Advances in Economics, Business and Management Research, volume 113

Factors of Regional Extensive Development (FRED 2019)

Reinforcement Learning Approach for Market-Maker Problem Solution

Lokhacheva K.A.
Department of Geometry and Computer Science
Orenburg State University
Orenburg, Russia
ksenia.lohacheva.97@mail.ru

Parfenov D.I.
Department of Applied Mathematics
Orenburg State University
Orenburg, Russia
parfenovdi@mail.ru

Bolodurina I.P.
Department of Applied Mathematics
Orenburg State University
Orenburg, Russia
ipbolodurina@yandex.ru

***Abstract*—The paper considers the implementation of machine learning technologies to algorithmic trading. The paper studies the process of the stock market trading and the role of the market maker in the trading process, methods of mathematical description of the market maker strategy, along with the possibility of applying reinforcement learning to implement the market maker strategy. The results of testing and evaluating the effectiveness of the developed algorithmic and software tools on the data of the Moscow Exchange are given.**

Keywords—reinforcement learning, machine learning, algorithmic trading, market make, market liquidity

I. INTRODUCTION

An Algorithmic trading (AT) is a way of trading where computer algorithms directly manage the trading process and all the operations using a built-in algorithm [3]. AT is a high-tech and it has become rapidly developing technology due to the interest of the major market participants in finding more effective trading algorithms and improving of existing solutions.

Machine learning can speed up and improve the efficiency of decision-making process, becoming the "technology of the century" for business. The use of the Machine learning approaches allows to capture the risks and to find features that can not be detected with the help of the traditional means of Analytics.

Reinforcement learning is an area of the Machine learning concerned with how software agents can learn while interacting with the environment. Effective algorithms that allow creating reliable predictive models based on large data sets (Big Data) are the main object of the Machine learning. That is why it may become a good practice to use Reinforcement learning algorithms for solving Algorithmic trading problems.

This issue is relevant since the observing trends in the Algorithmic trading stimulation demonstrate the tendency of the exchange markets to fully automate trading operations, and the use of Reinforced learning methods can increase the efficiency of the trading algorithms [4].

Lots of papers are devoted to the use of intelligent methods for solving Algorithmic trading problems. D. Montague [1] obtained a high-performing trading strategy on historical data. The paper contains the results of compression of four regression algorithms: linear and regularized (ridge) linear regression; Neural Networks; Random Forests; and gradient-boosted decision trees. According to this results gradient-boosted decision trees demonstrated the best performance. However, not all intellectual methods (in particular, Machine learning methods) have been studied in this paper. The authors of [2] consider Reinforcement learning methods regarding the problem of modeling the stock agent behavior. In particular, an idea of the necessary aspects of Reinforcement learning is given, some of the original automated trading agents based on various algorithms are described, and the application of these agents to artificial and real time series of daily set prices for financial assets are proposed. Nevertheless, the parameters of the systems in the proposed solutions remain standard, which indicates an incomplete study of the application of Reinforcement learning methods to the problems of the Algorithmic trading.

The problem statement considered in this paper is similar to the one studied in [5], however, the methods used to solve it are different from those used by Fernandez-Tapia J. This study is aimed at developing a software and algorithmic solution for modeling the market-maker behavior using Reinforcement learning methods in order to increase the market liquidity.

II. PROBLEM STATEMENT

Mathematical modeling of the market dynamics is rather complicated problem. Even complex and multicriterion models can not truly reflect real financial systems. It is more evident for high-frequency trading, which is rather new area of

great interest to both financial institutions and market regulators.

A. Market-Maker Problem

Due to the unstable economic situation in the world, the largest exchanges attract hundreds of market-makers to increase investor loyalty, who submit orders, i.e. quote tens of thousands of financial instruments. The goal of a market maker is to provide liquidity to quoted instruments and to reduce the spread (the difference between sell price and buy price). The market-maker earns the price-difference between these two orders. Thus, the market-maker's algorithm would like to maximize the number of pairs of buy/sell trades executed, at the larger possible spreads and by holding the smallest possible inventory at the end of the trading session [5]. Hence, the market-maker faces the following trade-off: it is expected that a large spread means a lower probability of execution while a narrower spread will mean a lower gain for each executed trade. On the other hand, if the trading algorithm only executes its orders on one side (because of price movements, for example), then its inventory moves away from zero, bearing the risk to eventually having to execute those shares at the end of the trading session at a worst price [5]. That is why market-maker problem has a multicriterion nature.

