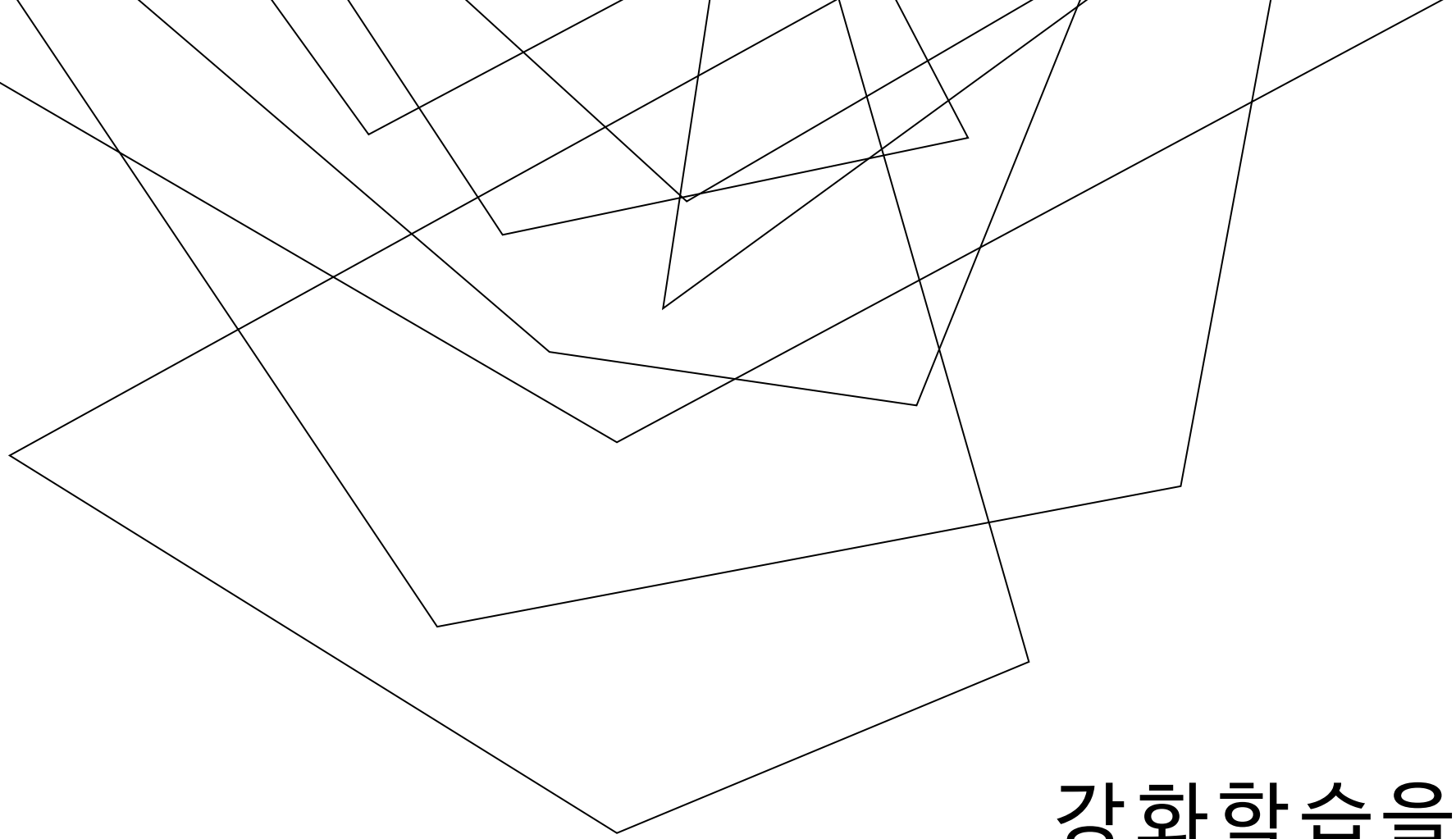
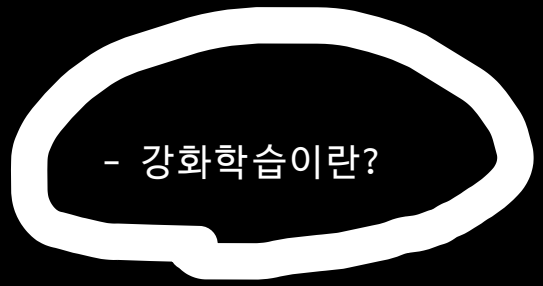


강화학습을 이용한 단타매매 봇

산업경영공학과 21학번 khuda 3기 임동휘

목차

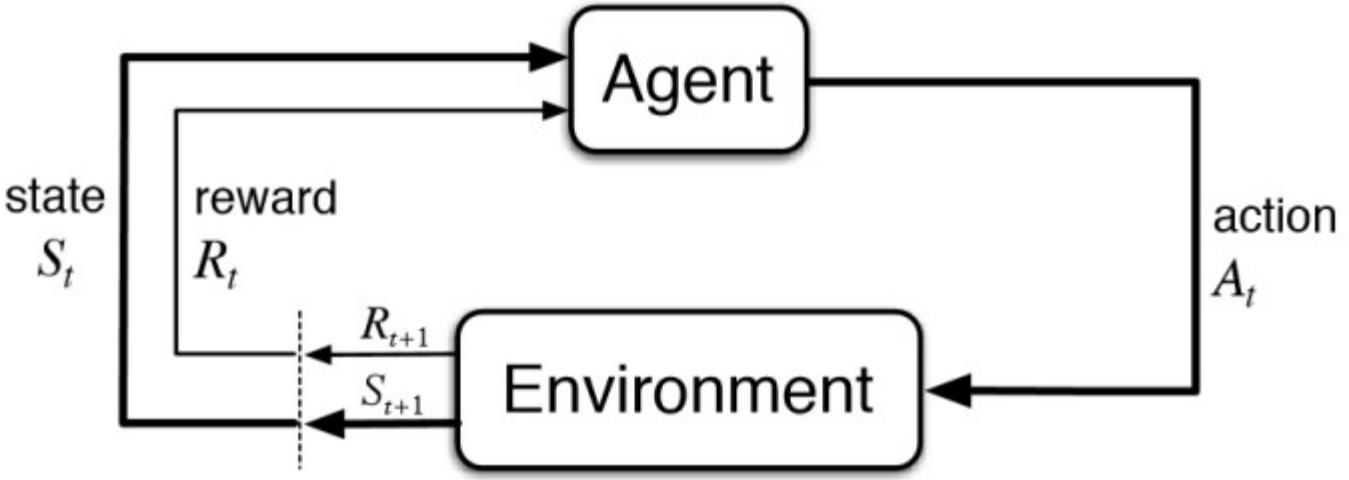


- 강화학습이란?

- 프로젝트 설명

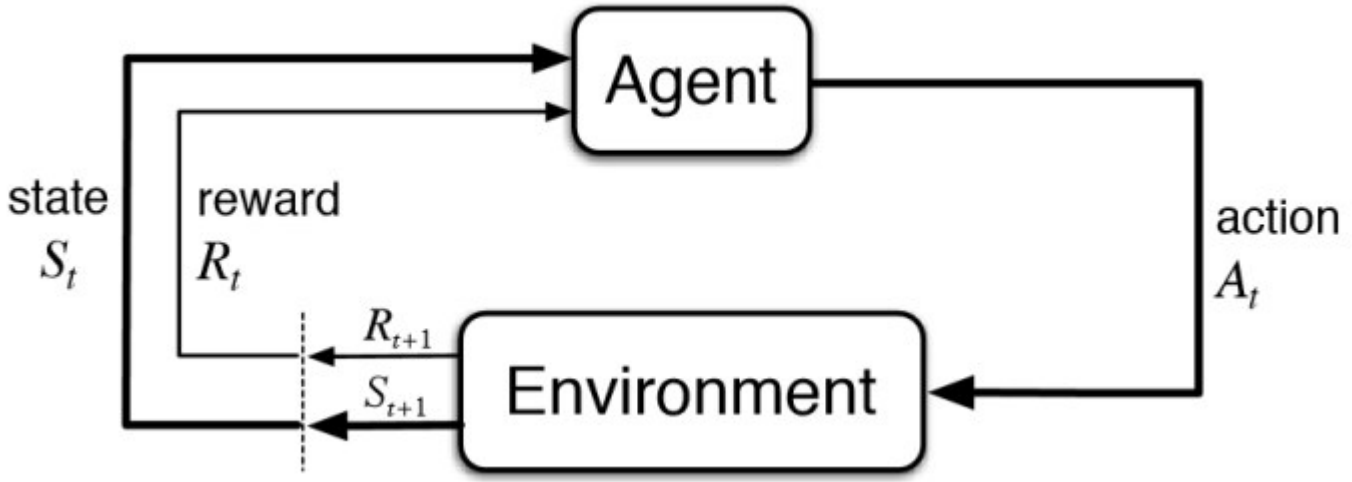
강화학습이란?

주변 환경과 Agent가 상호작용하면서
Agent가 학습하는 학습 방법



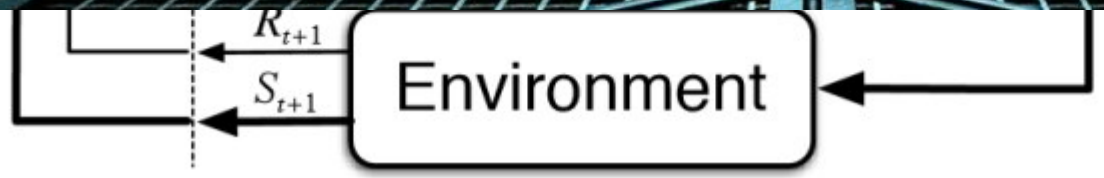
환경(ENV)이란?

