



강화학습 기반 부동산 경매 포트폴리오 입찰 전략 연구

서울 아파트 경매 데이터를 기반으로

서강대학교 AI·SW대학원

A70055 / 박은지

(GitHub Link) https://github.com/eumjil/RL_project



목차

- 01 주제 및 목표
- 02 활용데이터
- 03 State, Action, Reward 설계
- 04 RL 환경 및 의사결정구조
- 05 강화학습 알고리즘 및 hyperparameter
- 06 실험 셋업
- 07 실험 결과
- 08 토의 및 결론



1. 주제 및 목표

프로젝트 주제

- 예산 및 대출(LTV) 등 금융 제약조건을 고려한 아파트 경매 입찰 전략을 강화학습을 통해 학습하고, 순차적 매물 선택 과정에서 수익 극대화 정책을 도출

프로젝트 목표

- 부동산 특성, 자본 제약 등 다양한 상태정보를 고려하여 수익을 최대화하는 입찰 정책을 학습
- 강화학습이 경매 환경에서 합리적 전략을 스스로 발견/검증할 수 있는지 분석
- 입찰 전략 변화 및 목표 수익 수렴 추이를 검증, RL 정책의 타당성 평가

현실적 금융 제약 조건 속에서 수익률을 극대화하는 최적 경매
입찰 정책을 강화학습으로 탐색



2. 활용데이터 - 데이터 소개

서울시 부동산(아파트) 경매데이터 (1,959건)

- (데이터 출처) 세종옥션(SJAU) 경매정보, 국토부 부동산 실거래가 정보
- (수집방법) 웹크롤링, 공공데이터 다운로드
- (대상기간) 2024.1. 1. ~ 2025. 9. 30.(매각기일 기준)
- (대상지역) 서울특별시
- (물건 유형) 아파트 매각 완료건

(데이터 연계) 경매데이터와 부동산 실거래가정보의 실거래가, 건축년도, 실거래건수 연계

(변수 구성)

변수 그룹	내용	변수 그룹	내용
물건 특성	감정가, 면적, 층, 건물나이	시세 정보	실거래가
입찰 정보	최저가율, 유찰횟수	결과 변수	낙찰가, 낙찰가율
입지 정보	시군구, 실거래건수		



2. 활용데이터 - 데이터 전처리



수치형 변수 Log Transform

- > Scikit-learn 기반 HGBR 구현
- > 비선형성 자동 학습 가능
- > 대규모 데이터셋에 효율적

gradient 안정화

분포 정상화



파생변수 생성

- > 범주형 가격비율 변수 생성 ; 최저가율, 낙찰가율, 실거래가율, 저평가율
- > 건물나이 (= 2025 - 건축년도)
- > 실거래상대지수(local/global)



범주화 + One-hot Encoding

- > 면적구분 → 면적_소형/중형/대형
- > 층구분 → 층_low/mid/high

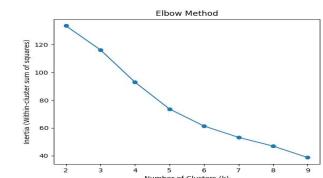


시군구 KMeans 군집화

- > 감정가, 실거래지수, 낙찰가율 등 구 단위 평균값 이용
- > KMeans 군집화 결과, 5개 그룹으로 구분(k=5)
- > 구_구분_0 ~ 구_구분_4 생성

입지 특성 요약

state dimension 축소



입찰 의사결정에 필요한 가치 정보(28개 feature)를 포함



3. State, Action, Reward 설계

1) State (상태)

