

강화학습 기반 국내 섹터 ETF 포트폴리오 최적화

PPO + LSTM을 이용한 섹터 로테이션 및 리밸런싱 전략

 학번 / 이름

A70083 / 임보라

 GitHub

https://github.com/boraelle/RL_prj_stock_portfolio

프로젝트 개요 및 목표

! 문제 정의

- 국내 섹터 ETF 12종과 현금을 대상으로 하는 주식 포트폴리오 최적화 문제
- 시장의 변동성에 대응하여 매일 자산 비중을 조절(Daily Rebalancing)함으로써, 리스크를 관리하면서도 시장 대비 초과 수익을 달성하는 것을 목표로 함

📊 핵심 목표 수치

Sharpe Ratio

≥ 1.5

연환산 수익률

$\geq 30\%$

Max Drawdown

$\leq -15\%$

🧠 해결 접근법: PPO + LSTM

시계열 특성 학습: LSTM(Long Short-Term Memory) 네트워크를 사용하여 과거 20일간의 가격 및 지표 패턴 분석

안정적인 정책 학습: PPO(Proximal Policy Optimization) 알고리즘을 적용하여 급격한 정책 변화를 방지하고 안정적인 학습을 유도

멀티모달 상태 표현: 가격 정보, 기술적 지표, 포트폴리오 상태를 결합하여 의사결정에 필요한 정보를 종합적으로 처리

포트폴리오 자산 구성

위험자산 : 섹터별 ETF (12종 선정)

- ⚡ Energy (에너지)
- 🏭 Industrials (산업재)
- 👕 Discretionary (자유소비)
- 🏦 Financials (금융)
- 💻 Semiconductor (반도체)
- 🛡 Defense (방산)
- 🏠 Materials (소재)
- 🛒 Staples (필수소비)
- 💊 Healthcare (헬스케어)
- 💡 IT (정보기술)
- 🔋 Battery (2차전지)
- 🧬 Bio (바이오)

무위험 자산 : 현금

- 💰 Cash (현금) 연 2.0% 고정 수익

포트폴리오 요약

총 자산 수	13개
위험 자산 (ETF)	12개
무위험 자산	1개 (현금)

데이터 기간 및 전처리

🕒 데이터 수집 기간

2024.01.01 ~ 2025.11.30

23

개월 (Months)

465

거래일 (Trading Days)

📊 데이터셋 분할 (80:20)

Training Set (80%)

Validation Set
(20%)

● 학습 데이터 (약 372일)

● 검증 데이터 (약 93일)

i 시계열 순서를 유지하여 미래 데이터를 학습하지 않도록 분할

✂ 데이터 전처리 프로세스

1. 데이터 정렬 및 정제

- 모든 자산의 OHLCV 데이터를 날짜 기준으로 정렬
- 중복 데이터 제거하여 일관성 확보

2. 결측치 처리 (Zero Padding)

- 상장 폐지나 거래 정지 등으로 인한 결측 구간은 0으로 채워 데이터 차원 유지 (Lookback Window 부족 시 등)

3. Feature 생성

- MA5, MA20, RSI, MACD 등 기술적 분석 지표를 사전에 계산하여 Feature로 활용

4. 정규화 (Normalization)

- 모델의 효율적인 학습을 위해 자산 데이터 가격대 스케일링 처리 :

State 설계 (Multi-Modal)

에이전트가 가격 추세, 기술적 지표, 그리고 현재 포트폴리오 상태를 종합적으로 고려하도록 3가지 다른 형태(Modality)의 정보를 동시에 관찰하도록 설계함

Price History

📊 Shape: (12, 20, 5)

12개 ETF 자산별 시계열 데이터

20일 룩백 윈도우 (과거 20일치)

OHLCV 5개 피처
: 시가, 고가, 저가, 종가, 거래량

Technical Indicators

📈 Shape: (12, 6)

MA5, MA20: 단기/중기 이동평균

RSI: 상대강도지수 (모멘텀)

MACD: 추세 및 발산 지표

Bollinger Bands: 변동성 상단

Volume Ratio: 거래량 변동

Portfolio State

📁 Shape: (3,)

현금 비율: 현재 자산 중 현금 비중

보유 종목 수: 분산 투자 정도

Max Drawdown (MDD)
: 현재 포트폴리오의 최대 낙폭 상태
(리스크 관리용)

Action & Reward 설계



Action Space (행동)

포트폴리오 내 각 자산의 목표 비중을 결정하는 연속적인 값입니다.