Agent(모델)이 학습하는 환경
모델은 env속에 들어가서 실시간으로 학습하게 됨
env는 Agent가 살아가는 시뮬레이션이라고 할 수 있음



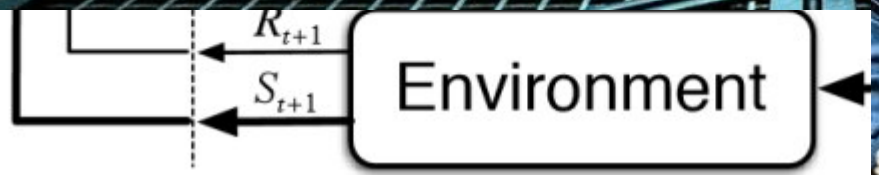
stat

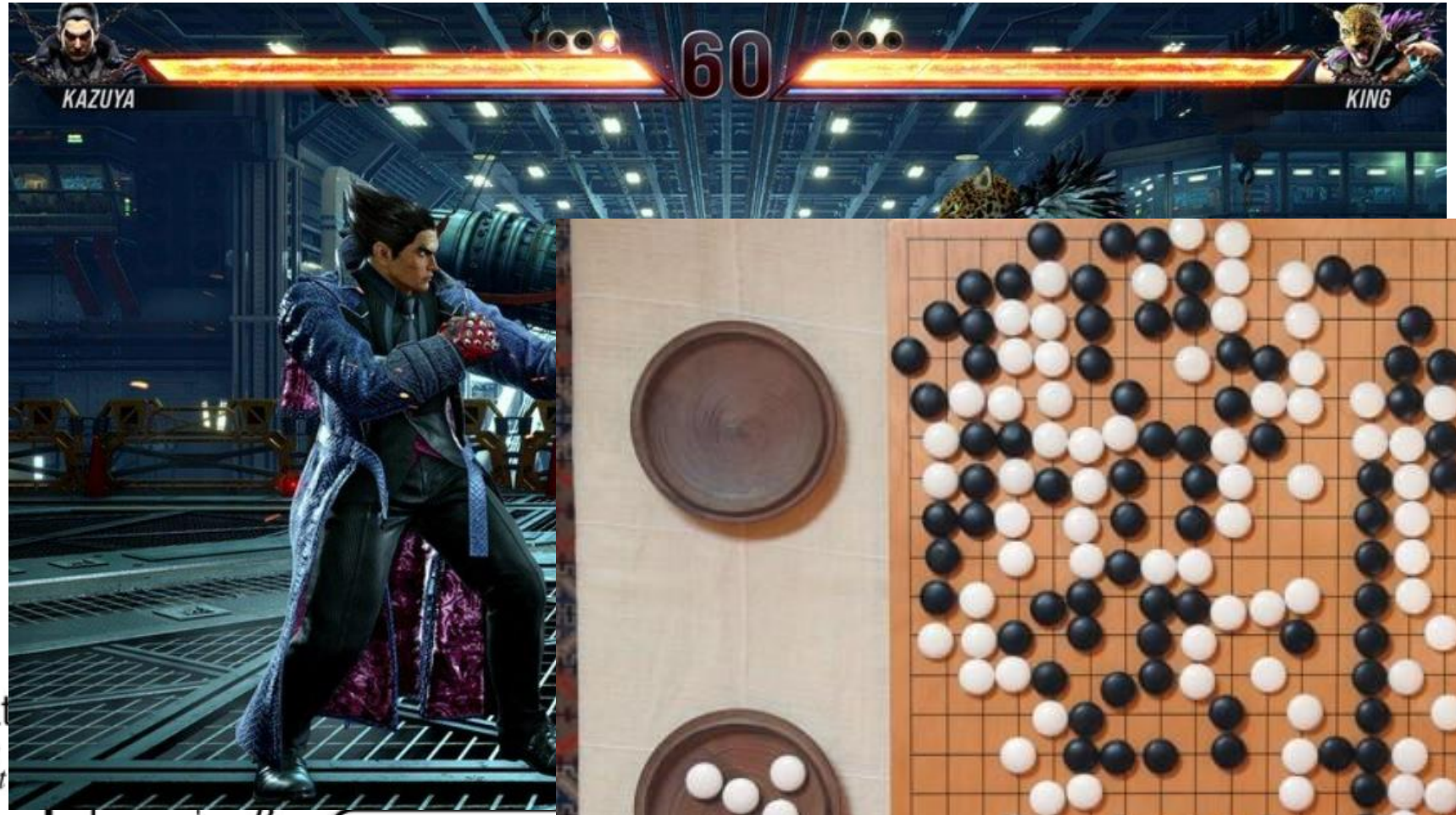
S_t



stat

S_t



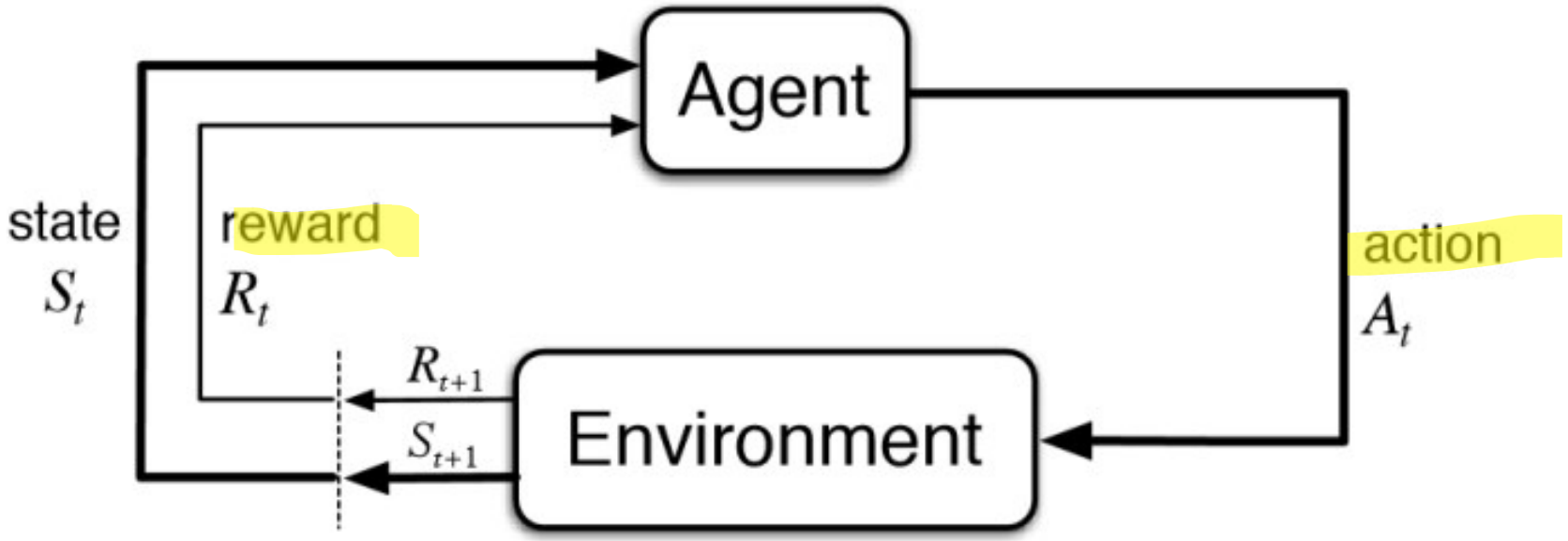


stat
S_t

R_{t+1}
S_{t+1}
Environ

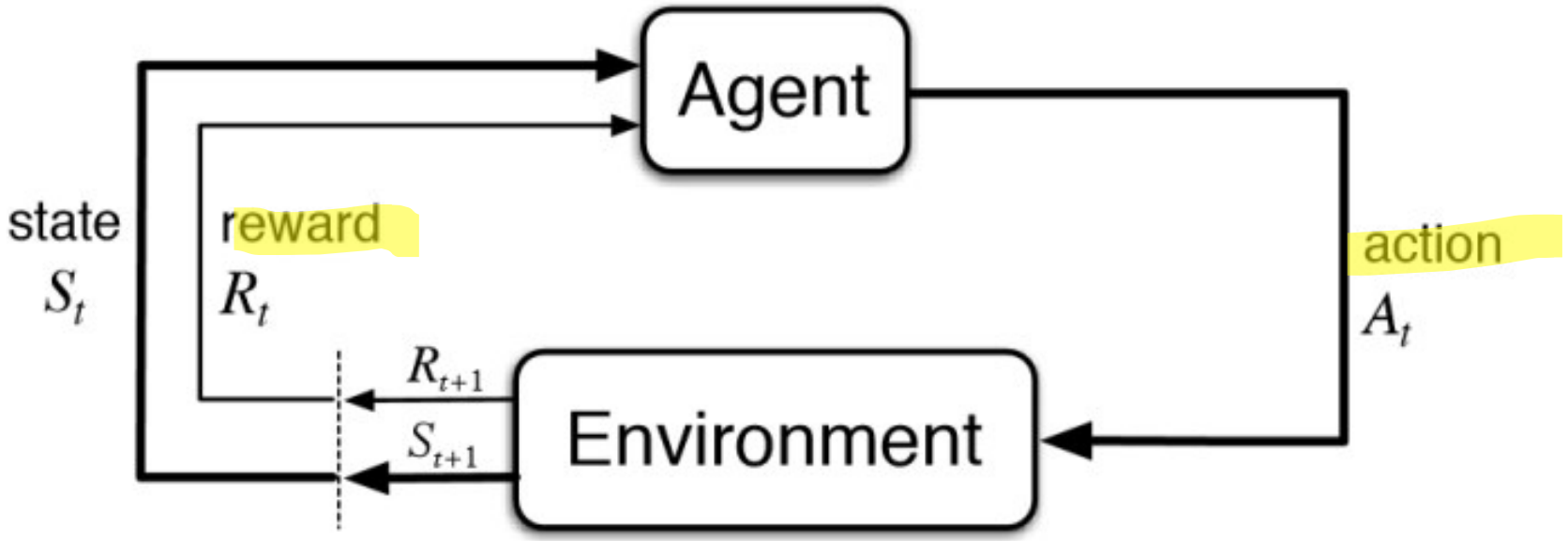
그래서 어떻게 학습하는데?

Action - Reward – State(상태) 순서를 반복
Reward를 받으면서, Reward를 Maximize하는
방식으로 action을 수행



그래서 어떻게 학습하는데?

