대한공간정보학회지
(Journal of Korean Society for Geospatial Information Science)
Vol.28 No.4 December 2020 pp.59-67

ISSN: 1598-2955 (Print)
ISSN: 2287-6693(Online)
http://dx.doi.org/10.7319/kogsis.2020.28.4.059

연구논문

# 대기시간 최소화를 위한 강화학습 기반 교차로 신호 모형

# Intersection Signal Model based on Reinforcement Learning to Minimize Waiting Time

구자운* · 김대균** · 이민혁*** · 전철민****
Gu, Ja Woon · Kim, Dea Gyun · Lee, Min Hyuck · Jun, Chul Min

要 旨

최근 교통혼잡 문제를 해결하기 위해 전통적인 교통이론 기반의 접근방식이 아닌, 기계학습을 이용한 데이터 기반의 접근방식이 활용되고 있다. 본 연구에서는 차량들의 대기시간 최소화를 위한 강화학습 기반 교차로 신호 모형을 제안한다. 본 모형은 Deep Q-Network 알고리즘을 기반으로, 실시간 교통량을 입력받고, 차량들의 정체현상을 감소시키는 신호패턴을 산출한다. 본 연구에서는 현실 적용성을 고려하여, 일반적인 교차로 신호 순서와 최소 녹색시간의 제약을 두고 모형의 학습을 수행하였다. 마이크로 교통 시뮬레이터를 이용하여 전통적인 최적 모형, 무작위 신호순서 모형과 본 모형의 성능을 비교한 결과, 대기시간, 정지횟수 등의 평가척도에서 본 모형이 더 우수한 결과를 도출하였다.

핵심용어 : 교차로 신호 모형, 강화학습, 교통 시뮬레이션, Deep Q-Network

Abstract

To solve the traffic congestion problem, recent studies are more using data-driven approaches by machine learning than traditional traffic theory-based approach. This study proposes a reinforcement learning-based intersection signal model to minimizing waiting time. Based on Deep Q-Network algorithm, RL model computes signal patterns that reduce traffic congestion by taking into account real-time traffic flow. For practical applicability, RL model learning was performed including signal sequences and minimum green time constraints. Using a micro-traffic simulator, we compared the performance of RL model with the traditional model, and we found that RL model had better results in the evaluation indicators such as waiting times and the number of stops.

Keywords : Intersection Signal Model, Reinforcement Learning, Traffic Simulation, Deep Q-Network

## 1. 서 론

도로 위의 교통혼잡 문제를 해결하기 위해 다수의 교통신호 최적화 연구들이 수행된 바 있다(Cai et al., 2009; Pandit et al., 2013; Mannion et al., 2016). 교통신호 최적화란, 차량들의 대기시간 및 정지횟수를 최소화하는 신호패턴을 찾는 연구이다. 신호패턴에는 어느 방향 먼저 녹색신호를 줄 것인가, 얼마동안 줄 것인가, 현재 신호 다음에는 어떤 방향에 녹색신호를 줄 것인가 등 다양한 고려사항이 있다.

현재 대부분의 도로에는 고정형 신호 모형이 적용되어 있다. 고정형 신호 모형은 사전에 계획된 신호패턴이 일정시간 동안 반복 운영되는 체계이다. 예를 들어, 직진 방향 녹색 45초, 좌회전 방향 녹색 15초로 하여 오전 7시부터 9시까지 반복적으로 지속하는 형태가 고정형 신호 모형이다. 이로 인해 실시간으로 변화하는 교통량에 유연하게 대처하기 어려운 한계가 있다(Youn and Ji, 2008).

