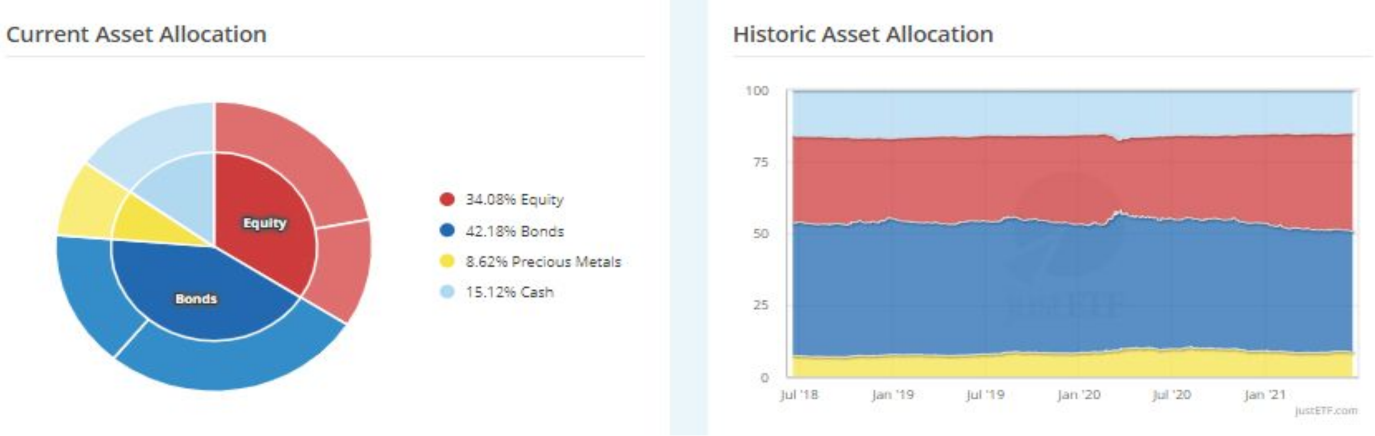

Figure 1: Portfolio selection.

Main Challenges

The non-stationary price series is hard to represent in PS.

- Price sequences of assets contain complicated sequential patterns, such as the long term trend and local oscillations.
- Financial assets in portfolios have complex correlations that may vary rapidly over time.

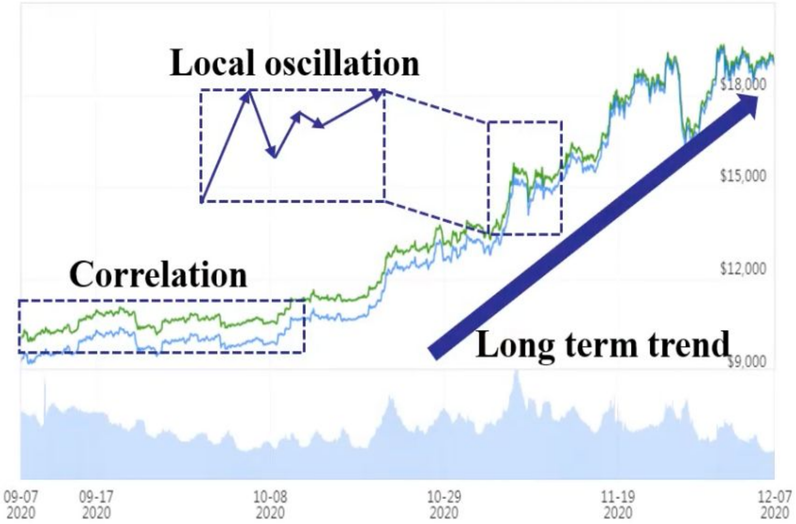

Figure 2: Price series.

Formulate PS as Markov Decision Process (MDP)

- The price series is defined as a **state.**
- The policy network $\pi(\mathcal{P}_t)$ generates a portfolio vector $\mathbf{a}_t$ as an **action**.
- The profit improvement is regarded as **reward**.
- As the market changes, the current state shifts to the next state.

Figure 3: MDP for PS.

Optimization for the policy network

- Given a PS task with m assets during a total number of n trading periods, we seek to train the policy network (RAT) by maximizing a reward function:

$$R(\mathbf{s}_0, \mathbf{a}_0, .., \mathbf{s}_n, \mathbf{a}_n) = \frac{1}{n} \sum_{t=0}^{n} \ln(\mathbf{a}_t^\top \mathbf{y}_t (1 - c_t)) \qquad (1)$$

- $\mathbf{a}_t = [a_{t,1}, a_{t,2}, ... a_{t,m}]^\top \in \mathbb{R}^m$. $a_{t,i}$ is the wealth proportion regarding asset i.
- $\mathbf{y}_t := \frac{\mathbf{p}_t^c}{\mathbf{p}_{t-1}^c} \in \mathbb{R}^m$ denotes the price change of all assets. $\mathbf{p}_t^c$ is the close price.
- c_t indicates the transaction cost.

