ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 19-04-38
266
j.inst.Korean.electr.electron.eng.Vol.23,No.4,1381~1386,December 2019
http://dx.doi.org/10.7471/ikeee.2019.23.4.1381

다중 교차로에서 협력적 교통신호제어에 대한 연구

A Study on Cooperative Traffic Signal Control at multi-intersection

김 대 호*, 정 옥 란*★

Dae Ho Kim*, Ok Ran Jeong*★

Abstract

As traffic congestion in cities becomes more serious, intelligent traffic control is actively being researched. Reinforcement learning is the most actively used algorithm for traffic signal control, and recently Deep reinforcement learning has attracted attention of researchers. Extended versions of deep reinforcement learning have been emerged as deep reinforcement learning algorithm showed high performance in various fields. However, most of the existing traffic signal control were studied in a single intersection environment, and there is a limitation that the method at a single intersection does not consider the traffic conditions of the entire city. In this paper, we propose a cooperative traffic control at multi-intersection environment. The traffic signal control algorithm is based on a combination of extended versions of deep reinforcement learning and we considers traffic conditions of adjacent intersections. In the experiment, we compare the proposed algorithm with the existing deep reinforcement learning algorithm, and further demonstrate the high performance of our model with and without cooperative method.

요 약

도시의 교통 혼잡 문제가 심각해지면서 지능형 교통신호제어가 활발하게 연구되고 있다. 강화학습은 교통신호제어에 가장 활발하게 사용되고 있는 알고리즘으로 최근에는 심층 강화학습 알고리즘이 관심을 끌고 있다. 또한 심층 강화학습 알고리즘이 다양한 분야에서 높은 성능을 보이면서 심층 강화학습의 확장 버전들이 빠른 속도로 등장했다. 하지만 기존 교통신호제어 연구들은 대부분 단일 교차로 환경에서 진행되었으며, 단일 교차로의 교통 혼잡만 완화하는 방법은 도시 전체의 교통 상황을 고려하지 못한다는 한계가 있다. 본 논문에서는 다중 교차로 환경에서 협력적 교통신호제어를 제안한다. 신호제어 알고리즘에는 심층 강화학습의 확장 버전들이 결합된 알고리즘을 적용했으며 다중 교차로를 효율적으로 제어하기 위해 인접한 교차로의 교통 상황을 고려하였다. 실험에서는 제안하는 알고리즘과 기존 심층 강화학습 알고리즘을 비교하였으며, 더 나아가 협력적 방법이 적용된 모델과 적용되지 않은 모델의 실험 결과를 보여줌으로써 높은 성능을 증명한다.

Key words : Traffic signal control, Deep reinforcement learning, Deep Q Network, Coordination, Multi-intersection traffic signal control

* Dept. of Software, Gachon University
★ Corresponding author
E-mail : orjeong@gachon.ac.kr, Tel : +82-31-750-5831

※ 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) (Nos.NRF 2019R1A2C1008412, 2015-0-00932).
Manuscript received Dec. 5, 2019; revised Dec. 26, 2019; accepted Dec. 29, 2019.

Ⅰ. 서론

최근에, 도심에서 교통 혼잡 문제가 심각하다.

도로의 용량은 제한되어 있고, 현재 교통신호체계는 고정형 교통신호기 때문에 늘어나는 교통량의 변화에 유연하게 대처하지 못 한다[1]. 이러한 문제를 해결하기 위해 적응형 교통신호제어에 대한 연구가 활발하게 진행되고 있다[2]-[5].

적응형 교통신호제어는 교통신호를 동적으로 전환하거나 녹색신호의 길이를 조절한다. 예를 들어, 교차로에서 각 차선의 교통량이 관측되었으면 적응형 교통신호제어는 교통량이 많은 차선에 녹색신호를 할당함으로써 교통 혼잡을 완화한다.

강화학습은 교통신호제어에 가장 많이 사용되는 알고리즘으로, 알파고(AlphaGo)의 등장 이후로 활발하게 연구되고 있다[6]-[7]. 강화학습은 관측 가능한 환경(Environment)에서 상태(State)가 주어지면, 상태를 기반으로 행동(Action)을 선택하고, 선택한 행동에 대한 보상(Reward)을 받는다. 강화학습 모델은 궁극적으로 보상을 최대화하는 정책을 찾도록 훈련 된다[6].

