Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 43, No. 2, 143-153, 2025
https://doi.org/10.7848/ksgpc.2025.43.2.143

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

현실 제약 조건을 반영한 다중 교차로에서의 협력적 강화학습 기반 교통 신호 제어

Multi Intersection Traffic Signal Control based on Cooperative Reinforcement Learning reflecting Real-World Constraints

박진혁[1] · 김혜민[2] · 이준우[3] · 전철민[4]
Park, Jin Hyuk · Kim, Hye Min · Lee, Joon Woo · Jun, Chulmin

Abstract

As extreme weather events such as heatwaves and heavy rainfall caused by climate change increase, the importance of reducing greenhouse gas emissions has grown. This study proposes a cooperative reinforcement learning-based traffic signal control model at multi intersection to alleviate traffic congestion, one of the causes of greenhouse gas emissions in the transportation sector. The proposed model configures the state of the model and the Q-Network so that the agent in reinforcement learning considers the state and actions of its neighbors. Additionally, real-world constraints such as signal sequence and minimum green time are included to enhance real-world applicability. For validation, the proposed model was compared with a general reinforcement learning model and a model without constraints. As a result, the model with added multi-agent reinforcement learning was shown to be more efficient in signal control at multi intersection, as it showed less vehicle waiting time and CO_2 emission compared to the general reinforcement learning model. Although the model without constraints showed relatively less vehicle waiting time and CO_2 emission than the proposed model, but showed a large number of vehicle stops.

Keywords: Traffic signal control, Cooperative reinforcement learning, Multi intersection, Real-World constraints, Intelligent transportaion system

초 록

최근 기후변화로 인한 폭염, 폭우 등의 극한 현상이 증가함에 따라 온실가스 감축의 중요성이 높아지고 있다. 본 연구는 수송 부문 온실가스 배출의 원인 중 하나인 교통체증 완화를 목적으로 다중 교차로에서의 협력적 강화학습 기반 교통 신호 제어 모델을 제안한다. 제안 모델은 강화학습 내 에이전트가 이웃의 상태 및 행동을 고려하도록 모델의 상태와 Q-Network를 구성하였다. 또한 신호 순서, 최소 녹색시간 등의 현실 제약 조건을 추가하여 현실 적용성을 높였다. 본 제안 모델의 검증을 위해 일반 강화학습 모델, 제약 조건 해제 모델과의 비교를 진행하였다. 결과적으로 다중 에이전트 강화학습을 추가한 모델이 일반 강화학습 모델과 비교하여 차량 대기 시간, CO_2 배출량 등이 적게 나타나, 다중 교차로에서의 신호 제어에 더 효율적인 것으로 나타났다. 현실 제약을 제거한 모델은 제안 모델보다 차량 대기 시간, CO_2 배출량이 비교적 적게 나타났지만 차량 정지 횟수가 크게 나타났다.

핵심용어: 교통 신호 제어, 협력적 강화학습, 다중 교차로, 현실 제약 조건

Received 2025.01.08, Revised 2025.02.25, Accepted 2025.03.05
1) Department of Geoinformatics, University of seoul (E-mail:jk0518@uos.ac.kr)
2) Department of Geoinformatics, University of seoul (E-mail:kimhm77@uos.ac.kr)
3) Department of Geoinformatics, University of seoul (E-mail:leejoon924@uos.ac.kr)
4) Corresponding Author, Member, Department of Geoinformatics, University of seoul (E-mail:cmjun@uos.ac.kr)

1. 서 론

최근 기후변화의 영향으로 폭염과 폭우, 가뭄 등의 극한 현상이 증가하고 있다(IPCC, 2023). 만약 온실가스를 현재와 동일한 양으로 배출한다면, 2040년에 지구온난화로 인한 평균 기온 상승이 1.5℃에 도달하고 기후변화로 인한 폭염은 8.6배, 집중호우는 1.5배, 가뭄은 2배 증가할 것으로 예상된다(IPCC, 2021). 이러한 기후변화 문제에 대응하기 위해, 국제 사회에서는 파리협정을 체결하고 한국에서는 2050 탄소중립을 선언하고, 국가 탄소중립·녹색성장 기본 계획을 수립하는 등의 대책을 마련하고 있는 실정이다(Park, 2016, Yi and Kang, 2022, Park *et al*., 2024). 기후변화 대응에 대한 전세계적인 관심이 높아짐에 따라 수송 부문 온실가스 감축 또한 중요한 과제로 대두되고 있다. 2024 GHG EMISSIONS OF ALL WORLD COUNTRIES에 따르면 2023년 한국의 온실가스 배출량은 653.846 Mt CO_2eq/yr이고 이 중 도로 수송 부문이 차지하는 비율은 약 15%이다(Crippa *et al*., 2024). 수송 부문의 온실가스 배출량 감소를 위한 대표적인 방법으로 도로 혼잡 완화가 있다. 도로 혼잡으로 인해 멈춘 차량은 이동이 없는 상태에서 차량의 동력이 유지되는 공회전이 발생하고 주행 시보다 4배 많은 대기오염물질을 배출한다(Jin and Jin, 2021; Green Transport, 2001). 도로 혼잡을 완화하여 도로에서 자동차 공회전이 줄어들면 1년간 약 1.96톤의 오염물질 배출량을 감소시킬 수 있다(Eom and Park, 2013).

도로 혼잡 문제를 완화하기 위해 국내에서는 혼잡도로 통행료 부여, 교통 신호 제어, 대중교통 개선 등 다양한 선행연구들이 진행되었다(Jung and Jung, 2007; Youn and Ji, 2008; Yang *et al*., 2009). 그 중 교통 신호 제어 연구는 교차로의 신호체계 조정을 통해 교통 혼잡 문제를 해결하고자 하는 연구이다. 최근에는 AI 기술을 신호체계에 활용한 지능형 교통 체계(ITS: Intelligent transportaion system)가 전 세계적으로 개발 및 사용되고 있다(Singh and Gupta, 2015). 한국에서는 2021년 국토교통부에서 지능형 교통체계 기본계획 2030을 발표하여 실시간 교통량에 따라 신호를 최적화하여 정체를 최소화하는 스마트 신호 운영시스템 확대 계획을 수립하였다.

ITS의 대표적인 효과로 그린 웨이브 현상이 있다. 그린 웨이브는 차량이 특정 도로를 따라 주행할 때, 연속적으로 녹색 신호를 받아 모든 교차로를 원활하게 통과할 수 있게 되는 현상이다(Ma and He, 2019). 그린 웨이브 현상은 각 교차로에서 차량의 지연시간을 줄이고 온실가스 및 대기오염 물질 배출량을 최대 60% 감소시킨다(Kiers and Visser, 2017). 연속된 신호등이 차량의 주행 속도에 맞춰 녹색 신호를 주어야 차량이 멈추지 않고 주행할 수 있다. 즉 신호등이 서로 협력할 때 자연스럽게 그린 웨이브 현상이 나타나고 교통흐름이 원활하게 유지된다(Warberg *et al*., 2008). 따라서 도로 위의 온실가스 감소를 위해 교차로 간의 협력이 반영된 신호체계가 필요하다.