최근에는 전통적인 교통이론 기반의 접근방식이 아닌, 머신러닝, 딥러닝 등의 데이터 기반 접근방식을 활

Received: 2020.10.05, revised: 2020.10.21, accepted: 2020.11.25

* 서울시립대학교 공간정보공학과 석사과정(Master's Student, Dept. of Geoinformatics, University of Seoul, umseakind2@uos.ac.kr)
** 서울시립대학교 공간정보공학과(Dept. of Geoinformatics, University of Seoul, june6723@uos.ac.kr)
*** 서울시립대학교 공간정보공학과 박사과정(Ph. D. Student, Dept. of Geoinformatics, University of Seoul, lmhll123@uos.ac.kr)
**** 교신저자 · 서울시립대학교 공간정보공학과 교수(Corresponding Author, Professor, Dept. of Geoinformatics, University of Seoul, cmjun@uos.ac.kr)

용하여 교통공학 비전공자들이 신호 모형에 관한 연구를 진행하고 있다. 인공지능이 실시간 교통흐름을 학습하여 상황에 적절한 신호패턴을 산출할 수 있도록 '알파고'와 같은 지능적인 신호 모형들이 연구되고 있다(Mousavi et al., 2017; Liang et al., 2018, Rasheed et al., 2020).

본 연구에서는 강화학습 기반 교차로 신호 모형을 제안한다. 강화학습은 레이블이 명시된 데이터를 학습하는 지도학습과 달리, 인공지능이 주어진 상황에 대하여 적절한 결과를 도출했을 때 상점을 주는 방식으로 학습을 수행한다. 본 연구에서는 실제 도로 환경과 유사한 마이크로 교통 시뮬레이터를 이용하여 인공지능에게 실시간 교통흐름을 제시하고, 인공지능이 차량들의 정체현상을 감소시키는 신호패턴을 산출했을 경우 상점을 부여하는 방식으로 학습 기반 신호 모형을 도출하였다.

## 2. 관련 연구 분석

Fig. 1은 교차로에서의 일반적인 신호 운영 예시이다. 총 4가지 신호로 구성되어 있고, 각 방향별 녹색시간은 20초, 40초, 20초, 40초이다. 현재 북쪽과 남쪽으로 향하는 좌회전에 녹색신호가 들어왔고, 이후에는 서쪽과 동쪽으로 향하는 직진 방향에 녹색신호가 들어온다. 서론에서 언급한 고정형 신호 모형은 좌회전 차선에 차량이 없더라도, 좌회전 방향에 20초간 녹색신호를 유지한다.

이러한 한계를 보완하기 위해 시시각각 변하는 교통량에 유연하게 대응하는 적응형 신호 모형이 제안되었다(Luyanda et al., 2003; Kim and Kim, 2019). 적응형 모형은 센서를 통해 실시간 교통량을 수집하고, 동적인 교통상황에 적합한 신호패턴을 산출할 수 있도록 고안되었다. 다만, 기존 연구들은 모형식 기반의 최적화 방법론을 취하고 있어, 네트워크 규모가 커지면 복잡도가 너무 높아져 연산에 한계가 있다(Al Islam and Hajbabaie, 2017).

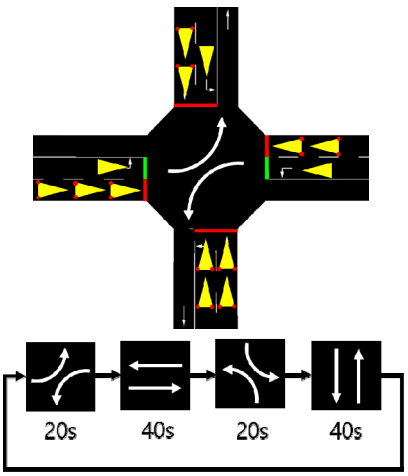


Figure 1. Intersection signal

최근에는 모형식 기반의 신호 최적화 한계를 극복하기 위해 강화학습 알고리즘이 활용되고 있다. 강화학습은 에이전트, 환경, 상태, 행동, 보상 등의 요소가 있다. 에이전트는 학습의 주체를 의미하고, 환경은 에이전트가 학습을 수행하는 공간을 뜻한다. 에이전트는 상태를 관찰하고 그에 따른 행동을 취한다. 그리고 행동에 따른 보상을 에이전트에게 제공한다. 강화학습은 에이전트가 자신이 받을 수 있는 보상을 최대화하도록 신경망을 학습시키는 알고리즘이다(Sutton and Barto, 1998).

