



MADDPG

Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments

Ryan Lowe*
McGill University
OpenAI

Yi Wu*
UC Berkeley

Aviv Tamar
UC Berkeley

Jean Harb
McGill University
OpenAI

Pieter Abbeel
UC Berkeley
OpenAI

Igor Mordatch
OpenAI

(Submitted on 7 Jun 2017 ([v1](#)), last revised 16 Jan 2018 (this version, [v3](#)))

7107018017 林祐陞



Schema

- Introduction
- DDPG
- MADDPG
- Summary
- Reference

1

Introduction

Policy-gradient
Q-learning

Actor-critic

+DQN

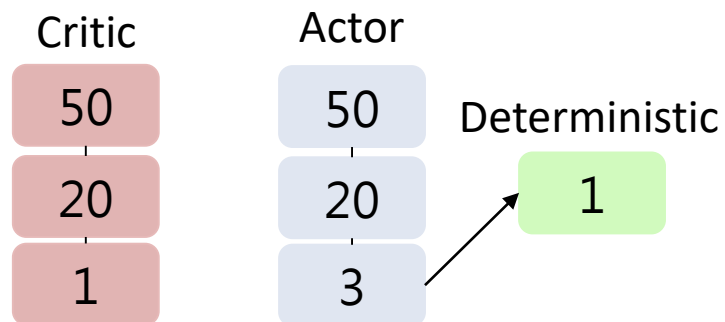
DDPG

MADDPG

2

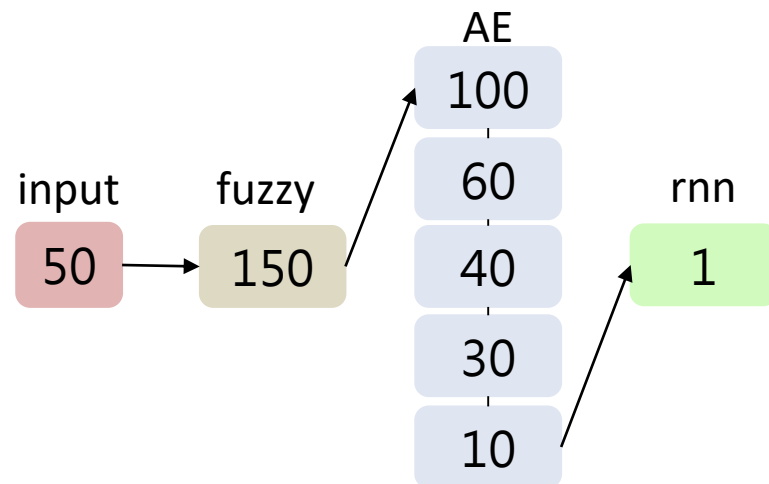
DDPG

- DDPG: Actor-critic + DQN
- Deep : memory pool + 雙網路
- Critic跟Actor要輪流train
- Deterministic : 概率→明確



