19^{th} AIAI 2023

A Sharpe Ratio Based Reward Scheme in Deep Reinforcement Learning for Financial Trading

Georgios Rodinos, Paraskevi Nousi, Nikolaos Passalis, and Anastasios Tefas

Presenter: G. Rodinos

E-mails: {grodinos, paranous, passalis, tefas}@csd.auth.gr

Computational Intelligence and Deep Learning Group (CIDL), AIIA Lab.

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Co-funded by Greece and the European Union

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Structure

- **Introduction**
- **Proposed Method**
- **Experimental Evaluation**
- **Conclusion**

DeepFinance

Introduction

- **Automated Financial Trading with Deep Learning**
  - The use of Deep Learning (DL) models aided the prediction of price movements
  - Based on the forecast direction, a trader can make an informed decision about whether to take a long or short position
- **Challenges**
  - Frequently, the generation of supervised labels is necessary
  - This task might be a challenge due to the unpredictability of the financial markets

Introduction

- **Deep Reinforcement Learning in Financial Trading**
  - Deep Reinforcement Learning (DRL) is an effective approach that addresses the challenges associated with supervised learning limitations
  - The integration with DL has enabled the direct optimization of trading policies to maximize expected profits even in the presence of volatility and uncertainty

Introduction

- **Reward Approaches**
  - **Profit and Loss (P&L):** most DRL systems for financial trading use P&L as a reward function
  - **Sharpe ratio:** captures the risk-related component of an agent's performance, and is used to evaluate a portfolio's risk-adjusted performance
- **Drawbacks**
  - **Profit and Loss (P&L):** Doesn't take into account the risk associated with the returns
  - **Sharpe ratio:** Limited samples are available during the training phase. To perform the calculation, it is necessary to consider the returns over a long period of time

Introduction

- **Goal:** Incorporating the Sharpe ratio into the reward function of a DRL agent aims to enhance the overall performance of the portfolio by mitigating the risk associated with the agent's decisions

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

Proposed Method

DeepFinance

Proposed Method

- Our study outlines a method for integrating the Sharpe ratio into the training procedure of a DRL agent
- A dynamic window is proposed which adjusts its size based on the returns obtained within an RL episode. Consequently, an estimation of the Sharpe ratio can be incorporated into the reward function

Proposed Method

- **PnL reward**

The profit-based reward is defined as:

$$r_t^{(PnL)} = \begin{cases} z_t & if\ agent\ going\ long \\ & if\ agent\ going\ short \\ 0 & if\ agent\ has\ a\ neutral\ position \end{cases}$$

where z_t is the return change and is defined as:

$$z_t = \frac{p_c(t) - p_c(t-1)}{p_c(t-1)}$$

which is also referred to as the change of the close price p_c.

Proposed Method

With the return definition, the profit-based reward can be written as:

$$r_t^{(PnL)} = e_t \cdot z_t$$

Where e_t is the current market position and $e_t \in \{1, 0, -1\} = \{\text{long}, \text{neutral}, \text{short}\}$

When the agent changes position is obligated to pay an extra fee. That is called the commission, in which case an additional reward is formulated as:

$$r_t^{(fee)} = -c \cdot |e_t - e_{t-1}|$$

where c denotes the commission. The total PnL reward can be defined as:

$$r_t^{(total)} = r_t^{(PnL)} + r_t^{(fee)}$$

Proposed Method

- **PnL and Sharpe ratio reward**

Reward, based on the approximated Sharpe ratio is defined as:

Number of steps

Returns during a DRL episode

The point at which $r_t^{(sr)}$ is incorporated

$$r^{(sr)} = \frac{E[\boldsymbol{z}]}{\sqrt{Var[\boldsymbol{z}]}} \cdot \alpha, t \in \{w, \dots, m\}, \boldsymbol{z} = (z_0, \dots, z_t)$$

where $w = m/2$, $\boldsymbol{z}$ is a vector with the returns, and α is a constant value.