최근에는 강화학습과 심층 신경망이 결합된 심층 강화학습 기반 교통신호제어가 활발하게 연구되고 있다[8]-[10]. 심층 강화학습은[11] 바둑이나 테트리스와 같은 충분히 관찰할 수 있는 환경에서 연구가 시작되었고 교통신호제어에도 적용되어 높은 성능을 보였다. 하지만 기존의 심층 강화학습 기반 교통신호제어 연구들은 대부분 단일 교차로 환경에서 진행되었기 때문에 실제 도로 규모에 적용하기 힘들다. 또한 다중 교차로를 효율적으로 제어하는 교통신호제어 연구들이 매우 부족하다.

본 논문에서는 다중 교차로에서 협력적 교통신호제어를 제안한다. 교통신호제어 알고리즘으로 심층 강화학습의 여러 가지 확장 버전들을 결합하여 성능을 높였으며, 다중 교차로를 효율적으로 제어하기 위해 인접한 교차로의 교통 상황을 고려하는 협력적 방법을 제안한다.

Ⅱ. 관련 연구

현재 운용되고 있는 교통신호체계는 신호 길이가 고정되어 있는 정적 교통신호체계에 속 한다[1]. 도심의 경우, 보통 출퇴근 시간에 교통이 혼잡하고 새벽시간에 한적한 경향을 보인다[12-13]. 정적 교통신호체계는 교통량이 갑자기 증가하는 시간대에도 이전과 동일한 교통신호 길이를 유지하기 때문에 교통 변화에 유연하게 대처하지 못한다는 한계가 있다[12]. 적응형 교통신호제어는 관측되는 교통량을 기반으로 신호를 동적으로 변환하기 때문에 교통 혼잡 문제를 해결할 수 있다[2-5].

강화학습은 적응형 교통신호제어 연구에 가장 많이 사용된 대표적인 알고리즘이다. 강화학습은 지도학습과 같이 학습에 정답이 표기되어 있는 데이터 셋이 필요하지 않고, 주어진 환경에서 스스로 탐험하며 학습한다는 장점이 있다[6]. 이러한 장점으로, 강화학습 기반 적응형 교통신호제어에 대한 연구가 활발하게 진행되었다[2-5]. 강화학습은 기본적으로 Q러닝을 기반으로 학습하며, Q러닝은 매 학습마다 발생한 상태와 행동을 Q테이블 형태에 저장한다[6]. 하지만 상태 또는 행동의 가지 수가 많아지면 Q테이블의 크기 또한 커지기 때문에, 계산 복잡도가 매우 높아진다는 한계가 있다[11].

심층 강화학습은 기존 강화학습과 심층 신경망이 결합된 알고리즘으로, 높은 차원의 상태 또는 행동을 심층 신경망을 통해 근사시킴으로써 기존 강화학습의 한계를 해결했다[11]. 심층 강화학습 기반 교통신호제어 연구들은[8-10] 대부분 시뮬레이터의 스냅샷에서 CNN(Convolutional Neural Network)을 사용하여 차량의 위치 혹은 속도 정보를 얻는 방법을 통해 높은 차원의 상태 데이터를 근사 할 수 있다. 하지만 대부분의 교통신호제어 연구들은 단일 교차로에서 진행되었고, 단일 교차로에서 교통신호제어는 제어할 수 있는 도로의 규모가 너무 작기 때문에 실제 환경에 적용하기 힘들다는 한계가 있다[14].

본 논문에서는 다중 교차로에서의 협력적 교통신호제어를 제안한다. 각 교차로는 인접한 교차로에서 들어오는 교통량에 영향을 받기 때문에, 각 교차로에서 관측된 상태를 인접한 교차로에 전달하는 방법을 소개한다. 각 교차로는 개별적으로 최적의 교통신호체계를 찾도록 학습되어지는 동시에 인접한 교차로를 고려함으로써 전체적으로 교통 혼잡을 완화할 수 있다.