Relation-Aware Transformer

RAT consists of an encoder and a decoder.

Encoder:

- A sequential attention layer is devised to capture sequential patterns for asset prices.
- A relation attention layer is devised for capturing asset correlations.

Decoder:

- The decoder has network modules similar to the encoder, apart from a new decision-making layer.

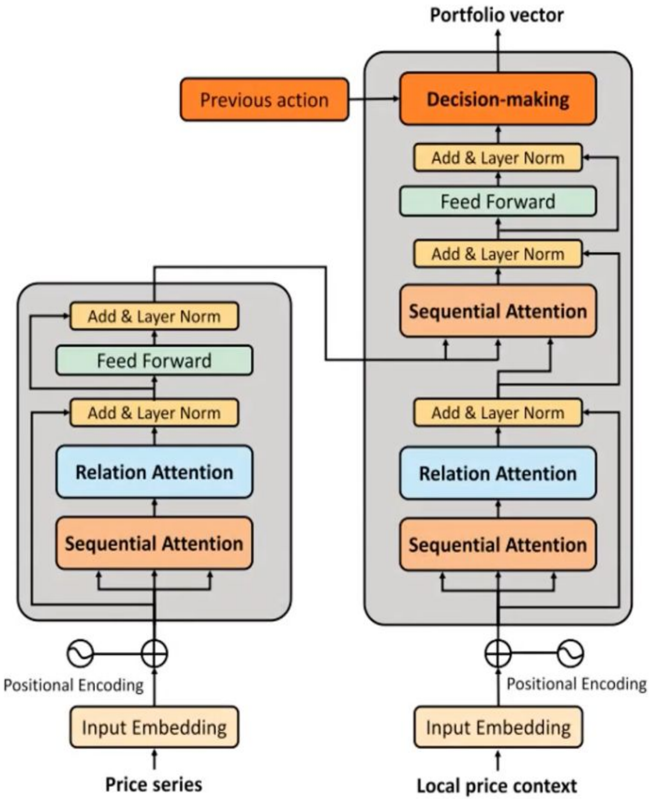

Figure 4: The scheme of RAT.

Sequential Attention Layer

- A price series is often affected by surrounding events.
- Local price context makes sequence modeling more robust to price noise.

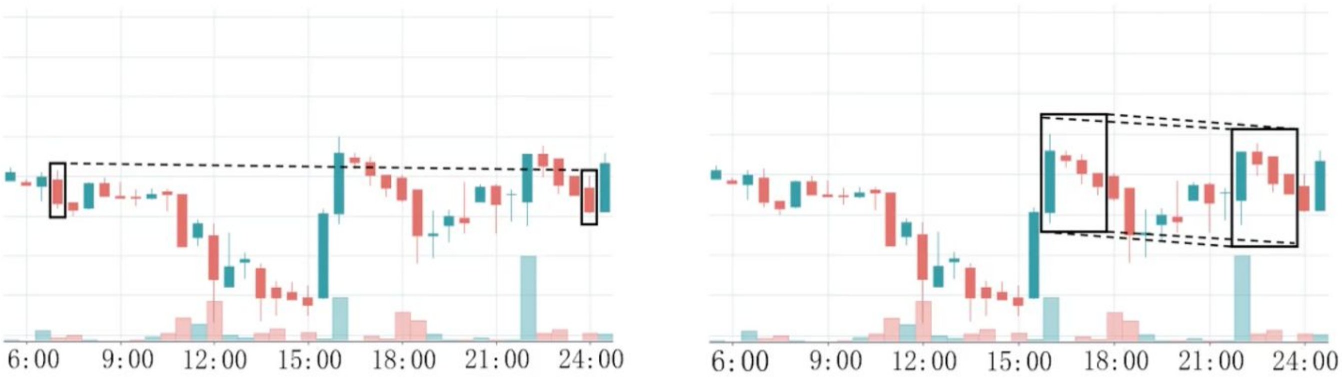

Figure 5: Local context contains information of short term pattern.

Sequential Attention Layer

- Standard self-attention cannot well exploit the context information.
- The query-key matching in standard self-attention is computed based on point-wise values.

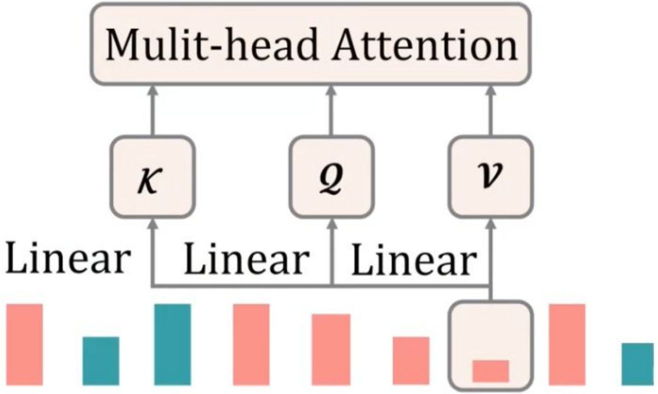

Figure 6: Standard self-attention.