강화학습 기반 신호 모형 연구들은 현실과 유사한 교통 시뮬레이션 환경을 구성하고, 신호제어기를 에이전트로 하여 교통정체를 최소화하는 신호패턴을 산출하도록 학습을 수행하였다. 다만, 대부분의 연구가 수학적인 최적화에만 초점이 맞추어져 있다(Kim and Jeong, 2020). 이는 교통량이 극히 적은 도로에는 녹색신호를 주지 않아, 해당 도로의 차량들은 한없이 대기하는 상황을 초래할 수 있다. 총 대기시간 측면에서는 최솟값일 수 있기 때문이다. 또한 신호의 순서가 완전히 교통량에만 반응하여 무작위로 나타날 수 있다. 기존 신호 패턴에 적응한 운전자들에게 무작위 신호는 큰 혼란을 유발한다.

본 연구에서 제안하는 강화학습 기반 교차로 신호 모형은 현실 적용성을 고려한 학습을 수행했다는 점에서 관련 연구와 차이가 있다. 우선, Fig. 1과 같이, 교차로에 적용된 일반적인 신호의 순서를 유지한 채 학습을 수행하였다. 그리고 신호별로 최소 녹색시간을 두어, 한 주기에 반드시 한번은 녹색신호가 켜지도록 하였다.

## 3. 방법론

### 3.1 개요

본 신호 모형은 정체현상 최소화를 목적으로, 실시간 교통상황을 고려한 신호패턴을 산출한다. Fig. 2는 본 모형의 강화학습 과정을 나타낸 것이다. 모형의 입력값, 즉, 상태는 차선별 정지 차량과 같은 실시간 교통상황이다(state). 상태가 입력되면, 모형은 현재 신호를 유지할지 다음 신호로 넘어갈지에 대한 행동을 결정한다(action). 모형이 결정한 행동에 따라 차량들이 통행하

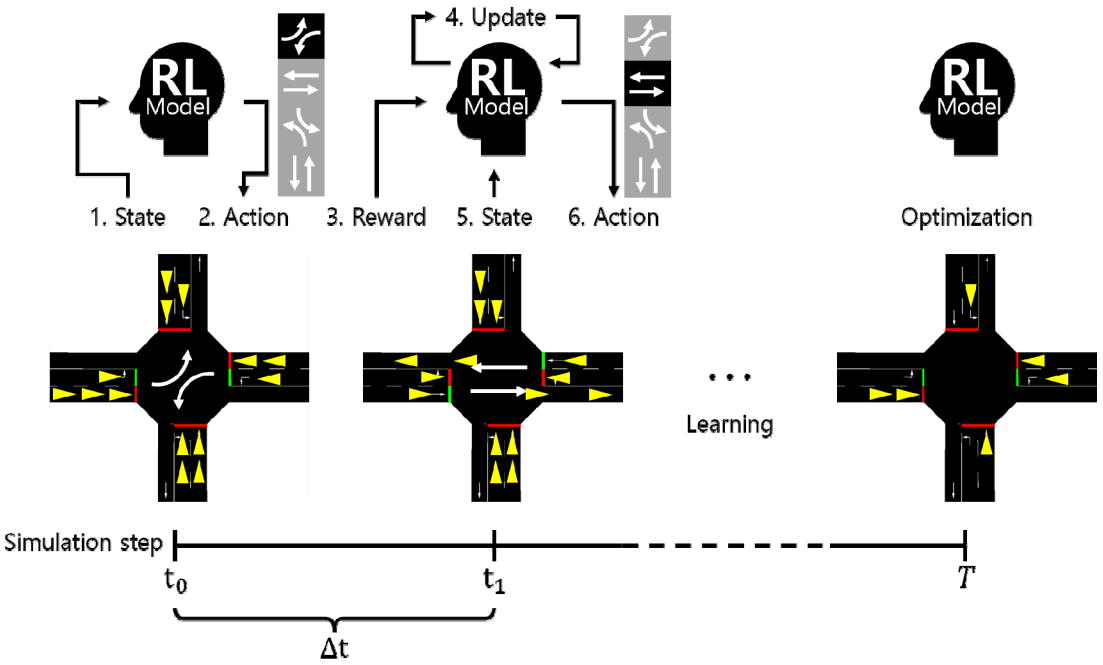



Figure 2. Learning process of the intersection signal model

고($\triangle t$), 줄어든 정체 차량만큼 모형에게 보상이 부여된다(reward). 보상에 따라 모형, 즉, 강화학습 신경망이 업데이트되고(update), 상태를 입력받는 단계부터 반복되면서 최적화를 위한 신호 모형의 학습이 수행된다.