Action - Reward – State(상태) 순서를 반복
Reward를 받으면서, Reward를 Maximize하는
방식으로 action을 수행



environment

agent

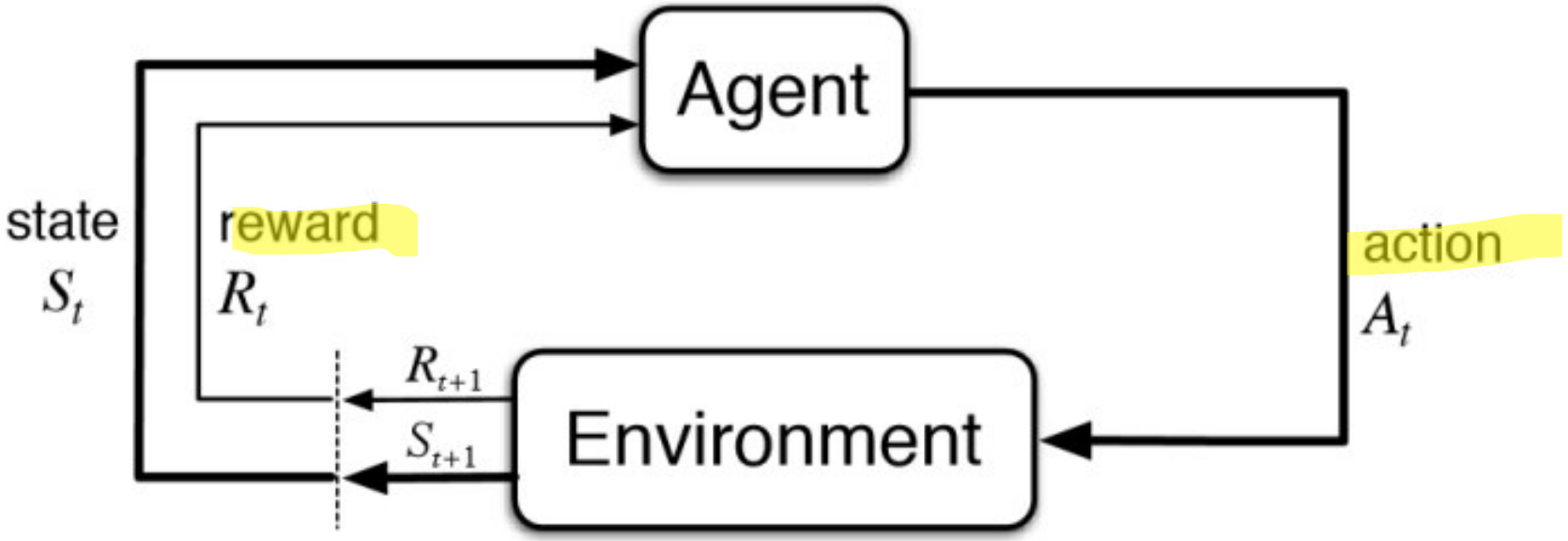
actions

rewards

observations

REWARD FUNCTION ~= LOSS FUNCTION

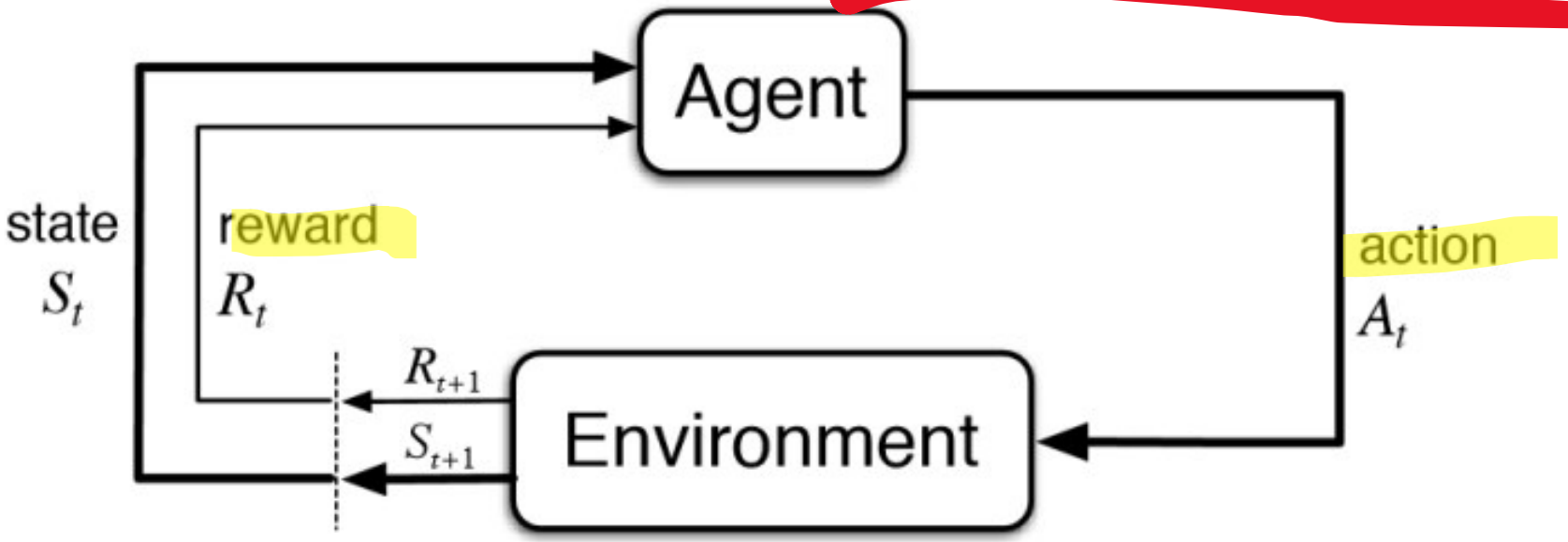
Action - Reward – State(상태) 순서를 반복
Reward를 받으면서, Reward를 Maximize하는
방식으로 action을 수행



REWARD FUNCTION ~= LOSS FUNCTION

Action - Reward – State(상태) 순서를 반복
Reward를 받으면서, Reward를 Maximize하는
방식으로 action을 수행

Loss를 Minimize



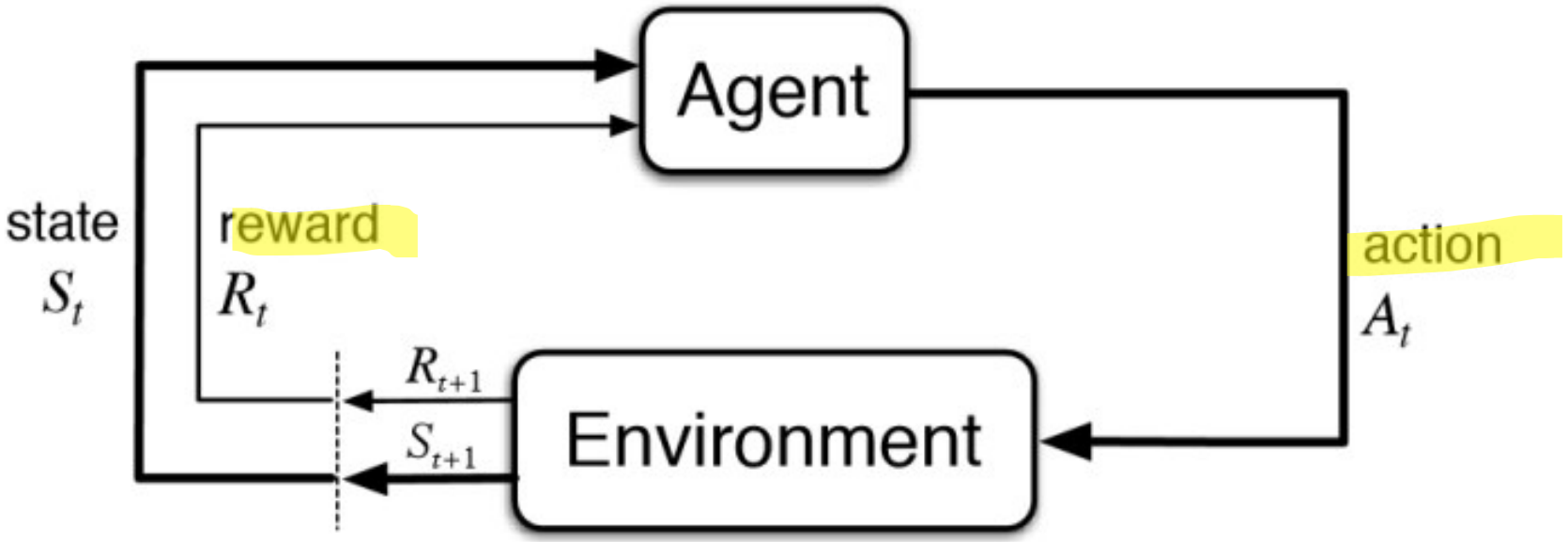
개발자가 의도에 맞게 설정!

Action - Reward – State(상태) 순서를 반복

Reward를 받으면서, Reward를 Maximize하는

방식으로 action을 수행

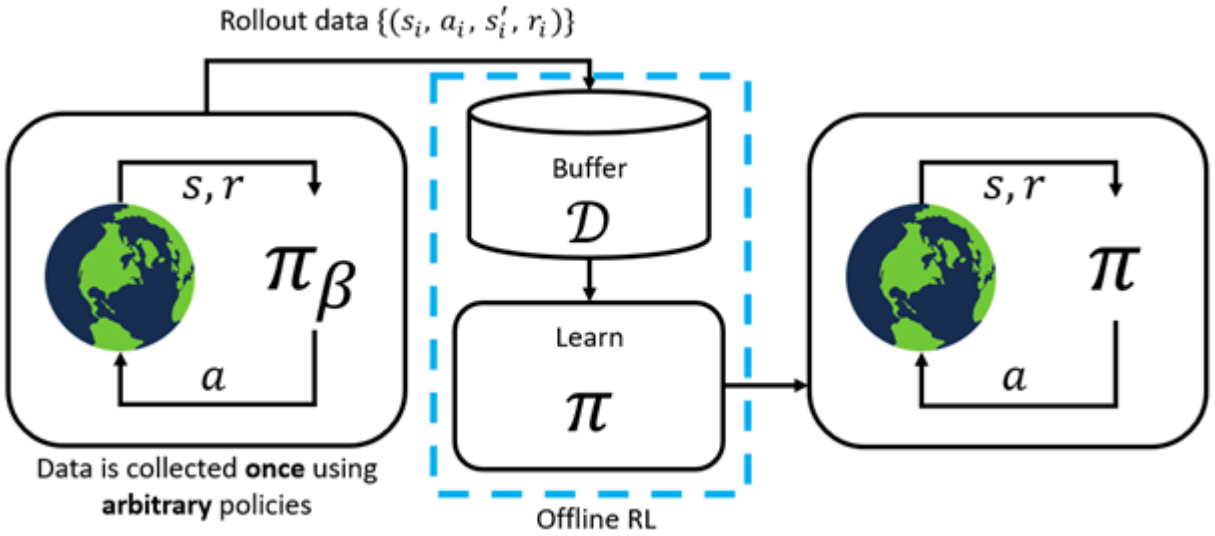
Loss를 Minimize



Agent가 환경(Env)과 상호작용
(action -> reward -> state)하며
실시간(online)으로 학습!

