Reinforcement learning and Inverse Reinforcement learning

Applications in Finance

진영봉

FRE LAB

Apr 12, 2024

Contents

Introduction

- Notation
  - Markov Decision process (MDP): Collection $(\mathcal{S}, \mathcal{A}, T, \mathcal{R}, \gamma)$
    - state space $\mathcal{S}$, action space $\mathcal{A}$, (short-term) reward $r_t \in \mathcal{R}$
    - discount factor $\gamma \in [0,1)$, policy $\pi \in \Pi$ mapping $\mathcal{S}$ to $\mathcal{A}$
    - transition function T: $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0,1]$, reward function $\mathcal{R}$: $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$

- Basic of Reinforcement Learning
  - 강화학습은 보상을 극대화하는 policy를 찾는 방법

$$V^{\pi}(s_t) = \mathbb{E}_{\mathbb{P}}\left(\sum_{k\in\mathbb{N}} \gamma^k r_{t+k} \,\middle|\, s_t, \pi\right), \text{ where } r_t = \mathbb{E}_a[\mathcal{R}(s_t, a, s_{t+1})],\ \mathbb{P}(s_{t+1}|s_t, a_t) = T(s_t, a_t, s_{t+1})$$

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{\mathbb{P}}\left(\sum_{k\in\mathbb{N}} \gamma^k r_{t+k} \,\middle|\, s_t, a_t, \pi\right)$$

$$Q^{\pi}(s_t, a_t) = \sum_{s'\in\mathcal{S}}\left[r(s_t, a_t, s') + \gamma Q^{\pi}\left(S', \pi(s')\right)\right]T(s_t, a_t, s') \qquad \text{(Bellman equation)}$$

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(r(s,a,s') + \gamma \max_{a'\in\mathcal{A}}\{Q(s',a')\})$$

  - 일반적인 기계학습 알고리즘과 달리, 강화학습은 학습 시스템과 해당 데이터에 대한 피드백 루프가 있음 [1]

Introduction

- Basic of Inverse Reinforcement Learning [2, 3]
  - 역강화학습은 optimal policy와 모델이 주어졌을 때, reward function을 찾는 방법

Find $\mathcal{R}^*$ s.t. $\mathbb{E}[\sum_{t=0}^{\infty}\gamma^t\mathcal{R}^*(s_t)|\pi^*] \geq \mathbb{E}[\sum_{t=0}^{\infty}\gamma^t\mathcal{R}^*(s_t)|\pi] \quad \forall\pi$

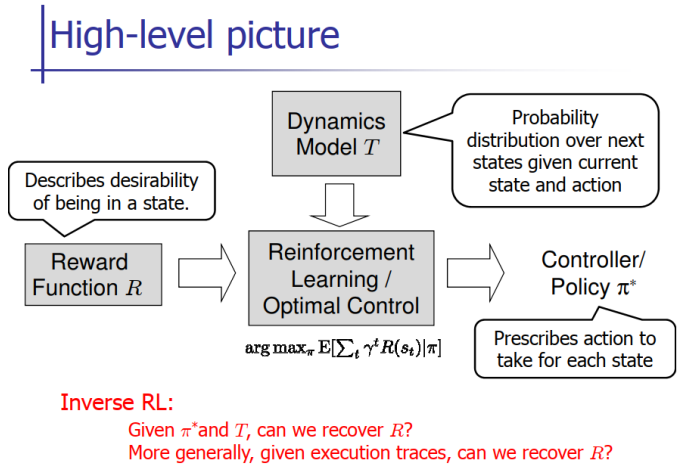
Let $\mathcal{R}(s) = w^{\mathrm{T}}\phi(s)$ where $w \in \mathbb{R}^n,\ \phi: \mathcal{S} \rightarrow \mathbb{R}^n$

$$\mathbb{E}[\sum_{t=0}^{\infty}\gamma^t\mathcal{R}(s_t)|\pi] = \mathbb{E}[\sum_{t=0}^{\infty}\gamma^t w^{\mathrm{T}}\phi(s)\,|\pi] = w^{\mathrm{T}}\mathbb{E}[\sum_{t=0}^{\infty}\gamma^t\phi(s)\,|\pi] = w^{\mathrm{T}}\mu(\pi)$$