The total PnL and Sharpe ratio reward is defined as:

$$r_t^{(total)} = \begin{cases} r_t^{(PnL)} + r_t^{(fee)}, & t < w \\ r_t^{(PnL)} + r_t^{(fee)} + r_t^{(sr)}, & for\ t \geq w \end{cases}$$

DeepFinance

Proposed Method

- **Proposed**

The total reward of the proposed scheme is defined as:

$$r_t^{(total)} = \begin{cases} r_t^{(PnL)} + r_t^{(fee)}, & for\ t < w \\ r_t^{(PnL)} + r_t^{(fee)} + r_t^{(sr)}, & for\ t = w \\ r_t^{(PnL)} + r_t^{(fee)} + r_t^{(sr)}, & if\ r_t^{(sr)} > r_{t-1}^{(sr)},\ for\ t > w \\ r_t^{(PnL)} + r_t^{(fee)} - r_t^{(sr)}, & if\ r_t^{(sr)} < r_{t-1}^{(sr)},\ for\ t > w \end{cases}$$

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Experimental Evaluation

DeepFinance

Dataset

- Crypto trading data
- The dataset provided by SpeedLab AG from 2017-08-17 up to 2022-02-12.
- Training set from 2017-08-17 to 2021-03-15, Test set from 2021-03-15 to 2022-02-12
- In total: 439.737 candles, where the train/test candles are 327.596 and 112.141 candles, respectively.
- 14 currency pairs such as the BTC/BUSD, BTC/USDT, and ETH/USDT among others
- The Open-High-Low-Close (OHLC) price level technique was used to preprocess the data. More specifically, OHLC values are:
  - open price (the first traded price of the set interval)
  - highest and lowest traded prices within the interval
  - close price (the last price that a trade did occur during the interval)
- The minute-price candles are resampled to hour candles

Input Features

1. $x_{t,1} = \frac{p_c(t) - p\_c(t-1)}{p_c(t-1)}$
2. $x_{t,2} = \frac{p_h(t) - p_h(t-1)}{p_h(t-1)}$
3. $x_{t,3} = \frac{p_l(t) - p_l(t-1)}{p_l(t-1)}$
4. $x_{t,4} = \frac{p_h(t) - p_c(t)}{p\ (\ )}$
5. $x_{t,5} = \frac{p_c(t) - p_l(t)}{p_c(t)}$
6. time-related features are created, including day, month, week, and year features.

The described features are concatenated into a feature vector $\boldsymbol{x}_t \in \mathbb{R}^{13}$ for each time t.

