Adaptive Confidence-Weighted LLM Infusion for Financial Reinforcement Learning

Emran Y. Alturki, Aydin Javadov, Qiyang Sun, Bjorn W. Schuller

IMPERIAL

ETH zürich

Introduction

Goal: Achieving higher returns on the FinRL Contest Task 1, in Nasdaq-100 trading using LLM prompting for tickers news.

Challenges:

- **Epistemic uncertainty in LLM-generated stock scores:** The model's confidence in its own recommendations or risk assessments is not explicitly quantified or communicated.

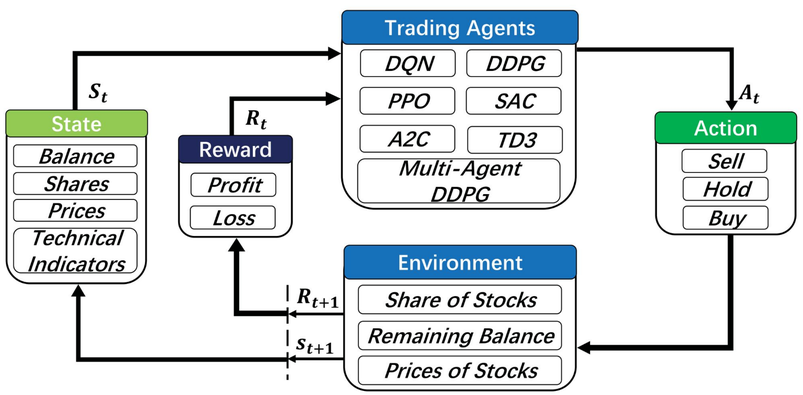
Contributions:

- Applying a confidence-weighted LLM infusion method.
- Introducing a new LLM infusion method.
- Achieving higher returns compared to the original paper [1]

Financial Reinforcement Learning (FinRL)

FinRL [2,3] is the model used with some modifications.
Two main trading agents are used:

- Proximal Policy Optimization (PPO)
- Conditional Value at Risk PPO (CVaR-PPO)



References

[1] M. Benhenda, "FinRL-DeepSeek: LLM-infused risk-sensitive reinforcement learning for trading agents." [Online]. Available: http://arxiv.org/abs/2502.07393

[2] Xiao-Yang Liu, Hongyang Yang, J. Gao, C. D. Wang. FinRL: Deep reinforcement learning framework to automate trading in quantitative finance. ACM International Conference on AI in Finance, 2021.

[3] Ben Hambly, Renyuan Xu, and Huining Yang. Recent advances in reinforcement learning in finance. Mathematical Finance, 2023.

Code is available here

Methodology

i. Improved LLM Prompt

System: You are a financial risk analyst assessing stocks as of {date}. Evaluate news for {Stock} and assign: - A risk score (1-5): 1=very low, 2=low, 3=moderate (default if unclear), 4=high, 5=very high. - A probability (0.00-1.00) reflecting confidence, based on news clarity, data availability, and task difficulty. Risk Factors to Consider: - Regulatory changes (e.g., antitrust lawsuits) → Higher risk (4-5). - Product launches (e.g., VisionPro release) → Moderate risk (3). - Market sentiment (e.g., price might decrease) → Score 2-4 depending on certainty. - Ambiguous statements (e.g., might get the Chinese effect) → Default to 3 with low probability. Risk 1: Very low risk - Strong positive news (e.g., 45% stock increase with no negatives). Risk 3: Moderate risk - Vague or neutral news (e.g., 'might decrease'). Risk 5: Very high risk - Clear negative news (e.g., major scandal or bankruptcy risk). Probability Guidance: - Probability = 0.9-1.0 if the news cites a definitive event (e.g., FDA approval delayed). - Probability = 0.6-0.8 if the risk is implied but not certain (e.g., may face fines). - Probability < 0.5 if the news is vague (e.g., could be impacted by market trends). Response Format: Risk: [risk score] Probability: [probability]