➔ Find w^* s.t. ${w^*}^{\mathrm{T}}\mu(\pi^*) \geq {w^*}^{\mathrm{T}}\mu(\pi) \quad \forall\pi$

$$\min_{w}\|w\|_2^2 + C\sum_i \xi^{(i)}$$

$$\text{s.t} \quad w^{\mathrm{T}}\mu(\pi^{(i)*}) \geq w^{\mathrm{T}}\mu(\pi^{(i)}) + m(\pi^{(i)*}, \pi^{(i)}) - \xi^{(i)} \qquad \forall i, \pi^{(i)}$$



- 해당 알고리즘은 전문가의 실제 보상함수를 올바르게 복구한다는 보장은 없지만, 해당 보상함수 만큼의 성능을 발휘하는 정책을 찾을 수 있음[2-4]

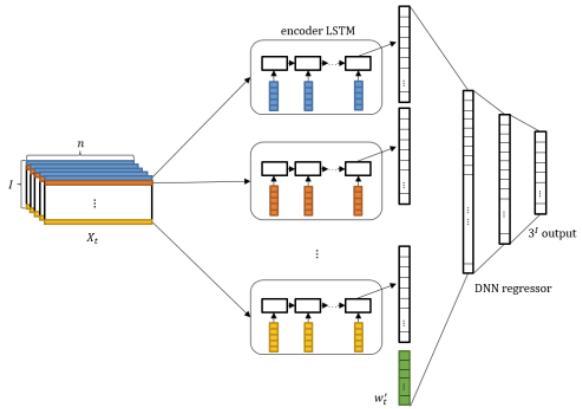
RL & IRL in Finance

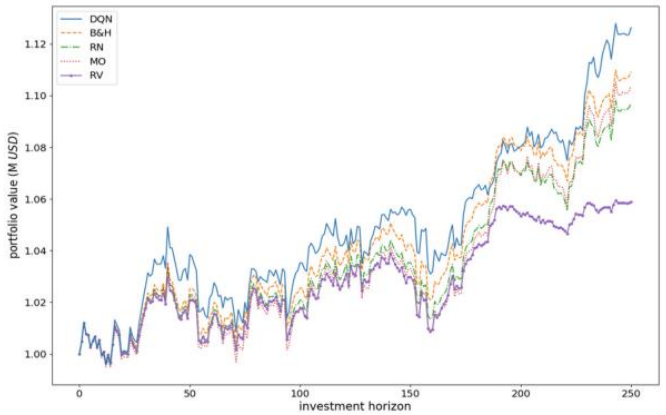
- Applications in RL & IRL

task	Ref.	RL algorithm	states	Reward function	Performance measure
Trading & portfolio	[6-8]	Value-based Policy-based Actor-critic	Asset price Asset past return Holding asset Balance	(Portfolio) return Sharpe ratio	(Portfolio) return Sharpe/Sortino ratio Cumulative Profit
Option pricing & hedging	[9,10]		Asset price Current position Option strikes Time remaining expiry	Expected return Option payoff Hedging cost	Hedging strategy (BSM, Heston) Hedging cost/error/loss
Market making	[11,12]		Bid/ask prices Holding asset order-flow imbalance	Profit and loss implementation shortfall	Profit and loss
Inverse RL	[13-16]	Trading: Transaction / sentimental LOB dynamics: GP IRL Asset allocation recommendations: T-REX			-