실시간 교통상황을 연출하기 위한 시뮬레이터로는 Simulation of Urban MObility (SUMO)를 이용하였다. SUMO에는 차선, 연속적인 차량의 움직임, 차량과 신호의 상호작용 등이 구현되어 있다. SUMO에서 구현된 실시간 교통상황이 Tensorflow 기반 강화학습 모형에 전달되고, 모형이 결정한 행동이 다시 SUMO에 전달되어 교통흐름을 변화시킨다.

### 3.2 상태(state)

Fig. 3은 임의의 학습 시점 $t$에 대한 교차로를 나타낸 것이다. 횡 방향 도로를 교통량이 많은 주방향으로 가정했을 때, 신호는 주방향 좌회전, 주방향 직진, 부방향 좌회전, 부방향 직진의 4가지 종류가 있다. 현재 녹색신호는 주방향 좌회전이다. 이에 Fig. 3에도 4, 8번 진입 차선에 녹색신호가 적용되어 있다.

$$
\begin{aligned}
S_t &= [Q_t, P_t, d_t] \\
Q_t &= [q_{1,t}, q_{2,t}, \cdots, q_{k,t}] \\
P_t &= [p_1, p_2, \ldots, p_n]
\end{aligned} \tag{1}
$$

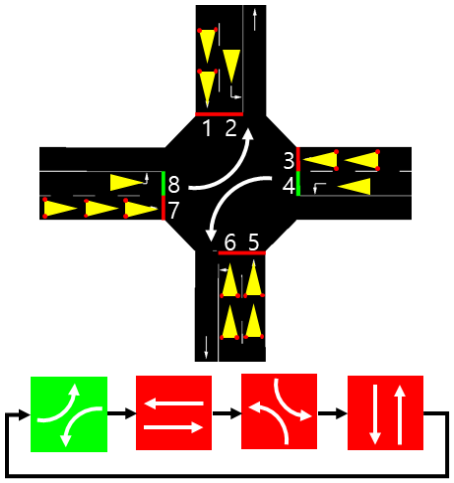



Figure 3. Simulation environment

이 시점에서 신호 모형에 전달되는 상태 $S_t$는 Eq. (1)과 같다. $Q_t$는 $t$시점의 대기행렬 벡터이다. 교차로에 진입하는 차선이 $k$개 있을 때, $q_{i,t}$는 $t$시점에 차선 $i$에 정지한 차량의 수를 의미한다($i \leq k$). Fig. 3에서 교차로에 진입하는 차선은 총 8개이다($k=8$). 따라서 $Q_t = [2,0,2,0,2,2,3,0]$ 이다.

$P_t$는 현재 어느 신호가 켜져 있는지 나타내는 벡터

로, 총 $n$개의 신호가 있을 때, $p_j$는 $j$번째 신호의 녹색 여부이다($j \leq n$). $j$번째 신호가 녹색이라면, $p_j = 1$, 그렇지 않다면 $p_j = 0$이다. 예를 들어 Fig. 3에는 주방향 좌회전에 녹색신호가 켜져 있다. 따라서 $P_t = [1, 0, 0, 0]$이다.

$d_t$는 현재 켜져 있는 신호의 경과 시간으로, Fig. 3에서는 주방향 좌회전 신호의 경과 시간이 적용된다. 결과적으로 상태는 현재 어느 차선이 정체가 심한지, 어떤 신호가 켜져 있는지, 그 신호가 얼마 동안 지속되었는지를 의미한다.

### 3.3 행동(Action)

$t$시점에서 상태를 입력받은 신호 모형은 현재 신호를 유지하거나($A_t = 0$), 다음 신호로 넘어가는($A_t = 1$) 행동 $A_t$를 취한다. Fig. 3을 예시로 들면, 주방향 좌회전을 유지할지, 혹은 다음 신호인 주방향 직진으로 녹색신호를 변경할지에 해당한다. 모형은 주어진 두 가지 선택지 중 더 큰 보상을 기대할 수 있는 행동을 확률적으로 선택하게 된다. 다만, 모형의 선택이 다음 신호로 변경($A_t = 1$)이더라도, 현재 신호가 최소 녹색시간을 만족하지 않았다면, 모형의 선택과는 별개로 현재 신호를 유지하게 된다.