현재 한국의 신호체계는 대부분 고정형 신호체계이다. 고정형 신호체계는 직진 신호 80초, 좌회전 신호 40초와 같이 정해진 신호 순서와 시간을 반복하는 방식이다. 이러한 방식은 주 신호 방향이 아닌 방향의 교통량 증가와 같은 교통량 변화에 유연하게 대처하는데 한계가 있다(Jo *et al*., 2014). 이러한 문제를 해결하기 위해 강화학습을 이용하여 신호체계를 최적화하는 연구가 진행되고 있다. 강화학습 기반 신호체계는 고정형 신호체계와 달리 실시간 교통량에 따라서 교통 신호 타이밍을 조정할 수 있다(Gao *et al*., 2017). Genders and Razavi(2016)은 DQN(Deep Q-Network)을 사용한 신호체계 최적화를 진행하여 기존의 얕은 neural network를 사용한 신호체계에 비해 평균 누적 지연시간을 약 82%, 평균 대기열 길이를 약 66% 개선하였다. Gao *et al*.(2017)은 DQN을 사용한 신호체계 최적화를 진행하여 고정형 신호체계에 비해 평균 차량 지연시간을 약 86% 개선하였다. Huo *et al*.(2018)은 A3C (Asynchronous Advantage Actor-Critic) 알고리즘을 사용하여 포화 상태의 교통 상황에서 신호 제어를 수행하고 고정 시간 신호 제어 모델에 비해 평균 대기열 길이를 약 71%, 평균 대기 시간을 약 21% 개선하였다. 이러한 기존 연구들은 하나의 교차로에 대하여 단일 에이전트를 학습하였다. 일반적으로 현실 세계의 도로 네트워크는 여러 개의 교차로가 이어져 있는 다중 에이전트 환경으로 볼 수 있고, 이때 교차로는 인접한 교차로의 교통량, 신호주기, 교차로 간의 거리 등에 따라 영향을 받는다(Joo and Lim, 2020). 하지만 단일 에이전트 강화학습의 경우 다중 에이전트 환경에서 효율적으로 정책을 학습하기 어려운 경우가 많다(Jung and Kim, 2021). 이러한 문제를 해결하기 위해 다양한 다중 에이전트 강화학습 기법이 등장하였고 교통 신호 제어 연구에도 활발히 사용되고 있다. Kim and Jung(2019)은 4x4의 다중 교차로 환경에서 인접한 교차로 간에 교통 상황을 주고받는 방식을 사용하는 협력적 교통 신호를 제안하였다. 교차로의 학습 과정에서 교차로가 인접한 교차로의 상태와 행동을 통해 계산된 Q 함숫값을 사용하여 학습하는 방식으로 교차로가 인접한 교차로를 고려하여 학습하고 행동할 수 있도록 하였다. Haddad *et al*.(2022)은 2x2, 2x3 두 가지 교차로 시나리오에 대해서 다중 에이전트 강화학습 방식을 사용하여 교통 신호를 제어

하였다. 강화학습 과정 중 Q함수 업데이트 과정에서 이웃 교차로의 보상을 공유하는 방식으로 교차로가 협력 활동을 하며 정책을 학습할 수 있도록 하였다. Tan *et al.*(2019)은 큰 규모의 다중 교차로에서의 신호 제어를 더 작은 규모의 다중 교차로에서의 신호 제어 문제로 분해하는 방식을 제안하였다. 큰 규모의 다중 교차로에 Global Agent를 두어 더 작은 규모의 다중 교차로에서의 행동을 결정해주고 그에 따른 결과를 단순히 합하여 사용하는 것이 아닌 전역 Q함수를 통해 학습함으로써 전체 교차로의 상호작용을 반영하여 학습할 수 있도록 하였다. 현실 세계의 신호등은 신호 순서나 신호 주기에 제약이 존재한다. 기존 신호체계에서 신호 순서의 변화는 교차로의 지체도와 안전도에 영향을 줄 수 있다(Park and Huh, 2023). 또한 신호 순서나 제약이 없으면 특정 신호에 무한히 신호를 유지하거나 특정한 두 신호만 반복하여 다른 신호가 켜지지 않는 현상이 발생할 수 있다. 하지만 대부분의 교통 신호 제어 연구에서 신호 순서나 최소 녹색시간 등의 제약 조건을 고려하지 않았다.

따라서 본 연구에서는 현실 제약 조건을 반영한 다중 교차로에서의 협력적 교통 신호 제어 방식을 제안한다. 다중 교차로에서 신호등의 협력적 상호작용을 반영할 수 있는 새로운 방식을 제안하고 이를 통해 불필요한 대기 시간과 공회전을 줄여 온실가스를 줄이는 것을 목표로 하였다. 또한 현실의 교차로와 교통량, 신호 제약 조건을 반영한 현실성 높은 신호 모형을 제안하여 기존 연구와의 차별성을 두었다.

2. 강화학습 알고리즘

강화학습 에이전트는 시간이 지남에 따라 환경과 상호작용하면서 받는 장기적인 보상을 최대화하는 것을 목표로 학습한다(Li, 2017). Eq. (1)은 에이전트의 한 단계 학습을 위한 학습 샘플을 나타내는 식이다. 에이전트는 주어진 환경에서 상태를 관찰하고 행동을 결정하여 행동에 따른 보상을 받고 다음 상태를 관찰한다. 해당 과정은 에이전트가 최종 상태에 도달할 때까지 반복된다(Li, 2017). 많은 강화학습 알고리즘에서 에이전트는 가치 함수 추정에 기반하여 행동을 선택한다. 가치 함수는 현재 상태에서 기대되는 보상 혹은 현재 상태에서 특정 행동을 수행했을 때 기대되는 보상을 추정한다(Sutton and Barto, 2018).

$$(S,\ A,\ R,\ S') \tag{1}$$

where, S: state, A: action, R: reward, S': next state.

Q함수를 정의하는 식은 Eq. (2)와 같다. Q함수는 행동-가치 함수라고 불리며 특정 상태에서 특정 행동을 했을 때 받을 수 있는 보상의 기댓값을 말한다(Sutton and Barto, 2018). 즉, Q함숫값이 높은 행동을 선택했을 때 높은 보상을 받을 가능성이 가장 높다는 것을 의미한다. 이렇게 Q값에 따라 행동을 선택하는 강화학습 알고리즘을 Q-learning이라고 한다(Qiang and Zhongli, 2011).

$$Q(s,a)= E_{\pi}\left[r_{t+1}+\gamma r_{t+2}+\gamma^2 r_{t+3}+\cdots \mid S_t=s,\ A_t=a\right] \tag{2}$$

where, s: state, a: action, r_{t+1}: reward at time $t+1$, γ: discount factor, S_t: state at time t, A_t: action at time t.