RL & IRL in Finance

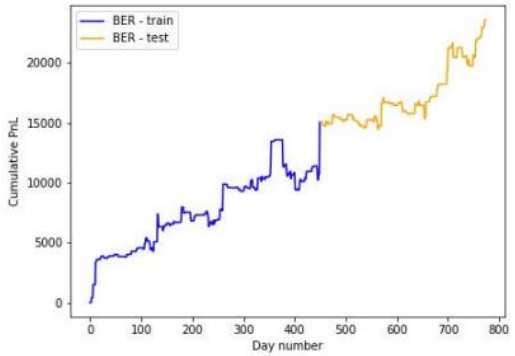
- Trading & Portfolio optimization
  - An intelligent financial portfolio trading strategy using deep Q-learning (2020) [8]
    - State space: 포트폴리오 비중, 자산 가격 및 지표
    - Action space: buy, hold, sell
    - Reward function: portfolio return
    - 자산군: 각각 미국, 한국 시장의 ETF 3종목
    - RL 모델: DQN
    - 불가능한 action이 가장 우수한 경우, 유사한 value를 갖는 다른 action을 선택



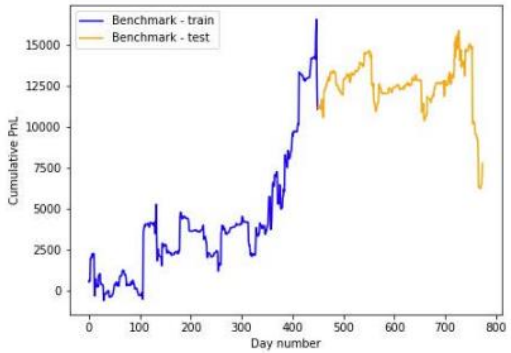


RL & IRL in Finance

- Market Making
  - High Frequency Automated Market Making Algorithms with Adverse Selection Risk Control via Reinforcement Learning (2021) [12]
    - State space: Book Exhaustion Rate (5min, 30min)
    - Action space: bid, ask (정해진 거래계약 개수 셋에서 각각 선택{1,5,0})
    - Reward function: cash balance와 position간의 선형 조합
    - 자산군: E-mini S&P 500 futures, 10-Year Treasury Note future

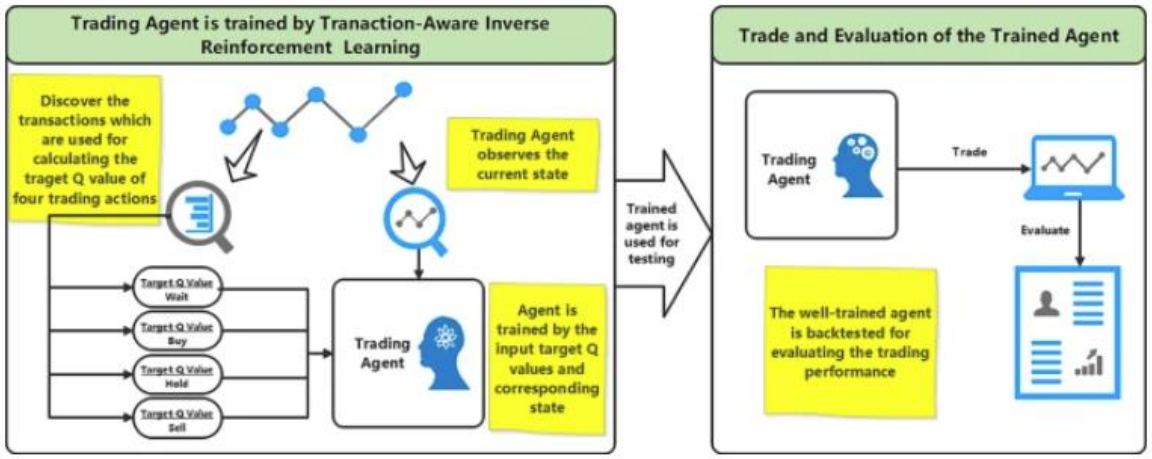


- 금융 분야에서 RL의 최근 연구 동향
  - State space의 확장 (candle chart 이미지, 뉴스 기사)
  - RL 모델의 feature extractor의 정교화
  - RL 에이전트 앙상블



RL & IRL in Finance

- Inverse Reinforcement Learning
  - Transaction-aware inverse reinforcement learning for trading in stock markets (2023) [13]