Ⅲ. 다중 교차로에서 협력적 교통신호제어

이번 섹션에서는 제안하는 교통신호제어 모델에

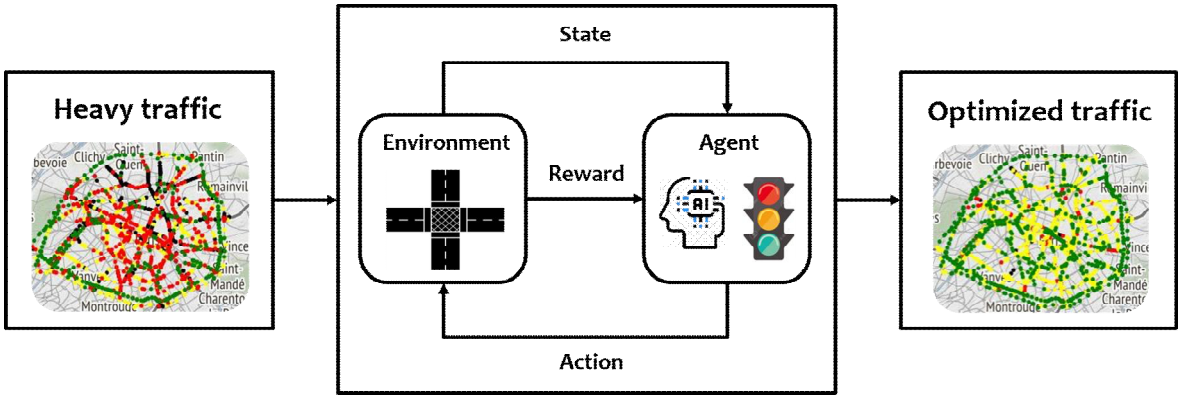


Fig. 1. Deep reinforcement learning based traffic signal control.
그림 1. 심층 강화학습 기반 교통신호제어

적용된 강화학습 알고리즘을 소개한다. 또한 제안하는 알고리즘이 어떻게 교통신호를 제어 하는지 소개한다. 마지막으로 교차로 간에 교통 상황을 공유함으로써 다중 교차로에서 효율적인 협력 방법을 소개한다.

1. 심층 강화학습 기반 교통신호제어 알고리즘

그림 1은 교통신호제어에서 강화학습의 학습 방식을 보여준다. 강화학습은 관측 가능한 환경에서 상태를 받고 행동을 취하면서 받는 보상을 최대화하도록 학습된다. 강화학습은 보통 행동-가치함수(Q-function)로 표현되며 다음 수식 (1)과 같다.

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

s는 상태, a는 행동, r은 보상, E는 기대 값, γ는 할인 계수, π는 정책을 의미한다. 즉, 상태와 선택된 행동을 통해 미래에 받을 수 있는 기대되는 보상을 예측한다.

우리는 도로 환경에서 관측되는 교통량을 상태, 교통신호의 전환을 행동, 현재 차량들의 평균 대기시간에서 이전 차량들의 평균 대기시간을 뺀 값을 보상으로 정의하였다. 즉, 현재 차량들의 대기시간이 이전보다 감소하였으면 보상 값이 양수고 반대 경우에는 음수로 주어진다. 강화학습은 보상을 최대화하는 방향으로 학습하기 때문에 보상을 어떻게 정의하는지가 매우 중요하다. 보상 함수는 다음 수식 (2) 와 같다.

$$R_{waiting\_time} = W_t - W_{t+1} \quad (2)$$

대표적인 심층 강화학습 알고리즘은 DQN (Deep Q Network)이 있다[11]. DQN은 강화학습에 심층 신경망이 결합된 알고리즘으로 주요 특징은 경험 리플레이와 타겟 네트워크가 있다. 경험 리플레이는 학습 시 발생한 상태, 행동, 보상, 다음 상태를 리플레이 메모리에 저장해두었다가 일정 기간 이후에 저장된 샘플을 추출하여 신경망을 학습시킨다. 타겟 네트워크는 메인 네트워크와 별도로 구축하여 안정적인 학습을 가능하게 한다. DQN이 다양한 분야에서 높은 성능을 보이면서 DQN의 여러 가지 확장 버전들[15-20]이 등장했으며, 우리는 교통신호제어에 DQN의 여러 가지 확장버전들을 결합한 알고리즘을 교통신호제어에 적용했다.