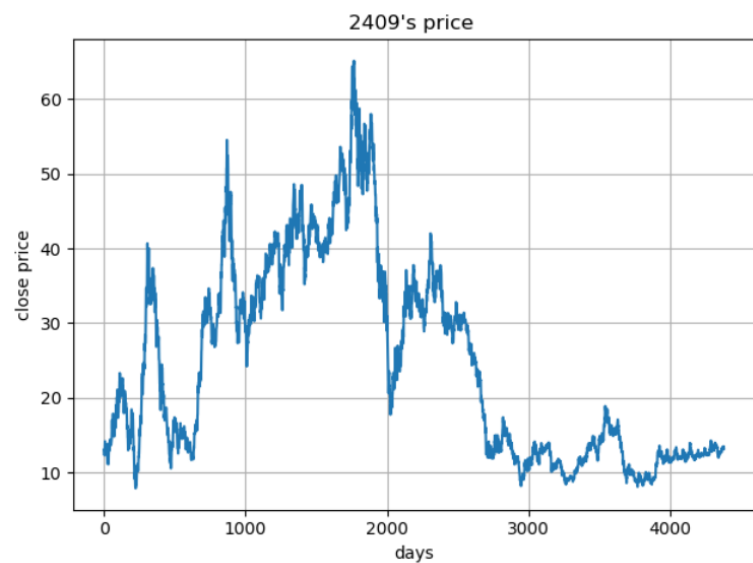
FDRNN

- Fuzzy部分讓資料分開一點
- AE部分加深網路
- RNN部分是模型主軸
- 以tanh方式壓輸出值

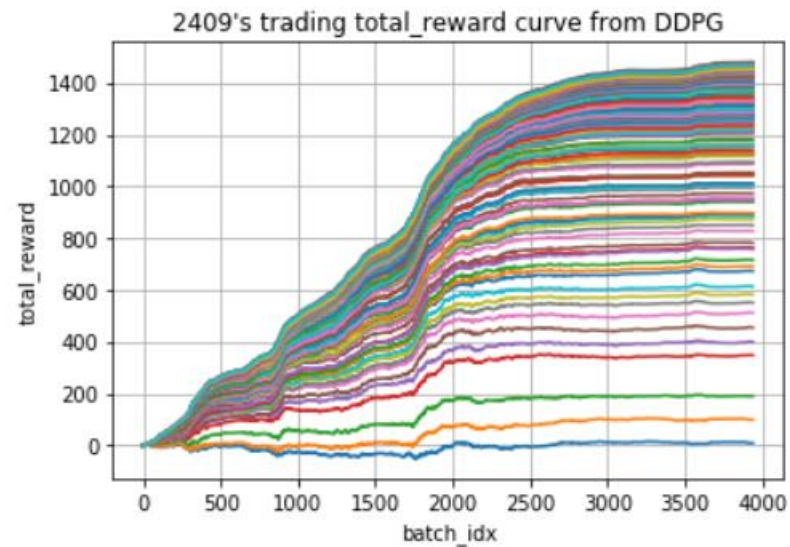


2

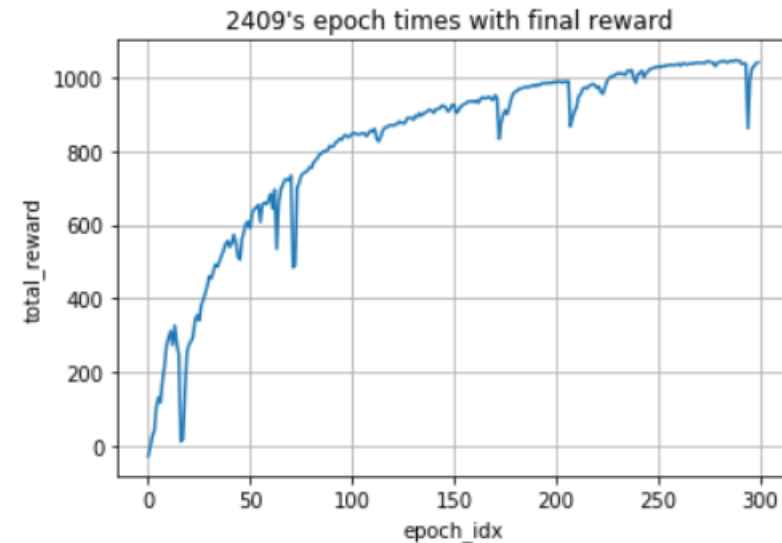
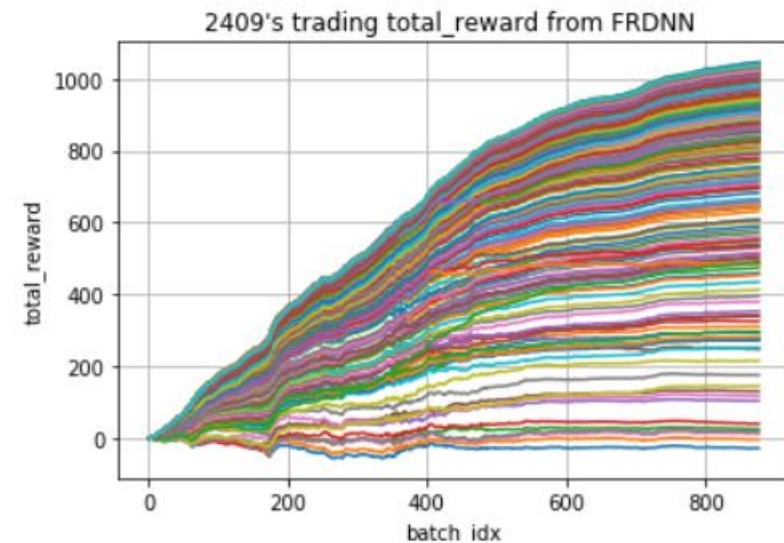
友達