- State dimension: 22

* 미래 수익에 영향을 미치는 시장·건물 특성과, 재정 상태를 표현하는 현금비율·진행도를 포함

구분	변수명	설명
재무상태 변수	Cash_ratio (현금 비율)	현재 보유 현금 ÷ 초기 예산 (재무 여력 정보 제공)
학습/에피소드 변수	Episode_progress	현재 물건 순서 ÷ 에피소드 총 물건 수 (시간 경과 정보)
Log 변환 변수	감정가_log 건물면적_log 실거래가_log	감정가의 log변환 건물면적(m^2) log변환 실거래가 log변환
비율/지표 변수	최저가율 유찰횟수_adj	최저가 / 감정가 유찰횟수 예상
시장/지역 변수	월 실거래 건수 실거래상대지수_local 실거래상대지수_global	해당 지역 월 거래량 지역 내 거래활성도 상대수준 서울 전체 대비 거래활성도 상대수준
건물 특성	건물나이	2025 - 건축년도
범주형(One-Hot)	면적_소형/중형/대형 층_low/mid/high 구_구분_0~4	면적 구분 더미 층 구분 더미 시군구 군집(Kmeans, k=5)



3. State, Action, Reward 설계

2) Action (행동)

- 입찰가 수준 5단계 : Action space = {입찰 없음, 70%, 80%, 90%, 100%}

Action	의미	감정가 대비
0	입찰하지 않음	0%
1	보수적 입찰	70%
2	중도 입찰	80%
3	공격적 입찰	90%
4	최고가 입찰	100%

- 낙찰조건 : 입찰가 \leq 예산(초기 10억) + 대출가능액(LTV 60% 고정)

3) Reward (보상)

- 낙찰 성공시, reward = (실거래가 - 입찰가) / 1,000 * 학습안정성을 위해 1,000으로 나눔
- 낙찰 실패시, reward = 0
- 무리한 투자 (현금 부족시) 패널티 부여 : -5,000(만원)



4. RL 환경 및 의사결정구조

환경 구성

- 한 에피소드는 여러 경매 물건(max 50개) 을 순차적 입찰로 구성
- step 순서 : 물건 선택 → 입찰 행동 → 낙찰 여부 → Reward

에피소드 구성

- 초기 현금 : `initial_cash = 10억 원`
- 에피소드당 반복횟수 : `max_items_per_episode = 50`
- 에피소드 종료 조건: ①50개 물건을 모두 처리 OR ②현금이 일정 수준(1,000만) 이하

입찰 의사결정 관점 해석

- 각 step의 reward : 개별 매물 선택의 즉각적 결과(수익 또는 패널티)
- 한 episode의 return : 제한된 자본 하에서 여러 입찰 선택이 누적된 총 수익
→ 순차적·제약 기반 의사결정 문제로 경매 입찰을 재구조화한 RL 시뮬레이션



5. 강화학습 알고리즘 및 hyperparameter

사용 알고리즘



DQN (Deep Q-Network)

- > 2층 MLP, ReLU 활성화
- > Replay Buffer + Target Network 사용
- > epsilon-greedy 탐색



SARSA (NN 기반 on-policy)

- > DQN과 유사한 네트워크 구조
- > On-policy 업데이트



Baseline

- > Random Policy
- > Conservative 70% Policy

Hyperparameters

항목	DQN	SARSA
네트워크 구조	128 → 128 → n_actions	
Optimizer	Adam ($\text{lr} = 1\text{e-}3$)	
Discount Factor (γ)	0.99	
탐색 방법 (Epsilon-Greedy)	epsilon_start = 1.0 epsilon_end = 0.05 epsilon_decay = 3000 (exp)	
Replay Buffer	capacity = 50,000	적용하지 않음
Batch Size	64	Online update (Batch 사용 X)
Target Network	target_update_freq = 200	적용하지 않음
학습 Episode 수	500	



6. 실험 셋업

데이터 및 환경 구성

- 실제 서울 아파트 경매 사례 1,959건
- 환경: **PortfolioAuctionEnv**
 - 초기 예산: **10억 원**, LTV = 60% 적용
 - Episode는 여러 물건을 무작위로 순차 입찰
- State 차원: **22 features**
Action: **5가지** 입찰 전략
- 학습 방식: 총 500 Episodes, ϵ -greedy 탐색, CPU 환경

Evaluation Metric

- Episode Return = 누적 총 수익
(Reward 합계, 만원 단위)
- 비교 기준: Episode Return의 분포 및 신뢰구간
 - 평균(mean), 표준편차(std),
$$95\% \text{ 신뢰구간 } CI = 1.96 \cdot \frac{std}{\sqrt{n}}$$