Sequential Attention Layer

- To enhance the queries and keys with locality, we use context attention to transform local price context into queries or keys by exploring dependencies between the current price and local price context.

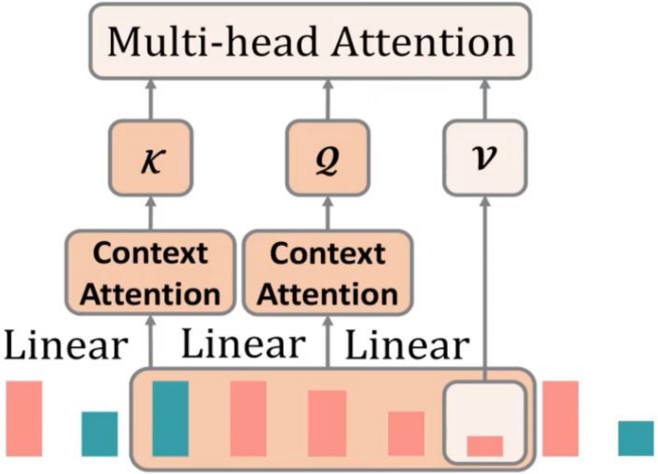

Figure 7: Context-aware self-attention.

Relation Attention Layer

- To capture the correlations, we use scaled self-attention to model asset correlations and enhances features after the sequential attention layer.

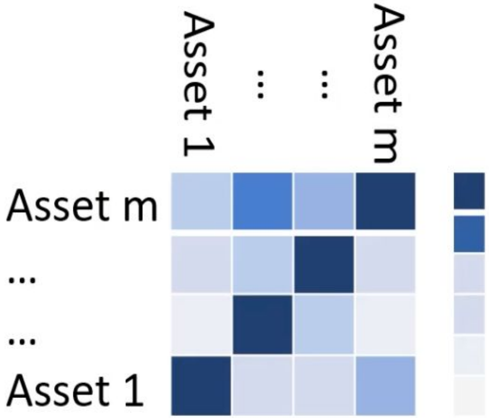

Figure 8: Relation attention weight matrix.

Decision-making Layer

- Existing RL based methods for PS decide the portfolio through a fully connected layer with softmax, which enforces the proportion of assets to be positive.
- The short sale is a transaction in which seller can first borrow some assets for sale, and then reinvest the liquidated money into other assets.

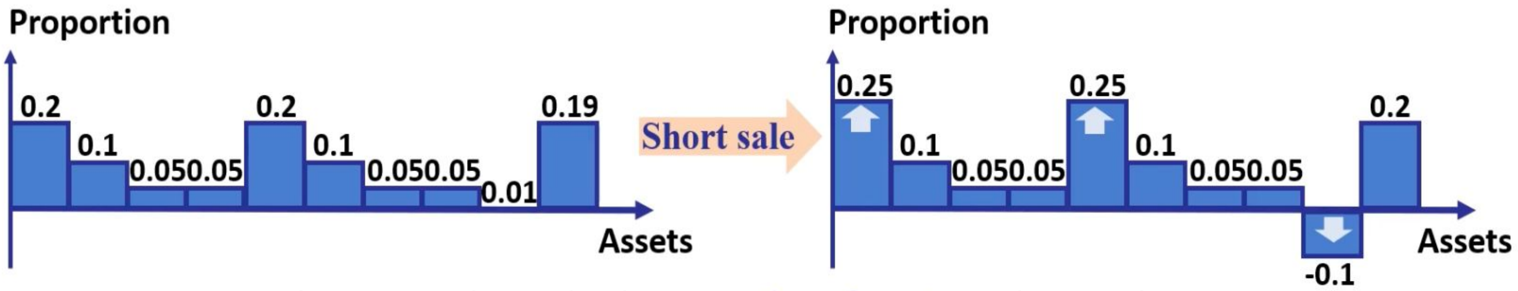

Figure 9: In short sale, the proportion of assets can be negative.

Decision-making Layer

- To introduce short sale, new decision-making layer is designed with three independent softmax heads.
  - One head outputs an **initial portfolio vector** $\hat{\mathbf{a}}_t$.
  - One head outputs a **short sale vector** $\hat{\mathbf{a}}_t^s$.
  - The last one outputs a **reinvestment vector** $\hat{\mathbf{a}}_t^r$.
- The final portfolio vector is decided by $\boldsymbol{a}_t = \hat{\mathbf{a}}_t - \hat{\mathbf{a}}_t^s + \hat{\mathbf{a}}_t^r$.
- The proportion regarding asset i becomes $a_{t,i} \in (-1,2)$, where $\sum_{i=1}^{m} a_{t,i} = 1$.