From a modeling standpoint, one iteration of a market-making tactic can be seen as the following: agent submits orders in order-book, then waits during time interval interacting ΔT. Market and limit orders arrive to the book during this period. After, this period the new state of the order book is counted, the market-maker update his orders and the process is repeated (fig. 1).

At the end of each iteration the payoff is represented by a random variable [5]:

$$\theta(\delta_a, \delta_b) = P(N_a(\delta_a, \xi), N_b(\delta_b, \xi), \xi), \quad (1)$$

where δ_a – the position of the sell-order with the respect to reference; δ_b - the position of the buy-order with the respect to reference; N_a – the number of handled sell-orders during time-window ΔT; N_b – the number of handled buy-orders during time-window ΔT; ξ – the exogenous variables influencing the payoff (e.g. price, spread).

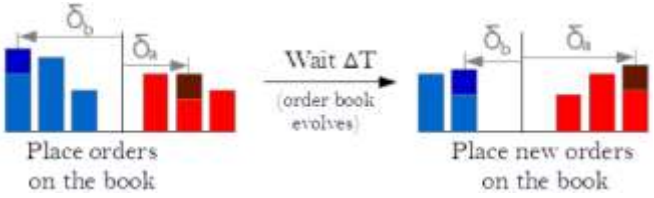

Fig. 1. Schema of the strategy [5].

The main bases for using Reinforcement learning methods for solving the market-maker problem is the reward postponement.

B. Reinforcement learning approach

Reinforcement learning is a challenge faced by an agent learning with "exploration and exploitation" methodology, continuously interacting with a dynamically changing environment. Basically, the Reinforcement learning model consists of

- the set of environment states S;
- the set of agent actions A;
- the set of reward signals R.

In a basic Reinforcement learning model agent obtains current environment state s_i at each step i, then chooses the action a_i as an output signal. This action changes the environment state and the agent receives the reward signal r as the result of this change. Agent should select an action that will increase the discounted value of the sum of the rewarding signals:

$$R_t = \sum_{i=0}^{\infty} \gamma^i r_i \quad (2)$$

where γ – discount factor.

Let us consider specific Reinforcement learning methods to implement the market-maker strategy.

C. Q-learning and environment indicators

The implementation of the Dynamic Programming algorithms to assess the usefulness of the used strategies is one of the principles of Reinforcement learning methodology. Q-learning algorithm, which implements this principle, uses Q-function. Current environment state and chosen action are the independent variables of this function. This allows to build the Q-function iteratively and thus to find the optimal control strategy. The expression of the Q-function updating [6] is

$$Q(s_i, a_i) \leftarrow r_i + \gamma V(s_{i+1}). \quad (3)$$

Since the aim of the system is to maximize the total reward it is possible to bring the recurrence relation (3) to the form (4), the iterative procedure of which is written in the form (5) [6]:

$$Q(s_i, a_i) \leftarrow r_i + \gamma \max_{a \in A} Q(s_{i+1}, a), \quad (4)$$

$$Q(s_i, a_i) \leftarrow Q(s_i, a_i) + \alpha(r_i + \gamma \max_{a \in A} Q(s_{i+1}, a) - Q(s_i, a_i)). \quad (5)$$

This algorithm performs the convergence of the Q-function to the optimal one regardless of the implemented strategy [6].

In practice, it is necessary to obtain the values of available technical indicators in real time to describe the state of the

order-book (the state of the environment). We considered 2 of them:

- exponential moving average (EMA).

This indicator is used to identify trends in the market and to identify the moment of entry into the market and the moment of exit from the market.

$$EMA_t = \alpha P_t + (1-\alpha) EMA_{t-1} \quad (6)$$

where α – coefficient, that shows the rate of reducing data relevance: the bigger the value of the coefficient, the more important new observable values of the random variable, the less important old values (generally, $\alpha = 2/(1+N)$, N – number of days in period); P_t – the value of reference price at time t.

The point of entrance, that is called the signal of the long position opening, appears when fast moving average (that has shorter period) crosses slow moving average (that has longer period) in the rising direction. The point of exit, that is called the signal of the long position closing, appears when fast moving average crosses slow moving average in the descending direction.

It is better to use EMA for short-period trading. This close satisfies restrictions of our problem.

- relative strength index (RSI).

