Advanced Science and Technology Letters
Vol.106 (Information Technology and Computer Science 2015), pp.61-64
http://dx.doi.org/10.14257/astl.2015.106.14

Q-value Additional Updating Method for Reducing Learning Time

Yunsick Sung[1,*]

[1] Faculty of Computer Engineering, Keimyung University, Daegu 704-701, South Korea,
yunsick@kmu.ac.kr

Abstract. In this paper, an update method of Q-value is proposed to increase the learning rate of Q-learning. When Q-value of executed action is small, even if it is an optimal action, the learning becomes longer because the frequency to be executed again becomes lower. The proposed method increased the execution frequency of optimal action by forcefully increasing the Q-value through the Q-value update method. In the test, up to 60% higher goal achievement rate was shown at a maximum.

Keywords: Q-learning, Q-value, Learning Time

1 Introduction

Because reinforcement learning conducts learning without defining an environment model, it is still widely used in various areas. For example, it is used as a technique for controlling a virtual agent [1] in a virtual environment or controlling a robot [2]. Also, there is a study that applied it in designing a controller for UAV [3].

In Q-learning, one of reinforcement learning, an optimal action-selection method is learned to achieve a goal through recursive learning with unsupervised learning algorithm. Particularly, to learn optimal action execution policy, arbitrary actions are recursively selected and executed in the early stage of learning. Thus, the learning results vary even if the same agent learns in an identical environment. When a selected arbitrary action is an optimal action, the learning time of Q-learning is shortened but otherwise, a problem occurs that the learning time becomes longer.

In this paper, a method of updating the updated Q-value in the previous Q-value again is proposed to reduce the learning time of Q-learning. By updating the Q-value again, the updating can be performed in such a way that the goal can be achieved quickly. Because this method can be used by integrating with various studies related to Q-learning, it is expected that its application in various areas as well as UAV will be possible.

This paper is organized as follows. In Section 2, a method is proposed for reducing the learning time. In Section 3, the proposed method is implemented and the test

*Corresponding Author: Yunsick Sung (yunsick@kmu.ac.kr)

ISSN: 2287-1233 ASTL

results are introduced. Lastly, in Section 5, a conclusion is given for the proposed method.

2 Q-Value Update Method

When an agent learns with traditional Q-learning, learning is performed in a state where the Q-values corresponding to each action have been initialized in the early stage of learning. Because arbitrary actions are repeatedly executed, not only an optimal action but also non-optimal actions are executed repeatedly. Even if an action a_t performed in the state s_t is an optimal action, if the frequency of arbitrary action performed in the state s_{t-1} is high, the possibility of performing the action a_{t-1} again in the state s_{t-1} becomes lower in general. Despite of action a_{t-1} and action a_t being optimal actions, if the possibility of execution becomes low, a problem occurs that the learning time becomes longer.

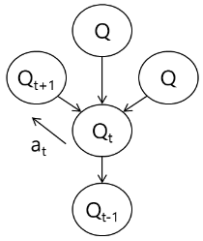


Fig. 1. Proposed Q-value update process

The traditional Q-learning updates only one Q-value, Q_t when performing an action a_t at time t. However, the proposed method updates the Q-value once more as shown in Equation (1) to solve the above problem. To perform the action a_{t-1} more in the state s_{t-1}, $Q(s_{t-1}, a_{t-1})$ is increased by assigning the product of $Q(s_t, a_t)$ and learning rate α. By increasing the Q-value, the execution frequency of a_{t-1} is increased.

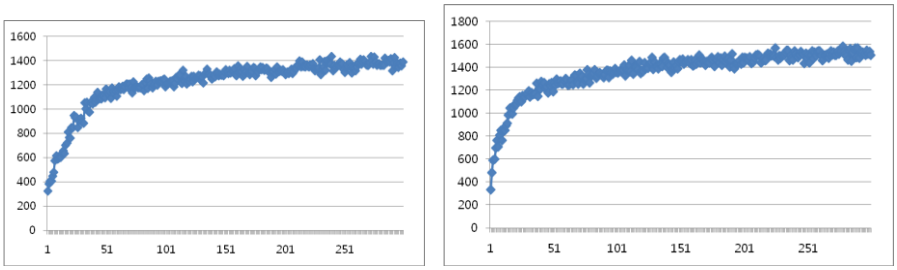
$$\begin{aligned} &IF\ \mathrm{Q}(s_{t-1}, a_{t-1}) < \mathrm{Q}(s_t, a_t) \times \alpha\ THEN \\ &\quad \mathrm{Q}(s_{t-1}, a_{t-1}) = \mathrm{Q}(s_t, a_t) \times \alpha \\ &END \end{aligned} \tag{1}$$

3 Experiment

In the experiment, an agent learned in multiple agent environments to reach an arbitrarily given position in each episode. If the agent collides with opponent agent while moving, the goal achievement fails. Therefore, the given position has to be reached while evading the opponent agent. The state is shown by the learning agent's position, opponent agent's position, and the position to be reached.

Fig. 2(a) shows the success frequency measured whenever episodes were learned 10,000 times in early stage of learning with the traditional Q-learning. The success

frequency increased sharply until learned 500,000 times and measured 50th time and then showed gradual increases. Fig. 2(b) shows the success frequency measured whenever episodes were learned 10,000 times with the proposed method. A sharp increase was shown up to about 25th measurement and then gradual increases were shown. It can be observed that a certain success frequency is reached faster compared to the traditional Q-learning. Therefore, it was found that the success frequency increased faster in the proposed Q-learning than the traditional Q-learning.



(a) Traditional Q-learning (b) Proposed Q-learning

Fig. 3. Success frequency of Q-learning

Fig. 4 compares the success frequency between the traditional Q-learning and the proposed Q-learning. Compared to the traditional Q-learning, the proposed Q-learning showed 60% more success frequencies at a maximum in the early stage of learning.

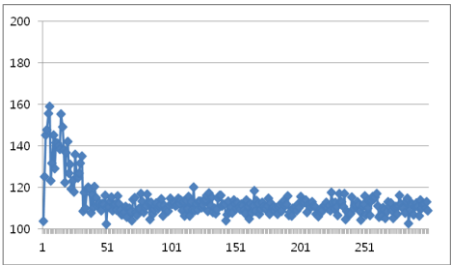


Fig. 4. The success rate proportion of proposed Q-learning vs. traditional Q-learning

5 Conclusion

In this paper, an update method of Q-value was proposed to increase the learning rate of Q-learning. The traditional Q-learning updates the Q-value once when an action is executed once. However, in the proposed method, the update was performed once more to repeatedly perform an optimal action. In the experiment, up to 60% higher result was shown in the early stage of learning.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1005955)

References

1. Sung, Y., Fong, S., Cho, S., Cho, K., Jeong, Y., Um, K.: Manipulated Motor-Primitive-Based Interactive Q-learning in Virtual Ubiquitous Computing Environments. The Second International Conference on Ubiquitous Context-Awareness and Wireless Sensor Network (UCAWSN-13), Lecture Notes in Electrical Engineering, 380-384, 15-17 July, 2013, Jeju Island, South Korea.
2. Sung, Y., Cho, S., Um, K., Jeong, Y., Fong, S. and Cho, K.: Human-Robot Interaction Learning Using Demonstration-Based Learning and Q-Learning in a Pervasive Sensing Environment. International Journal of Distributed Sensor Networks. 2013 (2013).
3. Bou-Ammar, H., Voos, H., Ertel, W.: Controller Design for Quadrotor UAVs using Reinforcement Learning, 2010 IEEE International Conference on Control Applications (CCA 2010), 2130-2135 (2010).