User: News to Stock Symbol -- AMZN: Amazon warehouse strike begins ### News to Stock Symbol -- GOOGL: Google fined $1B for data privacy ### News to Stock Symbol -- JPM: JP Morgan beats earnings forecasts
Assistant: Risk: 3 Probability: 0.55, Risk: 5, Probability: 0.80, Risk: 1 Probability: 0.95
User: News to Stock Symbol -- AAPL: Apple faces antitrust lawsuit in the EU ### News to Stock Symbol -- TSLA: Tesla's Q4 deliveries miss estimates
Assistant: Risk: 4 Probability: 0.75, Risk: 4 Probability: 0.60

USER: ### News to Stock Symbol -- {Stock}: On {date}, {text}

ii. Adaptive Factors Method

- Prompted DeepSeek with the above prompt along the summarized news article to produce the recommendation score of a stock and its confidence probability in the generated score
- Then, the score and its confidence are added to the general PPO model using the following formula

$$S_f^{\text{mod}} = 1 + (S_f - 1) * C_{\text{rec}} \cdot \alpha$$
$$a_t^{\text{mod}} = S_f^{\text{mod}} \cdot a_t \,,$$

$S_f = Recommendation\ Factor$
$C_{rec} = Confidence\ Probability\ of\ S_f$
$\alpha = Scaling\ Parameter$

- Same steps are repeated here with risk instead of recommendation score

$$(R_f^i)^{\text{mod}} = 1 + \left(R_f^i - 1\right) * C_{\text{risk}}^i \cdot \beta$$
$$R_f^{\text{mod}} = \sum w_i \cdot \left(R_f^i\right)^{\text{mod}},$$
$$D_{R_f^{\text{mod}}}(\pi_\theta) = R_f^{\text{mod}} \cdot D(\pi_\theta).$$

$R_f^i = Risk\ Factor$
$C_{risk}^i = Confidence\ Probability\ of\ R_f^i$
$\beta = Scaling\ Parameter$

iii. Entropy Regularization Method

$$\beta = coef \cdot (5 - S_{\text{LLM}}) \cdot C_{\text{rec}} \quad \text{Or} \quad \beta = S_{\text{LLM}} \cdot C_{\text{rec}}^2$$

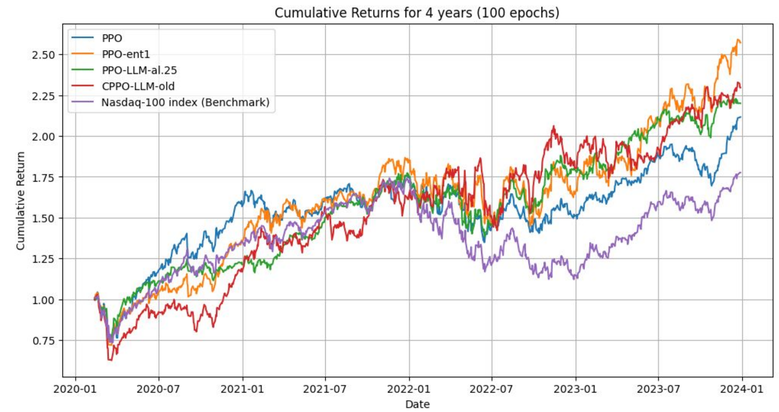
Results

i. Early Stopping

- Applied 400K training steps for shorter period of training and testing
- Experimented diverse options to choose among; variations included:
  - alpha: 0.25, 0.5, 0.75
  - beta: 0.25, 0.5, 0.75
  - coefficient: 0.005, 0.01, 0.1

ii. Full Training

- Applied 2 Million training steps on a handful of methods that scored best in the early stopping stage
- Presented are best 2 models along the old paper winners



Model	Cumulative Returns	Rachev Ratio	Max Drawdown	Outperf. Frequency	Downturn Outperf.
PPO-ent1	1.5715	1.0367	-0.2807	48.7701	74.3468
CPPO-LLM (old)	1.296	0.9698	-0.3742	48.7701	70.3088
PPO-LLM al0.25	1.1996	0.9567	-0.2419	47.2727	78.8599
PPO (old)	1.1156	0.9982	-0.2571	48.7701	79.3349

Conclusion & Future Work

It was found that the entropy regularization method got the highest result with coefficient = 0.01
For future:

- More variations could be experimented
- Expand news sources
- Pipeline Optimisation