Performance of Deep Reinforcement Learning for High Frequency Market Making on Actual Tick Data

Extended Abstract

Ziyi Xu
Peking University
Beijing, China
ziyixu@stu.pku.edu.cn

Xue Cheng
Peking University
Beijing, China
chengxue@math.pku.edu.cn

Yangbo He
Peking University
Beijing, China
heyb@math.pku.edu.cn

ABSTRACT

High frequency market making is a trading strategy in which agent quotes passively at both ask and bid sides. In this paper, by applying Dueling Double Deep Q Network (D3QN) and a novel reward function, we develop market making agents who can balance profit and inventory robustly, flexibly and full-automatically. Thanks to the actual tick data of stock, we are able to train and test D3QN agents in a relatively realistic environment. To further explore the agent's performance, we also consider a double-agent situation, where the agent competes with a special-designed market maker. And we find in this case, with further training, the D3QN agent learns to quote more narrowly to increase transaction probability. Furthermore, we analyze the impact of high frequency market making on market quality in both single-agent and double-agent cases.

KEYWORDS

High Frequency Market Making; Dueling Double Deep Q Network; Reward; Competition; Market Quality

ACM Reference Format:
Ziyi Xu, Xue Cheng, and Yangbo He. 2022. Performance of Deep Reinforcement Learning for High Frequency Market Making on Actual Tick Data: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

High frequency trading (HFT) has become a potent force in many markets, representing between 40% and 70% of the trading volume in US futures and equity markets, and slightly less in European, Canadian and Australian markets [1]. Compared with the other two main strategies of HFT, cross venue arbitrage and short-horizon directional speculation [2], market making quotes passively at both ask and bid sides, through which provides liquidity for the market.

In this paper, Dueling Double Deep Q Network (D3QN) is applied to develop agents who make the market. By designing a novel reward function, the market maker is able to make stable profit and control inventory risk full-automatically without setting an exogenous inventory bound as in existing literature [3][5]. The market maker in this paper is also able to step forward when encountering a competitor. To our knowledge, we are the first to design RL market makers through a realistic trading matching engine on stock's tick data which enables us to discuss the impacts of high frequency market making.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

Table 1: Action Space

Action ID	0	1	2	3	4	5
θ_{ask}	0.01	None	0.01	0.01	0.02	0.02
θ_{bid}	None	0.01	0.01	0.02	0.01	0.02

2 THE AGENT

2.1 Trading Strategy

The agent quotes every 3-seconds. The action space is listed in Table 1. Action 0 (action 1) represents that the agent only sells (buys) and other actions represents that the agent quotes at both sides. The ask and bid prices are specified by Equation (1) and Equation (2):

$$p_{ask}(t_i) = m(t_i) + \theta_{ask}(t_i), \tag{1}$$

$$p_{bid}(t_i) = m(t_i) - \theta_{bid}(t_i), \tag{2}$$

where $m(t_i)$ is the mid-price at t_i. We define quoting spread $\theta = \theta_{ask} + \theta_{bid}$, it represents how wide the market maker quotes. As for action 0 and action 1, we define $\theta = 0.01$.

2.2 Reward Functions

Consider the reward in $[t_i, t_{i+1}]$, $t_{i+1} - t_i = 3s$. Denote the executed quantity at ask side by $q_{ask}(t_i)$, at the bid side by $q_{bid}(t_i)$, change of inventory is $\Delta I(t_i) = I(t_{i+1}) - I(t_i)$.

A prevailing reward in existing literature is the realistic PnL:

$$\text{PnL}(t_i) = \text{Trading Profit}(t_i) + \text{Inv Loss}(t_i), \tag{3}$$

where

$$\text{Trading Profit}(t_i) = \theta_{ask}(t_i)q_{ask}(t_i) + \theta_{bid}(t_i)q_{bid}(t_i), \tag{4}$$

$$\text{Inv Loss}(t_i) = I(t_{i+1})(m(t_{i+1}) - m(t_i)). \tag{5}$$

This reward provides an intuitive interpretation. However when $I(t_{i+1})$ is large (usually happens when the order flow is extremely unbalanced), the reward will be mainly determined by the change of mid-price. In this case, the agent would become a trader who aims at predicting the price trend. In order to balance Trading Profit(t_i) and $I(t_{i+1})$, an artificial inventory bound is necessary.

To deal with above problems, we define the reward function as:

$$\text{reward}(t_i) = \text{Trading Profit}(t_i) - \lambda_1 \Delta I(t_i) - \lambda_2 \theta(t_i) 1_{\text{No trades}}. \tag{6}$$

The second term in Equation (6) is the penalty for inventory, which is inspired by [4].

