ICIC Express Letters
Volume **14**, Number **8**, August **2020**

ICIC International ©2020 ISSN 1881-803X
pp. 799–805

A STUDY ON THE TRAFFIC SIGNAL CONTROL USING THE EXTENDED DEEP Q NETWORK

Dae Ho Kim, Ji Hye Kim and Ok Ran Jeong*

Department of Software
Gachon University
1342, Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do 13120, Korea
{ ikimdh91; kimjihae28 }@gc.gachon.ac.kr; *Corresponding author: orjeong@gachon.ac.kr

Received February 2020; accepted May 2020

Abstract. *Recently, problems such as traffic congestion and traffic accident rate are socially serious. In order to alleviate these problems, many researches on intelligent transportation systems are actively being carried out. Especially, traffic signal control method can mitigate serious traffic congestion problem through reinforcement learning model by using data such as traffic flow. Reinforcement learning algorithms evolved from deep reinforcement learning to various extensions and showed improved performance. In this paper, we applied an algorithm combining various extensions of deep reinforcement learning to traffic signal control research area. Our model learns to derive the optimal traffic signal system having the minimum traffic flow while adjusting the traffic signal length. We demonstrated that our algorithm has higher performance and faster learning speed than existing algorithms. Also, compared with the current traffic signal system, our algorithm shows the possibility of contributing to traffic alleviation when applied to real environment.*

Keywords: Deep Q network, Traffic signal control, Intelligent transportation system, Reinforcement learning

1. **Introduction.** Traffic congestion happens mostly at center of a city which has a high population density. In particular, this problem is severe at specific time zone like commute time. For relieving it, [1-5] about intelligent traffic system like a traffic prediction and traffic signal control are progressing actively. A research about traffic prediction is the technology to analyze the pattern of traffic variation based on traffic data observed in the past and to predict the traffic flow happening in the future. However, this research has no immediate means which can relieve the traffic congestion. Whereas the traffic signal control method can relieve it by controlling traffic signal interval based on the data like traffic flow observed in real time or queue length. It is difficult to deal with traffic variation changing every moment flexibly because current traffic signal which is operated on the road has static interval. Mostly traffic signal control researches consist of inputting traffic flow observed at an intersection through reinforcement learning and controlling the traffic signal interval. The goal is deduction of traffic signal system which has minimum traffic flow in an intersection by continuous learning. Figure 1 shows a process of the reinforcement learning in the traffic signal control field. Main elements of the reinforcement learning are environment (intersection), agent (model), state (traffic flow), action (duration of green light) and reward (+ or −). An agent takes an action based on the state observed from the environment in every learning. It sets a reward based on the action through learning. The reward is the criteria to judge whether the model is trained well or not. Since deep reinforcement learning is applied to the AlphaGo [6], the reinforcement learning algorithm is researched actively and several extended algorithms [7-13] are introduced. Most recently, [14] combined six extended algorithms of DQN

DOI: 10.24507/icicel.14.08.799

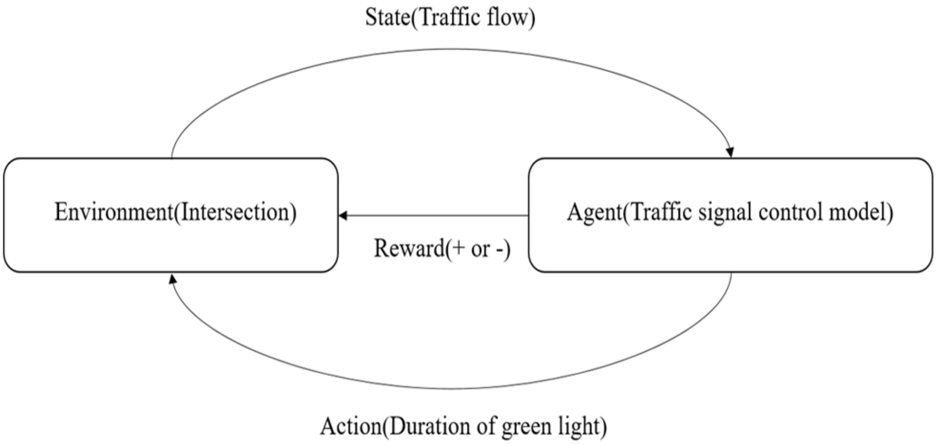
State(Traffic flow)
Environment(Intersection)
Agent(Traffic signal control model)
Reward(+ or -)
Action(Duration of green light)

Figure 1. Traffic signal control model based on reinforcement learning

which are introduced and prove a high performance compared with previous one at an Atari game environment.