Q-learning에서 Q함수의 업데이트는 Eq. (3)으로 이루어진다. 벨만 방정식에 근거하여 최적 Q함수의 값은 $R+\gamma \max_{a} Q(s',a')$의 기대값과 같다. 따라서 Eq. (3)을 따른 반복적인 업데이트를 통해 Q함수는 최적의 Q함수로 수렴한다(Mnih, 2013).

$$Q(s,a)\leftarrow Q(s,a)+\alpha\left(R+\gamma \max_{a} Q(s',a')- Q(s,a)\right) \tag{3}$$

where, s: state, a: action, α: learning late, R: reward, γ: discount factor.

Q-learning은 Q함숫값을 상태-행동 축으로 이루어진 테이블에 저장하고 업데이트한다. 하지만 대부분 현실의 사례에는 테이블에 저장할 수 있는 것보다 훨씬 많은 상태가 존재한다(Sutton and Barto, 2018). 문제 해결을 위해 신경망의 함수 근사 속성을 활용한 딥러닝을 적용하였고 심층 강화학습은 기존 강화학습과 달리 차원의 저주를 효율적으로 처리하였다(Arulkumaran, 2017). 딥러닝을 사용한 심층 강화학습의 한 종류인 DQN의 업데이트는 Eq. (4)로 이루어진다. Q-learning과 달리, DQN 알고리즘은 Q함수 업데이트 시 Q함숫값 대신 Q함수의 가중치가 업데이트된다. 본 연구에서는 이러한 DQN을 사용하여 교통 신호 제어를 위한 강화학습 에이전트를 학습하였다. 또한 더욱 높은 성능을 위해 Dueling DQN, Prioritized experience replay 등의 DQN 확장 버전들을 결합한 강화학습 알고리즘을 연구에 적용하였다.

$$MSE(\theta)= \frac{1}{m}\sum_{t=1}^{m}\left\{\left(R+\gamma \max_{a} Q(s',a',\theta^*)- Q(s,a,\theta)\right)\right\}^2 \tag{4}$$

where, m: the batch size, R: reward, γ: discount factor,

s:state, a: action, s': next state, a': next action, θ: parameters of the Q-network which is an approximated Q function, θ^*: parameters of the target network.

3. 방법론

3.1 상태

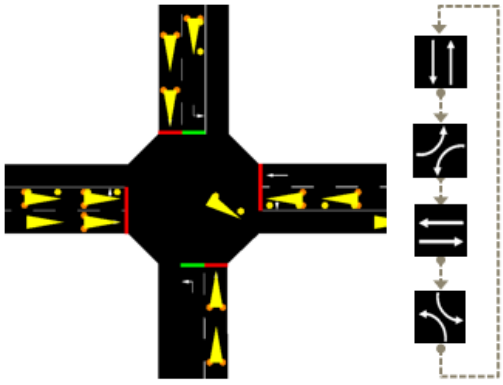

Fig. 1. Example of intersection and intersection signal phase

Fig. 1은 특정 시간대 교차로의 모습을 나타내는 그림이다. Eq. (5)는 교차로의 상태를 나타내는 식이다. Pt는 이진법 형태로 각 신호의 현시 여부를 나타내는 벡터이며, Fig. 1에서 P_t=[0, 0, 0, 1]로 나타난다. d_t는 현재 신호의 현시 이후 현재까지의 시간을 의미하며, 현재 신호의 현시 이후 3초가 지났다면 d_t=3으로 나타난다. N_t는 신호 차선별 차량 수를 나타내는 벡터이다. 각 신호에 해당하는 차선에 존재하는 차량의 수를 집계한 것으로 Fig. 1에서 N_t=[4, 4, 2, 1]로 나타난다. V_t는 신호 차선별 차량 평균 속도를 나타내는 벡터이다. N_t에서 집계한 차량들의 현재 속도를 평균 내어 계산하였으며, V_t=[1.2, 4.5, 2.7, 6.0]과 같은 형태로 나타날 수 있다.

본 논문에서 '신호 차선별'은 특정 신호와 관련 있는 차선별로 집계했다는 것을 의미한다. 강화학습에서 상태는 현재 상태를 나타내는 충분한 정보를 가지고 있어야 한다(Scherer *et al.*, 2018). 따라서 중요한 정보는 유지하되 불필요한 차원을 줄이기 위해 특정 신호와 관련 있는 차선별로 집계하는 방식을 사용하였다.

$$S_t = \{P_t, d_t, N_t, V_t\} \tag{5}$$

where, S_t: state at time t, P_t: vector representing in binary which signal is on at time t, d_t: time the current signal was on at time t, N_t: vehicle queue by signal lane at time t, V_t: Average vehicle speed per signal lane at time t.

3.2 행동

본 연구에서 에이전트의 행동은 0 또는 1로 선택된다. 에이전트의 행동이 0일 경우 현재 신호가 유지되고, 1일 경우 정해진 신호현시 순서에 따라 다음 신호로 변경된다. 하지만 최소 녹색시간과 최대 녹색시간에 따라 행동과 관계없이 신호가 유지되거나 변경될 수 있다. 예를 들어 최소 녹색시간을 3초, 최대 녹색시간을 10초로 가정하면, d_t가 3초 미만일 경우 행동이 1이어도 신호를 변경하지 않는다. 또한 d_t가 10초 초과일 경우 행동이 0이어도 신호를 변경하게 된다. 본 연구에서는 신호현시 순서 유지 및 최소, 최대 녹색시간과 같은 현실 제약 조건을 반영하여 현실 적용성을 높였다.

3.3 보상

에이전트가 행동 후에 받는 보상 함수는 Eq. (6)와 같다. 에이전트는 많은 보상을 얻기 위해 행동하고 학습하므로 목적에 맞는 보상 함수를 설정하는 것은 중요하다. 본 연구의 목적은 차량 대기 시간을 줄이는 것이므로 보상 함수를 차량 대기 시간의 음수로 설정하였다. 제안된 보상 함수는 도로에 많은 차량이 대기하고 있을수록 에이전트가 적은 보상을 받게 하여 에이전트가 차량 대기 시간을 줄이는 방향으로 학습하도록 한다.

$$R_t = -\sum W_t \tag{6}$$

where, R_t: reward at time t, W_t: waiting time for stopped vehicle at time t.

