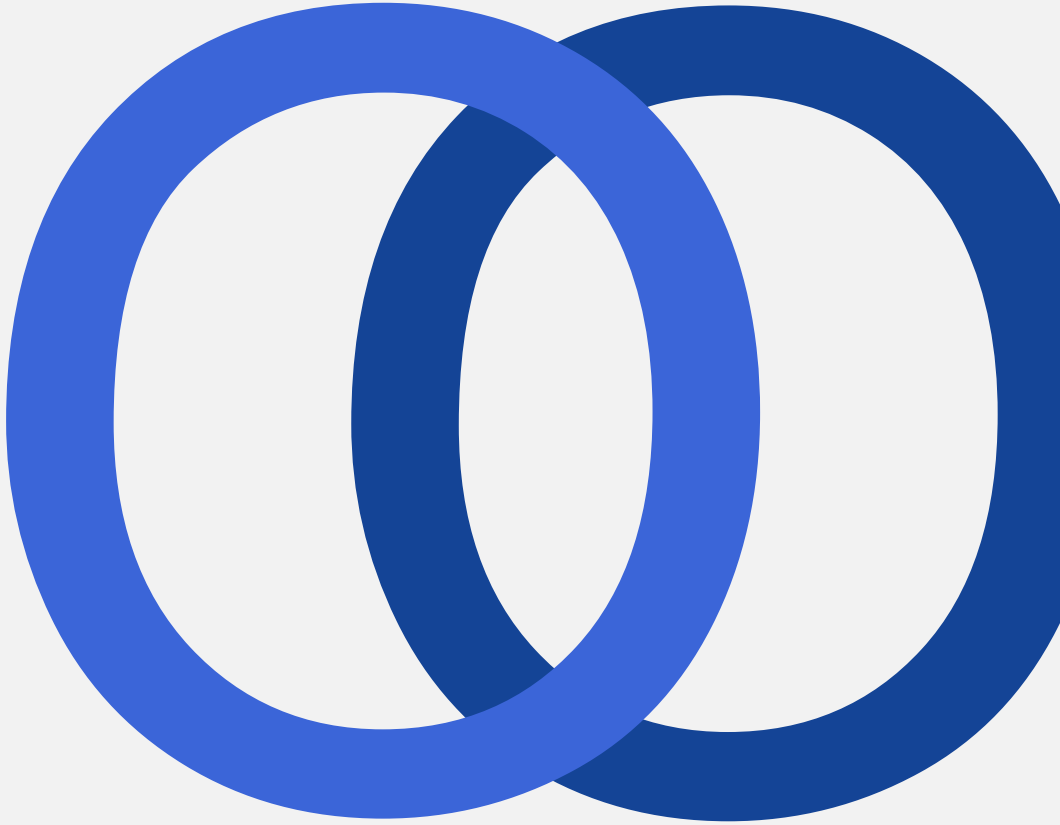
OPTIMIZED ORDER EXECUTION WITH REINFORCEMENT LEARNING

강화학습을 이용한 최적 주문 집행 전략

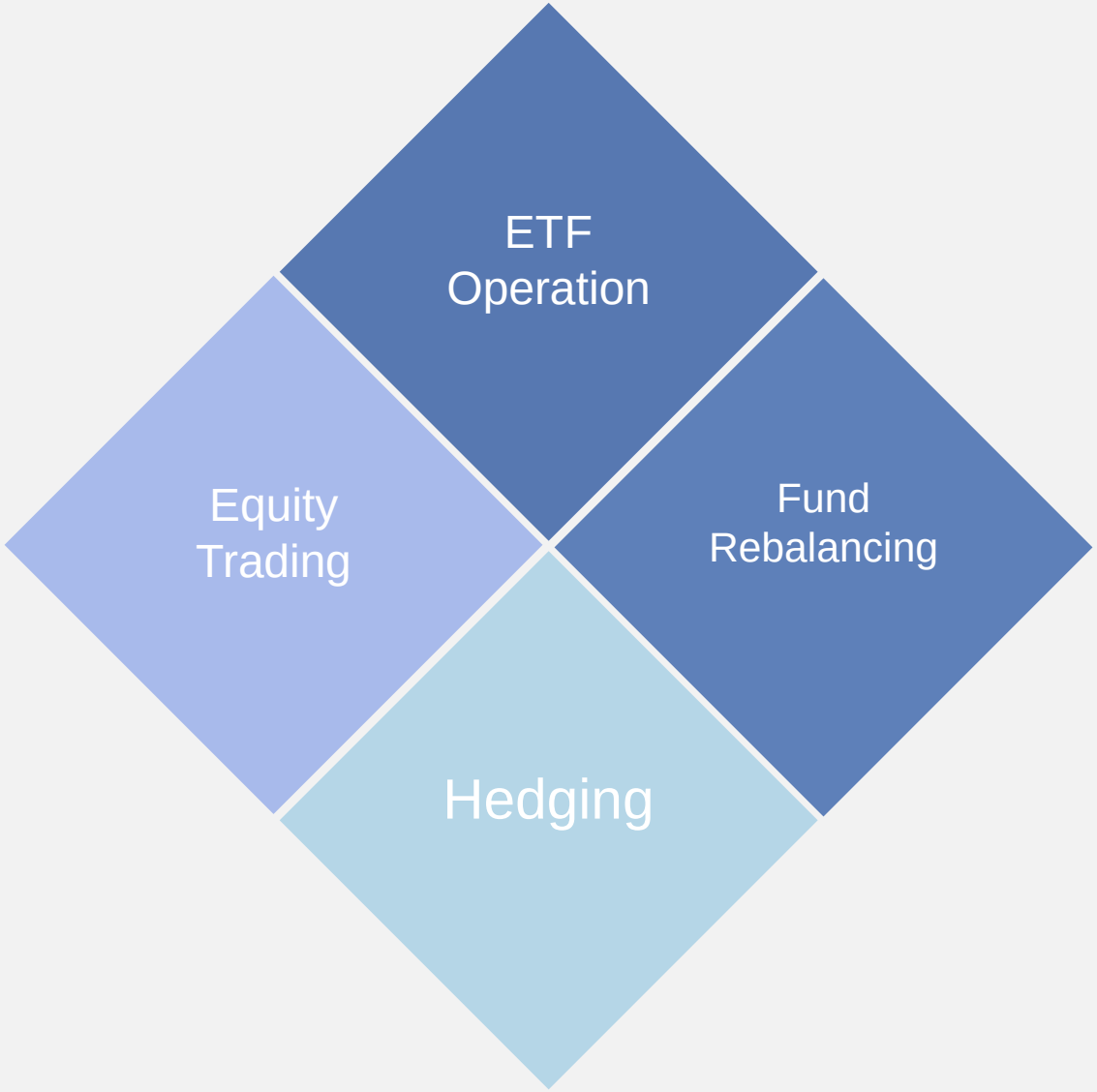
김 성민
QRAFT TECHNOLOGIES, INC.