In this paper, we apply an algorithm combining those to the traffic signal control research at the first and optimize it. We prove this algorithm has the fastest learning speed, highest performance, and stablest learning.

2. **Related Work.**

2.1. **Traditional traffic signal control method.**

Static traffic signal control method [15] exploits fixed signal interval. So it cannot reflect the variation of the traffic flow flexibly. Also it depends on human experience to maintain the system. For solving this problem, two methods are introduced. First, adaptive traffic signal control method [16-18] focuses on reflecting the variation of the traffic flow in the real environment by using diverse sensors. However, it is difficult to apply this system to real environment because of difficulty of installing a lot of sensor and sensor malfunction due to lights or weathers and so on. So researchers research reinforcement traffic signal control method [19]. It exploits Q learning without a model and a simulator like SUMO [20]. For that reason, it is easy to design and optimize a reward function. However, either applying it in the real world or dealing with many features is hard.

2.2. **Deep Q network based traffic signal control method.**

Deep Q network based traffic signal control method (DQN) [21,22] consists of a Q learning of reinforcement method and deep neural network to approximate an action value in the state with high dimension. There are two weaknesses in this method. It takes long learning time and is not easy to design a reward function. For overcoming these limitations, various methodologies extending DQN are proposed [7-13]: double Q-learning, dueling networks, prioritized relay, noisy nets, distributional reinforcement learning, and multi-step learning. The Rainbow [14] announced recently is a methodology combining DQN and those extended DQNs. Liang et al. [23] proposed the new traffic signal control method combining double DQN, dueling DQN, prioritized DQN and performed better than existing DQN-based traffic signal control studies.

In this paper, we propose new traffic signal control method applying the all extensions referenced above at the first.

3. **Reinforcement Learning.**

The main elements of reinforcement learning are agent, environment, state, behavior, reward, and policy. The components are shown in Table 1. Since the agent at the beginning of learning does not have any information, agent explores the environment with arbitrary action. If we select an action through the policy function

Table 1. Main elements of reinforcement learning

Term	Description
Agent	The object of learning that interacts with the environment.
Environment	A virtual world that changes depending on agent behavior.
State	The appearance of the environment observed by the agent. The state is generally expressed as s.
Action	A behavior that an agent can take in any state. The action is generally expressed as a.
Reward	A function that calculates a value that an agent can take after performing a certain action. The reward is generally expressed as r.
Policy	A function that determines what action the agent will take in a particular state. The policy is generally expressed as π.

π for which action we choose in a given state, the state of the environment changes, and the agent newly observes the changed state of the environment. The ultimate goal of reinforcement learning is to find the best policy that will maximize the sum of the rewards an agent gets from interacting with the environment. This can be expressed as follows,

$$Q^{\pi}(s,a) = E\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k}\middle| s_t = s, a_t = a, \pi\right] \tag{1}$$

The state in this paper is set as traffic flow, and traffic flow can be obtained from simulator. The reason for setting the state as traffic flow is that traffic flow is directly related to traffic congestion. The action is set to the duration of the green signal, and the agent increases or decreases the duration. Reward is set to scalar or distribution by comparing the current state with the next state after taking action. If the traffic flow at the next state is decreased, it is set to positive value, and if it is increased, it is set to a negative value.

4. DQN and Extensions.

4.1. DQN.

DQN [6] is a combination of existing Q-learning [5] based reinforcement learning method and deep neural network. Traditional reinforcement learning has a limitation in that it cannot be applied to real environment because the computational complexity increases when the state or action dimension is high. To solve this problem, DQN approximates high dimensional state or action space by applying deep neural network. The main features of DQN are experience replay and target network. Experienced replay stores state, action, reward, and next state for each training in replay memory in the form of one transition. And when the replay memory is full, it updates the neural network by sampling the transitions randomly. This method is more stable than existing Q-learning because it can eliminate the temporal correlation between the transitions in each training. The target network is a method of updating a neural network using two networks, a main network and a target network. In the conventional method, when the Q value in the current state and the Q value in the next state are calculated by one network, the DQN can reliably update the neural network using the two networks.