3.4 협력적 교통 신호 제어

Eq. (7)은 인접한 교차로 간의 상호작용을 반영하기 위해 협력적 강화학습 방식으로 정의된 Q함수 업데이트 식이다. 기존 Q함수 업데이트 식에 이웃 교차로의 이전 시간 보상을 추가하여 교차로가 인접한 교차로를 반영하여 학습할 수 있도록 하였다. 이때 이웃 교차로는 현재 교차로와 직접 연결된 교차로를 의미한다. 또한 현재 교차로의 상태에 이웃 교차로의 상태 및 행동을 추가하여 에이전트가 행동을 선택할 때 이웃 교차로의 상태, 행동을 고려하여 선택하도록 하였다. Fig. 3은 교통 신호 제어 과정에서 협력적 강화학습과 일반 에이전트 강화학습의 차이점를 나타낸 그림이다.

$$Q_{t+1}^{i}(s_t^i,\ a_t^i) \leftarrow Q_t^i(s_t^i,\ a_t^i;\theta_i) + \alpha(t)\Big[r_t + \ \gamma \max_{a'} Q_t^i(s_{t+1}^i, a';\theta_i^*) - Q_t^i(s_t^i,\ a_t^i;\theta_i)\Big] + \frac{1}{N_{adj}} \sum_{j \in N_{adj}} r_{t-1}^j \quad (7)$$

where, s: state, a: action, $\alpha(t)$: learning late at time t, r_t: reward at time t, γ: discount factor, θ: parameters of the Q-network, N_{adj}: number of adjacent intersections

Eqs. (8) and (9)는 본 연구에서 제안된 방법론을 적용하여 현재 교차로와 이웃 교차로의 상태를 나타낸 식이다. 이때 이웃 교차로의 상태는 현재 교차로의 상태와 차이점이 존재한다. 이웃 교차로 상태의 $N_t^{'}$, $V_t^{'}$는 각각 이웃 교차로 내에서 현재 교차로로 진입할 수 있는 차선에 대한 차량 수와 평균 속도를 집계한 값이다. 이웃의 교차로에서 현재 교차로에 직접적인 영향을 줄 수 있는 차선들만 현재 교차로의 상태에 추가하여 효율적인 학습을 할 수 있도록 하였다.

$$S_t^i = S_t^i + \sum_{j \in N_{adj}} \{S_t^j, a_{t-1}^j\} \quad (8)$$

where, S_t: state at time t, a_t: action at time t, N_{adj} :number of adjacent intersections

$$S_t^j = \{P_t, d_t, N_t^{'}, V_t^{'}\} \quad (9)$$

where, S_t: state at time t, P_t: vector representing in binary which signal is on at time t, d_t: time the current signal was on at time t, $N_t^{'}$: vehicle queue by signal lane at time t, $V_t^{'}$: Average vehicle speed per signal lane at time t

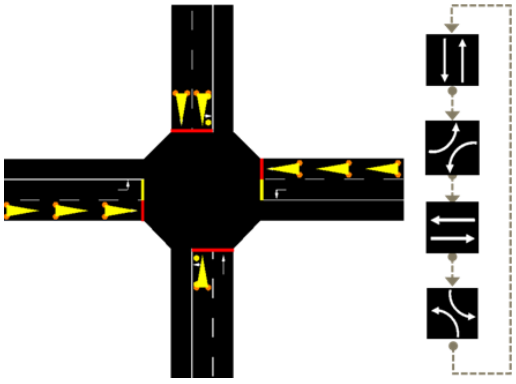

Fig. 2. Example of neighbor intersection and intersection signal phase

Fig. 2는 Fig. 1의 오른쪽에 인접한 교차로를 나타낸 예시 그림이다. 해당 교차로에서 Fig. 1에 진입할 수 있는 도로는 오른쪽 2차선 도로와 아래 1차선 도로이다. 해당 교차로의 신호현시 순서가 Fig. 1의 신호현시 순서와 같을 경우, Fig. 2에서 $N_t^{'}$=[0, 0, 3, 1], $V_t^{'}$=[0, 0, 0.4, 0.2]로 나타날 수 있다. Eq. (5)의 상태 계산 방식을 적용하면 N_t=[1, 0, 6, 2]로 나타나지만 $N_t^{'}$를 계산할 때는 Fig. 1에 진입할 수 없는 도로의 차량은 집계하지 않아 $N_t^{'}$=[0, 0, 3, 1]로 나타난다.

구분	협력적 강화학습	일반 에이전트 강화학습
상태	현재 교차로 상태 + 이웃 교차로 상태, 행동	단일 교차로 상태
정책 업데이트	이웃 교차로의 보상을 고려한 Q-network 업데이트	단일 교차로의 보상만 고려한 Q-network 업데이트

Fig. 3. Differences between cooperative reinforcement learning and general reinforcement learning in policy learning process

3.5 협력적 강화학습 기반 교통 신호 모형 학습

본 단락에서는 협력적 강화학습을 기반으로 교통 신호 모형을 학습하는 과정에 대하여 설명한다. 먼저 각 교차로의 에이전트는 자신과 주변 교차로의 상태를 관찰한다. 상태는 Eq. (8)의 형태로 입력을 받게 되며, Fig. 1이 현재 교차로이고 인접한 교차로가 Fig. 2의 교차로만 존재할 때 현재 상태는 S_t =[0, 0, 0, 1, 3, 4, 4, 2, 1, 1.2, 4.5, 2.7, 6.0, 0, 1, 0, 0, 3, 0, 0, 3, 1, 0, 0, 0.4, 0.2, 1]로 나타낼 수 있다. 리스트 내 1~13번째 원소는 현재 교차로의 상태를 나타내고, 14~26번째 원소는 이웃 교차로의 상태, 마지막 27번째 원소는 이웃 교차로의 이전 시간의 행동을 나타낸다. 본 상태는 협력적 강화학습을 위해 이웃 교차로의 정보를 추가한 형태이다. 이러한 상태로 인해 에이전트는 자신 교차로의 정보가 동일하더라도 이웃 교차로의 정보에 따라 다른 행동을 선택할 수 있게 된다. 이는 에이전트가 이웃 교차로의 정보(상태, 행동)를 고려하여 행동을 결정하는 것을 의미한다.

이후 에이전트는 현재 상태에서의 행동의 가치를 나타내는 Q-network와 정책에 따라서 현재 상태에서 취할 행동을 결정하게 된다. 에이전트가 행동을 결정한 후, 에이전트의 행동에 따라 신호가 변경된다. 3.2에서 언급한 대로 행동이 0일 경

우 신호를 변경하지 않고, 1일 경우 신호를 다음 순서로 변경한다.