$Shape: (13,)$

$12\ ETFs + 1\ Cash$

형태: Continuous Action Space

범위: $[0, 1]$ (Softmax 활성화 함수 적용)

제약조건: $\sum w_i = 1$ (비중 합은 항상 1)

빈도: Daily Rebalancing (매일 조정)



Reward Function (보상)

에이전트가 학습할 목표 함수로, 거래 비용을 차감한 순수익률을 최대화합니다.

$$R_t = r_{\{portfolio\}} - C_{\{transaction\}}$$

Portfolio Return		Transaction Cost		Final Reward
$r_{\{portfolio\}}$	$-$	$C_{\{transaction\}}$	$=$	R_t

* Realistic Transaction Cost (현실적 거래비용)

비현실적인 잦은 매매를 방지하고 실제 시장 환경을 모사하기 위해 거래 비용을 반영함.
포트폴리오 비중 변화량에 비례하여 비용이 발생하며, 증권사/개인별 상이함을 무시하고 static 값으로 고정함

Transaction Fee Rate

0.15%

PPO 알고리즘 개요



On-Policy 학습

- 현재 정책(Policy)으로 수집한 데이터만을 사용하여 학습을 진행
- 샘플 효율성은 낮지만 안정적인 학습 수렴을 보장



Clipped Surrogate Objective

- 정책 업데이트 비율(Ratio)을 일정 범위(예: 0.8~1.2)로 제한(Clipping)함
- 이를 통해 급격한 정책 변화로 인한 성능 붕괴를 방지



Actor-Critic 구조

- Actor: 최적의 행동(포트폴리오 비중) 결정
- Critic: 상태 가치(V-value)를 평가하여 학습 방향 가이드



GAE (Generalized Advantage Estimation)

- Advantage 추정 시 편향(Bias)과 분산의 트레이드오프 조절
- 안정적인 경사 하강법 학습 지원

$$L^{CLIP}(\theta) = E [\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)]$$



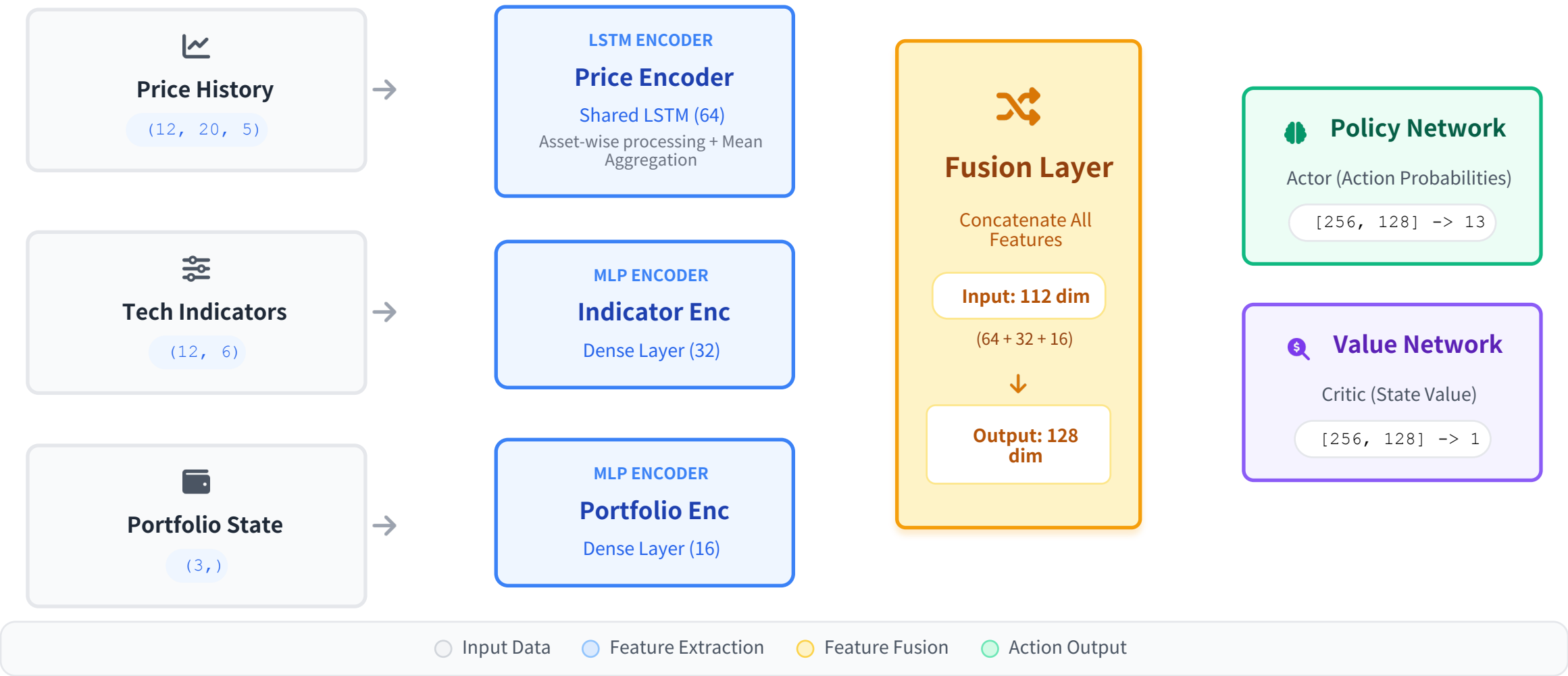
안정성(Stability): Trust Region Policy Optimization(TRPO)의 장점을 유지하면서 구현 복잡도를 낮춤