Double DQN[15]은 DQN에 double Q 러닝을 결합한 방법으로 행동의 선택을 평가에서 분리함으로써 기존 DQN의 과대추상의 문제를 해결하였다. Dueling DQN[16]은 신경망의 구조를 상태-가치 함수(state-value function)와 어드밴티지 함수(advantage function)로 분할한다. 분할 된 어드밴티지 함수가 오직 행동에만 집중할 수 있기 때문에 학습 속도가 빠르다. Prioritized experience replay DQN[17]은 기존 DQN의 경험 리플레이가 학습 시 샘플을 랜덤하게 추출하는 것을 개선한 방법으로, 학습이 더 필요한 샘플에 우선순위를 높게 매겨 학습 데이터의 효율성을 증가시켰다. Multi-step learning[18]은 매 학습마다 발생한 즉각적인 보상 함수를 계산하는 것 대신, n번의 학습 이후에 보상 함수를 계산함으로써 모델의 손실 값(loss)을 최소화 할 수 있다. Distributional-RL[19]은 기존 스칼라(scala) 형태의 보상 값 대신, 보상의 확률 분포를 학습하도록 해서 보상 함수가 복잡할 때 유용하다. 마지막으로 NoisyNet[20]은 기존 신경망에 잡음(Noise)을 추가한다. 잡음이 추

가된 신경망은 시간이 지남에 따라 잡음을 무시하면서 학습을 진행하게 되고, 행동의 차원이 높을 때 효율적이다. 수식 (3)은 앞서 설명한 DQN의 확장 버전들이 결합된 제안하는 알고리즘의 수식이다.

$$Q^i_{t+1}(s^i_t,a^i_t) = Q^i_t(s^i_t,a^i_t;\theta_i) + \eta * (R_{t+1} + \gamma \max Q^i_t(s'_{t+1,}a';\theta'_i) - Q^i_t(s^i_t,a^i_t;\theta_i)) \quad (3)$$

θ_i는 평가 신경망의 파라미터, θ'_i는 타겟 신경망의 파라미터를 의미한다. 수식 (3)은 다음 시간(t+1)의 보상을 예측한다.

2. 협력적 교통신호제어

다중 교차로 환경에서 각 교차로는 하나의 강화학습 에이전트로 할당된다. 모든 에이전트는 자신이 할당된 교차로에서 관측되는 교통상황을 기반으로 최적의 행동을 계산한다. 하지만 다중 교차로 환경에서 각 교차로는 인접한 교차로의 교통 상황에 영향을 받기 때문에, 독립적인 학습은 비효율적이다. 따라서 인접한 교차로의 교통 상황을 전달하는 협력적 방법을 소개한다. 그림 2는 다중 교차로 환경에서 각 교차로 간 통신하는 그림을 보여준다. t는 현재시간이고 t-1은 이전 시간을 의미한다. 즉, 인접한 교차로에서 이전 시간대에 관측된 상태를 전달함으로써 현재 시간대에 최적의 행동을 선택할 때 적용한다. 수식 (4)는 제안하는 알고리즘의 최종 수식을 보여준다.

$$Q^i_{t+1}(s^i_t,a^i_t) = Q^i_t(s^i_t,a^i_t;\theta_i) + \eta * (R_{t+1} + \gamma \max Q^i_t(s'_{t+1,}a';\theta'_i) - Q^i_t(s^i_t,a^i_t;\theta_i)) + \sum_{j\in N_{adj}} Q^i_{t-1}(s^j_{t-1},a^j_{t-1};\theta_j) \quad (4)$$

N_{adj}은 인접한 교차로의 수를 의미하고, θ_j는 인접 교차로 신경망의 파라미터다. 이전 수식에서 추가된 부분은 이전 시간대에 발생한 인접한 교차로의 Q함수를 공유하는 부분이다.