신호가 변경되고 특정 시간(초)이 지난 후 보상함수에 따라 보상을 계산한다. 에이전트에게 보상이 주어지면, 에이전트는 Eq. (7)과 보상값에 따라서 자신의 Q-network를 업데이트한다. 보상이 높으면 에이전트는 현재 상태에서 현재 행동을 선택한 것을 높게 평가할 것이고, 다음에 비슷한 상황이 왔을 때 같은 행동을 할 확률이 높아질 것이다. 보상이 낮을 경우, 에이전트는 현재 상태에서 현재 행동을 선택한 것을 낮게 평가할 것이며, 다음에 비슷한 상황이 왔을 때 같은 행동을 피하려 할 확률이 높아질 것이다. 본 연구에서는 차량 대기 시간의 음수값을 보상함수로 설정하였다. 따라서 에이전트는 차량 대기 시간의 음수값을 늘리는 방향, 즉 차량 대기 시간을 줄이는 방향으로 학습하고 행동하게 된다. 제안된 보상함수를 설정함으로써 차량 대기 시간과 공회전 시간을 줄이고 이에 따른 온실가스 배출량 감소 목표를 달성할 수 있도록 하였다. 협력적 강화학습을 적용한 모델은 이전 시간대의 보상값을 자신의 Q-network 업데이트에 반영한다. 이는 에이전트가 더 이상 자신의 보상값만을 통해 현재 상태에서의 현재 행동을 평가하지 않고, 이웃의 보상까지 고려하여 평가하게 된다는 것을 의미한다. 즉, 현재 상태에서의 현재 행동으로 인해 높은 보상값을 받았다고 하더라도 이웃 교차로의 보상값이 낮으면 그 행동은 좋은 행동으로 평가되지 않는다. 따라서 협력적 강화학습을 적용한 모델은 이웃 교차로들의 보상을 함께 높일 수 있는 방향으로 학습하게 된다. 해당 과정을 통해 Q-network가 업데이트되면 에이전트는 다시 상태를 관찰하고 Q-network를 업데이트하는 일련의 과정을 종료 조건을 만족할 때까지 반복한다.

4. 실험 및 결과

4.1 실험 환경

실험에 사용된 교차로 환경은 경기도 이천시의 일부 연속된 6개의 교차로를 대상으로 한다. 6개의 교차로는 일렬로 이어져 있는 형태이다. 각 교차로는 3지 또는 4지 교차로로 구성되어 있으며 하나의 주 방향 도로가 존재하고 교차로마다 부 방향 도로를 가지고 있는 형태이다. Fig. 4은 첫 번째 교차로부터 각 교차로의 형태와 방향별 교통량을 나타낸 그림이다. 각 교차로는 고유의 신호현시 순서, 최소 녹색시간, 최대 녹색시간, 교통량 등을 가진다. 신호현시 순서는 실제 교차로에서 운영되고 있는 신호현시 순서를 그대로 적용하였다. 최소 및 최대 녹색시간은 현재 교차로에서 운영되고 있는 신호주기를 반영하여 산출하였다.

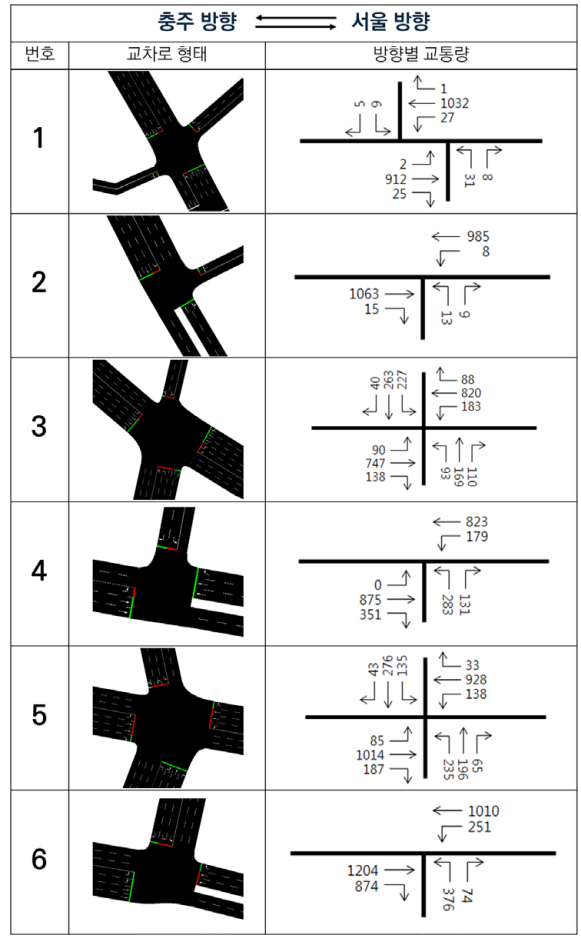

Fig. 4. The shape of the intersections, road direction and traffic volume by direction at the intersections used in the experiment. They are arranged in a single row from the left

교통량은 실제 교차로에서 수집된 교통량이 사용되었다. 차량이 많은 첨두 시간대(18:00~19:00시)에 수집된 교통량을 적용하여 혼잡한 도로 상황에서의 신호 제어를 학습하였다.

본 연구의 시뮬레이션 및 도로 네트워크는 Sumo (Simulation of Urban MObility)를 통해 구현되었다. 다중 에이전트 기반 강화학습 신호 제어 모델은 PyTorch를 활용하여 구현되었다. 심층 강화학습의 신경망은 완전연결방식으로 입력층, 3개의 은닉층, 출력층으로 구성된다. 시뮬레이션에서 교통량은 Fig. 4의 교통량에 따라서 일정 시간동안 생성되며 한 번의 시뮬레이션은 모델이 다중 교차로 환경에서 정해진 양의 교통량을

처리했을 경우 종료된다. 모델은 시뮬레이션을 150회 반복하여 학습하였다. 에이전트의 탐험을 도와주는 파라미터인 ε은 0.9에서 0.03까지 Decay 파라미터에 따라 지수적으로 감소하도록 하였다. 리플레이 버퍼의 메모리 크기는 1000, 배치 크기는 32로 설정하였다. 감쇠율은 0.9, 학습률은 0.005로 설정하였다. 최적화 함수와 손실함수는 각각 Adam (Adaptive Moment Esimation), MSE (Mean Square Error)가 적용되었다. DQN의 Target Network는 매 200번 학습마다 Network와 동기화하도록 설정하였다. 파라미터들은 강화학습 기반 신호 최적화 선행 연구(인용들)의 파라미터들과 임의의 값을 바탕으로 학습을 진행하여 안정적으로 수렴하고 가장 높은 보상 값을 받은 학습의 파라미터들로 선정하였다. 실험에 사용된 파라미터들은 Table 1.와 같다.

Table 1. Parameters used in reinforcement learning-based signal optimization model

파라미터	값
에피소드	150 (회)
학습 종료 조건	8400 (대)
Delta Time	3 (초)
Epsilon Start	0.9
Epsilon Decay	20000
Epsilon End	0.03
배치 크기	32 (개)
버퍼 크기	1000 (개)
감쇠율	0.9
학습률	0.005
최적화 함수	Adam
Loss 함수	MSE
Network와 Target Network의 동기화 기준	매 학습 200번

4.2 실험 결과

본 연구에서는 제안 모델인 다중 에이전트 강화학습 기반 신호 모델의 성능을 평가하기 위해 두 가지 비교 모델을 설계하였다. Fig. 5는 설계한 2가지 모델과 제안 모델의 차이점을 나타낸 그림이다. 일반 모델은 협력적 강화학습의 효과를 비교하기 위해 신호 제약 조건은 유지하되, 협력적 강화학습을 위한 이웃 상태 추가, Q함수 업데이트식을 적용하지 않은 단일 에이전트 강화학습 기반 신호 모형이고, 비교 모델은 신호 제약 조건의 효과를 비교하기 위해 협력적 강화학습은 유지하되, 신호현시 순서 및 최소 녹색시간 등의 신호 제약 조건을 해제한 다중 에이전트 강화학습 기반 신호 모형이다. 이외에 실험에 사용된 조건은 모두 동일하다.