단순성(Simplicity): 1차 미분만 사용하여 계산 효율성이 높고 하이퍼파라미터에 덜 민감함

모델 아키텍처 (Multi-Modal LSTM)

Multi-Modal LSTM 구조로 가격 시계열, 기술 지표, 포트폴리오 상태를 독립 처리 후 통합하여 PPO 기반 Actor-Critic 네트워크로 최적 자산 배분을 학습함



주요 하이퍼파라미터

Parameter	Value	Description
PPO Training Configuration		
total_timesteps Total interaction steps for training	100,000	Standard training duration
learning_rate Step size for optimization	3e-4 (0.0003)	Adam Optimizer default
n_steps Steps per environment update	2,048	Batch collection size
batch_size Mini-batch size for gradient update	64	Optimization batch size
n_epochs Epochs per update	10	Reuse data 10 times
PPO Hyperparameters		
gamma (γ) Discount factor	0.99	Focus on long-term return
gae_lambda (λ) GAE smoothing parameter	0.95	Balance bias/variance
clip_range (ϵ) PPO clipping range	0.2	Conservative updates

환경 설정 (Environment)

1.0 억 초기 자본 (KRW)	0.15% 거래 비용
20일 Lookback Window	2.0% 현금 수익률 (연)

설정 근거

- 총 스텝 수(100k)는 약 23개월 데이터(465일) 대비 충분한 epoch 반복을 위해 설정
- 거래 비용(0.15%)은 실제 국내 주식 시장 수수료 및 세금을 반영한 현실적인 수치로 설정

평가 지표

평가 지표 (Evaluation Metrics)



총 수익률
Total Return



연환산 수익률
Annualized Return



Sharpe Ratio
Risk-Adjusted Return



Max Drawdown
Maximum Loss



변동성 (Volatility)
Standard Deviation of Returns

비교 대상 (Baseline Strategies)



Equal Weight (균등 비중)

모든 자산에 동일한 비중($1/N$)을 할당하고 정기적으로 리밸런싱하는 전략으로, 분산 투자의 가장 기본적인 벤치마크



Buy & Hold (매수 후 보유)

초기 시점에 자산을 매수하고 매도 없이 끝까지 보유하는 소극적 전략으로, 거래 비용이 발생하지 않는 것이 특징이나, 이 경우에는 위와 유사한 케이스임

실험 결과 비교 (Performance Comparison)



Total Return
108.03%
↑ vs Baseline +80.9%p



Annualized Return
51.55%
High Performance



Max Drawdown
-16.56%
Target: -15.00%

Strategy	Total Return	Sharpe Ratio	Max Drawdown
RL Agent (PPO) Proposed	108.03%	1.17	-16.56%
Equal Weight	27.05%	2.46 Stable	-6.05%
Buy & Hold	20~25% (Est.)	~2.0	~-8.0%