(a) Trading actions recorded on AMZN stock

(b) Total value recorded during the testing period for AMZN stock

- Trading task에서 지연보상 문제에 초점을 맞추어 reward function을 새롭게 디자인
- 모든 transaction을 고려하여 전문가(RL optimal policy) 시연을 생성
- 벤치마크 대비 우수한 성능 달성

Summary

- Summary
  - 금융 산업의 발전과 디지털화로 인해, 다양한 리스크 관리 및 투자기회가 창출되고 있음
  - RL은 비교적 가벼운 가정으로 financial decision-making problems을 해결할 수 있음
  - 기존 classical model을 RL에 단순히 적용하는 것을 넘어 복잡한 금융 시장 특성을 반영할 수 있도록 발전하고 있음
  - 다양한 task에서 우수한 성능을 보여주고 있으며, 학계 및 금융 기관에서의 RL에 대한 관심도가 높아지고 있음
  - 아직까지 금융시장에 IRL이 적극적으로 채택되고 있지는 않지만, 점차 그 효용을 보여주고 있음

- Discussion & Future plan
  - 최근 RL을 이용한 투자 전략의 경우, 추가적인 state 확보 및 feature extractor 구조를 정교하게 만드는 것에 집중
  - RL의 reward function은 중요한 요소이지만, 금융시장에 명확한 reward는 밝혀지지 않았음
  - IRL을 통해 금융 시장에 적합한 reward를 찾아 RL 정책을 개선해 나가고자 함
  - Self-supervised learning과 stochastic control theory등을 추가적으로 공부하고 IRL에 적용하여 RL-IRL 피드백(상호개선) 구조를 구축하고 금융시장에 적용하는 것을 목표로 하고 있음

References

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[2] Abbeel, P., & Ng, A. Y. (2010). Inverse Reinforcement Learning.

[3] Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning*. In Proceedings of international conference on Machine learning*.

[4] Ratliff, N. D., Bagnell, J. A., & Zinkevich, M. A. (2006). Maximum margin planning. *In Proceedings of international conference on Machine learning*.

[5] Hambly, B., Xu, R., & Yang, H. (2023). Recent advances in reinforcement learning in finance. *Mathematical Finance*.

[6] Du, X., Zhai, J., & Lv, K. (2016). Algorithm trading using q-learning and recurrent reinforcement learning. *Positions.*

[7] Pendharkar, P. C., & Cusatis, P. (2018). Trading financial indices with reinforcement learning agents. *Expert Systems with Applications*.

[8] Park, H., Sim, M. K., & Choi, D. G. (2020). An intelligent financial portfolio trading strategy using deep Q-learning. *Expert Systems with Applications*.

[9] Cao, J., Chen, J., Hull, J., & Poulos, Z. (2021). Deep Hedging of Derivatives Using Reinforcement Learning. *The Journal of Financial Data Science*.

[10] Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2019). Deep hedging. *Quantitative Finance*.

References

[11] Spooner, T., & Savani, R. (2020). Robust market making via adversarial reinforcement learning. *arXiv preprint arXiv:2003.01820.*

[12] Zhao, M., & Linetsky, V. (2021, November). High frequency automated market making algorithms with adverse selection risk control via reinforcement learning. *In Proceedings of the second ACM international conference on AI in finance*

[13] Sun, Q., Gong, X., & Si, Y. W. (2023). Transaction-aware inverse reinforcement learning for trading in stock markets. *Applied Intelligence.*

[14] Yang, S. Y., Yu, Y., & Almahdi, S. (2018). An investor sentiment reward-based trading system using Gaussian inverse reinforcement learning algorithm. *Expert Systems with Applications*.

[15] Roa-Vicens, J., Chtourou, C., Filos, A., Rullan, F., Gal, Y., & Silva, R. (2019). Towards Inverse Reinforcement Learning for Limit Order Book Dynamics. *arXiv preprint arXiv:1906.04813.*

[16] Halperin, I., Liu, J., & Zhang, X. (2022). Combining reinforcement learning and inverse reinforcement learning for asset allocation recommendations. *arXiv preprint arXiv:2201.01874.*