Seed 기반 반복 실험

- Seed: {42, 11, 99}
- Seed당 30 Episode 평가 → 총 90 회 측정
- 목적: “Random seed 변경에 따른 성능 안정성 및 신뢰구간 검증”

비교 대상 정책

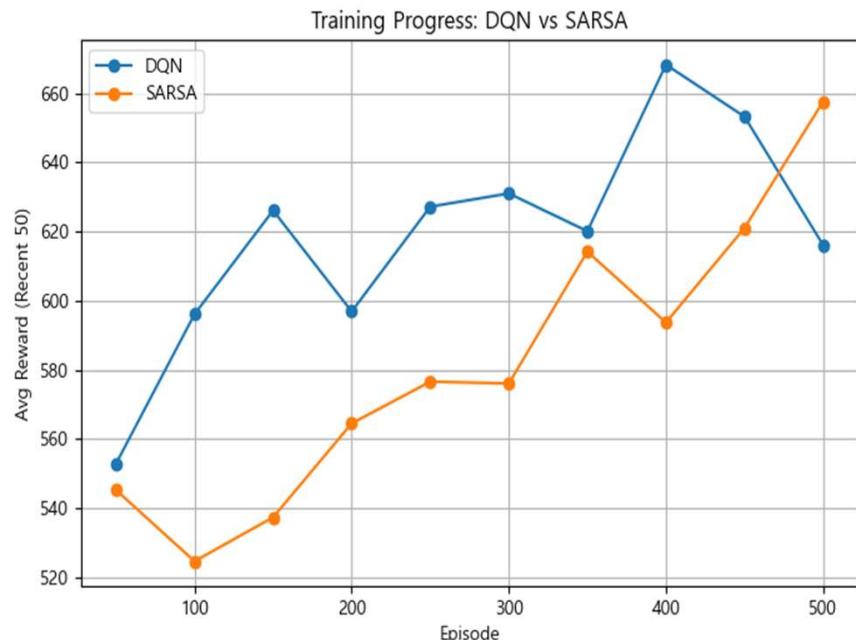
정책	설명
Random	무작위 입찰 전략
Conservative 70%	항상 보수적 비율로 고정 입찰
DQN (Greedy)	학습된 Q값 최대 행동
SARSA (Greedy)	On-policy 업데이트 기반 탐색 반영 전략



7. 실험 결과 - 학습곡선

알고리즘별 학습 로그

Episode	DQN		SARSA	
	평균 Reward	Epsilon	평균 Reward	Epsilon
50	552.8	0.984	545.2	0.984
200	596.8	0.939	564.5	0.939
400	668.3	0.881	614.2	0.895
500	616.1	0.854	657.6	0.854



- DQN, SARSA 모두 초기(50 episode 전후) 대비 평균 Return이 증가하며 학습이 진행됨
- DQN은 중간 구간에서 높은 Return에 도달 후 변동
- SARSA는 다소 보수적이지만, 최종 500 episode 기준 마지막 50 평균이 더 높게 나타남



7. 실험 결과 - 정책별 성능 비교

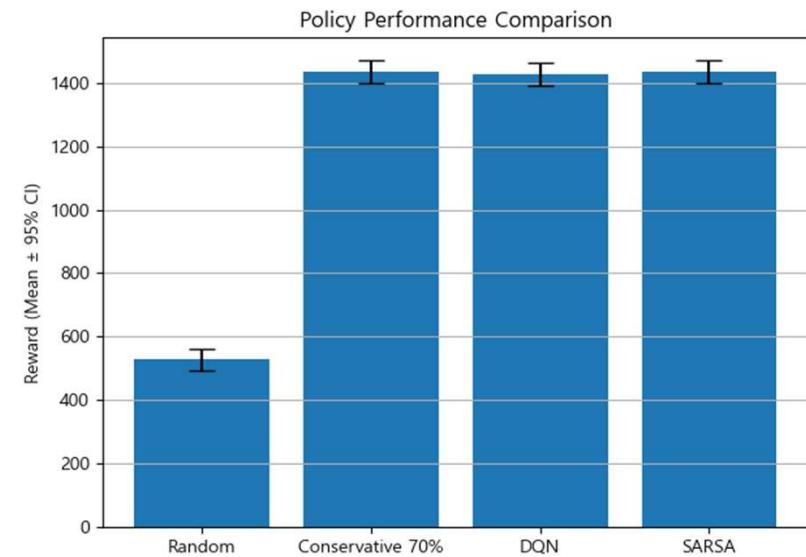
정책별 평가 결과 (n = 90)