DDPG



FDRNN

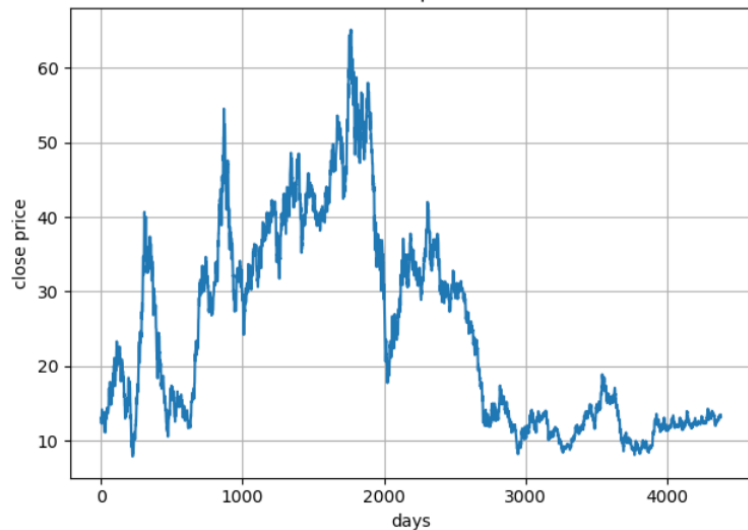


2

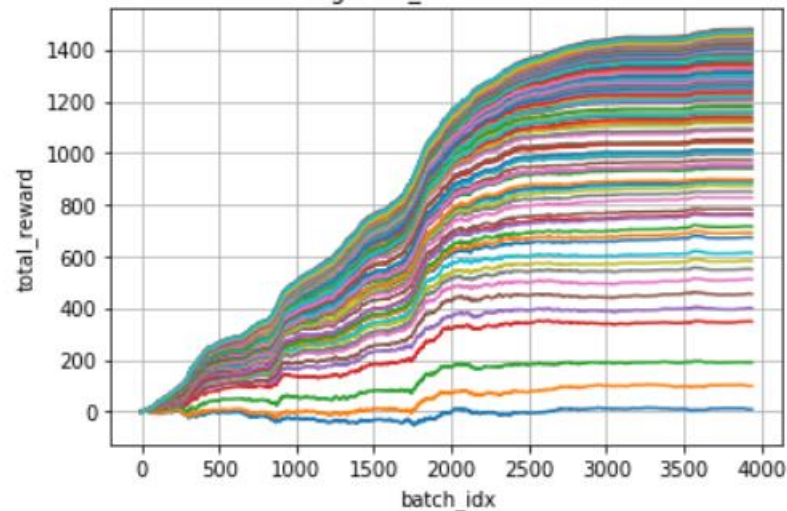
DDPG友達

友達

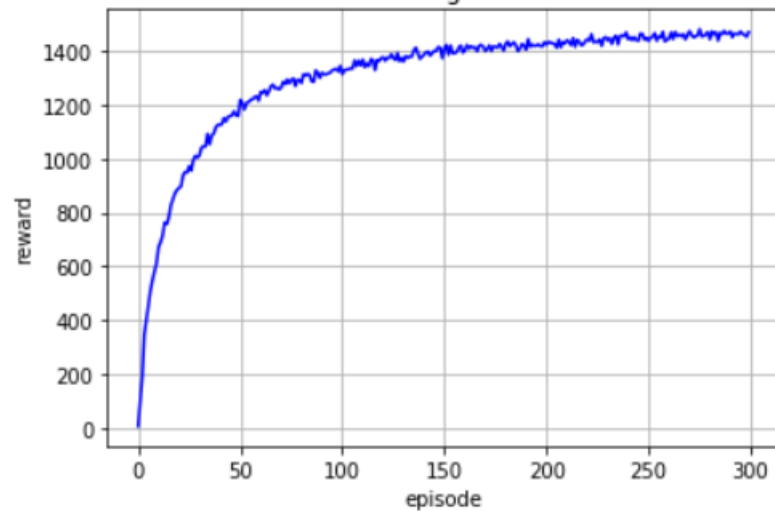
2409's price



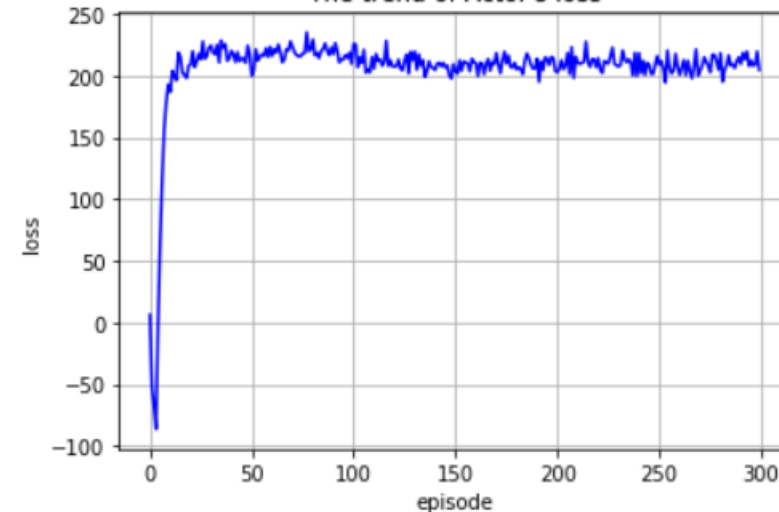
2409's trading total_reward curve from DDPG