강화학습 분야 – OFFLINE 강화학습

Agent가 환경(Env)과 상호작용 (action -> reward -> state)했던 데이터로 비실시간(offline) 학습



강화학습 분야 – 비지도 강화학습

Reward function(loss function) 없이
스스로 학습하는 방법

강화학습 분야 – MULTI AGENT 강화학습

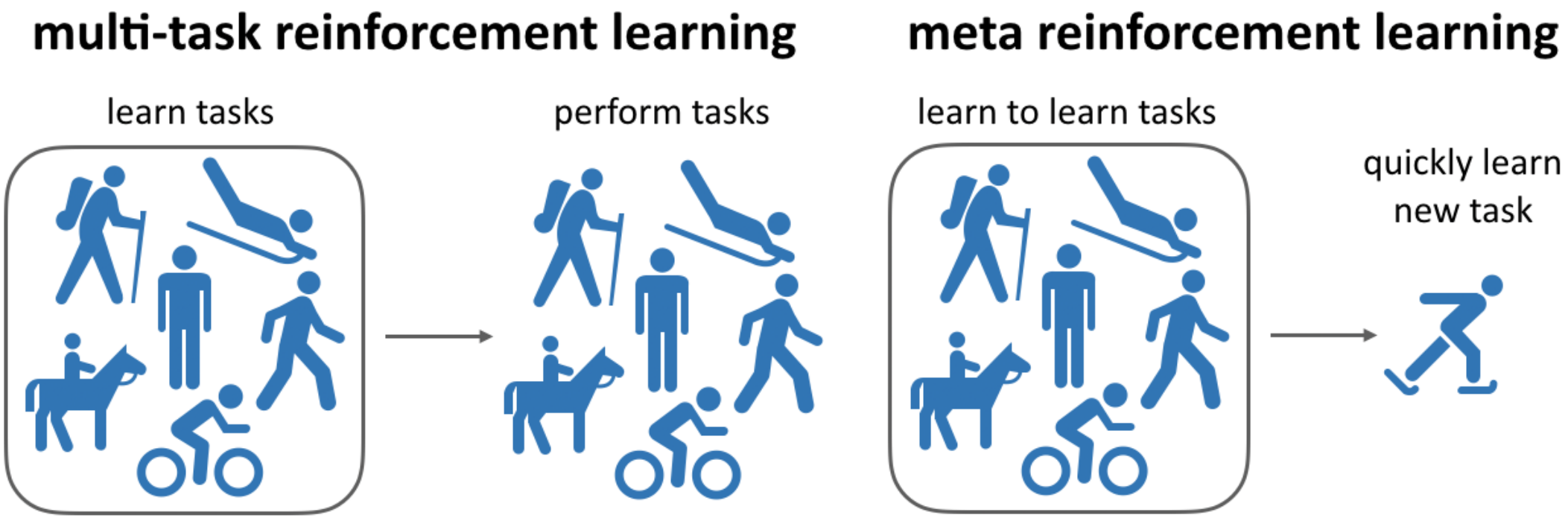
https://youtu.be/kopoLzvh5jY?si=HEfLnq2pY5BJ4gER

여러명의 Agent를 학습.

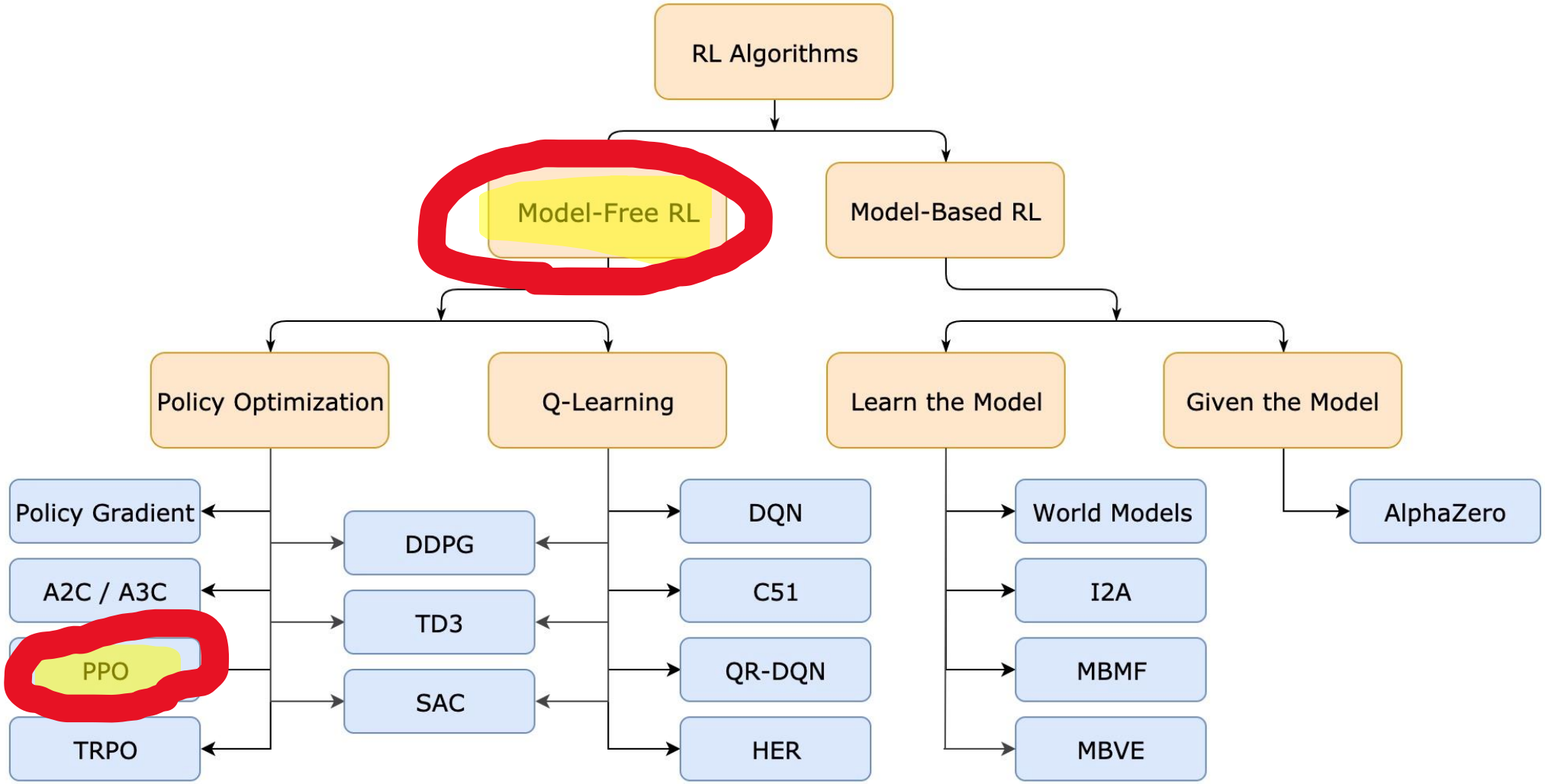
협력적 / 적대적 학습

강화학습 분야 – META 강화학습

여러가지 일을 잘하는 강화학습 모델



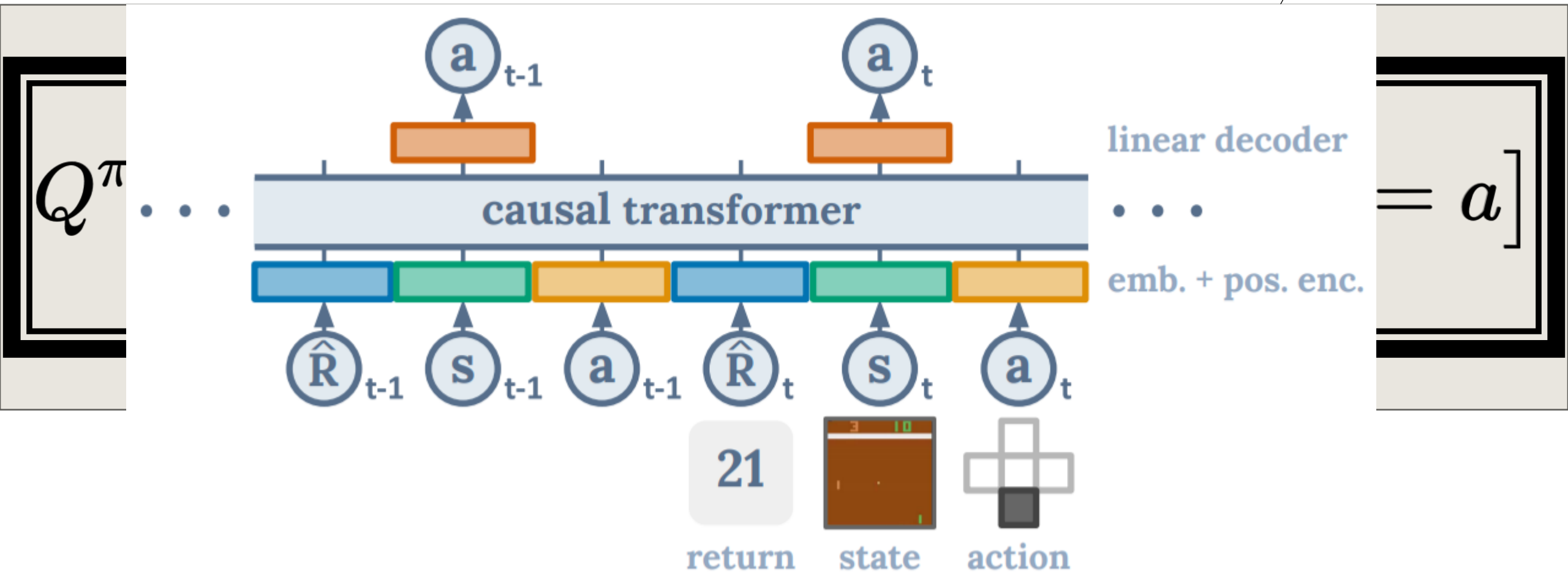
강화학습 알고리즘


RL Algorithms
Model-Free RL
Model-Based RL
Policy Optimization
Q-Learning
Learn the Model
Given the Model
Policy Gradient
A2C / A3C
PPO
TRPO
DDPG
TD3
SAC
DQN
C51
QR-DQN
HER
World Models
I2A
MBMF
MBVE
AlphaZero