However, with the inventory penalty, the agent may quote wide deliberately to avoid trading when one-side transaction happens, so we add the last term. The performances of agents trained with reward (6) and (7) will be compared:

$$\text{reward}'(t_i) = \text{Trading Profit}(t_i) - \lambda_1 \Delta I(t_i). \tag{7}$$

3 METHODS

We use tick data of 000333.XSHE during the time period of 2020.08.06-2020.12.31 (100 trading days). The data is divided into training, validation and testing set using a 64/16/20 split.

We call the agent trained in the single-agent case market maker 1 and the agent trained in the double-agent case market maker 2. In double-agent case, the competitor also quotes every 3-seconds and the trading strategy is naive:

(1) The inventory bound is 500, when the position is within the bound, the competitor quotes at ask and bid side:

$$p_a(t_i) = m(t_i) + 0.01, \tag{8}$$

$$p_b(t_i) = m(t_i) - 0.01. \tag{9}$$

(2) When the position is greater than 500, the competitor only sells at $p_a(t_i)$, and vice versa.

At the beginning of second training, the initial trading strategy of market maker 2 is inherited from market maker 1, which borrows the idea of "pre-train" in deep learning.

4 RESULTS

4.1 Compare Reward Functions

Figure 1 shows the PnL of market maker 1 trained with reward (6) or (7) in the single-agent case. It is worth mentioning that when they adopt action 5, the transaction probability of the latter one is 54%, compared with that of the former one, 70%. It means that if the quoting spread is not properly punished, market makers may become "lazy" and adopt useless wide quotations.

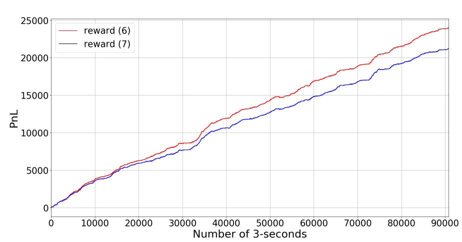

Figure 1: The PnL of market maker 1 using different rewards

In addition, the inventory of market maker 1 during testing period is within $[-400, 400]$, no matter how the price changes.

4.2 Double Market Makers

From Figure 2, market maker 2 outperforms market maker 1 in the double-agent case. What's more, market maker 2 still keeps inventory within $[-400, 400]$.

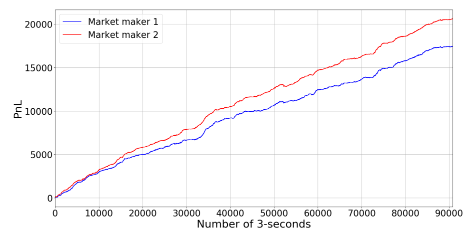

Figure 2: The PnL of market maker 1 and 2 in the double-agent market

The reason why market maker 2 wins is shown in Figure 3. In the presence of a competitor, market maker 2's action distribution skews more to the actions 2-4 but market maker 1 still quotes widely, which indicates that the D3QN agent learns to compete.

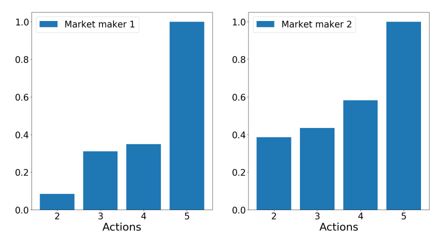

Figure 3: Empirical distribution functions (actions 2-5) of market maker 1 and 2 in the double-agent market

4.3 Market makers' impact on market

Figure 4 shows the daily mean of effective spread in the testing set, which is a metric for traders' trading costs. It is obvious that investors benefit from the competition among market makers, which is also the result of D3QN agent's ability to adjust quotation when there is more than one market maker.

Figure 4: The daily effective spread in different markets

ACKNOWLEDGMENTS

Financial supports from the Fintech Joint Laboratory of Wisecotech and Peking University, National Natural Science Foundation of China under Grants No. 11971040 and 11871010 are gratefully acknowledged.

REFERENCES

[1] Yacine Ait-Sahalia and Mehmet Saglam. 2017. High frequency market making: Implications for liquidity. *Available at SSRN 2908438* (2017).

[2] Ekkehart Boehmer, Dan Li, and Gideon Saar. 2015. Correlated high-frequency trading. *Manuscript, Cornell University, Ithaca, NY* (2015).

[3] Bruno Gašperov and Zvonko Kostanjčar. 2021. Market Making With Signals Through Deep Reinforcement Learning. *IEEE Access* 9 (2021), 61611–61622.

[4] Ye-Sheen Lim and Denise Gorse. 2018. Reinforcement learning for high-frequency market making. In *ESANN 2018-Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. ESANN, 521–526.

[5] Thomas Spooner, John Fearnley, Rahul Savani, and Andreas Koukorinis. 2018. Market making via reinforcement learning. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 434–442.