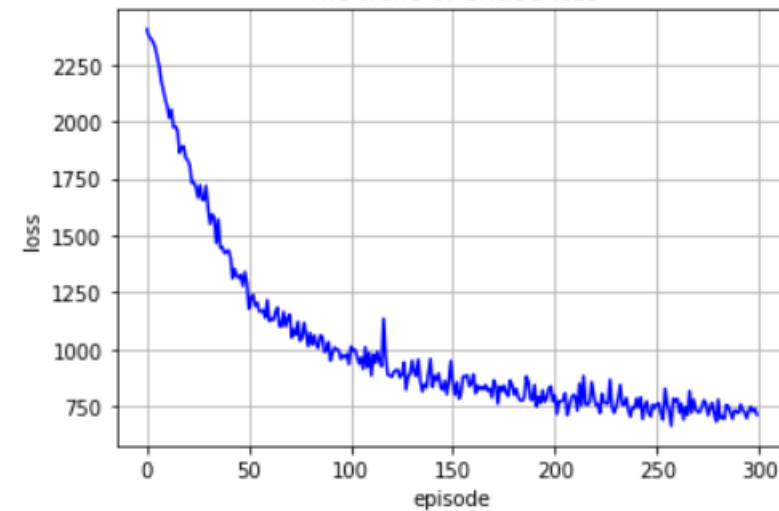
The trend of Agent's reward



The trend of Actor's loss



The trend of Critic's loss

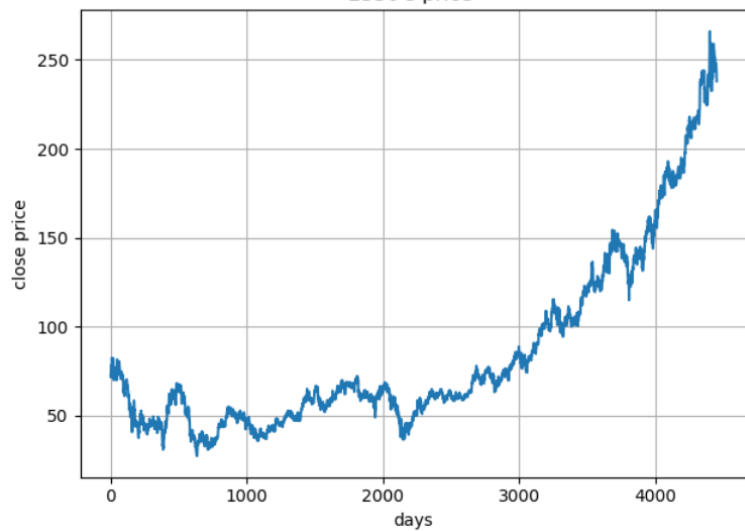


2

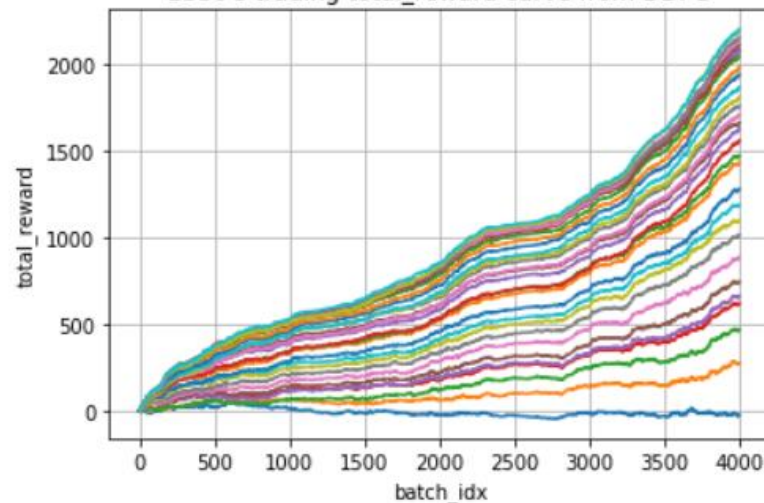
DDPG台積電

台積電

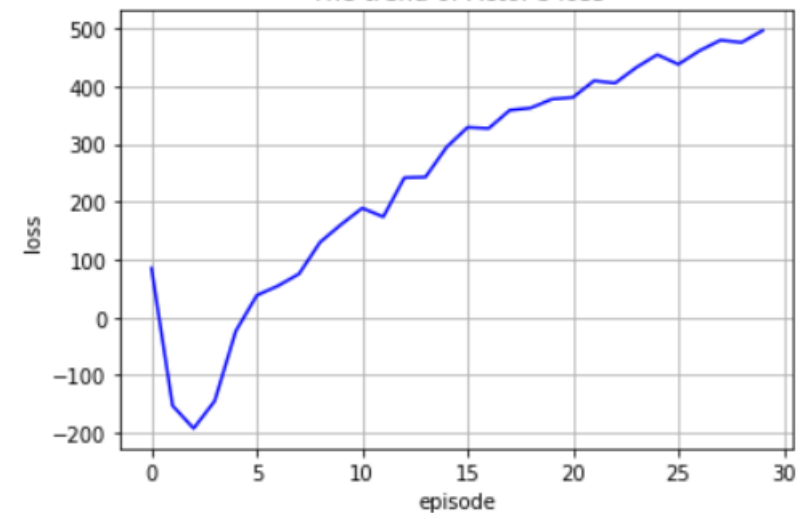
2330's price



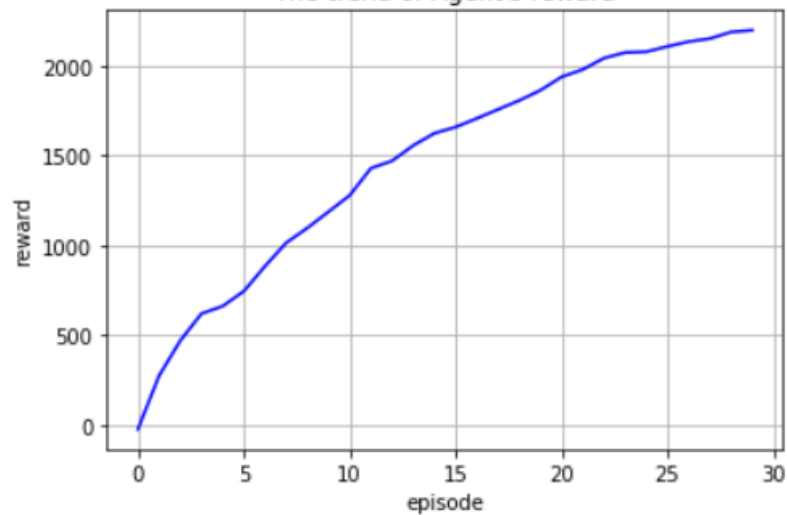
2330's trading total_reward curve from DDPG