구분	협력적 강화학습 O	협력적 강화학습 X
신호 제약 조건 O	제안 모델	일반 모델
신호 제약 조건 X	비교 모델	-

Fig. 5. Comparison of signal models designed to evaluate the performance of the proposed model

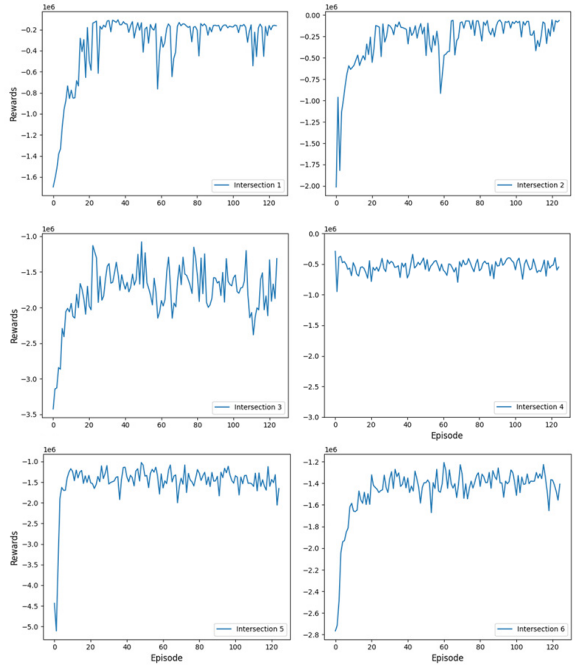

Fig. 6. Intersection-by-intersection reward graph received during the training process of the proposed model

Fig. 6는 제안 모델이 125개의 에피소드 동안 학습하면서 받은 에피소드별 보상의 총합을 각 교차로에 대해 나타낸 그래프이다. 에피소드가 증가함에 따라 보상이 증가하고 수렴하는 것을 볼 수 있다. 따라서 제안 모델이 본 연구의 다중 교차로 환경을 잘 학습하였다는 것으로 판단하였다.

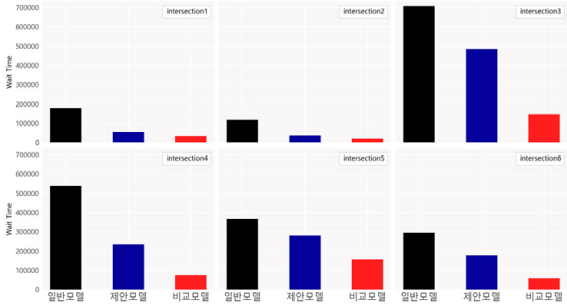

Fig. 7. Graph comparing simulation results (vehicle waiting times) for each model at each intersection

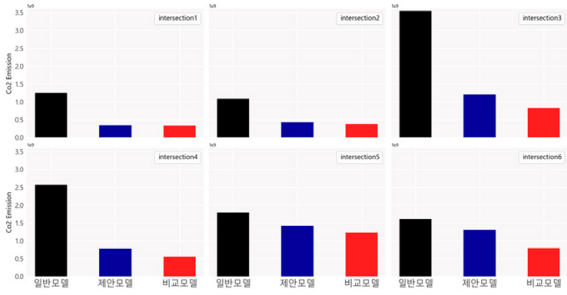

Fig. 8. Graph comparing simulation results (vehicle CO2 emission) for each model at each intersection

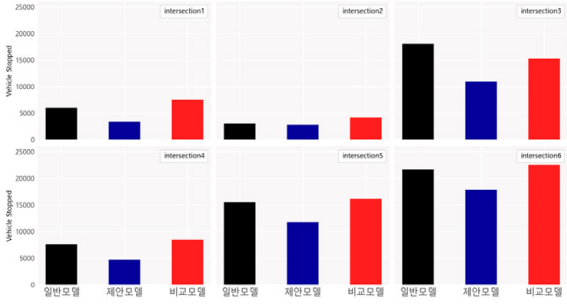

Fig. 9. Graph comparing simulation results (number of vehicle stops) for each model at each intersection

Fig. 7, 8, 9은 학습된 3가지 모델로 학습 환경과 같은 시뮬레이션 환경에서 신호를 제어한 결과를 교차로별로 나타낸 그래프이다. 각각 차량 대기 시간, CO_2 배출량, 차량 정지 횟수를 나타낸 그래프이다. 차량 대기 시간과 정지 횟수는 Python을 통해 직접 계산하였고, CO_2 배출량은 Sumo의 내장함수를 통해 계산하였다. 차량 대기 시간은 모든 교차로에서 일반모델, 제안모델, 비교모델 순으로 높게 나타났다. 총합으로는 제안모델이 일반모델에 비해 차량 대기 시간을 약 43% 감소시켰고, 비교모델이 제안모델에 비해 차량 대기 시간을 약 61% 감소시킨 것으로 나타났다. CO_2 배출량 또한 모든 교차로에서 일반모델, 제안모델, 비교모델 순으로 높게 나타났다. 총합으로는 제안모델이 일반모델에 비해 CO_2 배출량을 약 44% 감소시켰고, 비교모델이 제안모델에 비해 CO_2 배출량을 약 25% 감소시킨 것으로 나타났다. 차량 정지 횟수는 대부분의 교차로에서 비교모델, 일반모델, 제안모델 순으로 높게 나타났다. 총합으로는 제안모델이 일반모델에 비해 차량 정지 횟수를 약 28% 감소시켰고, 비교모델에 비해 차량 정지 횟수를 약 31% 감소시킨 것으로 나타났다. 차량 정지 횟수의 증가는 차량의 가감속 횟수의 증가로 볼 수 있다. 차량의 가감속은 CO_2 배출량과 양의 상관성이 있어(Oh *et al*., 2013) 차량 가감속 횟수의 증가는 CO_2 배출량의 증가라고 할 수 있다. 또한 차량 정지 횟수가 늘어나면 차량 연료 소비량이 증가한다(Zheng *et al*., 2017). 차량 연료 소비량이 증가하면 CO_2 배출량 또한 증가하게 된다. 결과적으로 차량 정지 횟수의 증가는 CO_2 배출량 증가로 이어질 수 있다. 이러한 이유로 제안 모델에 대한 비교 모델의 CO_2 배출량 감소폭이 차량 대기 시간 감소폭에 비해 적게 나타난 것으로 보인다.