SEP 09 2019

ORDER EXECUTION



1 Order Execution

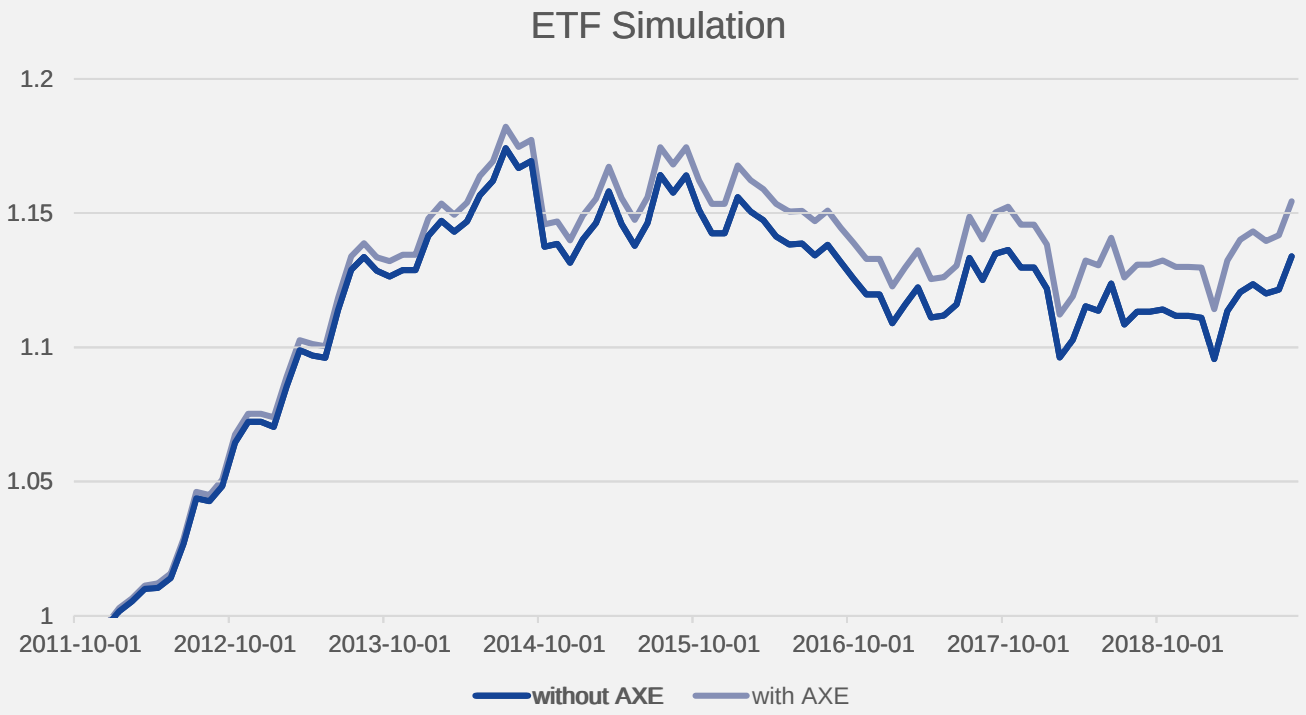


Part of Every Investment Decision

Order execution is implied in every investment process and it takes cost (slippage)

Optimized Order Execution by AXE

2 Hidden Cost of Order Execution



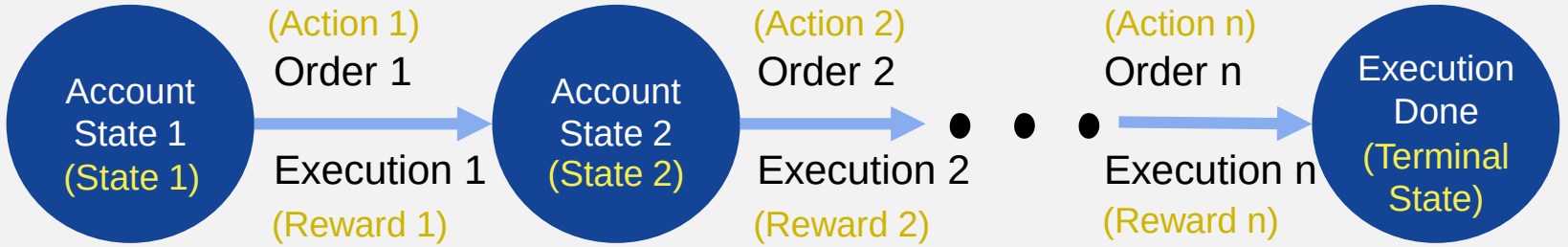
- Quarterly Rebalancing
- 70% of Average Turnover-rate (FnIndex Methodology Book)
- 8 bp cost reduction using AXE (AXE real market test average)

