Enhanced Trading Strategies and Reward Integration through Inverse Reinforcement Learning

진영봉
FRE LAB

Feb 11, 2025

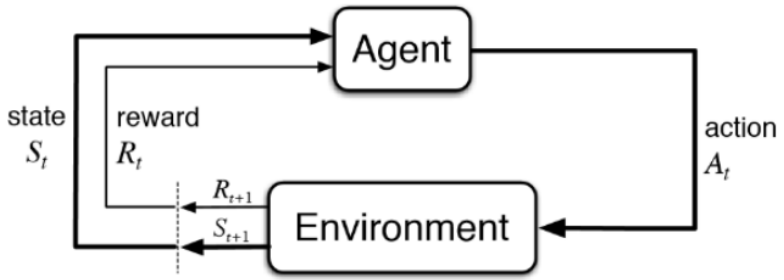
Introduction

- **연구의 필요성**
  - 금융 시장의 복잡성과 동적 특성으로 인한 알고리즘 트레이딩의 중요성 증대
  - 강화학습의 트레이딩 적용 시 보상 함수 설계의 어려움:
    - 금융 시장은 다수의 변수와의 상호작용으로 이루어져 있어 명시적인 보상함수 정의가 어려움
    - 게임, 로보틱스 분야와 달리 금융 분야에서의 보상함수는 시장의 구성요소에 변동적임
  - 역강화학습의 잠재적 해결책:
    - 전문가의 행동 데이터 분석을 통한 보상함수 추정 가능
    - 강화학습-역강화학습 피드백 루프로 안정적이고 신뢰할 수 있는 거래 전략 기대
- **연구 목적**
  - 다양한 보상함수를 통합하는 트레이딩 전략 개발
  - 구체적 목표:
    - 금융시장에서 전문가의 거래 행동 분석 및 보상함수 추정
    - 추정된 보상함수를 통해 에이전트가 전문가와 유사한 거래 전략을 학습하도록 유도
    - 실제 금융 데이터 적용 및 기존 강화학습 방법과 성능 비교

Related works

- RL in finance (Trading)
  - 강화학습은 에이전트가 환경과 상호작용하여 보상을 최대화하는 정책을 학습하는 방법
  - 강화학습의 기본 프레임워크는 Markov Decision Process(MDP)를 기반으로 함

$$\pi^* = \text{argmax}_\pi \, \mathrm{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right]$$



  - 트레이딩 환경에서의 MDP 적용 [1,2]:
    - 상태 S: 주가, 거래량, 기술적 지표 등
    - 행동 A: 매수, 매도, 홀딩 등의 트레이딩 결정
    - 보상 R: 수익률이나 샤프 비율 등의 성과지표 등

  - 최근 RL trading 연구 동향:
    - Multi-modal[3,4,5], 다중 스케일링 (multi timeframe)[6], 리스크 관리 강화[7]

Related works

- Reward design for RL finance
  - 금융 시장의 복잡성과 리스크를 고려한 리워드 함수를 설계하고자 하는 연구가 다수 존재[7~9]
    - 포트폴리오 관리에서 내재적 보상과 외부 보상함수를 결합한 방법론 제시[7]
    - 리스크를 명시적으로 고려한 리워드 함수를 통해 알고리즘 트레이딩 전략 향상[8]
    - Order book 데이터를 활용한 하이브리드 리워드로 market making 성능 최적화[9]

  - 수익과 리스크의 균형을 맞춤과 동시에, 다양한 시장 상황에 적응할 수 있도록 하는 것이 주요 목적
  - 그러나, 설계된 리워드가 여전히 주관적이며, 유연하지 못하다는 한계가 존재함

  - 역강화학습은 전문가의 행동에서 내재된 리워드 함수를 추출하고, 다양한 시장 상황에 대한 적응적 리워드 함수 학습이 가능

Related works

- Inverse RL
  - IRL은 전문가의 행동 데이터로부터 내재된 보상함수를 추론하는 기법으로, 기본 목표는 다음과 같음

$$r^* = \text{argmax}_r \min_{\pi \in \Pi} \left(\mathrm{E}_{\pi_E} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] - \mathrm{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] \right)$$

  - 이때, π_E는 전문가의 정책, Π 는 가능한 모든 정책을 의미함
  - 최근, Adversarial Inverse Reinforcement Learning (AIRL) 등의 기법[10]이 제안되어 더욱 효과적인 보상 함수 추론이 가능
    - AIRL은 생성적 적대 신경망의 아이디어를 IRL에 적용한 기법으로 목적함수는 다음과 같음:

$$\min_{\theta} \max_{\omega} L(\theta, \omega) = \mathrm{E}_{\pi_E}[\log D_{\omega}(s,a)] + \mathrm{E}_{\pi_\theta}[\log(1 - D_{\omega}(s,a))], \quad D_{\omega}(s,a) = \frac{\exp(f\_\omega(s,a))}{\exp\left(f_{\omega(s,a)}\right) + \pi_\theta(a|s)}$$

  - 여기서 f_{ω}는 보상함수를 나타내는 신경망
  - AIRL은 복잡한 도메인에서 전문가의 전략을 모방하고 내재된 보상 구조를 파악하는 데 유용할 수 있다고 알려져있음.