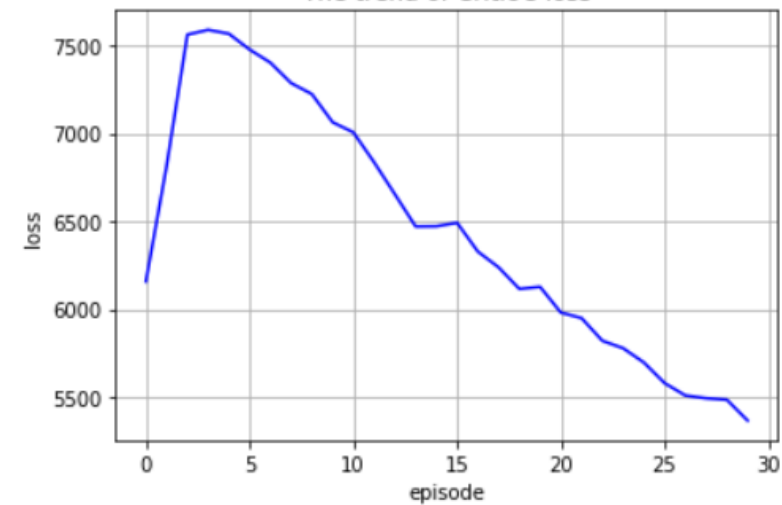
The trend of Actor's loss



The trend of Agent's reward



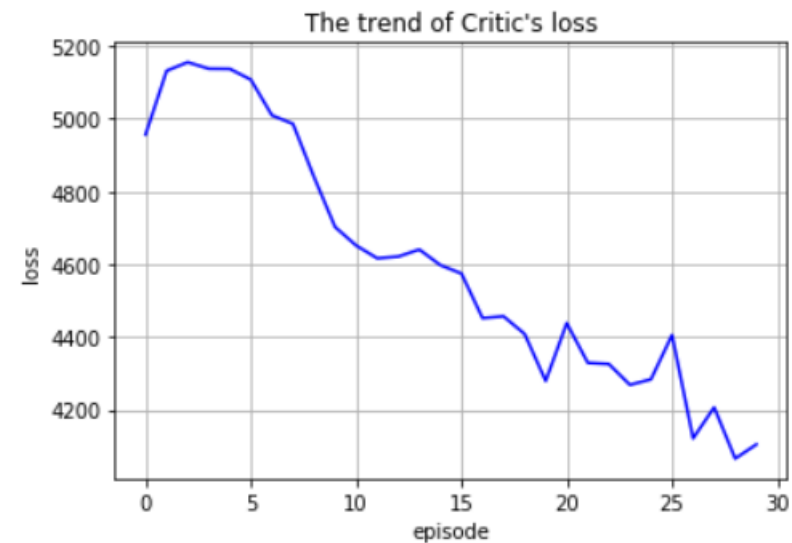
