Training Deep Reinforcement Learning Agents for Portfolio Management with a Sharpe ratio-based Reward

George Rodinos

E-mail: grodinos@csd.auth.gr

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

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

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

DeepFinance

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

DeepFinance

Structure

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

Introduction

- **Reward Shaping**
  - Is an important aspect of Deep Reinforcement Learning (DRL) and adapting it to a specific domain may significantly improve the performance of the DRL agent
  - Reward shaping techniques are designed with the objective of offering extra intermediate rewards or modifying the existing rewards in order to equip the agent with more informative feedback.

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

DeepFinance

Proposed Method

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:

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

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:

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

Proposed Method

- **PnL and Sharpe ratio reward**

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

PnL rewards during a DRL episode

Number of steps

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

$$r_t^{(sr)} = \frac{E[\boldsymbol{r}^{(\boldsymbol{PnL-total})}]}{\sqrt{Var[\boldsymbol{r}^{(\boldsymbol{PnL-total})}]}}, \boldsymbol{r}^{(\boldsymbol{PnL-total})} = \left(r_0^{(PnL-total)}, \dots, r_t^{(PnL-total)} \right), t \in \{w, \dots, m\}$$

where $w = m/2$, $\boldsymbol{r}^{(\boldsymbol{PnL-total})}$ is a vector with PnL rewards during a DRL episode. The total PnL and Sharpe ratio reward is defined as:

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

Where α is a constant value.

Proposed Method

- **Proposed**

The total reward of the proposed scheme is defined as:

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

The reward statistic lie on a very small scale, which would slow down the training of the employed RL estimators that's why after each step the $\boldsymbol{r^{(total)}}$ is simply divided by its standard deviation: $\boldsymbol{r^{(total)}} = \frac{\boldsymbol{r^{(total)}}}{\sigma_{\boldsymbol{r^{(total)}}}}$

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

DeepFinance

Experimental Evaluation

Dataset

- Crypto + Sentiment (CryptoSentiment dataset)
- Training set from 2020 to 2021-07-25, Test set from 2021-07-25 to 2022-02-12
- In total: 259.351 data points, where the train/test candles are 191.492 and 67.859 candles, respectively.
- 14 USDT currency pairs such as BTCUSDT, and ETHUSDT among others.
- We used price features, mostly percentage differences between OHLC candles, sentiment score, and time related features
- The features are concatenated into a feature vector $\boldsymbol{x}_t \in \mathbb{R}^{33}$ for each time t.