RSI is a technical analysis indicator that measures the magnitude of recent price changes in order to evaluate overbought or oversold conditions of this particular instrument. One can determine RSI as

$$RSI(n) = 100 - \frac{100}{1 + RS(n)} \quad (7)$$

where

$$RS(n) = \frac{\sum_{i=0}^{13} \max(C(n-i) - C(n-i-1), 0)}{\sum_{i=0}^{13} \max(C(n-i-1) - C(n-i), 0)} \quad (8)$$

RSI value is in range of 0 to 100. We say that the instrument is overbought, if RSI >70. This means it is better to sell it. We say that the instrument is oversold, if RSI < 30, this is the signal to buy the instrument.

The described technical indicators were used to determine and describe the environment state.

The next section presents the descriptions of a developed system prototype based on approaches discussed above

III. Description of the Developed System

As a result of applying of the Reinforcement learning principle to the market-maker problem, we obtain the following problem statement. Discretize the time of the algorithm in seconds. The market is the environment in which the agent (market maker) operates (fig. 2). We take the description of the order-book at a given time as the state of the environment. The action of an agent is the decision of the order submitting.

In addition, the agent will have to decide at what distances ($\delta_a^{(i-1)}$ and $\delta_b^{(i-1)}$) from the reference price S it should post its quotes. Let us note $N_b^{(i)}$ and $N_a^{(i)}$ the respective number of orders the market-maker executes at the bid and at the ask during this i-th period. Thus, we obtain the problem of the expected reward maximization:

$$\theta_i = \underbrace{N_a^{(i)}(S_{(i-1)\Delta T} + \delta_a^{(i-1)})}_{\text{sell}} + \underbrace{N_b^{(i)}(S_{(i-1)\Delta T} + \delta_b^{(i-1)})}_{\text{buy}} \quad (9)$$

The market maker should monitor the environment states, that is, the order-book, for some time to get information about what actions are accompanied by the maximum reward. As soon as there is enough data for independent activity of the agent, he starts to submit his own orders at the end of each period of time ΔT.

Implementing Q-algorithm approach to the described problem we obtain a Q-table. All possible combinations of the mentioned technical indicators and agent actions (buy/sell) are inputs of this table.

Moreover, since the market-maker strategy takes into account the number of executed orders at each distance δ from the reference price S, we add the third detention of the Q-table – the distance δ (note, that the Q-table is the 3D matrix in computer representation).

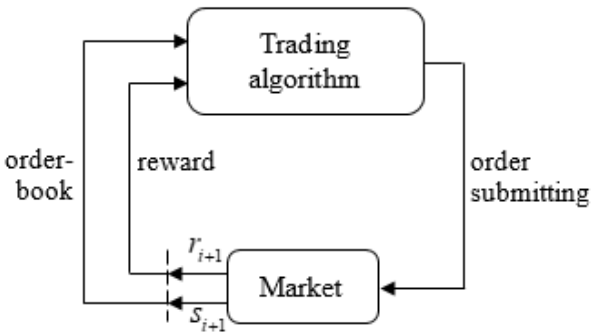

Fig. 2. Schema of the Reinforcement learning implementation to Algorithmic trading.

Thus, in order to find the optimal market-maker strategy we had to complete the following steps:

- create the 3D Q-table with the following inputs: "distance δ from the reference price" (that is found at order-book), "possible combinations of the technical

indicators" (values of which are counted as (6) – (8)) and "agent action" (that is buy/sell);

- train the model using real historical data according to Q-learning method (5) and the reward (9).

The market maker monitors the environment state for some time, that is, the order book, in order to get information about what actions are accompanied by the maximum reward. As soon as there is enough data for independent agent activity, it begins to submit its own orders at the end of the period of time ΔT, during which the state of the glass should change.

IV. Experimental Results

To test and evaluate the effectiveness of the developed algorithmic and software solutions, we used the data of the Moscow exchange (namely, the "stock Market") [7].

Testing of the developed system was carried out for daily data of Gazprom lists on 09/01/2014 – 09/05/2014, 5.046.709 transaction records were there in the test sample. The period is 10.00am - 5.40pm, the testing period is 5.40pm – 6.40pm.

The software tool, written in the C#, combines both functions of modeling the environment and the implementation of developed algorithm.

Figure 3 shows the possible average rewards that a market maker would receive if he did not submit orders on his own during the testing period, but continued to observe and record the changes in the order-book (highlighted in blue on the chart), and the possible average rewards that a market maker would receive if he submitted orders on his own., that is, using the developed algorithm (highlighted in red on the chart). Before plotting, the data obtained from the samples was centered, standardized, and then the average values of the centered-standardized values for every 2 minutes were found, these average values are the points of the chart.