2% of alpha with

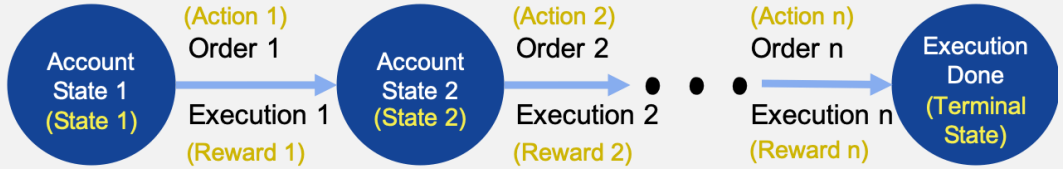
Optimized Order Execution

AXE Reinforcement Learning

3 The Need for Reinforcement Learning

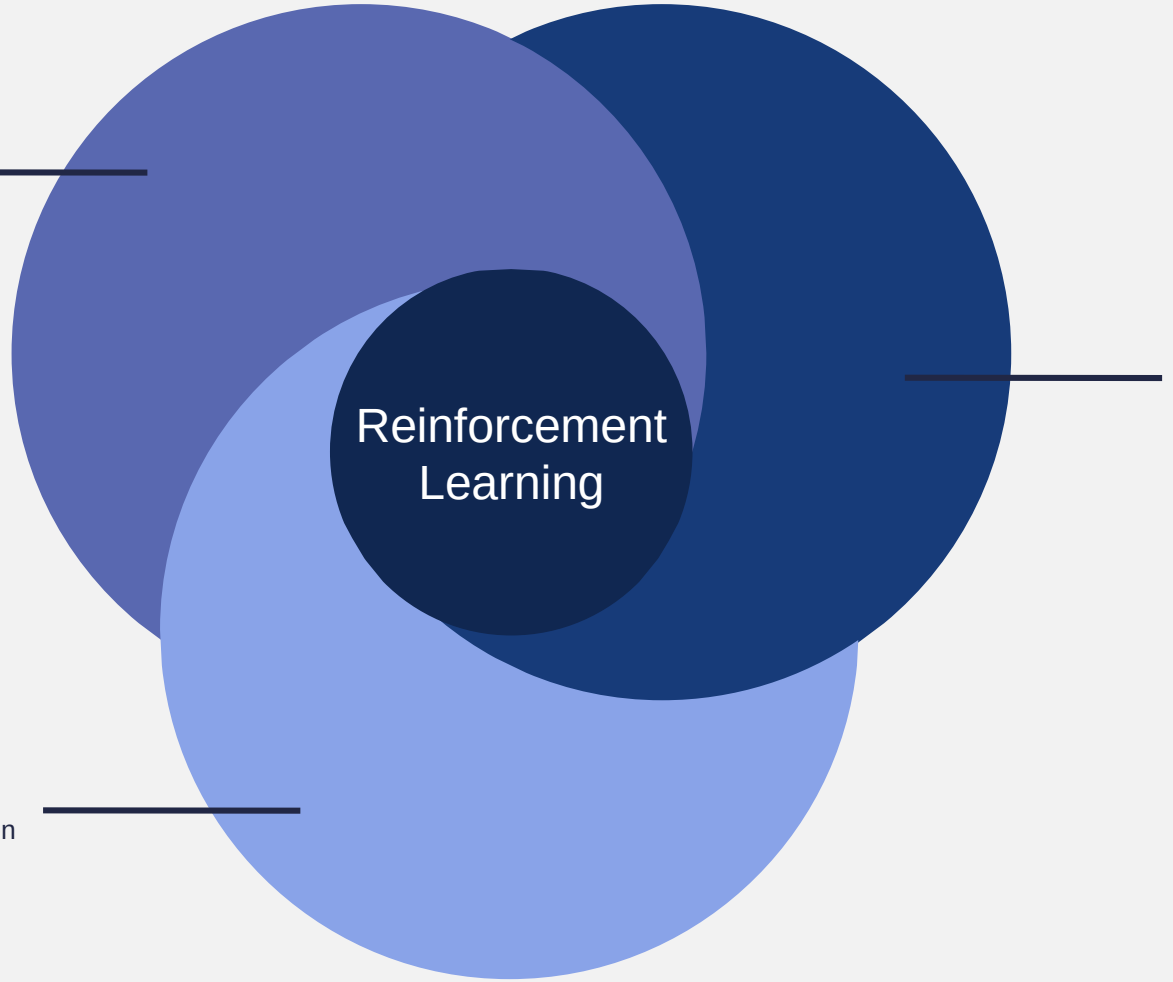


4 The Need for Reinforcement Learning



Deferred Reward

- Order may not be executed ever.
- Reward(Label) is not immediate.