DeepFinance

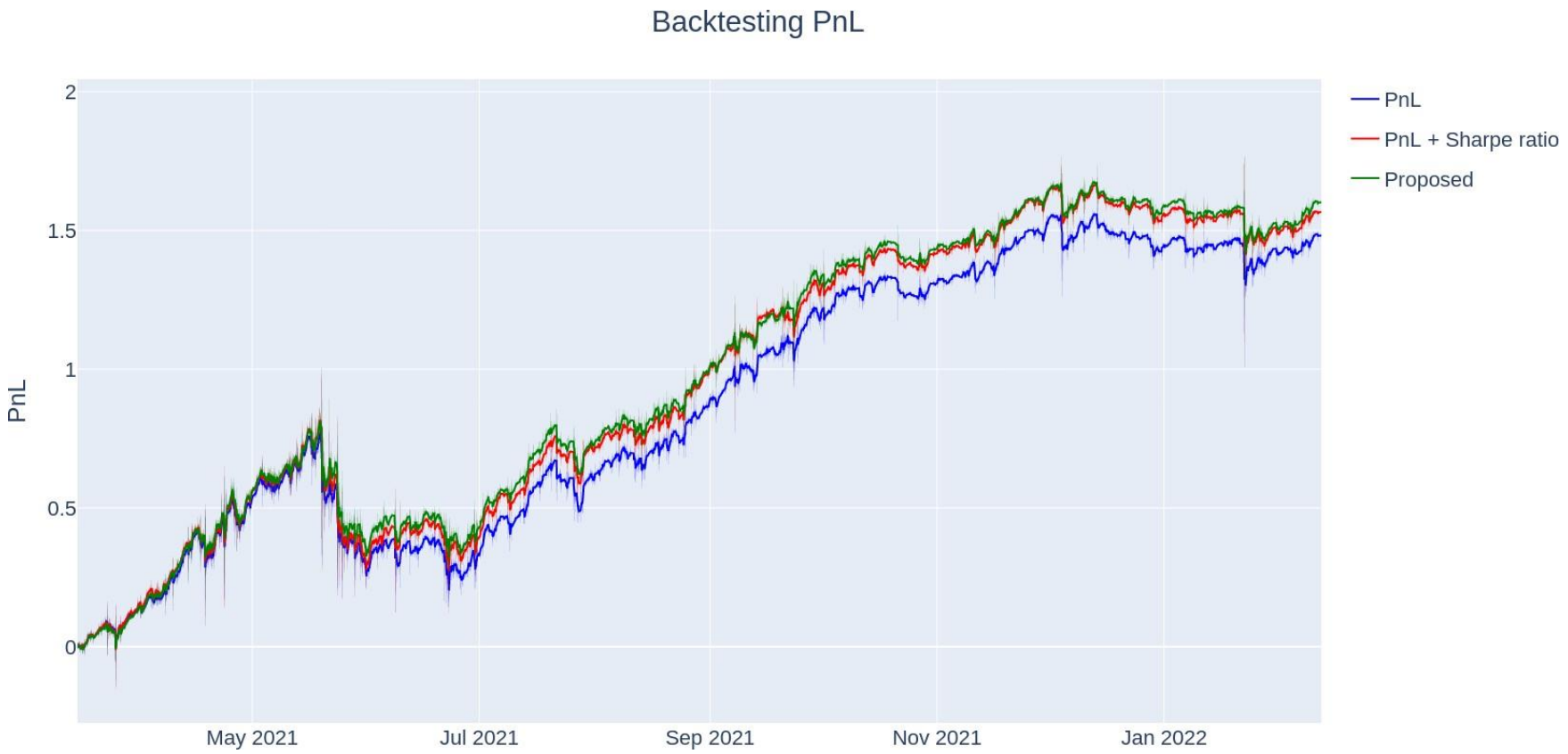
DRL setup characteristics

- The DRL agent is trained using the Proximal Policy Optimization (PPO) approach
- Neural Network architecture is Long-Short Term Memory
- The number of steps that an DRL episode consists of is equal to 100
- Each experiment is executed 10 times, with each instance using a different random seed
- The PnLs presented, are averaged throughout the 10 experiments as well as the Annualized Sharpe ratios

Proposed Reward Evaluation

Reward type	*Annualized Sharpe ratio Monthly Returns*	*Annualized Sharpe ratio Hourly Returns*
PnL	1.462 ± 0.055	2.374 ± 0.079
PnL + Sharpe ratio	1.499 ± 0.060	2.484 ± 0.090
Proposed	**1.617 ± 0.056**	**2.641 ± 0.083**

Proposed Reward Evaluation



ARISTOTLE
UNIVERSITY OF
THESSALONIKI

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Conclusion

DeepFinance

Conclusion

Developed and evaluated **a Sharpe ratio-based reward shaping scheme** for training DRL agents that are capable of decreasing the risk that often occurs in agents' trading decisions and improving the overall performance of a portfolio

ARISTOTLE UNIVERSITY OF THESSALONIKI

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Acknowledgements

This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T2EDK-02094).

Co-funded by Greece and the European Union

DeepFinance

ARISTOTLE UNIVERSITY OF THESSALONIKI

AIAI ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS

Thank you!

Questions?

DeepFinance

European Union
European Regional Development Fund

EPAnEK 2014–2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Co-funded by Greece and the European Union