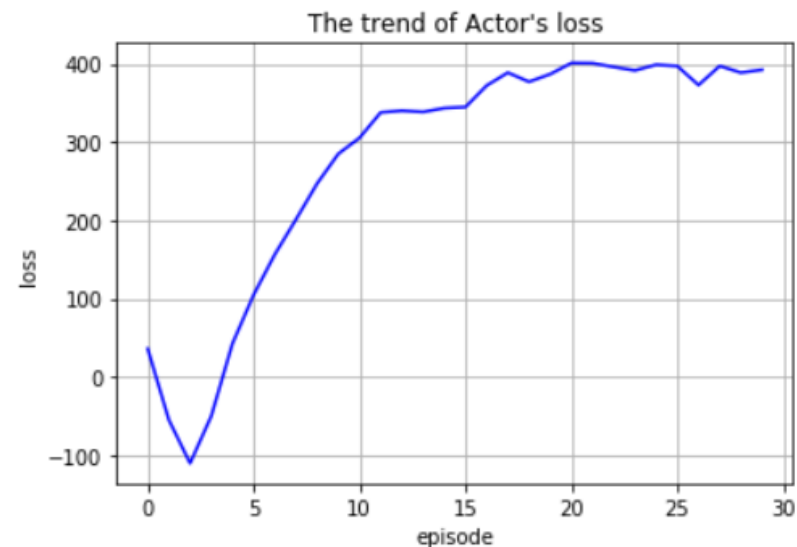
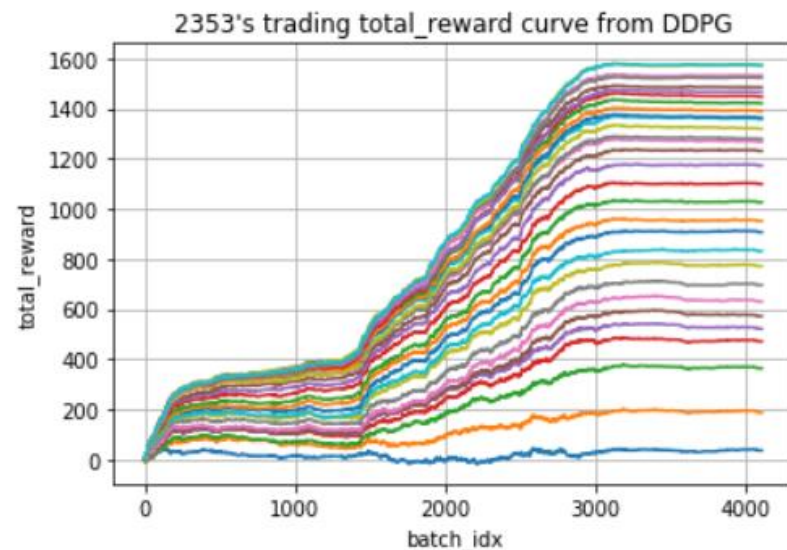
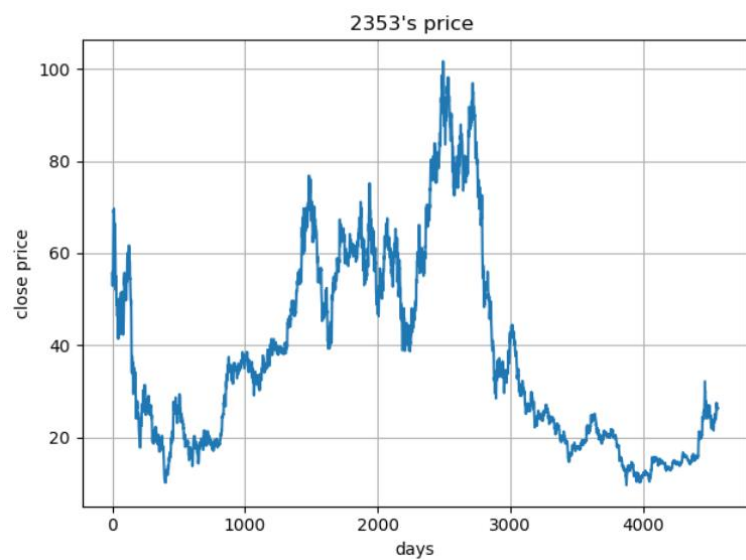
The trend of Critic's loss