Infinite State

- Not all states are relevant
- Combinations of technical indicator

Path Dependent

- Total reward is result of all actions taken
- Trajectory handling
- Credit assignment problem

AXE Specific RL Solution

5 AXE Specific RL Solution

State Modeling

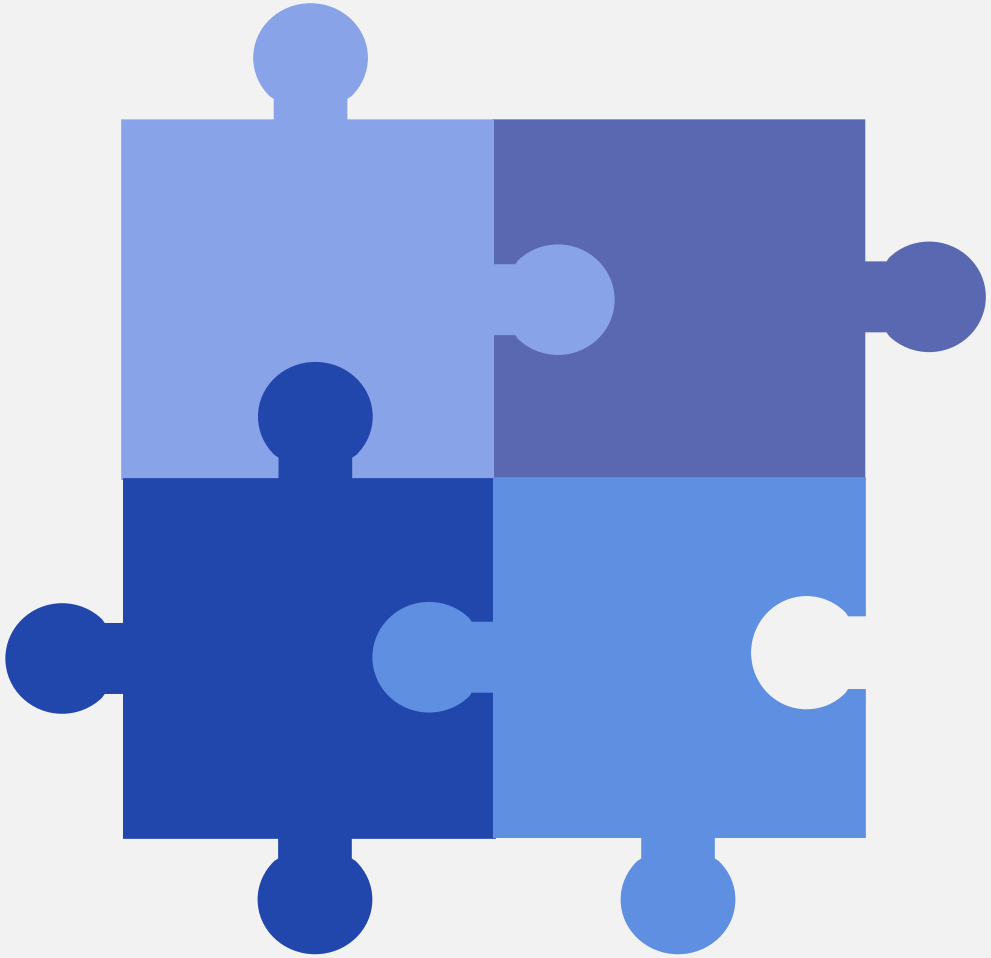
- Mapping Observations to State
- Approximate to Assumption of Markov Decision Process

Imitation Learning

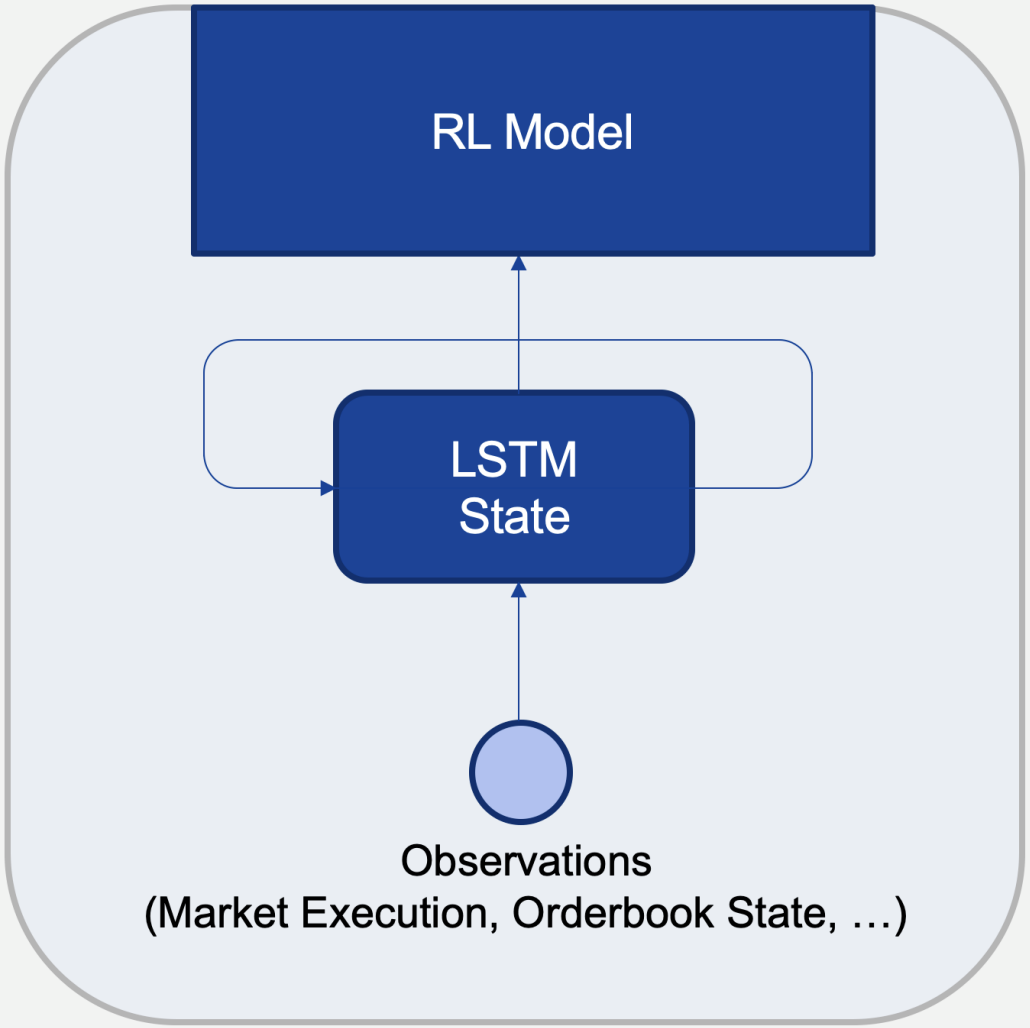
- Providing Guideline for the Model
- Training with Demonstration.