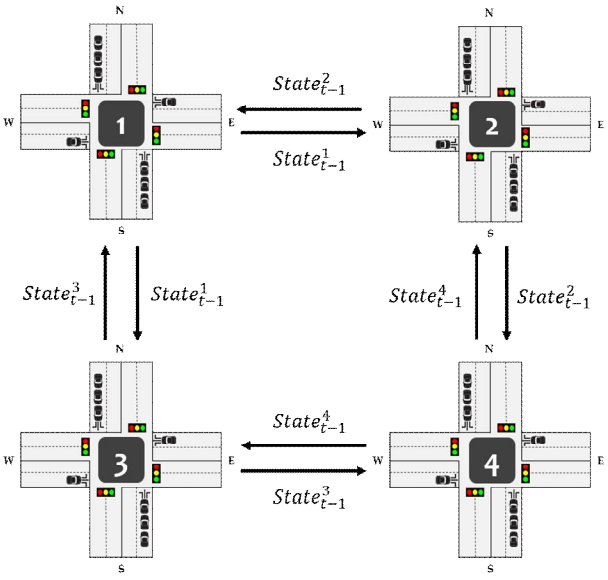


Fig. 2. Cooperative method at multi-intersection.
그림 2. 다중 교차로에서 협력적 방법

Ⅵ. 실험

이번 섹션에서는 제안하는 모델의 성능이 얼마나 높은지 증명하기 위해, 실험 환경 및 실험 방법에 대해 설명하고 실험 결과를 보여준다.

1. 실험 환경 및 방법

실제 환경과 유사한 도로 환경을 구성하기 위해, 교통 시뮬레이터로 가장 많이 사용되는 SUMO 시뮬레이터[21]를 사용했다. SUMO 시뮬레이터를 사용해서 16개의 다중 교차로 환경을 구성 했으며, 각 교차로의 모든 차선에서 발생한 교통량, 대기시간 등을 계산할 수 있다. 또한 학습된 신경망으로부터 선택된 행동으로 각 교차로의 교통신호를 제어할 수 있으며, 선택된 행동에 의해 교통신호가 전환되고 난 뒤 차량들의 평균 대기시간을 통해 보상 함수를 계산한다. 제안하는 모델은 보상을 최대화하는 최적의 학습 정책을 찾도록 학습된다. 또한 각 교차로는 이전 시간대의 인접한 모든 교차로의

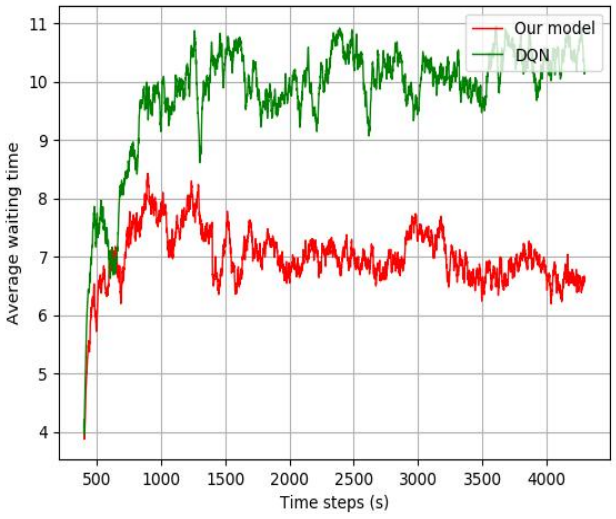


Fig. 3. Comparative analysis of average waiting time.
그림 3. 평균 대기시간 비교 실험

교통 상황을 전달 받는다.

실험은 크게 2가지로 진행되었다. 첫 번째 실험에서는 심층 강화학습의 확장 버전들이 결합된 제안하는 알고리즘과 기존 심층 강화학습인 DQN의 평균 대기시간을 비교하였다. 강화학습은 보상을 최대화하는 방향으로 학습을 진행하고, 본 논문에서의 보상은 대기 시간이기 때문에 평균 대기시간을 비교하였다. 두 번째 실험에서는 다중 교차로에 협력적 방법을 적용한 모델과 적용하지 않은 모델의 누적 보상을 비교함으로써 효율적인 다중 교차로 제어 성능을 증명한다.