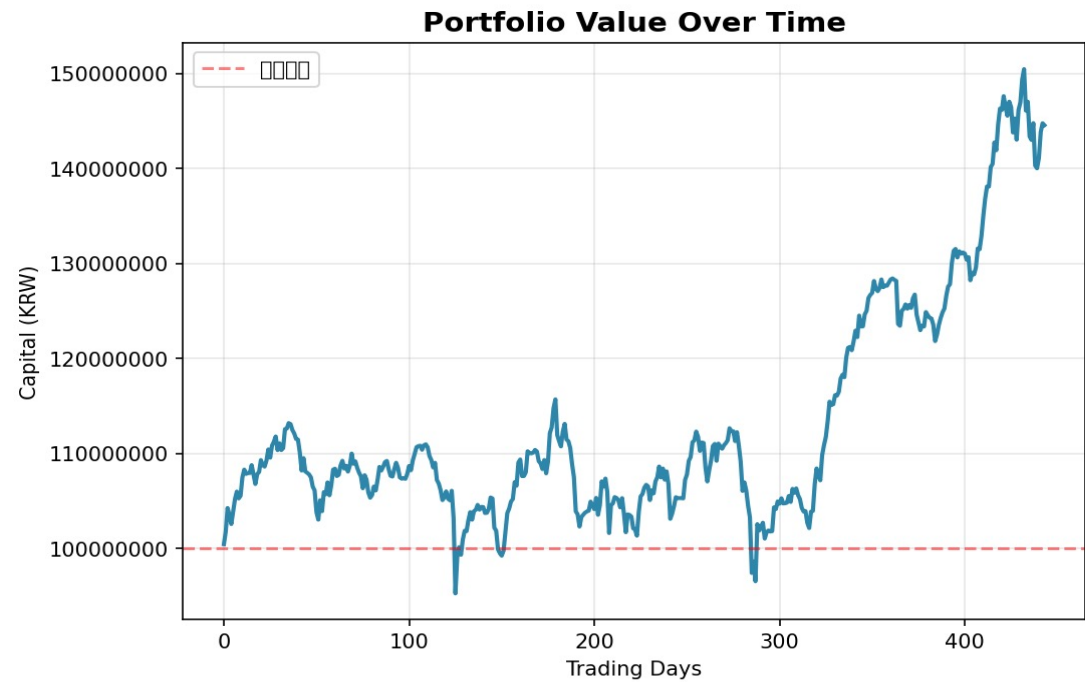
💡 결과 해석

수익성 우위
: RL Agent는 Equal Weight 대비 약 4배 높은 총 수익률을 기록

리스크 트레이드오프 존재
: 공격적인 투자 성향으로 인해 Sharpe Ratio(1.17)는 Baseline(2.46)보다 낮았으며, MDD(-16.56%) 또한 목표치(-15%)를 소폭 초과하여 안정성 보완이 필요



포트폴리오 가치 및 수익률 추이



🚩 초기 자본 (Initial)

100,000,000

KRW (Start)

🎯 최종 자본 (Final)

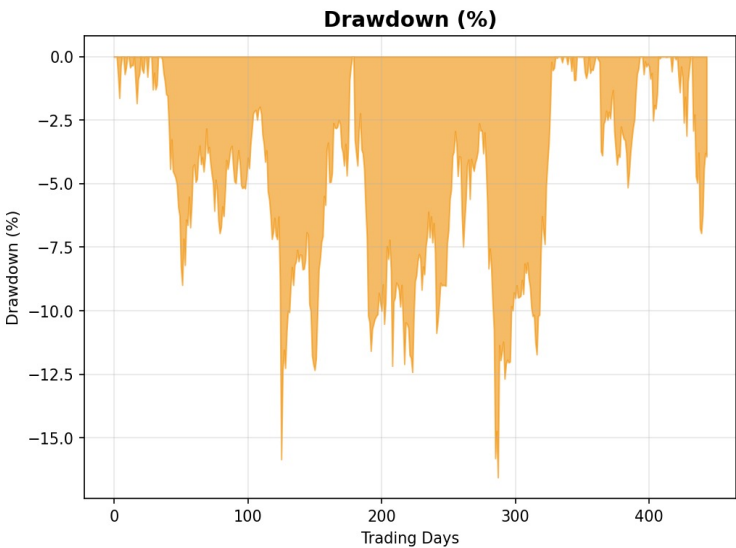
144,536,098

↑ 성장 달성

💰 성과 분석

- 초기 자본 1억원에서 시작하여, 약 23개월 간의 운용 결과 최종 1.44억원에 도달함.
- 전반적인 우상향 추세를 보임
- 시장 변동성 구간에서도 효과적인 자산 배분을 통해 자본을 방어하고 증식함