Dedicated Structure Modelling

- Adversarial Action
- Relative Positioning and Orderbook Allocator



State Modelling

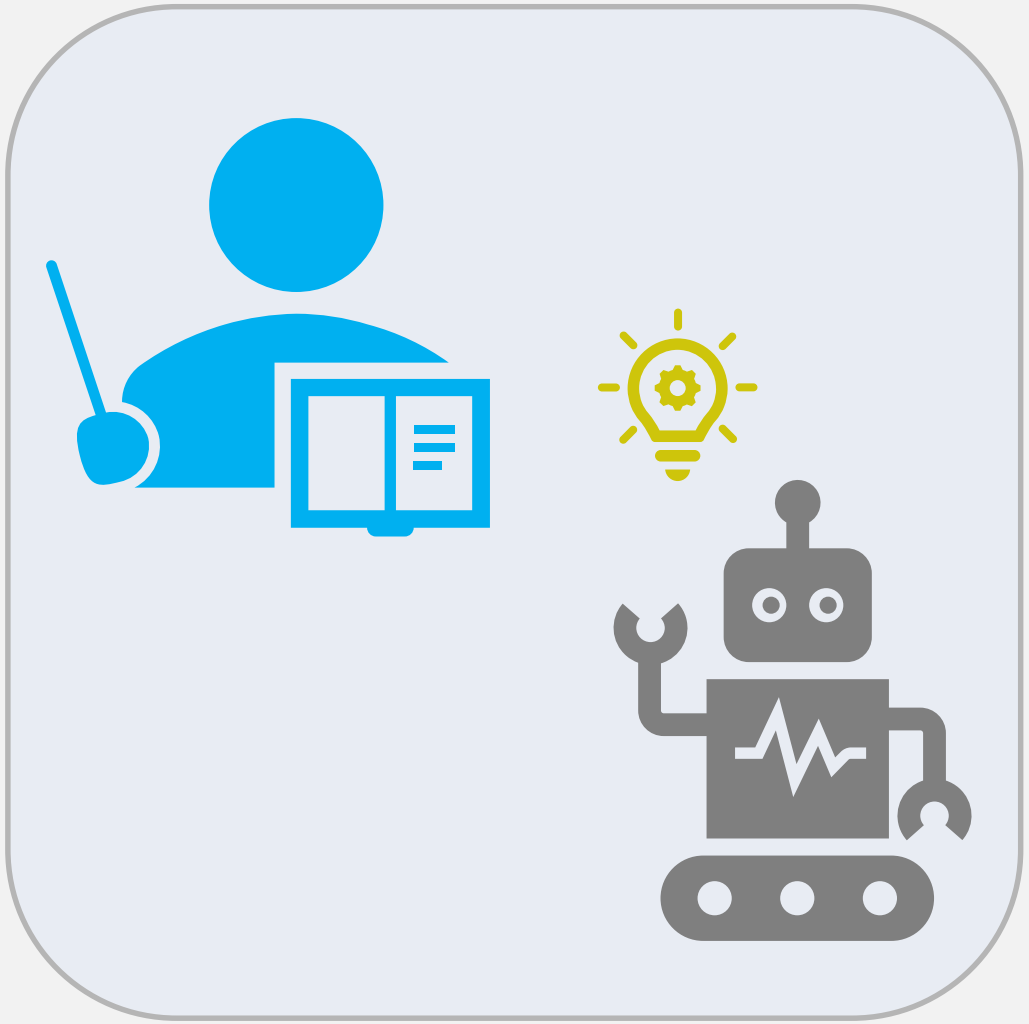


Using LSTM to encode Prior observations to state

Reinforcement Learning follows Markov Decision Process. it assumes the effe ct of action taken in a state depend only on that state not on the prior history w hich characteristic order execution's observations is not follow. So that we use LSTM to approximate.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 1997), 1735-1780. DOI=http://dx.doi.org/10.1162/neco.1997.9.8.1735