2

DDPG宏碁

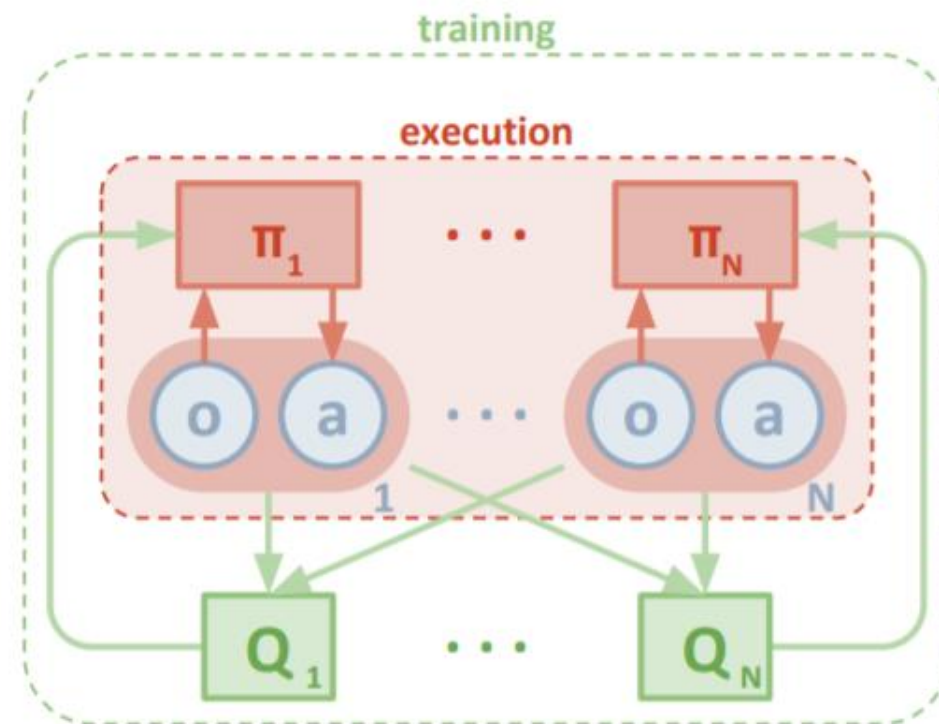
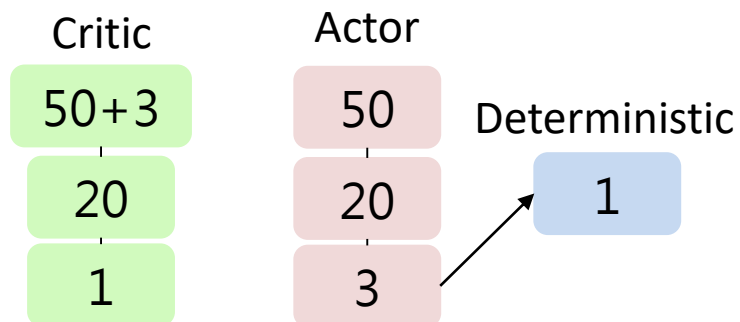
宏碁