강화학습 알고리즘 – Q FUNCTION

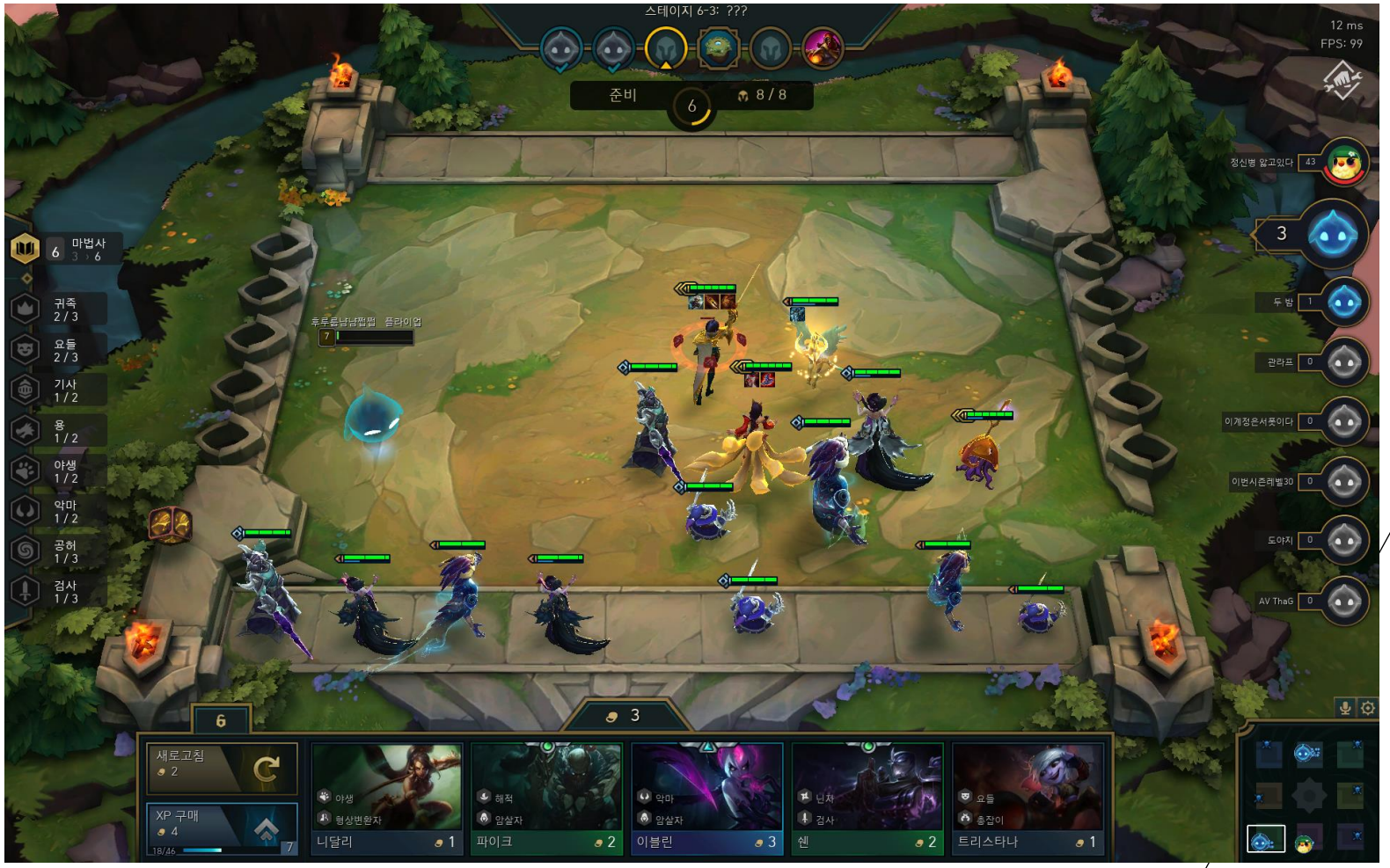
$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \middle| s_t = s, a_t = a\right]$$

강화학습 알고리즘 – DECISION TRANSFORMER



강화학습 그래서 어디 쓰이는데? – 게임

https://youtu.be/_84yVfk2NpA?si=FYtkZOX8utPOU3Kn



게임 봇을 통해
사용자에게 긍정적인
게임 경험 제공

게임 봇 데이터 통한
밸런스 패치

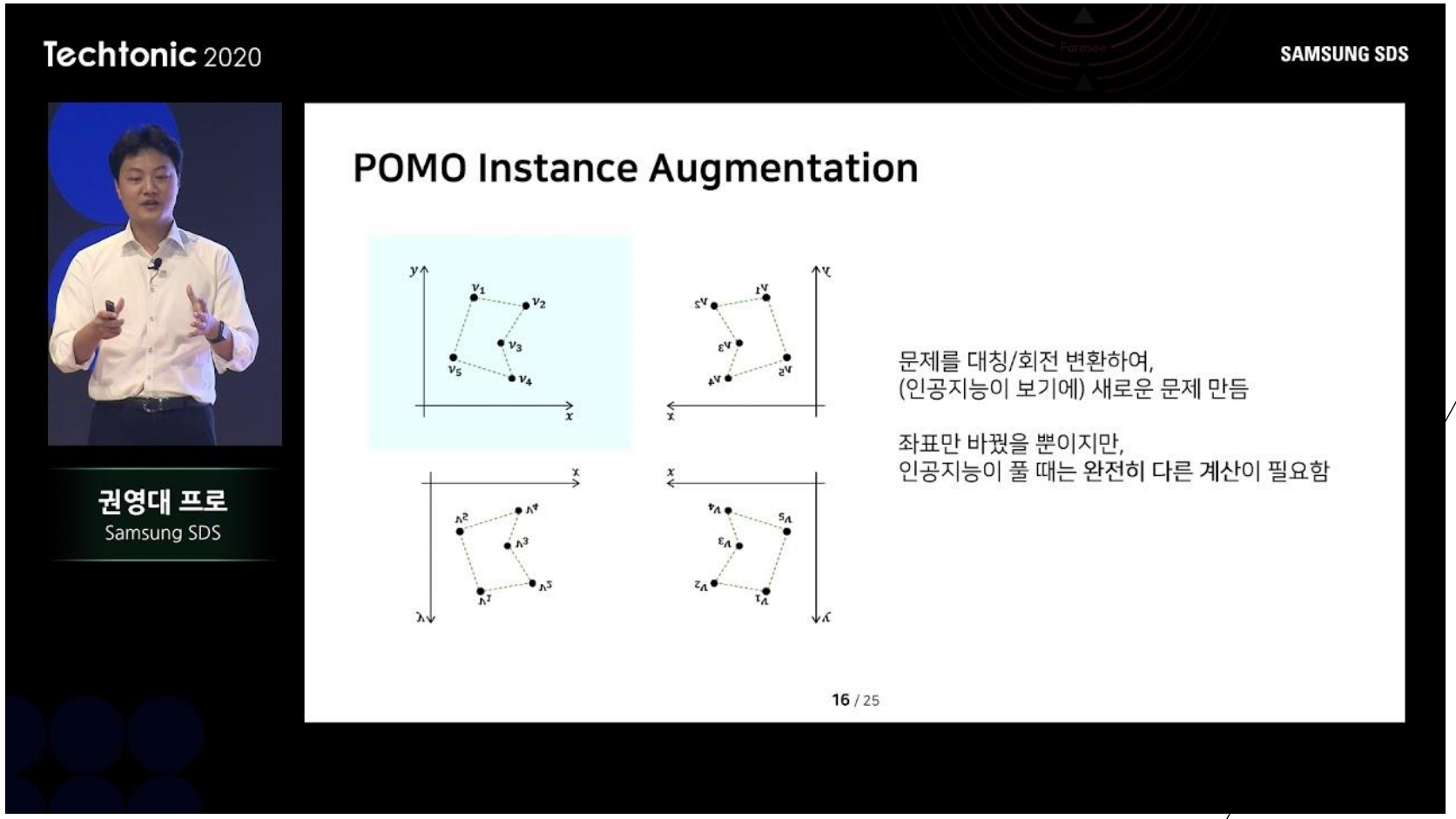
강화학습 그래서 어디 쓰이는데? - 로보틱스

로봇 분야에 적극 활용

강화학습을 통한 로봇 제어

강화학습 그래서 어디 쓰이는데? - 조합최적화

https://youtu.be/Dyp9lQpVgCs?si=RChMHulTD8l6PNUw

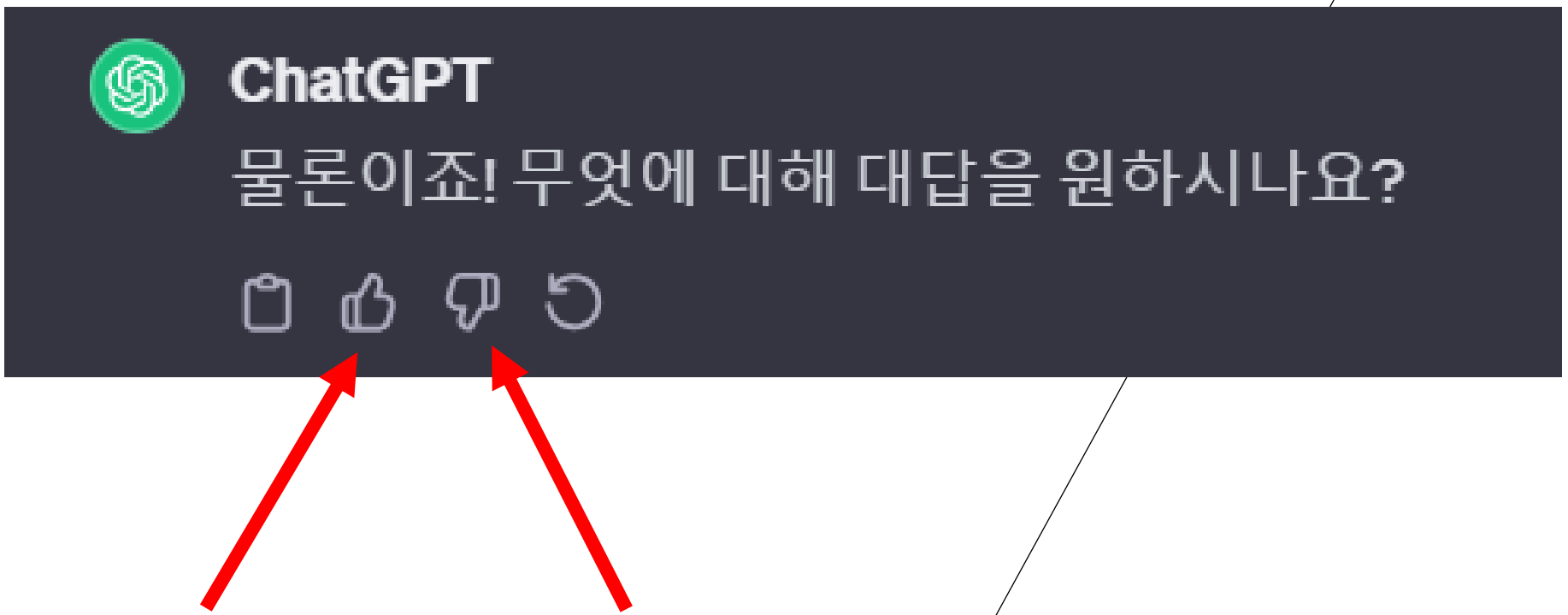


여러 조합 최적화 분야에서 사용 가능

TSP문제, 최적관리 문제, 길찾기 문제, 자원할당 문제

강화학습 그래서 어디 쓰이는데? – CHATGPT 보상(RLHF)