Imitation Learning



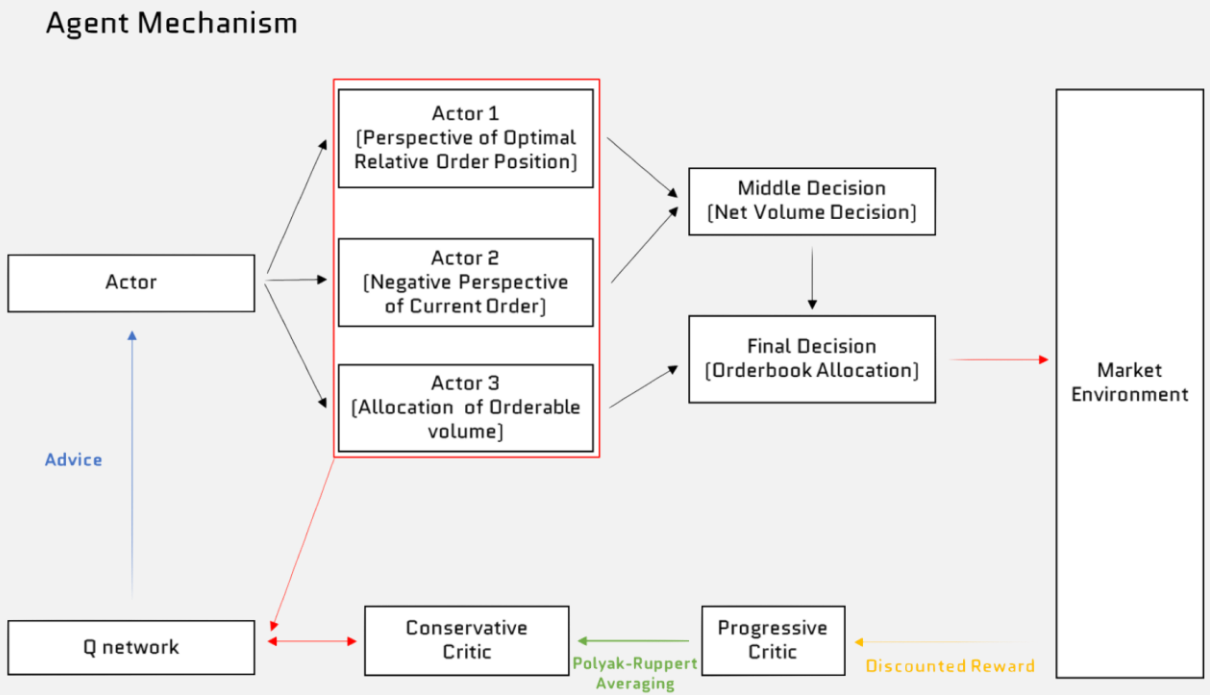
Guide
Network with Episodes (Trajectories)

Even reinforcement learning is theoretically converge to or wandering around optimum, in reality, it just doesn't work. In that case giving guideline to network can be helpful.

- Make optimal trajectories using planning.
- train network with generated trajectories by behavioral cloning.
- Behavioral cloning's objective is mimic the given trajectories as similar as p ossible

Sammut C. (2011) Behavioral Cloning. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA

Dedicated Structured Modelling



Dedicated Structure Modelling for Order Execution

Dedicated Structure is used for order execution process to converge easier

- Adversarial actions to cancel current orders and pending new orders
- Make final order quantities using order allocator
- and others… (Polyak averaging… Double q network…)

AXE Challenge Result

First Place at the AXE Challenge

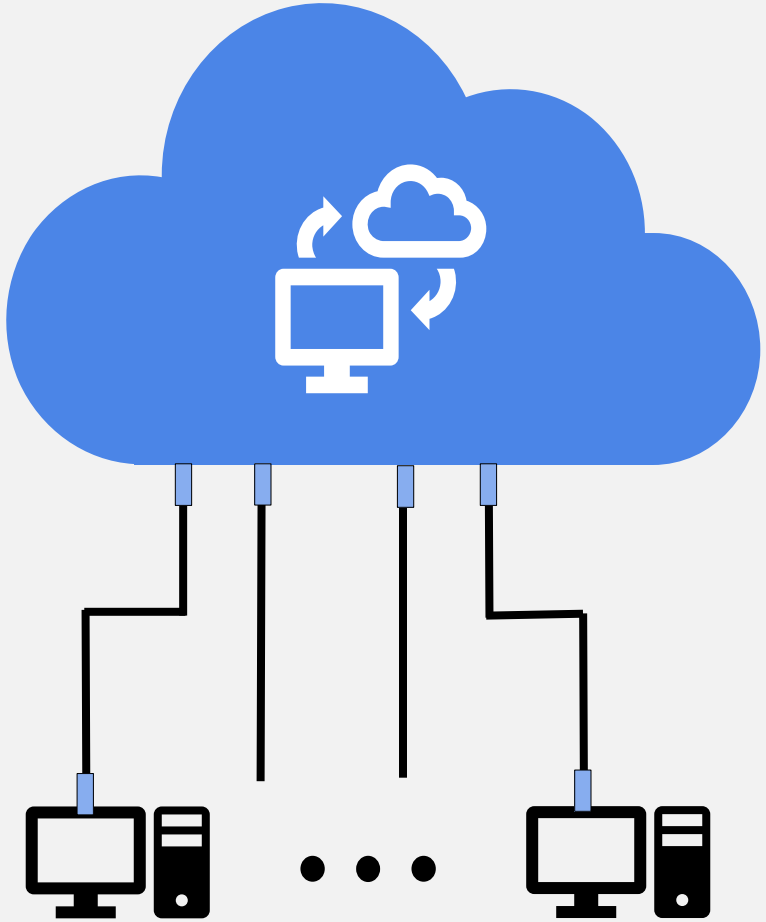
Previous version of AXE takes first place at the AXE Challenge

(from 2018.11.14~2018.11.22 Supervised by PWC)

- Winner takes ₩ 100,000,000
- Randomly select 50 stocks in KOSPI200
- Total Trading Amount: ₩ 2,536,275,830
- AXE's Average Saving Compared to VWAP: 5bp