3

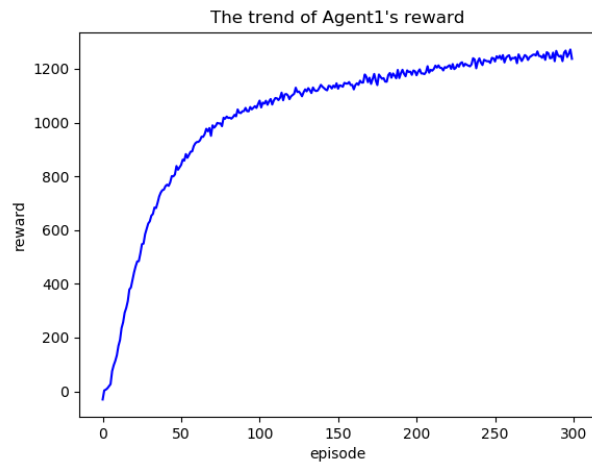
MADDPG

- 每個Agent的訓練和DDPG類似
- 不同點在於Critic的input, 增加了額外信息
 - 其他Agent的action
- 集中學習 + 分散執行

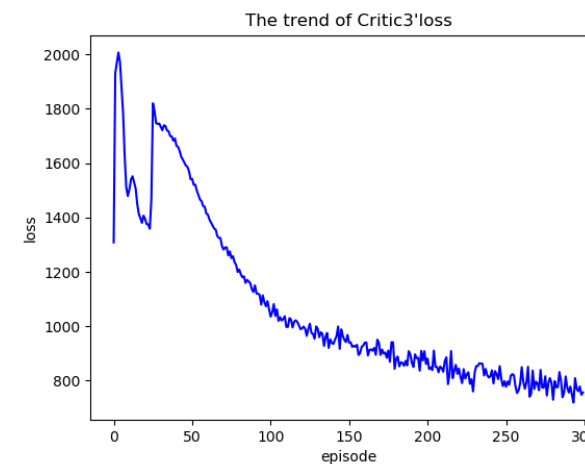
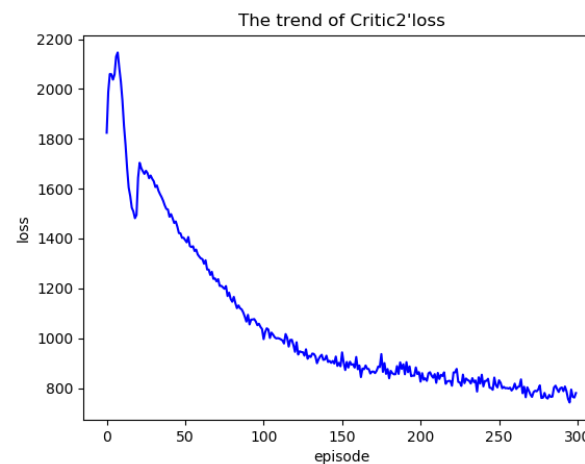
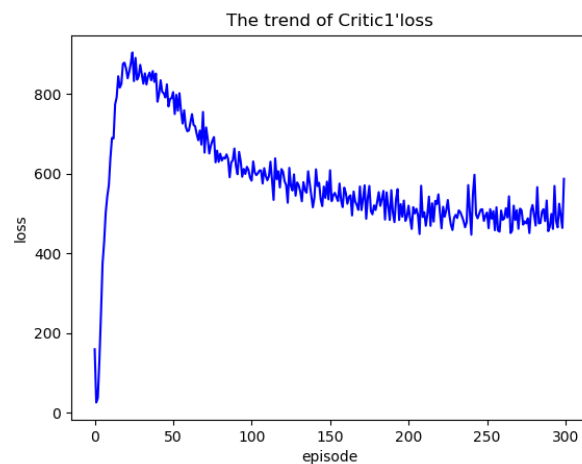


3

Reward



Critic



Actor