2. 실험 결과

가. 평균 대기시간 비교

그림 3은 제안하는 모델과 기존 심층 강화학습 모델의 평균 대기시간을 비교한 결과이다. 학습 초기에는 평균 대기시간이 비슷한 경향을 보이지만, 학습이 진행되면서 차이가 확연하게 드러난다. 학습이 끝날 시점에 제안하는 모델의 평균 대기시간이 기존 심층 강화학습 모델보다 약 2배 정도 짧았다. 제안하는 모델의 평균 대기시간이 더 짧다는 것은 보상 함수의 설계가 적절했으며, 학습이 잘 되었다는 것을 의미한다.

나. 누적 보상 비교

그림 4는 다중 교차로에 협력적 방법이 적용된 모델과 적용되지 않은 모델의 누적 보상 값을 비교한다. 강화학습에서는 보상 값이 클수록 학습이 잘 되었다는 것을 의미한다. 학습이 진행되면서, 전체적으로 제안하는 모델의 누적 보상이 크게 나타난다. 즉, 본 논문에서 제안한 협력적 방법이 다중 교차로 교통신호제어에 효과적임을 뒷받침 할 수 있다.

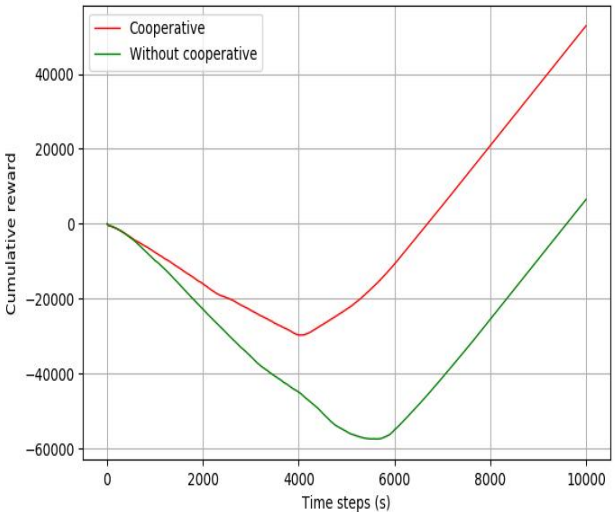


Fig. 4. Comparative analysis of cumulative reward.
그림 4. 누적 보상 비교 실험

V. 결론

본 논문에서는 협력적 교통신호를 제안하기 위해, 심층 강화학습의 여러 가지 확장버전들을 결합한 알고리즘을 교통신호제어에 적용했다. 실험을 통해 제안하는 알고리즘이 기존 심층 강화학습 알고리즘에 비해 높은 성능을 보이는 것을 증명하였다. 또한 다중 교차로를 효율적으로 제어하기 위해 협력적 방법을 제안한다. 각 교차로가 독립적으로 학습하는 모델에 비해, 다중 교차로 간 교통 정보를 공유함으로써 높은 성능을 보였다. 현재는 16개의 다중 교차로 환경에 적용되었지만, 앞으로 더 넓은 지역을 커버하는 도시 레벨의 다중 교차로 교통신호제어 연구로의 확장 가능성을 보였다.

References

[1] J. H. Youn, Y. G. Kang, “Simulation of Traffic Signal Control with Adaptive Priority Order through Object Extraction in Images,” *Journal of Korea Multimedia Society,* Vol.11, No.8 pp.1043-1193, 2008.

[2] Cai, Chen, Chi Kwong Wong, and Benjamin G. Heydecker. “Adaptive traffic signal control using approximate dynamic programming,” *Transportation Research Part C: Emerging Technologies,* Vol.17, No.5 pp.456-474, 2009. DOI: 10.1016/j.trc.2009.04.005

[3] Pandit, Kartik, et al. “Adaptive traffic signal control with vehicular ad hoc network,” *IEEE Transactions on Vehicular Technolog, Vol.*62, No.4 pp.1459-1471, 2013.
DOI: 10.1109/TVT.2013.2241460