4.2. Extensions.

As the high performance of DQN has been proved, various extensions [7-13] of DQN have appeared. Double DQN is an algorithm that applies double Q-learning to DQN, which overestimates existing DQN. Dueling DQN is an algorithm that applies a dueling network to existing DQNs and divides $Q(s,a)$ into state value functions $V(s)$ and advantage function $A(s,a)$ and then merges them. The method of calculating the state-focused state value function $V(s)$ and action-focused advantage function $A(s,a)$ separately learns fast and stable. Prioritized experience replay DQN is a method of improving the experience replay method of existing DQN. Unlike DQN, which randomly extracts transitions from replay memory, the prioritized experience replay method updates neural networks by preferentially extracting transitions that require more learning. Noisy DQN is related to action policy that selects the best action for learning. Noisy DQN is a useful algorithm when training a state and an action in a noisy network and selecting a variety of actions when the dimension of the action is high. Noisy DQN is a useful algorithm for high-scale behavioral patterns by selecting various actions. Distributional DQN is a technique related to reward, which sets the reward in terms of a random variable. Rather than defining a future reward simply as a scalar, this technique is useful when the future reward value is complex and has multimodal features by expressing the reward in the form of a distribution.

In this paper, we propose a technique that combines all the extensions mentioned above into traffic signal control. With double DQN, we construct the separate target network from the main network and these two networks make updating the neural network stable. With dueling network, the Q function of the output layer was re-constructed to divide into state value function and advantage function, which enable fast learning speed. With prioritized experience replay DQN, priority calculation logic among transitions was added to existing replay memory, which establish the robust learning policies. With noisy DQN, we add noisy stream to the existing network. Over time, the network can learn to ignore the noisy stream, which establishes the best action policy in the noisy network. Finally, with distributional DQN, logic which transforms an existing scalar return value into a distribution is added and this distributional return value can be efficiently applied to complex environment with randomness feature.

5. Experiment.

5.1. Experiment setting.

To get the traffic data such as traffic flow, we use SALT simulator [24]. The simulator can obtain information like traffic flow, speed, density, queue length observed on the road but we consider only traffic flow. Traffic data from the simulator is the real traffic flow measured at an intersection in Seoul. We set the time from 00:00∼06:00 without congestion to 07:00∼10:00. We progress this experiment at Ubuntu for establishing the suitable environment to the simulator and implement the traffic signal control algorithm by Pytorch [25] which is a Python deep learning library.

5.2. Experiment methodology.

We compare both the state variation and cumulative reward value according to the training of each algorithm. The goal of this experiment is to find the best traffic signal system with minimum traffic flow at the fastest learning speed. In the first experiment, we demonstrate that how quickly our proposed model reaches the minimum traffic flow. In the second experiment, we demonstrate the high performance of our model by showing how much cumulative reward which is the most important measure at the reinforcement learning is.

5.3. Experiment results.

Figure 2 shows the comparison of state variation between algorithms. x-axis represents epoch and y-axis represents the state. Comparison targets are our proposed model, DDP DQN (Double Dueling Prioritized Experience Replay), DQN and current traffic signal system (Fixed 50, Fixed 45). The reason why DDP DQN and DQN were chosen as comparison models is to demonstrate that the proposed model is superior to existing DQN or DDP DQN based traffic signal control studies. To enhance the performance of existing DQN-based traffic signal control research, we properly combined various DQN extensions and finally proved fast learning speed and stable learning. We can see that our proposed model outperforms the other models. First, our model reaches the lowest traffic flow with a minimum training time compared to other algorithms, and the state change is stable even after reaching the minimum traffic flow. Early traffic flow