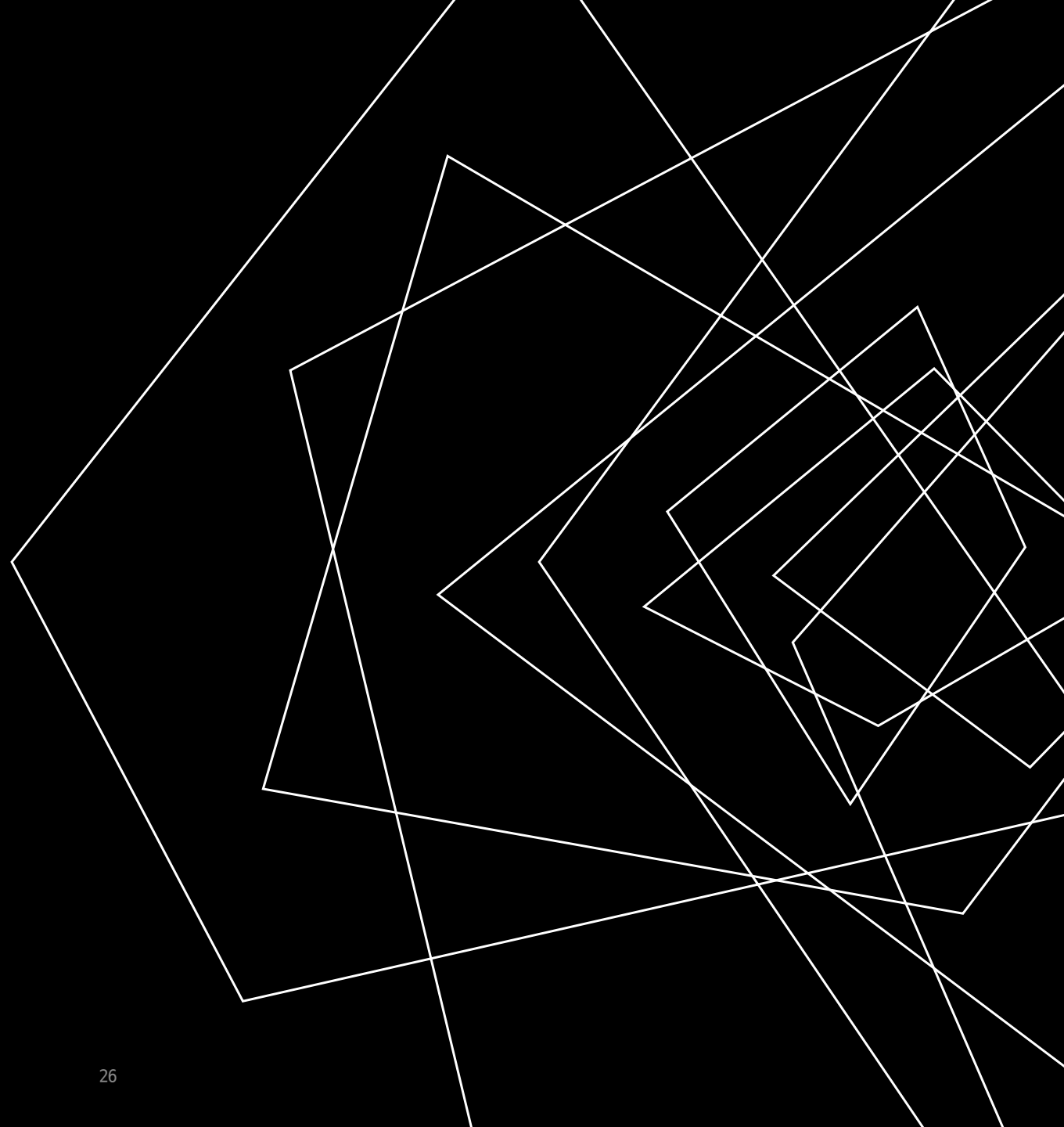
Reinforcement learning from human feedback



목차

- 강화학습이란?

- 프로젝트 설명



알고리즘 트레이딩이란?

컴퓨터와 알고리즘을 이용해서
트레이딩 하는 방법

왜 단타로 했어요?

여러 경제/정치적 변동 리스크에 대해서
공부하고 싶지 않습니다 ㅠㅠ 너무 어려워요

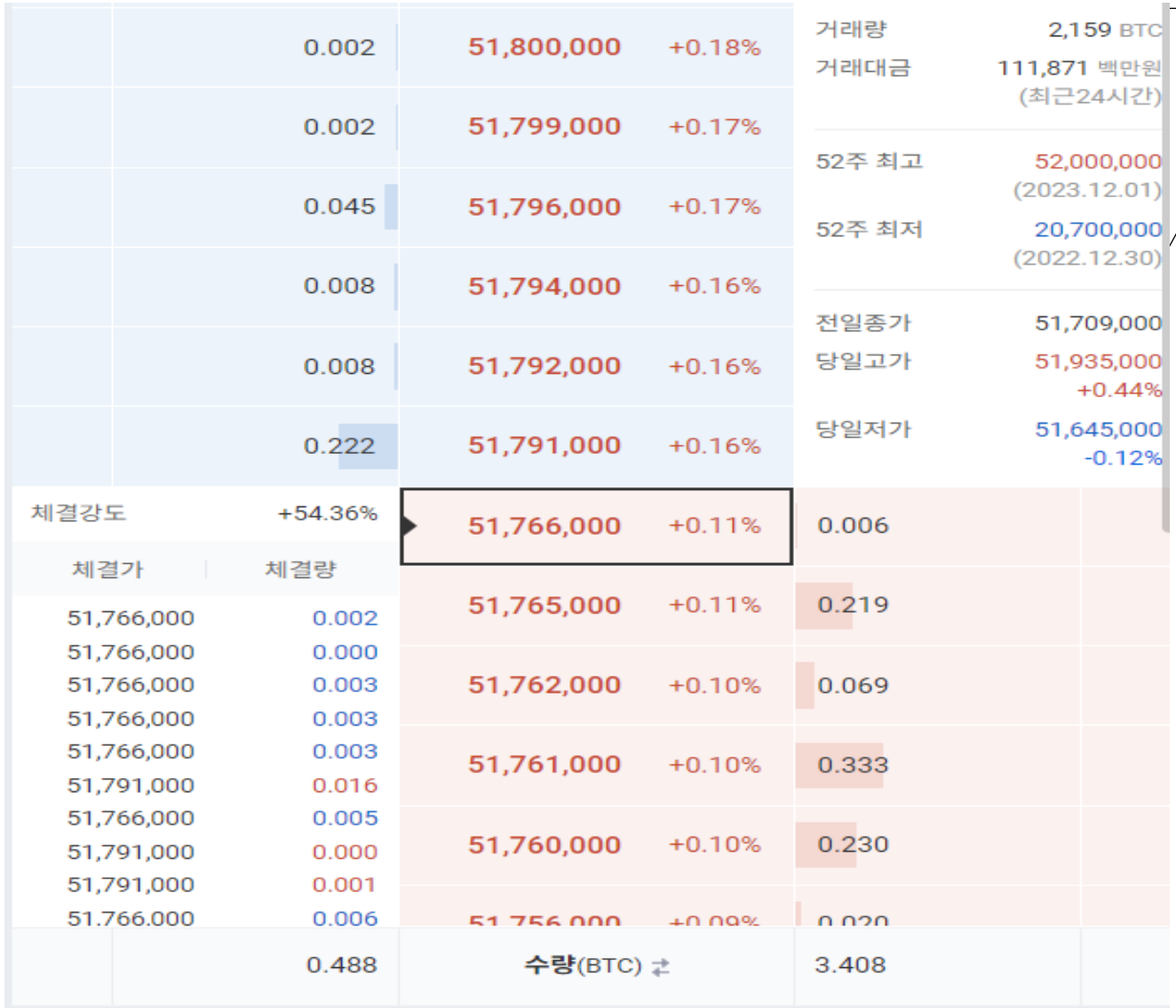
단타와 호가창 – 프로젝트 아이디어

단타와 호가창 – 체결이란?

체결 | 일별

체결시간	체결가격(KRW)	체결량(BTC)	체결금액(KRW)
12.03 01:19	51,766,000	0.00044193	22,877
12.03 01:19	51,766,000	0.00187876	97,256
12.03 01:19	51,771,000	0.00051663	26,746
12.03 01:19	51,766,000	0.00215385	111,496
12.03 01:19	51,766,000	0.02272028	1,176,138
12.03 01:19	51,766,000	0.01016500	526,201
12.03 01:19	51,768,000	0.00000861	446
12.03 01:19	51,768,000	0.03967565	2,053,929
12.03 01:19	51,768,000	0.00239483	123,976
12.03 01:19	51,768,000	0.00018654	9,657
12.03 01:19	51,770,000	0.01000000	517,700

단타와 호가창 – 호가창이란?