Policy	Mean Return	Std	95% CI (\pm)
Random	527.34	163.79	33.84
70% 고정 정책	1435.06	170.10	35.14
DQN (greedy)	1428.33	171.52	35.44
SARSA (greedy)	1435.06	170.10	35.14

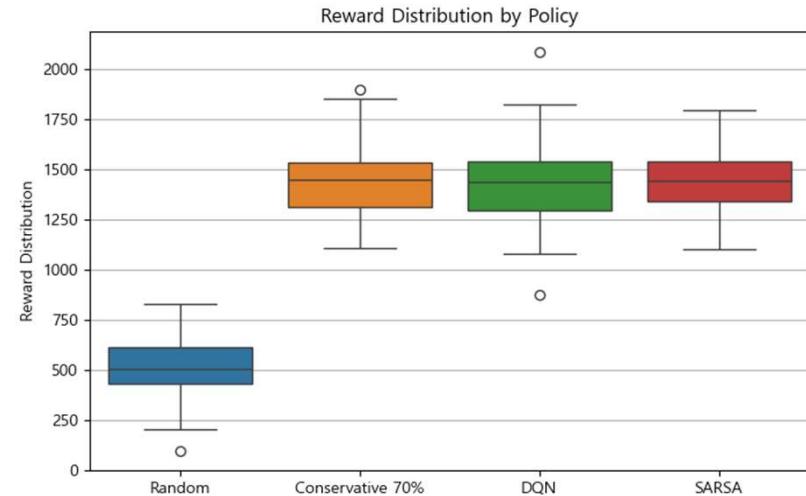
※ 모든 Episode Return 값은 '(수익 / 1,000)' 처리된 값이며, 단위는 "만원"

- Random 대비, 모든 정책 기반에서 수익이 크게 향상(약 2.7배)
- 70% 고정 정책은 단순한 규칙임에도 높은 성능을 보임
- DQN/SARSA는 여러 상태 변수와 재무 제약을 고려해 학습한 결과로, 70% 수준의 보수적인 입찰 전략에 수렴