Method

- Framework

Method

- Framework - detailed

Method

- Data

1. 대상 종목은 Dow jones 30 종목 중 대표종목

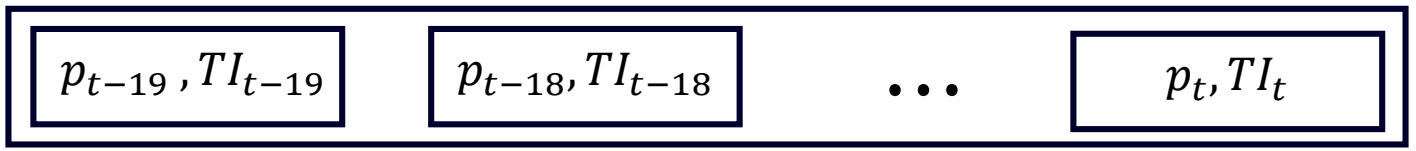
   : train: 2005.01.01~2017.12.31, valid: 2018.01.01~2020.12.31, test: 2021.01.01~2024.12.31

2. Data processing (Indicator & states)

   : open, high, low, close, volume, SMA, MACD, RSI, 현재 포지션 정보 (종목 보유 여부)

   : 20거래일의 데이터가 하나의 states를 구성

S_t :



Method

- Reward function
  - 수익률 기반

$$R_C = asset\ - previous\ asset$$

$$R_{CV} = (asset\ - previous\ asset) - volatility$$

$$R_p = \frac{\sum returns\,[returns > 0]}{|\sum returns\,[returns < 0]|}$$

  - 샤프 비율

$$R_{shr} = \frac{E[returns]}{\sigma_{returns}}$$

$$R_{sor} = \frac{E[returns]}{\sigma_{returns[returns<0]}}$$

  - 승률 기반

$$R_{Kelly} = count[returns > 0] - (\frac{count[returns < 0]}{E[returns > 0]/E[returns < 0]})$$

Method

- Inverse RL – Adversarial Inverse RL (AIRL) [10]

  1. 목표: 전문가 정책 π^*와 비슷한 정책 π_θ를 학습하는 것이 목표이며, 이 과정에서 보상함수 $\hat{R}$을 직접적으로 학습
  2. 주요 구성요소*

     : **생성자 (Generator; Policy network)** – 정책 네트워크로 에이전트가 환경에서의 행동을 생성 (거래 신호 생성)

     : **판별자 (Discriminator)** – 전문가의 데이터와 생성자의 데이터를 구분 (+ RL agents)

     : **보상함수 (Reward function)** – state-action pair에 대한 보상을 할당

Results - Test

- Trading rules
  - RL agent는 PPO
  - 거래 수수료: 0.1%, Short 포지션 불가능 (buy, hold, sell)

Conclusion

- **Summary**
  - 여러 리워드를 융합할 수 있는 방법을 제시하여 금융 시장의 복잡성과 변동성을 반영한 보상 함수를 도출
  - 학습된 리워드 함수를 통한 새로운 정책 학습
  - 유동성이 풍부한 몇몇 종목에서 우수한 성능을 달성

- **Limitations**
  - 모델의 복잡성 및 과적합 위험성

- **Future works & Plan**
  - Standard RL 결과와의 비교 분석 (수익률, MDD, sharpe ratio 등)
  - 과적합을 줄이기 위한 방법 적용 및 다른 시장에 적용
  - Sliding window 방식 적용

References

1. Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. IEEE transactions on neural networks and learning systems, 28(3), 653-664.
2. Moody, J., & Saffell, M. (1998). Reinforcement learning for trading. Advances in Neural Information Processing Systems, 11.
3. Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020). Deep reinforcement learning for automated stock trading: An ensemble strategy. In ACM International Conference on AI in Finance.
4. Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. R. (2021). Deep Attentive Learning for Stock Movement Prediction from Social Media Text and Company Correlations. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 3088–3101.
5. Liu, X., Huang, D., Ren, Y., & Li, H. (2022). FinRL-Meta: Market Environments and Benchmarks for Data-Driven Financial Reinforcement Learning. Advances in Neural Information Processing Systems, 35, 16763-16777.
6. Wu, X., Chen, H., Wang, J., Troiano, L., Loia, V., & Fujita, H. (2020). Adaptive stock trading strategies with deep reinforcement learning methods. Information Sciences, 538, 142-158.
7. Lim, M. H. Q., Lesmana, N. S., & Pun, C. S. (2024, November). Autoregressive DRL with Learned Intrinsic Rewards for Portfolio Optimisation. In Proceedings of the 5th ACM International Conference on AI in Finance (pp. 353-360).
8. Gao, Y., Lui, K. Y. C., & Hernandez-Leal, P. (2021). Robust risk-sensitive reinforcement learning agents for trading markets. arXiv preprint arXiv:2107.08083.
9. Guo, H., Lin, J., & Huang, F. (2023, June). Market Making with Deep Reinforcement Learning from Limit Order Books. In 2023 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
10. Fu, J., Luo, K., & Levine, S. (2017). Learning robust rewards with adversarial inverse reinforcement learning. arXiv preprint arXiv:1710.11248.