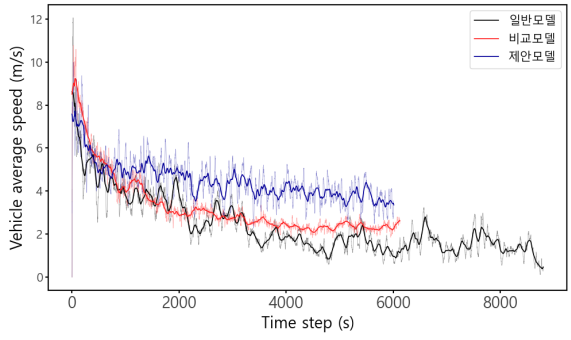

Fig. 10. Graph comparing vehicle average speed of each model during simulation time

Fig. 10은 각 모델이 시뮬레이션하는 동안 차량의 평균 속도를 시간대별로 나타낸 그림이다. 시뮬레이션은 시간이 지남에 따라 교차로가 수용할 수 있는 차량수 이상의 차량이 생성된다. 이에 따라 Fig. 10의 차량 평균 속도가 점차 줄어드는 것을 볼 수 있다. 제안 모델은 세 모델 중 가장 높은 평균 속도를 보이며 시간이 지남에 따라서도 평균 속도가 비교적 천천히 줄어드는 것을 볼 수 있다. 일반 모델은 초반부터 속도가 빠르게 떨어져 세 모델 중 가장 낮은 평균 속도를 보이며 정해진 교통량을 빠르게 처리하지 못해 다른 모델들과 비교하여 오랜 시간 동안 시뮬레이션이 진행된 것으로 나타났다. 비교 모델은 초반에 비교적 높은 평균 속도를 보였으나 속도가 점차 떨

어져 제안 모델과 비교하여 낮은 평균 속도를 유지하였다. 비교 모델은 Fig. 7, 9에서 나타난 것처럼 차량 대기 시간이 적고 차량 정지 횟수가 높게 나타났다. 이는 차량이 정지해 있는 시간 자체는 적지만 높은 차량 정지 횟수로 인해 가감속을 반복하여 최고 속도에 도달해 있는 시간이 짧다는 것으로 볼 수 있다. 이러한 결과가 비교 모델의 낮은 차량 평균 속도로 나타난 것으로 보인다. 본 실험에서 제안 모델은 비교적 원활한 속도로 차량이 주행할 수 있도록 신호를 제어하였다.

5. 결 론

본 연구는 교차로 간의 협력적 신호 제어를 위한 새로운 방법론과 신호현시 순서, 최소 녹색시간 등의 현실 제약을 통해 현실성을 높인 다중 교차로 교통 신호 제어 모델을 제안한다. 제안 모델은 일반 강화학습 모델과 비교하여 차량 대기 시간, 정지 횟수, CO_2 배출량 등 모든 면에서 다중 교차로 교통 신호 제어에 더 효과적인 것으로 나타났다. 신호 제약을 해제한 비교 모델은 신호 제약을 반영한 제안 모델과 비교하여 차량 대기 시간과 CO_2 배출량이 적게 나타났지만, 차량의 정지 횟수가 높게 나타났다. 이는 신호 제약 해제로 인해 신호의 순서가 없고 신호가 자주 변경될 수 있기 때문으로 판단된다. 이러한 신호체계는 운전자의 혼란을 초래하고 사고 발생률을 높일 가능성이 있다(Park and Huh, 2023). 따라서 비교 모델을 실제 교차로에 적용하기 위해서는 추가적인 검토가 필요할 것으로 보인다. 또한 세 모델 중 제안 모델에서 차량 평균 속도가 가장 높게 나타났으며 다중 교차로에서 원활한 교통 흐름을 생성하였다. 본 연구의 제안 모델은 현실 제약 조건을 통해 현실성을 높이면서 협력적 신호 제어를 통해 다중 교차로에서 효과적으로 교통 신호를 제어하였다. 본 연구에서 제안된 방법론을 활용하면 실제 다중 교차로에서 차량 대기 시간과 온실가스 배출량을 감소시킬 수 있을 것으로 기대된다. 하지만 본 연구는 돌발 상황(추돌 사고, 차량 고장 등), 극한 날씨(우천, 폭설), 대규모 행사 등의 현실에서 일어날 수 있는 다양한 변수들을 고려하지 않았다는 한계점이 있다. 향후 연구에서 돌발 상황이나 대규모 행사 등의 시나리오를 추가하고 극한 날씨로 인한 도로 상태 변화를 반영한다면 비정상적인 상태 및 교통량 변화에 대응할 수 있는 신호 최적화 모델이 될 것으로 예상된다. 또한 보상함수에 차량의 속도, 정지 횟수 등을 추가로 고려하거나 회전 교차로, 비보호 신호가 포함된 교차로, 2차원으로 연결된 교차로 등의 더욱 복잡한 다중 교차로 환경에서 최적화를 진행한다면 더욱 현실적인 신호 모형이 될 것으로 기대된다.

감사의 글

"본 연구는 환경부「기후변화특성화대학원사업」의 지원으로 수행되었습니다."

References

Arulkumaran, K., Deisenroth, M.P., Brundage, M., and Bharath, A.A. (2017), Deep reinforcement learning: A brief survey, *IEEE Signal Processing Magazine*, Vol. 34, No. 6. pp. 26-38.
https://doi.org/10.1109/MSP.2017.2743240

Crippa, M., Guizzardi, D., Pagani, F., Banja, M., Muntean, M., Schaaf, E., Monforti-Ferrario, F., Becker, W.E., Quadrelli, R., Risquez Martin, A., Taghavi-Moharamli, P., Köykkä, J., Grassi, G., Rossi, S., Melo, J., Oom, D., Branco, A., San-Miguel, J., Manca, G., Pisoni, E., Vignati, E., and Pekar, F. (2024), *GHG emissions of all world countries*, Publications Office of the European Union, Luxembourg, JRC134504, 278p.

Eom, D. L. and Park, S. W. (2013), Effectiveness of Idling Restriction in Signalized Intersections, *Journal of Transport Research*, Vol. 20, No. 2. pp. 153-161. (in Korean with English abstract)
https://doi.org/10.34143/jtr.2013.20.2.153

Gao, J., Shen, Y., Liu, J., Ito, M., and Shiratori, N. (2017), Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network, *arXiv preprint arXiv:1705.02755*.
https://doi.org/10.48550/arXiv.1705.02755

Genders, W. and Razavi, S. (2016), Using a deep reinforcement learning agent for traffic signal control, *arXiv preprint arXiv:1611.01142*.
https://doi.org/10.48550/arXiv.1611.01142

Green Transport. (2001), *White Paper on the Survey on the Status of Idling Vehicles and the Legislation of Prohibition of Idling Vehicles Project*, Technical report, Green Transport, Republic of Korea, 79p. (in Korean)

Haddad, T. A., Hedjazi, D., and Aouag, S. (2022), A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control, *Engineering Applications of Artificial Intelligence*, Vol. 114, 105019.

https://doi.org/10.1016/j.engappai.2022.105019

Huo, Y., Hu, J., Wang, G., and Chen, J. (2018), A traffic signal control method based on asynchronous reinforcement learning, *In 18th COTA International Conference of Transportation Professionals*, 5-8 July, VA: American Society of Civil Engineers, Reston, pp. 1444-1453.