	0.002	51,800,000 +0.18%	
	0.002	51,799,000 +0.17%	
	0.045	51,796,000 +0.17%	
	0.008	51,794,000 +0.16%	
	0.008	51,792,000 +0.16%	
	0.222	51,791,000 +0.16%	
체결강도	+54.36%	51,766,000 +0.11%	0.006
51,766,000	0.002	51,765,000 +0.11%	0.219
51,766,000	0.000		
51,766,000	0.003	51,762,000 +0.10%	0.069
51,766,000	0.003		
51,766,000	0.003	51,761,000 +0.10%	0.333
51,791,000	0.016		
51,766,000	0.005		
51,791,000	0.000	51,760,000 +0.10%	0.230
51,791,000	0.001		
51,766,000	0.006	51,756,000 +0.09%	0.020
	0.488	수량(BTC)	3.408

거래량	2,159 BTC
거래대금	111,871 백만원 (최근24시간)
52주 최고	52,000,000 (2023.12.01)
52주 최저	20,700,000 (2022.12.30)
전일종가	51,709,000
당일고가	51,935,000 +0.44%
당일저가	51,645,000 -0.12%

단타와 호가창 – 차트의 모양?

단타와 호가창 – 프로젝트 아이디어

rule-based model을 data-based model로
이길 수 있지 않을까?

심지어, data-based model은
rule-based모델이 형성한 데이터로 학습

Data-based model: Nosie에 취약하지만 robust하게 학습된다면
훨씬 유연하게 작동할 수 있음.

단타와 호가창 – 프로젝트 아이디어

주문 -> 호가창 -> 주식가격 -> 분봉

1. 나름의 예측 가능
2. 인간의 심리가 큰 영향
3. 이미 있는 트레이딩봇(룰 베이스)들이 가격을 형성 ->
   봇들의 가격 형성 규칙이 가격에 반영(데이터에 반영) ->
   규칙이 가격을 통해서 역으로 내가 수익을 창출

프로젝트 가정

1. 현재 상태에서 low가격과 high 가격의 평균으로 거래
2. 모든 거래가 다음 거래 이전까지 체결됨.

데이터 설명 - RAW DATA

index	open	high	low	close	volume	mid_price
2023-01-01 09:00:00	21079000.0	21082000.0	21061000.0	21081000.0	4.852014	21071500.0
2023-01-01 09:01:00	21080000.0	21080000.0	21061000.0	21061000.0	2.099329	21070500.0
2023-01-01 09:02:00	21062000.0	21062000.0	21055000.0	21062000.0	2.486127	21058500.0
2023-01-01 09:03:00	21057000.0	21078000.0	21055000.0	21064000.0	1.744549	21066500.0
2023-01-01 09:04:00	21063000.0	21077000.0	21056000.0	21064000.0	1.433399	21066500.0

데이터 설명 - 가공

index	60min_ago_volume	60min_ago_change	59min_ago_volume	59min_ago_change	58min_ago_volume	58min_ago_change
2023-01-01 09:59:00	4.852014	0.000000	2.099329	-0.000047	2.486127	-0.000570
2023-01-01 10:00:00	2.099329	-0.000047	2.486127	-0.000570	1.744549	0.000380
2023-01-01 10:01:00	2.486127	-0.000570	1.744549	0.000380	1.433399	0.000000
2023-01-01 10:02:00	1.744549	0.000380	1.433399	0.000000	1.662411	-0.000712
2023-01-01 10:03:00	1.433399	0.000000	1.662411	-0.000712	3.212477	-0.000095

ACTION 설명

간단하게 설명하면 Long / Short을 할 수 있음.

Long: 가상화폐 가격 상승을 예측하고 매수

Short: 가상화폐 가격 하락을 예측하고 매도

Leverage 비율을 정해 놓고, 공매도/공매수 가능

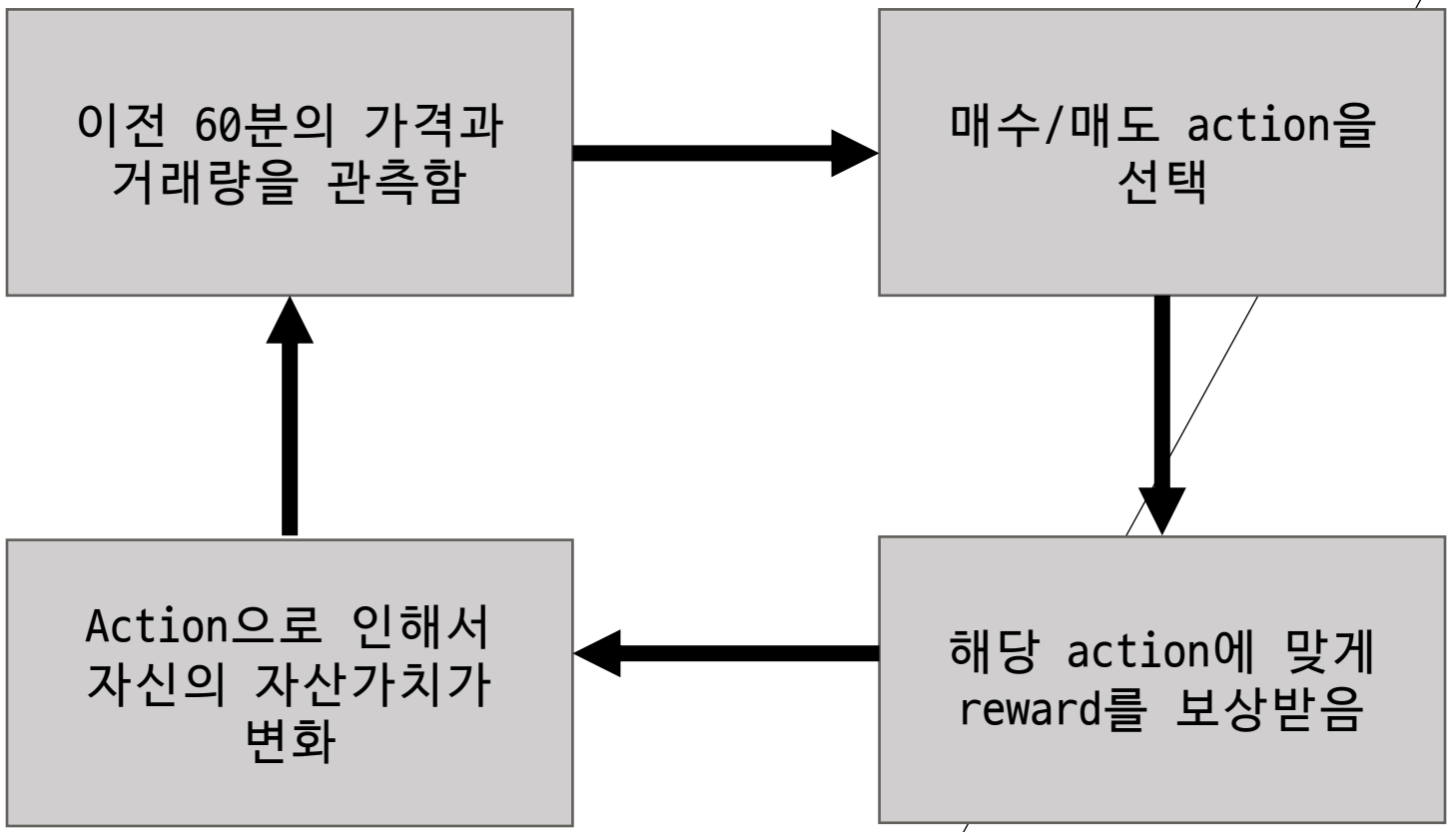
REWARD FUNCTION 설명

거래량 x (거래 이후 자신의 전체 자산의 가치 변화량)

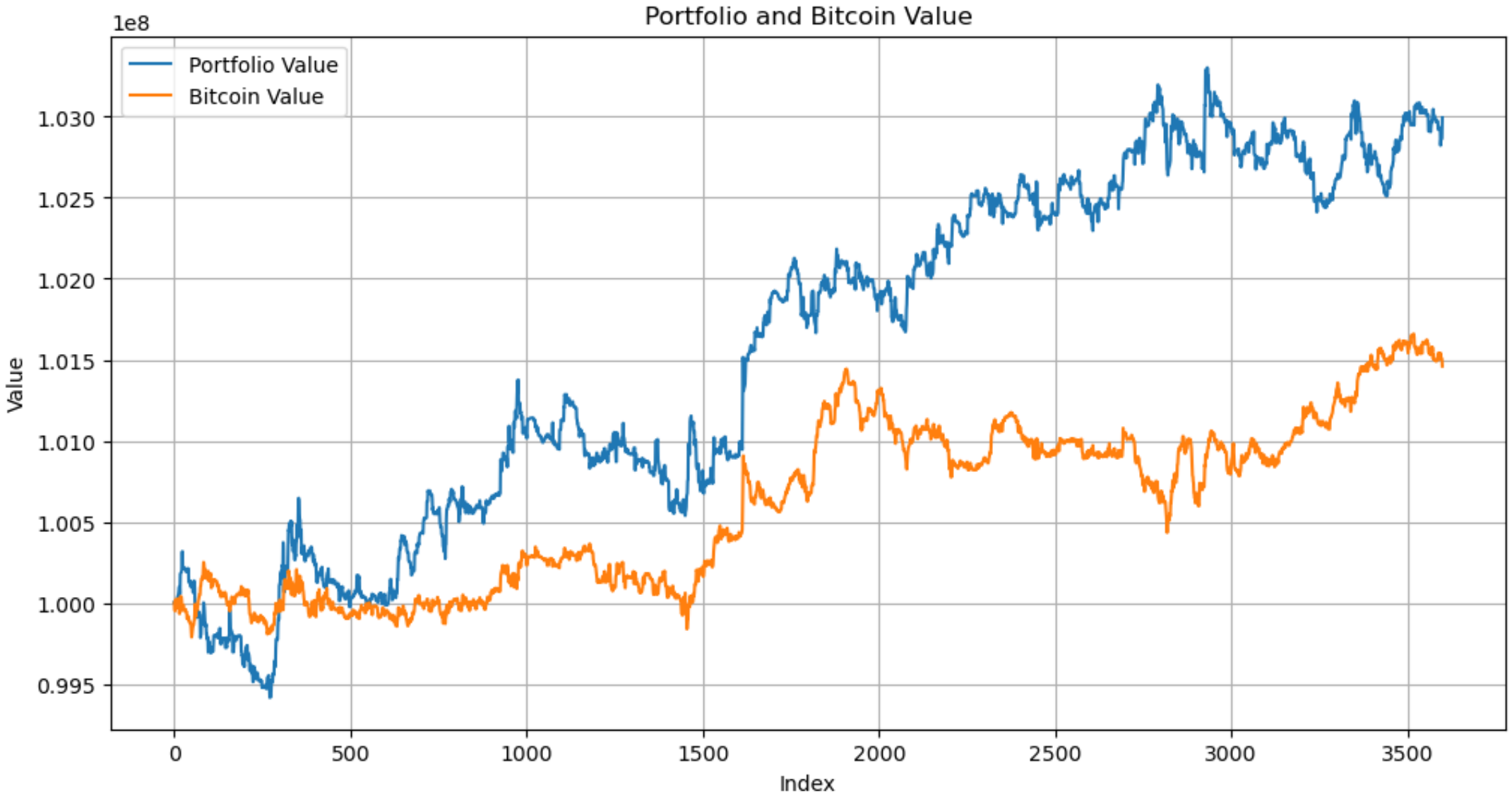
매수 action 수행 시:
비트코인 가격이 상승해야 (+)reward를 받음.