AXE
Work-in-Progress
& Todo

WIP: Cloud Based Distributed Learning



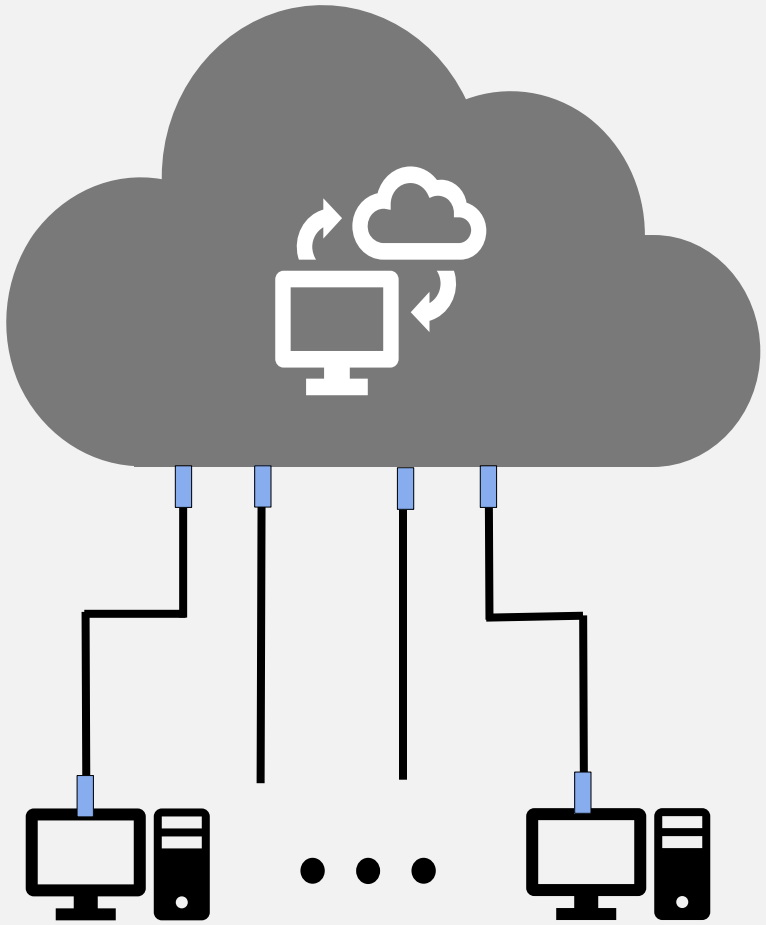
Scalable

- Scales up easily as you want
- Scales up immediately, little setup time required.

Cost Efficient

- Zero or little setup cost
- Cost efficient with proper plan
- Pay as you use

WIP: Cloud Based Distributed Learning



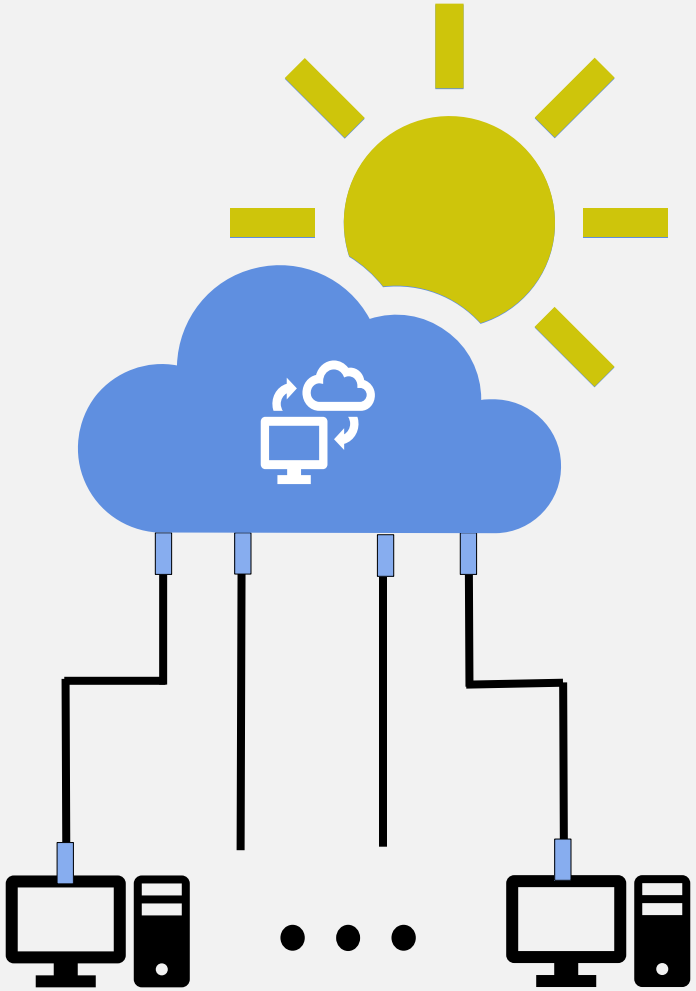
Scalable Designing is Not Easy

- You can’t scale up your experiment without modification.
- Make code scalable can takes lots of effort.

Cost is Still Expensive

- Running massive instances in cloud is still quiet expensive.

WIP: Cloud Based Distributed Learning



Scalable RL Framework

- Our team currently building scalable reinforcement framework.
- Code using framework will be ready to scale up