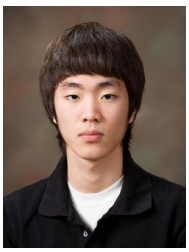
[4] Maslekar, Nitin, et al. “VANET based adaptive traffic signal control,” *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring).* IEEE, 2011. DOI: 10.1109/VETECS.2011.5956305

[5] Mannion, Patrick, Jim Duggan, and Enda

Howley. “An experimental review of reinforcement learning algorithms for adaptive traffic signal control,” *Autonomic Road Transport Support Systems*. Birkhäuser, Cham, pp.47-66, 2016.
DOI: 10.1007/978-3-319-25808-9_4
[6] Sutton, Richard S., and Andrew G. Barto. *Introduction to reinforcement learning*. Vol.2. No.4. Cambridge: MIT press, 1998.
[7] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. “Reinforcement learning: A survey,” *Journal of artificial intelligence research* 4, pp.237-285, 1996.
[8] Li, Li, Yisheng Lv, and Fei-Yue Wang. “Traffic signal timing via deep reinforcement learning,” *IEEE/CAA Journal of Automatica Sinica,* Vol.3, No.3 247-254, 2016. DOI: 10.1109/JAS.2016.7508798
[9] Genders, Wade, and Saiedeh Razavi. “Using a deep reinforcement learning agent for traffic signal control,” *arXiv preprint arXiv:1611.01142,* 2016.
[10] Gao, Juntao, et al. “Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network,” *arXiv preprint arXiv:1705.02755,* 2017.
[11] Mnih, Volodymyr, et al. “Human-level control through deep reinforcement learning,” *Nature* 518, 7540 pp.529-533, 2015. DOI: 10.1038/nature14236
[12] YoungTae Jo, JinSup Choi, and InBum Jung, “Intersection Traffic Signal Control based on Traffic Pattern Learning for Repetitive Traffic Congestion,” *Journal of Computing Science and Engineering,* Vol.20, No.8, pp.450-465, 2014.
[13] Kwang-baek Kim, “Intelligent Traffic Light Control using Fuzzy Method,” *Journal of The Korea Society of Computer and Information,* Vol.15, No.9, pp.19-24. 2010.
DOI: 10.6109/jkiice.2012.16.8.1593
[14] Chen, Ruey-Shun, Duen-Kai Chen, and Szu-Yin Lin. “ACTAM: Cooperative multi-agent system architecture for urban traffic signal control,” *IEICE transactions on Information and Systems* Vol.88, No.1, pp.119-126, 2005.
[15] Van Hasselt, Hado, Arthur Guez, and David Silver. “Deep reinforcement learning with double q-learning,” *Thirtieth AAAI conference on artificial intelligence*. 2016.
[16] Wang, Ziyu, et al. “Dueling network architectures for deep reinforcement learning,” *arXiv preprint arXiv:1511.06581,* 2015.
[17] Schaul, Tom, et al. “Prioritized experience replay,” *arXiv preprint arXiv:1511.05952* (2015).
[18] Sutton, Richard S. “Learning to predict by the methods of temporal differences,” *Machine learning* Vol.3, No.1, pp.9-44, 1988.
DOI: 10.1007/BF00115009
[19] Bellemare, Marc G., Will Dabney, and Rémi Munos. “A distributional perspective on reinforcement learning,” *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017.
[20] Fortunato, Meire, et al. “Noisy networks for exploration.” *arXiv preprint arXiv:1706.10295,* 2017.
[21] Krajzewicz, Daniel, et al. “SUMO (Simulation of Urban MObility)-an open-source traffic simulation,” *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002),* 2002.

BIOGRAPHY

Dae Ho Kim (Member)



2017 : BS degree in oftware, Gachon University.
2019 : MS degree in Software, Gachon University.
2019~present : PhD student in Software, Gachon University

Ok Ran Jeong (Member)

2005 : PhD degree in omputer Science and Engineering, Ehwa Womans University.
2006 : Postdoctoral Researcher, Seoul National University.
2007 : Postdoctoral Researcher, Univ. of Illinois at Urbana-Champaign
2008~2009 : Research Professor, Sunkyunkwan University.
2009~2019 : Associate Professor, Gachon University.