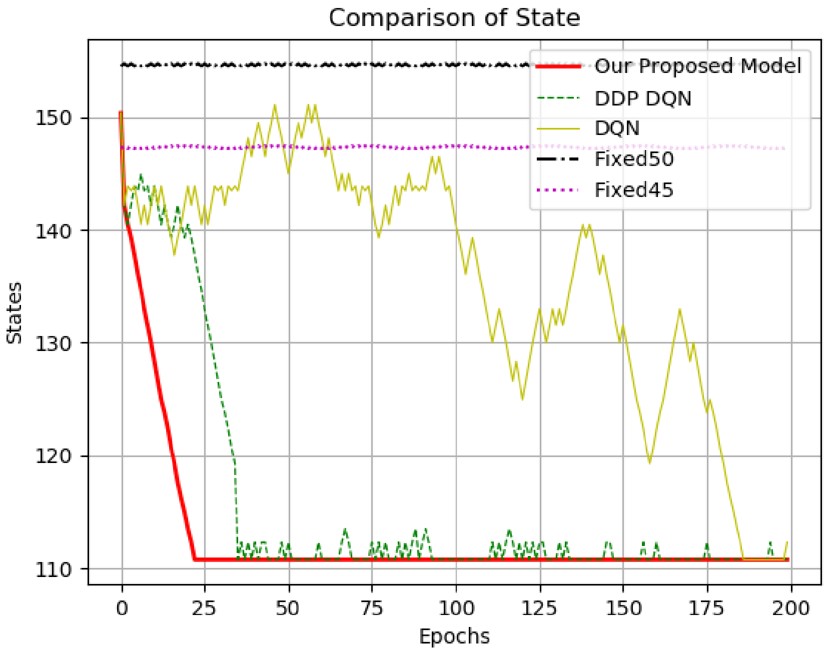

Figure 2. Comparison of state

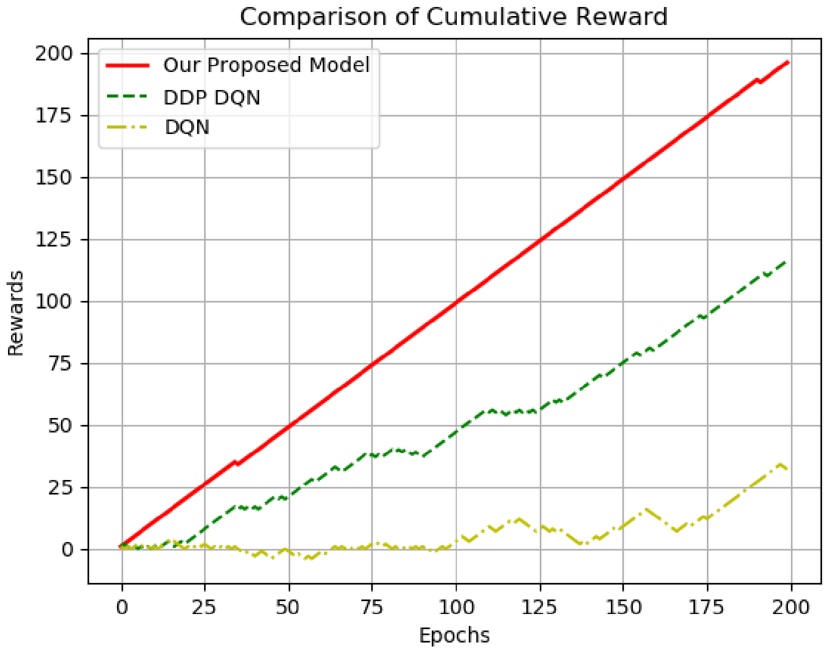

Figure 3. Comparison of cumulative reward

is about 150 vehicles but traffic flow is reduced up to about 110 vehicles at the end of the learning. And compared with current traffic signal system, we can estimate the value when the proposed model is applied to actual traffic environment.

Figure 3 compares cumulative reward between models. x-axis and y-axis represent epoch and cumulative reward. Comparison targets are our proposed model, DDP DQN, and DQN. The reason for comparing the cumulative reward is that the learning policy of the reinforcement learning learns to maximize the reward. In other words, the more reward is, the higher performance the model is. We can clearly confirm that the proposed model has a higher cumulative reward than other models, followed by DDP DQN and DQN.

6. **Conclusion.** Deep reinforcement learning is the most commonly used algorithm for traffic signal control research. As the performance of deep reinforcement learning has improved, various deep reinforcement learning extensions have emerged. In this paper, we first apply algorithms combined with various extensions to traffic signal control research. Through experiments, our model showed higher performance and faster learning speed than existing algorithms. And it shows the possibility of contributing to mitigation of traffic congestion problem in real environment by comparing with present traffic signal system. In the future, we will expand to multiple intersections and propose an efficient coordination method between intersections.

Acknowledgment. This research was supported by Basic Science Research Program through the NRF (National Research Foundation of Korea), and the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute for Information & communications Technology Promotion), and Copyright Technology R&D through the MCST (Ministry of Culture, Sports and Tourism), Korea and the KCC (Korea Copyright Commission) (Nos. NRF 2019R1A2C1008412, 2015-0-00932, 2019-CONTEXT-9500).