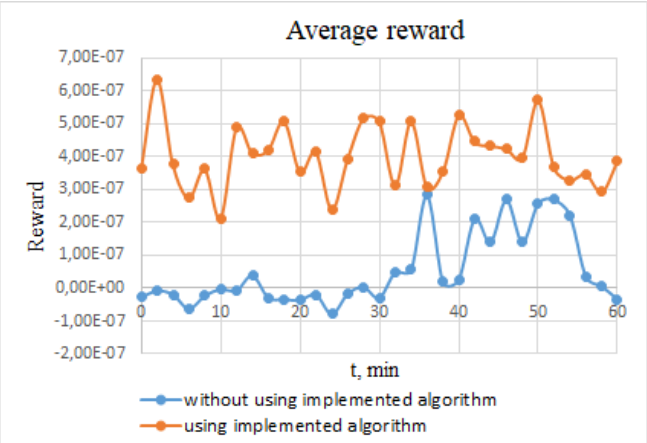

Fig. 3. The rewards obtained at 09/01/2014.

TABLE I. Accumulated reward

Date	Possible reward without using algorithm	Possible reward using algorithm
09/01/2014	29317222	6228467213
09/02/2014	132341250	8179061862
09/03/2014	209562379	3687779980
09/04/2014	4624416310	4663140298
09/05/2014	18243331	4663140298

As it clear from the chart (fig. 3) the amplitude of the average rewards obtained without using implemented algorithm is quite large (varies within two digits).

Moreover, every 2 minutes the average reward can be either more or less than the average reward for the entire test period. This means that the instant liquidity on this instrument is unstable. At the same time, the amplitude of the average of the average rewards obtained using implemented algorithm is quite small (varies within one digit). And every 2 minutes the average reward is always greater than the average reward for the entire test period. This means that the instant liquidity for this instrument is high and stable.

Similar results were obtained while testing the system on the data of September 2, 3, 4 and 5.

Moreover, the total accumulated reward for all test periods is greater in the case when the implemented algorithm of orders submitting is used (tab. 1).

V. Conclusion

The study on the topic "Reinforcement Learning Approach for the Market-Maker Problem" allows us to draw the following conclusion.

The mathematical apparatus of the process of trading in the market is studied; the problem of the market maker and the principle of reinforcement learning for training are considered.

The Q-learning algorithm for implementation of reinforcement learning principle is considered; technical indicators for assessing the state of the environment are selected.

The developed system that implements the considered algorithm is described, as well as the results of testing and evaluation of the effectiveness of the developed algorithmic and software tools on the data of the Moscow exchange are presented.

According to the test results, if the market-maker submits orders on its own, the amplitude of the average of the average rewards obtained using implemented algorithm is quite small (varies within one digit). At the same time, every 2 minutes the average reward is always greater than the average reward for the entire test period. This means that the instant liquidity for this instrument is high and stable, which satisfies the goal of the market-maker, as well as the research goal.

References

[1] Algorithmic Trading of Futures via Machine Learning at: http://cs229.stanford.edu/proj2014/ David%20Montague,%20Algorithmic%20Trading%20of%20Futures%20via%20Machine%20Learning.pdf, accessed: 11.10.2018

[2] Bertoluzzo, F. Reinforcement Learning for automatic financial trading: Introduction and some applications.Working Papers Department of Economics Ca'Foscari University of Venice. – 2012. – №33WP. – pp. 1-15

[3] A. Chaboud, B. Chiquoine, E. Hjalmarsson, C. Vega, "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market", Journal of Finance, 2013, Vol. 980, pp. 2045-2084.

[4] S. Volodin, A. Jakubov, "Development of algorithmic trading on the world financial markets: reasons, trends, prospects", Finance and credit, 2017, Vol. 23, pp. 532-548 ["Rasvitie algoritmicheskoy torgovli na mirovyih finansovyih ryinkh: prichinyi, tendencii, perspectivyi", Finance and credit, 2017, Vol. 23, pp. 532-548] (In Russian).

[5] J. Fernandez-Tapia, High-Frequency Trading Meets Reinforcement Learning. Exploiting the iterative nature of trading algorithms" at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594477.

[6] A. Grishko, S Udovichenko, L. Chalaya "Intelligent system of trading strategies formation using combined indicators", Intellectual bionics, 2011, pp. 9-17 ["Intellectualnaya sistema formirovaniya torgovyih strategii s ispolzovaniem combinirovannyih indikatorov", Bionika itellecta, 2011, pp. 9-17] (In Russian).

[7] Moscow Exchange at https//www.moex.com/ru/orders?historicaldata, accessed: 24.04.2019