<정책별 평균 Return Bar Plot>



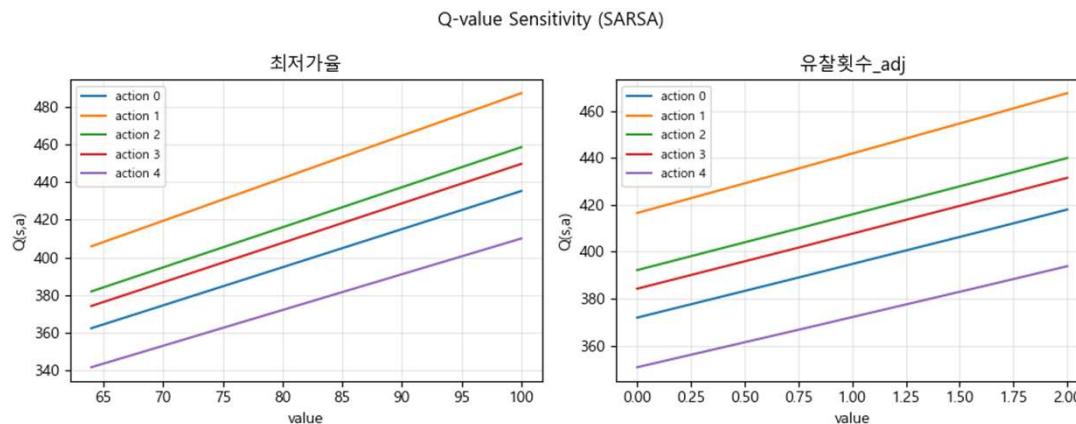
<정책별 Return 분포 Boxplot>





7. 실험 결과 - (심화분석) 결정 요인 분석

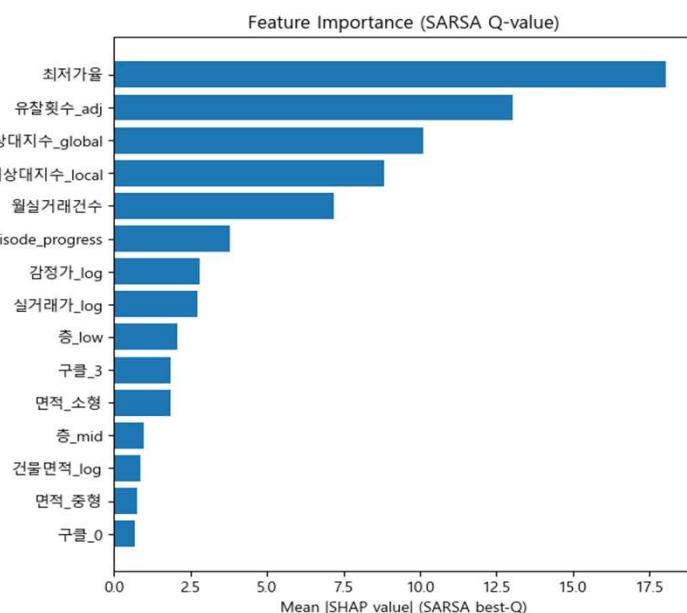
Q-value 민감도 분석 결과 (Policy Sensitivity)



매물 특성 변화에 따라 모든 행동의 기대수익(Q-value)은 증가하지만, 항상 '보수적 입찰(Action 1: 감정가 70%)'이 가장 높은 Q-value를 유지

- 최저가율↑·유찰횟수↑·시장 유동성↑ → 매수 기대 수익 증가
- 공격적 입찰(100%)은 수익 증가 폭 제한적 → 위험 대비 효율 ↓
- Action 1(70%)의 Q-value 우위 → 안정적 수익 기대

SHAP 기반 Q-value 변수 중요도 분석



Q-value 기반 변수 중요도 분석에서는
최저가율(저평가 정도), 유찰횟수, 시장 유동성
지표(지역/전체)가 수익 판단에 가장 큰 기여도를 보임.
이 요인들이 낙찰 가능성과 저평가 매물 선별의 핵심 지표

"저평가 + 유동성 높은 매물 선별" 전략의 자동화



8. 토의 및 결론

결론

- 실제 서울 아파트 경매 데이터(1,959건)를 기반으로, 예산·대출 제약, 낙찰 여부, 과도한 레버리지 패널티 등을 반영한 환경을 구성하고 강화학습 (DQN·SARSA) 을 적용하여 학습함
- 실험 결과, 두 알고리즘 모두 감정가 대비 약 70% 수준의 보수적 입찰을 선택했으며, 이는 고정 규칙의 모사가 아니라 저평가·유동성·유찰정보가 유리한 조건에서 선택된 전략임
- 따라서 실제 시장에서 경험적으로 사용되던 '저점 입찰 전략'(70%)이 데이터 기반으로도 가장 효율성이 높은 정책임을 정량적으로 검증한 것으로 해석됨
- Q-value·SHAP 분석을 통해 '저평가 매물&시장 유동성'이 수익 극대화를 결정하는 핵심 요인임이 확인됨
- 강화학습은 새로운 전략을 무작정 생성하기보다, 시장 경험적 전략의 타당성을 '정량 검증 및 보완'하는 분석도구로 활용될 수 있음을 시사함

보완 및 개선방안

- Reward에 취득세·중개수수료·보유비용 등의 현실 비용요소 추가
- 경쟁 참여자 의사결정을 반영한 Multi-agent RL 모델링
- 연속 입찰가학습(Continuous RL) 기반 Action 확장(DDPG/SAC 등 적용)