IPCC. (2021), 2021: Summary for Policymakers, Research report, IPCC, USA, 32p.

IPCC. (2023), CLIMATE CHANGE 2023: Synthesis Report, Research report, IPCC, Switzerland, 184p.

Jin, J. K. and Jin, J. I. (2021), A Study on the Effect of Traffic Congestion on Particulate Matter Concentration in Seoul: Big Data Approach, *Journal of Korea Planning Association*, Vol. 56, No. 1. pp. 121-136. (in Korean with English abstract)
https://doi.org/10.17208/jkpa.2021.02.56.1.121

Jo, Y., Choi, J., and Jung, I. (2014), 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. (in Korean with English abstract)

Joo, H. J. and Lim, Y. J. (2020), Distributed Traffic Signal Control at Multiple Intersections Based on Reinforcement Learning, *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 45, No. 2. pp. 303-310. (in Korean with English abstract)
https://doi.org/10.7840/kics.2020.45.2.303

Jung, I. H. and Jung, S. Y. (2007), Plan to introduce toll fees on congested roads to ease traffic congestion in metropolitan areas, *Korea Research Institute For Human Settlements Policy Brief*, Vol. 133, pp. 1-8. (in Korean)

Jung, K. Y. and Kim, I. C. (2021), C-COMA: A Continual Reinforcement Learning Model for Dynamic Multiagent Environments, *KIPS Transactions on Software and Data Engineering*, Vol. 10, No. 4. 4p. (in Korean with English abstract)

Kiers, M. and Visser, C. (2017), The effect of a green wave on traffic emissions, FH-Forschungsforum "Research – Innovation – Value", *in researchgate*, https://www.researchgate.net/publication/321145958_The_Effect_of_a_green_wave_on_traffic_emissions (last date accessed: 26 December 2024).

Kim, D. H. and Jung, O. R. (2019), A Study on Cooperative Traffic Signal Control at Multi-intersection, *Journal of IKEEE*, Vol. 23, No. 4. pp. 266-271. (in Korean with English abstract)
https://doi.org/10.3745/KTSDE.2021.10.4.143

Li, Y. (2018) Deep reinforcement learning: An overview, arXiv preprint arXiv:1701.07274.
https://doi.org/10.48550/arXiv.1810.06339

Ma, C. and He, R. (2019), Green wave traffic control system optimization based on adaptive genetic-artificial fish swarm algorithm, *Neural Computing and Applications*, Vol. 31, pp. 2073-2083.
https://doi.org/10.1007/s00521-015-1931-y

Mnih, V., 2013, Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602.
https://doi.org/10.48550/arXiv.1312.5602

Oh, H. U., Lee, Y. S., and Yoo, H. M. (2013), Greenhouse gas emission patterns at intersections by drivers, *International Journal of Highway Engineering*, Vol. 15, No. 4. pp. 147-154. (in Korean with English abstract)
https://doi.org/10.7855/IJHE.2013.15.4.147

Park, J. H. and Huh, J. S. (2023), Particle Swarm Optimization and SUMO based Multi-Intersection Traffic Signal Optimization, *Proceedings of KIIT Conference, KIIT*, 23-25 November, Jeju, Korea, pp. 347-350. (in Korean with English abstract)

Park, S. H., Lee, C. H., Jung, M. K., and Yeom, S. C. (2024), Diagnostic study on the status of implementation of 2030 National Determined Contribution through analysis of a local energy master plan and greenhouse gas emissions, *Journal of Climate*, Vol. 15, No. 3. pp. 327-341. (in Korean with English abstract)
https://doi.org/10.15531/KSCCR.2024.15.3.327

Park, S. W. (2016), Post-2020 Climate Regime and Paris Agreement – Key Issues and Agreed Results of UNFCCC COP 21 -, *Environmental Law and Policy*, Vol. 16, pp. 285-322. (in Korean with English abstract)
https://doi.org/10.18215/envlp.16..201602.285

Qiang, W. and Zhongli, Z. (2011), Reinforcement learning model, algorithms and its application, *2011 International Conference on Mechatronic Science, Electric Engineering and Computer*, IEEE, 19-22 August, Jilin, China, pp.

1143-1146.

Scherer, W. T., Adams, S., and Beling, P. A. (2018), On the practical art of state definitions for Markov decision process construction, *IEEE Access*, Vol. 6, pp. 21115-21128.
https://doi.org/10.1109/ACCESS.2018.2819940

Singh, B. and Gupta, A. (2015), Recent trends in intelligent transportation systems: a review, *Journal of Transport Literature*, Vol. 9, No. 2. pp. 30-34.
https://doi.org/10.1590/2238-1031.jtl.v9n2a6

Sutton, R. S. (2018), *Reinforcement learning: An introduction*, The MIT Press, England. 398p.

Tan, T., Bao, F., Deng, Y., Jin, A., Dai, Q. and Wang, J. (2019), Cooperative deep reinforcement learning for large-scale traffic grid signal control, *IEEE Transactions on Cybernetics*, Vol. 50, No. 6. pp. 2687-2700.
https://doi.org/10.1109/TCYB.2019.2904742

Warberg, A., Larsen, J., and Jørgensen, R. M. (2008), Green wave traffic optimization – a survey, Technical Report No. 2008-01, Informatics and Mathematical Modelling, D T U Compute, Denmark, 24p.

Yang, J., Son, G. M., and Chon, K. S. (2009), Centrality indicators as an instrument to evaluate transit network, *KOR-KST Conference*, Korean Society of Transportation, 5-6 November, Korea, Vol. 61, pp. 961-965. (in Korean)

Yi, D. G. and Kang, S. H. (2022), The Effects of Carbon Tax on the Transport Sector in Achieving the National Greenhouse Gas Reduction Goals by 2030, *Korean Energy Economic Review*, Vol. 21, No. 2. pp. 1-32. (in Korean with English abstract)
https://doi.org/10.22794/keer.2022.21.2.001

Youn, J. H. and Ji, Y. K. (2008), Simulation of Traffic Signal Control with Adaptive Priority Order through Object Extraction in Images, *Journal of Korea Multimedia Society*, Vol. 11, No. 8. pp. 1051-1058. (in Korean with English abstract)

Zheng, F., Li, J., Van Zuylen, H. J. and Lu, C. (2019), Influence of driver characteristics on emissions and fuel consumption, *IET Intelligent Transport Systems*, Vol. 13, No. 12. pp. 1770-1779.
https://doi.org/10.1049/iet-its.2018.5562