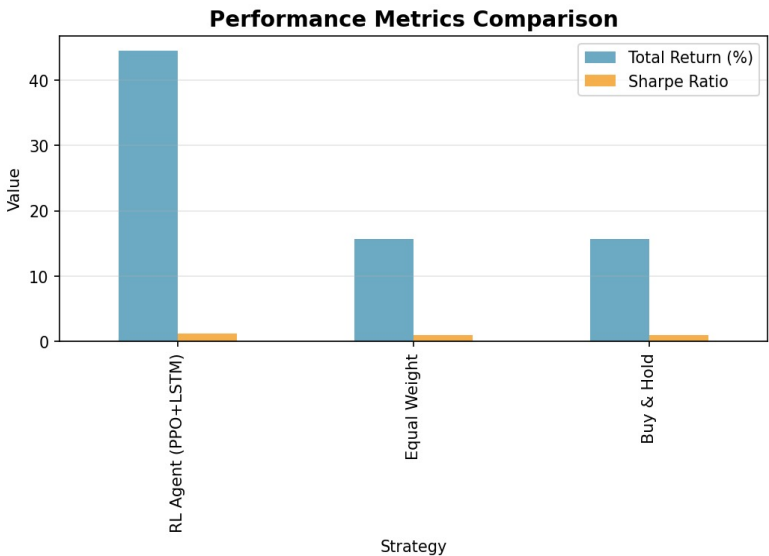
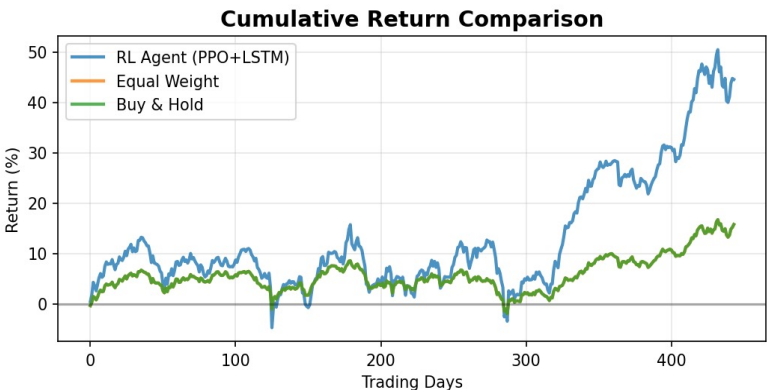
리스크 관리 및 벤치마크 비교



Performance Metrics

Total Return: 44.54%
Annualized Return: 23.25%
Sharpe Ratio: 1.17
Max Drawdown: -16.56%
Volatility: 19.23%

Final Capital: 144,536,098 KRW
Initial Capital: 100,000,000 KRW



MAX DRAWDOWN

-16.56%

최대 손실폭을 제한하며 안정적인 자산 운용을 유지

SHARPE RATIO

1.17

위험 대비 수익률이 우수함 (기준치 1.0 상회)

VOLATILITY (RISK)

19.23%

연간 변동성 수준

총 수익률 비교 (Total Return)

- RL Agent +44.54%
- Equal Weight ~15.00%
- Buy & Hold ~15.00%

Generated from RL Portfolio Project Results

강화학습 모델이 벤치마크 대비 약 3배 높은 수익률 달성

포트폴리오 비중 변화

Sector Allocation Over Time

13개 자산(12 ETF + 현금)의 일일 비중 변화



● 현금 ● IT/반도체 ● 2차전지 ● 금융 ● 헬스케어 ● 에너지/소재 ● 방산/산업재

🔄 섹터 로테이션 (Sector Rotation)

Dynamic

시장 주도주 변화에 따른
능동적인 비중 조절 패턴 관찰

🛡️ 현금의 역할 (Cash Role)

Risk Buffer

하락장 및 변동성 확대 구간에서
현금 비중 확대로 방어적 운용

🔍 패턴 분석

RL Agent는 단순히 특정 섹터를 보유하는 것이 아니라, 시장 상황 변화에 따라 주도 섹터로 비중을 옮기는 로테이션 전략을 구사했습니다. 특히 급격한 하락 구간에서는 현금 비중을 늘려 자산 가치를 방어하고 MDD를 관리하는 리스크 회피 동작이 확인되었습니다.