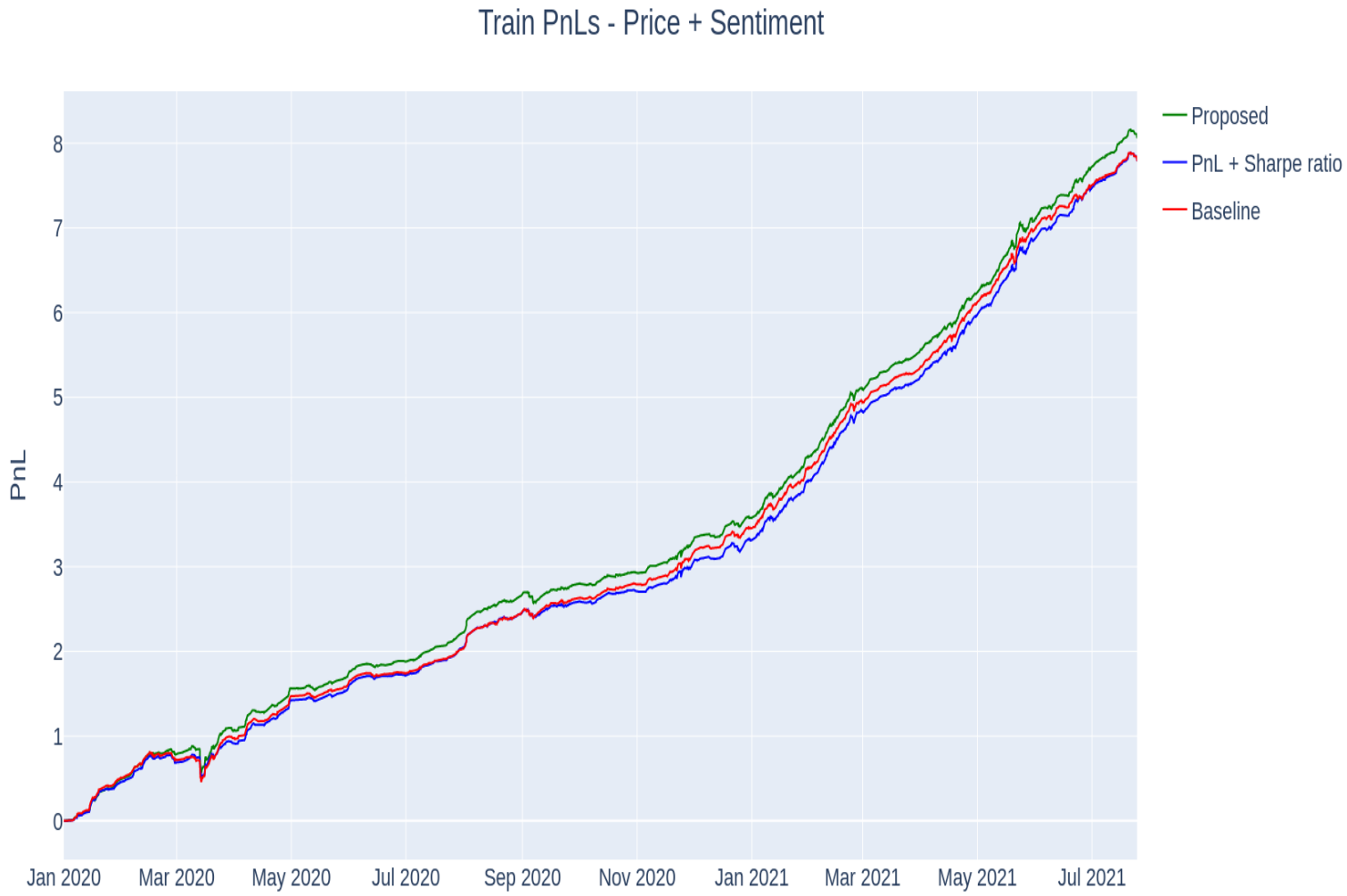
DRL setup characteristics

- 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	*Train Annualized Sharpe ratio*	*Test Annualized Sharpe ratio*
PnL	4.44 ± 0.062	1.662 ± 0.092
PnL + Sharpe ratio	4.36 ± 0.057	1.656 ± 0.029
Proposed	**4.54 ± 0.065**	**1.711 ± 0.097**

Proposed Reward Evaluation

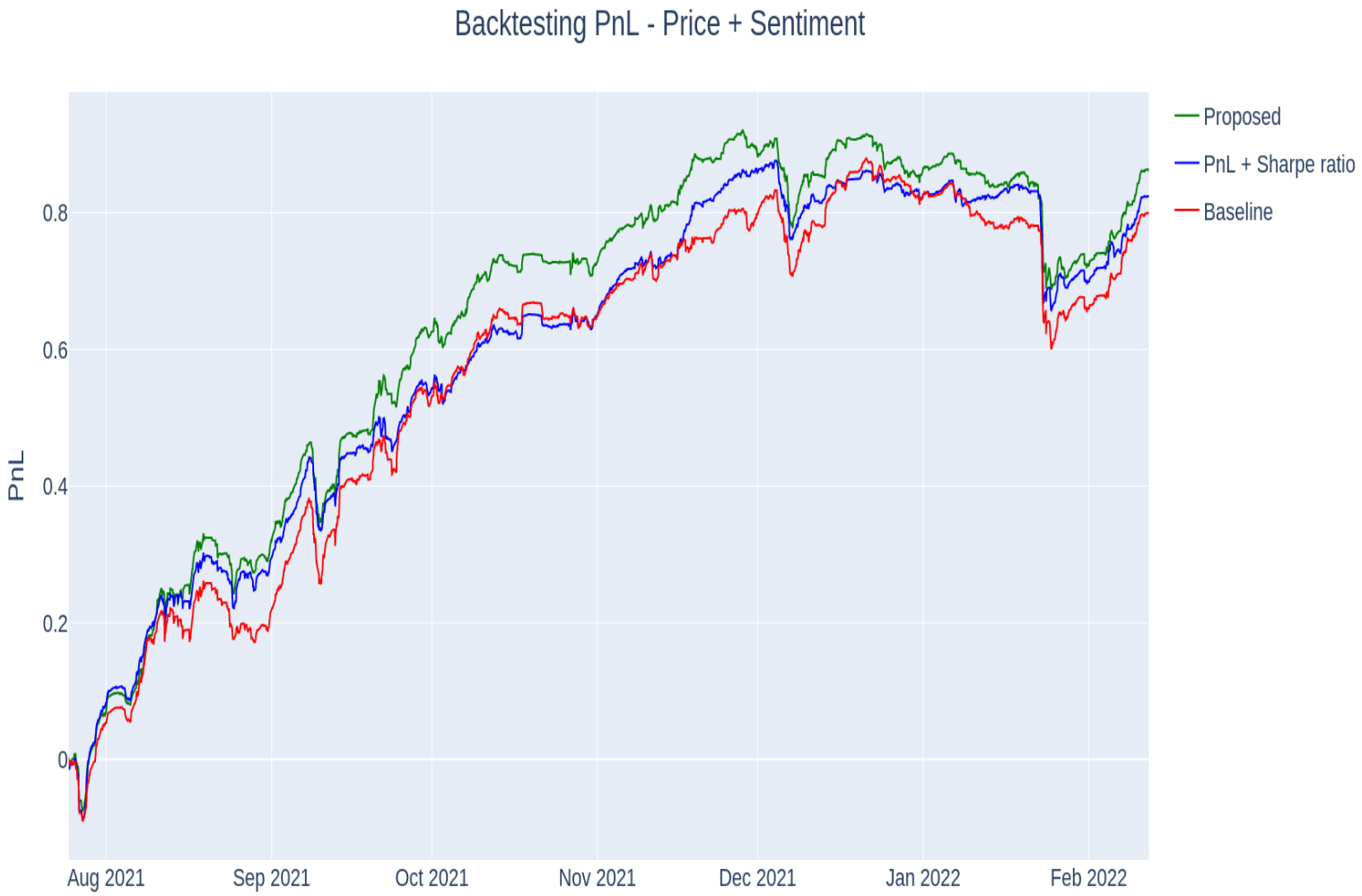


Reward type	*Train PnLs*
PnL	7.8067 ± 0.218
PnL + Sharpe ratio	7.8053 ± 0.215
Proposed	**8.0604 ± 0.222**

ARISTOTLE UNIVERSITY OF THESSALONIKI

DeepFinance

Proposed Reward Evaluation



Reward type	*Test PnLs*
PnL	0.7997 ± 0.025
PnL + Sharpe ratio	0.8239 ± 0.024
Proposed	**0.8627 ± 0.026**

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

DeepFinance

Conclusion

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

Thank you!

Questions?