매도 action 수행 시:
비트코인 가격이 하락해야 (+)reward를 받음.

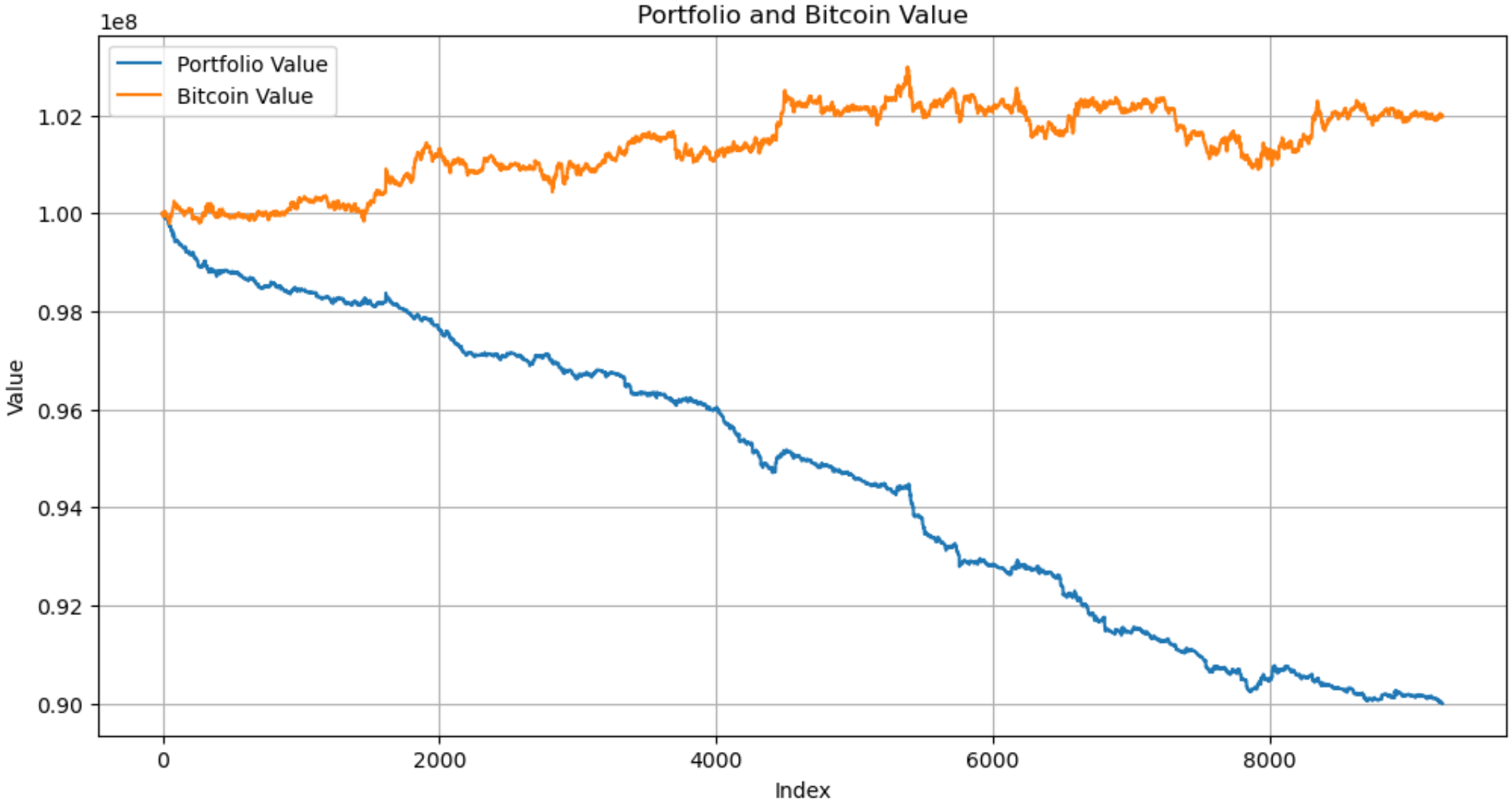
전체 프로젝트 설명


이전 60분의 가격과 거래량을 관측함
매수/매도 action을 선택
해당 action에 맞게 reward를 보상받음
Action으로 인해서 자신의 자산가치가 변화

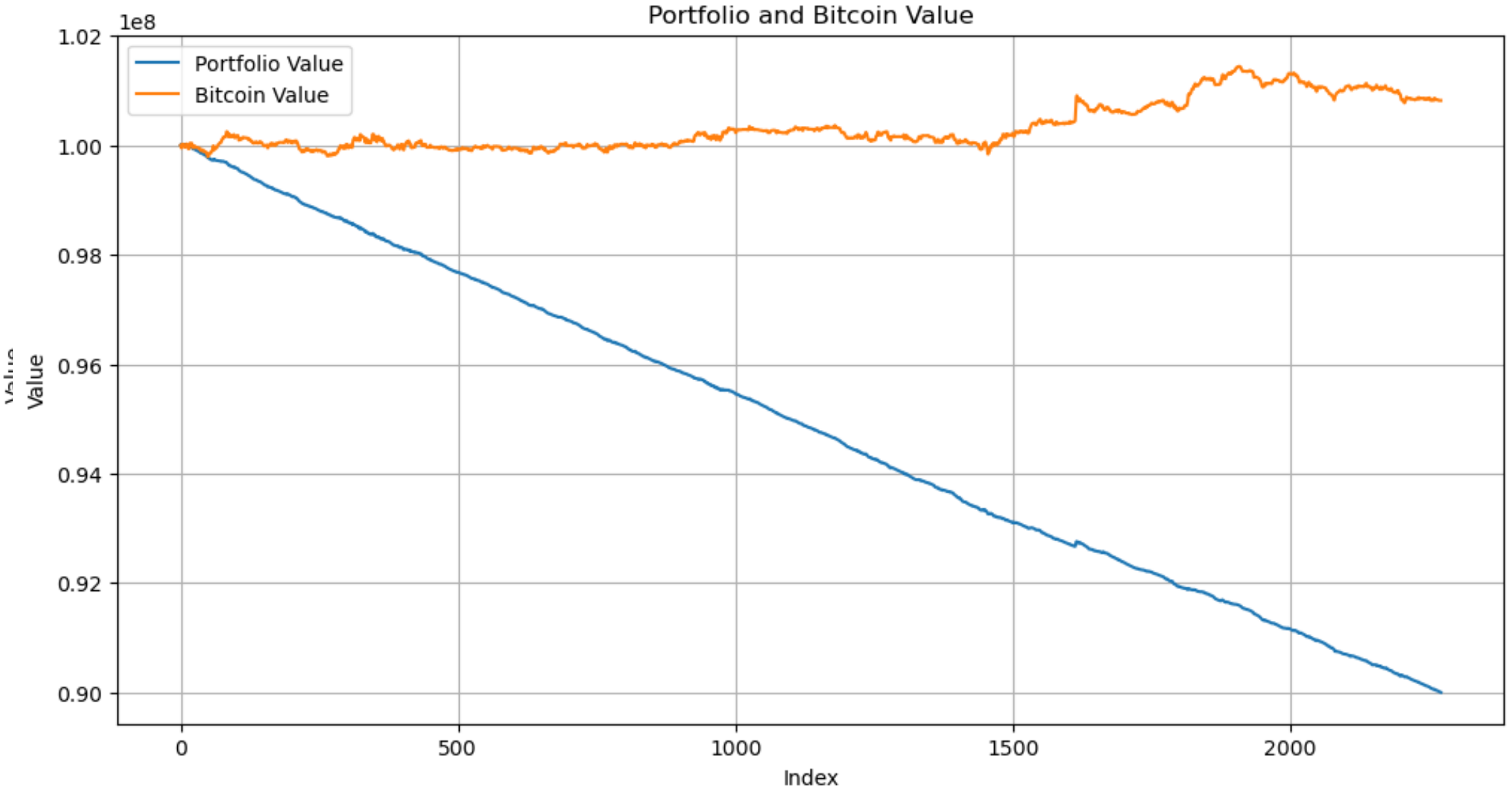
수익률 설명


Portfolio and Bitcoin Value
1e8
Portfolio Value
Bitcoin Value
Value
1.030
1.025
1.020
1.015
1.010
1.005
1.000
0.995
0
500
1000
1500
2000
2500
3000
3500
Index

수익률 설명


Portfolio and Bitcoin Value
1e8
Portfolio Value
Bitcoin Value
1.02
1.00
0.98
0.96
0.94
0.92
0.90
Value
0
2000
4000
6000
8000
Index

수익률 설명


Portfolio and Bitcoin Value
1e8
1.02
1.00
0.98
0.96
0.94
0.92
0.90
Portfolio Value
Bitcoin Value
Value
0
500
1000
1500
2000
Index

한계점

1. 단순 가격/거래량의 변화량만을 데이터로 사용함.

2. 보통 CNN이 첫층에 들어가는데 Dense Layer를 사용함

3. ENV 구성에 있어서 수수료를 과도하게 부여하는 Logic 사용

발전 방향

1. 여러가지 기술적 지표들을 사용해서 데이터로 사용

2. 첫 층을 CNN으로 변환

3. ENV의 수수료부과 로직 변경

4. 학습 데이터를 1분봉 -> 5분봉으로 변경
   (1분봉에는 많은 Noise가 존재할 수 있음)

향후 계획

1. 첫 층을 CNN층으로 바꿀 것입니다.
   (Trading Signal Agent: TSA)
2. Binance 실거래 데이터로 사용할 것입니다.
3. 체결을 담당하는 강화학습 model을 만들 것입니다.
   (Trading Execution Agent: TEA)
4. 백테스팅 프로그램을 만들 것입니다.
5. TSA-TEA 모델을 연결해서 백테스팅을 진행한 후에, API연결 후 실거래 까지 해볼 것입니다.