결과 분석

🏆 PROFITABILITY

108.03%

Baseline (27.05%) 대비 약 4배 높은 수익률 달성

전략 성과 요약

🤖 RL Agent

108.03%
Total Return

⚖️ Equal Weight

27.05%
Baseline

📈 Annualized

51.55%
CAGR

📊 Sharpe Ratio

Target Miss

1.17

Goal: 1.50

높은 수익률에도 불구하고 변동성이 동반되어 리스크 조정 수익률은 목표에 미치지 못함.

🛡️ Max Drawdown

Slight Miss

-16.56%

Limit: -15.00%

목표치(-15%)를 소폭 초과(-1.56%p). 하락장에서 현금 비중 확대가 다소 지연되는 경향 관찰.

💡 Key Insight

"수익성은 입증되었으나, 안정성 강화 필요" 단순 보유 전략 대비 압도적인 초과 수익을 달성하며 섹터 로테이션의 유효성을 입증했습니다. 다만, 목표한 Sharpe Ratio 달성을 위해서는 Reward Function에서 MDD 패널티 가중치를 높이는 추가 튜닝이 필요합니다.

토의 및 개선 방안



Reward Function 개선

- MDD 패널티 강화: 현재 단순 차감 방식에서 MDD 크기에 비례하는 지수적 패널티 도입 (예: $e^{k \times \text{MDD}}$)
- 변동성 제어: 포트폴리오 변동성(Standard Deviation)을 Reward 식에 직접 포함하여 Sharpe Ratio 최적화 유도

Goal: Sharpe Ratio > 1.5



제약 조건 (Constraints) 추가

- 최소 현금 보유: 하락장 방어력 확보를 위해 포트폴리오 내 현금 비중 하한선(Min 10%) 강제 설정
- 섹터 편중 방지: 단일 섹터 최대 비중 제한(Max 30%)을 통해 과도한 집중 투자 리스크 관리

Goal: MDD < -15%



학습 전략 고도화

- Hyperparameter Tuning: Optuna 등을 활용하여 Learning Rate, Clip Range, GAE Lambda 등 핵심 파라미터 최적화
- 데이터 확장: Bull/Bear 마켓 시나리오 데이터를 증강(Augmentation)하여 다양한 시장 국면 학습

Goal: Generalization



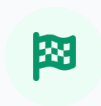
모델 아키텍처 확장

- Transformer 적용: LSTM의 Long-term dependency 한계를 극복하기 위해 Self-Attention 메커니즘 도입
- 앙상블 (Ensemble): 서로 다른 시드(Seed)나 구조를 가진 여러 에이전트의 결정을 조합하여 예측 안정성 향상

Goal: Robustness



결론 및 Future Work



프로젝트 성과

멀티모달 PPO+LSTM 모델을 통해 국내 섹터 ETF 시장에서 108.03%의 높은 수익률 달성

Baseline(균등 분배) 대비 4배 이상의 초과 수익 창출로 알파 창출 능력 입증

시장 하락기 현금 비중 확대를 통한 리스크 관리 동작 확인



한계점

Sharpe Ratio(1.17)가 목표(1.5)에 미치지 못하여 리스크 대비 수익성 개선 필요

MDD(-16.56%)가 목표(-15%)를 소폭 초과, 극단적 시장 상황에서의 방어력 보완 필요

23개월의 제한된 데이터로 인한 장기 사이클 검증 부족



향후 연구 방향

Transformer 기반 모델 도입으로 장기 시계열 패턴 학습 강화

Reward Function 고도화 (MDD 패널티 강화 및 변동성 제어)

제약 조건 최적화 및 실거래 환경 연동 테스트 진행