REFERENCES

[1] X. Feng, X. Ling, H. Zheng, Z. Chen and Y. Xu, Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction, *IEEE Trans. Intelligent Transportation Systems*, vol.20, no.6, pp.2001-2013, 2019.

[2] Y. Tian et al., LSTM-based traffic flow prediction with missing data, *Neurocomputing*, vol.318, pp.297-305, 2018.

[3] Y. Wu, H. Tan, L. Qin, B. Ran and Z. Jiang, A hybrid deep learning based traffic flow prediction method and its understanding, *Transportation Research, Part C: Emerging Technologies*, vol.90, pp.166-180, 2018.

[4] L. Shen, R. Liu, Z. Yao, W. Wu and H. Yang, Development of dynamic platoon dispersion models for predictive traffic signal control, *IEEE Trans. Intelligent Transportation Systems*, vol.20, no.2, pp.431-440, 2019.

[5] G. Lafferriere, *Mitigating Automobile Congestion through Urban Traffic Signal Control*, 2019.

[6] V. Mnih et al., Playing Atari with deep reinforcement learning, *arXiv Preprint, arXiv:1312.5602*, 2013.

[7] H. van Hasselt, A. Guez and D. Silver, Deep reinforcement learning with double Q-learning, *The 30th AAAI Conference on Artificial Intelligence*, 2016.

[8] Z. Wang et al., Dueling network architectures for deep reinforcement learning, *arXiv Preprint, arXiv:1511.06581*, 2015.

[9] T. Schaul, J. Quan, I. Antonoglou and D. Silver, Prioritized experience replay, *Proc. of ICLR*, 2015.

[10] M. Fortunato, M. G. Azar, B. Piot, J. Menick, I. Osband, A. Graves, V. Mnih, R. Munos, D. Hassabis, O. Pietquin, C. Blundell and S. Legg, Noisy networks for exploration, *arXiv Preprint, arXiv:1706.10295*, 2017.

[11] M. G. Bellemare, W. Dabney and R. Munos, A distributional perspective on reinforcement learning, *ICML*, 2017.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, MA, 1998.

[13] R. S. Sutton, Learning to predict by the methods of temporal differences, *Machine Learning*, vol.3, no.1, pp.9-44, 1988.

[14] M. Hessel, J. Modayil, H. van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar and D. Silver, Rainbow: Combining improvements in deep reinforcement learning, *arXiv Preprint, arXiv:1710.02298*, 2017.

[15] D. Zhao, Y. Dai and Z. Zhang, Computational intelligence in urban traffic signal control: A survey, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.42, no.4, pp.485-494, 2012.

[16] L. Y. Deng, N. C. Tang, D. Lee and C. T. Wang, Vision based adaptive traffic signal control system development, *The 19th International Conference on Advanced Information Networking and Applications (AINA'05)*, 2005.

[17] C. S. Jhaveri, J. Perring and P. Martin, *SCOOT Adaptive Signal Control: An Evaluation of Its Effectiveness over a Range of Congestion Intensities*, The Transportation Research Board, 2003.

[18] A. Arif and R. Rupali, Image processing based adaptive traffic control system, *IOSR Journal of Electronics and Communication Engineering*, 2013.

[19] C. J. C. H. Watkins and P. Dayan, Q-learning, *Machine Learning*, vol.8, nos.3-4, pp.279-292, 1992.

[20] D. Krajzewicz et al., Recent development and applications of SUMO – Simulation of Urban MObility, *International Journal on Advances in Systems and Measurements*, vol.5, no.3, pp.128-138, 2012.

[21] M. Bowling and M. Veloso, *An Analysis of Stochastic Game Theory for Multiagent Reinforcement Learning*, apps.dtic.mil, 2000.

[22] J. Gao, Y. Shen, J. Liu, M. lto and N. Shiratori, Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network, *arXiv Preprint, arXiv:1705.02755*, 2017.

[23] X. Liang et al., A deep reinforcement learning network for traffic light cycle control, *IEEE Trans. Vehicular Technology*, vol.68, no.2, pp.1243-1253, 2019.

[24] H. Song and O. Min, Statistical traffic generation methods for urban traffic simulation, *International Conference on Advanced Communications Technology (ICACT)*, 2018.

[25] A. Paszke et al., *Automatic Differentiation in Pytorch*, 2017.