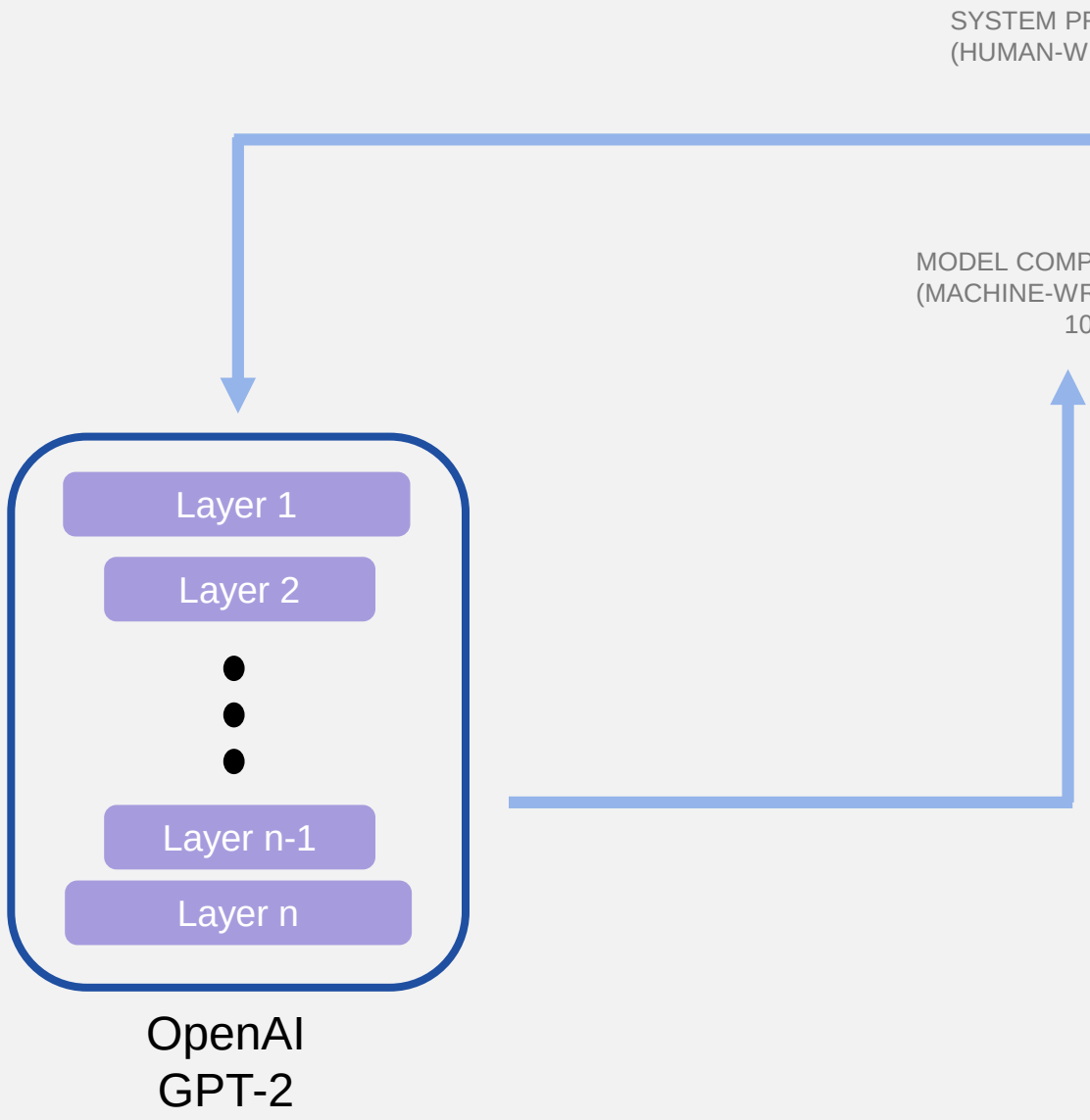
Fault Tolerant Design

- Preemptible instance is much cheaper (about 3 times)
- By designing framework fault tolerant. We can save out budget

Hyperparameter Search

- Reinforcement Learning’s hyperparameter search normally take much longer time than supervised learning
- Automate hyperparameter search between multiple nodes of computer

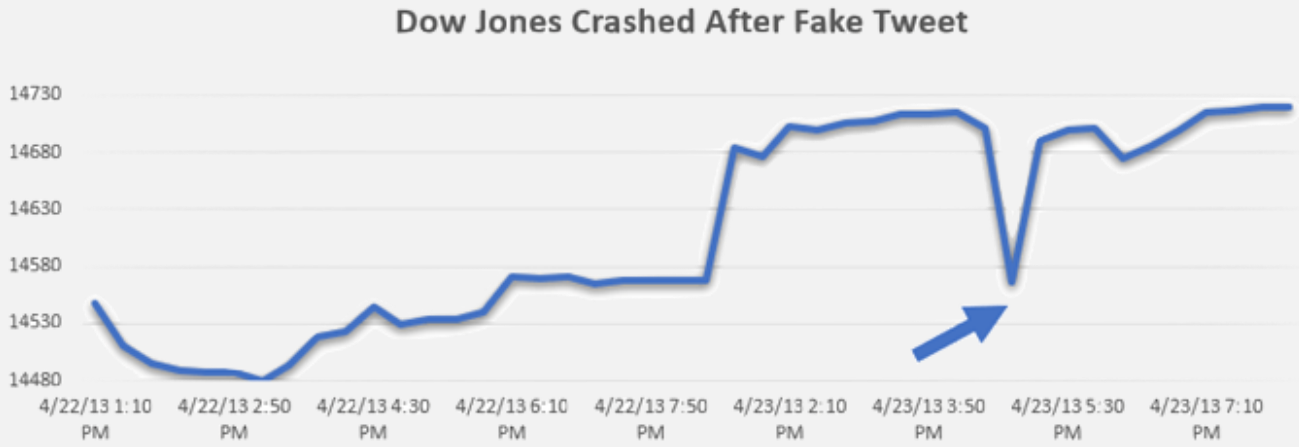
TODO: NLP Based Media Analysis



In a shocking finding, scientist discovered a herd of u nicorns living in a remote, previously unexplored vall ey, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

The scientist named the population, after their distin ctive horn, Ovid's Unicorn. These four-horned, silver- white unicorns were previously unknown to science.
Now, after almost two centuries, the mystery of wha t sparked this odd phenomenon is finally solved.
Dr. Jorge Pérez, an evolutionary biologist from the U niversity of La Paz, and several companions, were ex ploring the Andes Mountains when they found a sm all valley, with no other animals or humans. Pérez no ticed that the valley had what appeared to be a natu ral fountain, surrounded by two peaks of rock and sil ver snow.
Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," s aid Pérez.
Pérez and his friends were astonished to see the uni corn herd. These creatures could be seen from the ai r without having to move too much to see them – th ey were so close they could touch their horns.

TODO: NLP Based Media Analysis



Source: Bloomberg.

Sentiment

SNS

NEWS

BLOG

FORUM

THANK YOU