### 3.4 보상(Reward)

모형이 행동을 결정한 후, $\triangle t$ 시간만큼 시뮬레이션을 수행하고, 행동 $A_t$에 따른 보상 $R_{t+\triangle t}$를 모형에게 적용한다. 보상 $R_{t+\triangle t}$는 Eq. (2)와 같다. $q_{i,t}$는 $t$시점에 진입 차선 $i$에 정지한 차량의 수를 말한다.

$$R_{t+\triangle t} = -\sum_{i=1}^{k}(q_{i,t+\triangle t}) \tag{2}$$

따라서 $R_{t+\triangle t}$는 $t+\triangle t$ 시점의, 모든 진입 차선의 정지 차량의 총합에 음의 부호를 곱한 값이다. 모형은 더 큰 보상을 얻도록 학습을 수행하기 때문에, 정지 차량을 더 많이 줄일 수 있는 행동을 학습하게 된다. 즉, 모형은 교차로에 정지하고 있던 차량들을 최대한 많이 통과시킬 수 있는 신호를 선택하는 것이다.

### 3.5 업데이트

Fig. 4는 모형의 업데이트 과정을 나타낸 것이다. 신호 모형의 학습이란, 주어진 상태에서 선택할 수 있는 행동들에 대한 보상값을 잘 예측하도록 학습시키는 것을 의미한다. $t$에서 $T$까지 시뮬레이션이 진행되는 동안, 모형은 누적되는 보상을 최대화하도록 $\triangle t$ 주기로 학습을 수행한다. 따라서 목적함수는 Eq. (3)과 같고, 이는 정체 차량의 수를 최소화하는 것과 동일한 의미이다.

$$\max \sum_{t=t_0}^{T}\left\{-\sum_{i=1}^{k}(q_{i,t})\right\} = \min \sum_{t=t_0}^{T}\left\{\sum_{i=1}^{k}(q_{i,t})\right\} \tag{3}$$

모형은 Deep Q-network(DQN) 알고리즘을 적용하였다. 입력층과 출력층은 각각 상태벡터, 행동벡터의 차원과 동일한 개수로 노드를 구성하였다. 은닉층은 400개의 노드를 fully connected 방식으로 5층을 연결하여 구성하였다. 은닉층의 활성화 함수는 rectified linear unit(ReLU)를 적용하였고, 출력층의 활성화 함수는 이진 분류에 적합한 sigmoid를 적용하였다. 손실함수는 cross entropy, 학습률은 $2.5 \times 10^{-4}$를 적용하였다.

## 4. 시뮬레이션

### 4.1 학습 시나리오

Table 1은 강화학습 기반 신호 모형의 학습 시나리

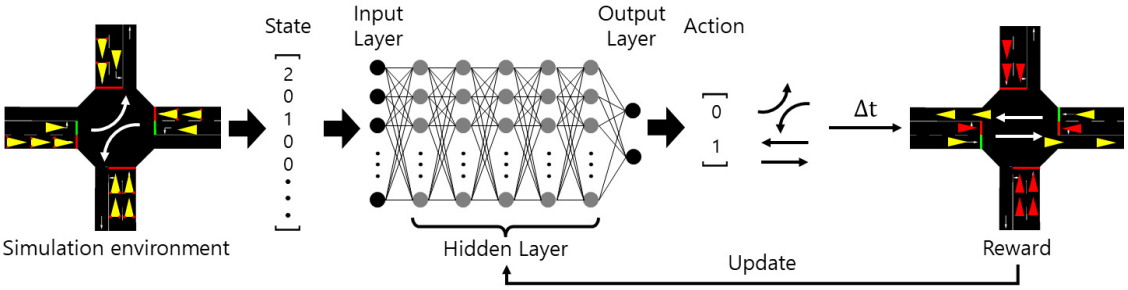
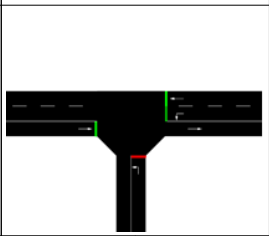
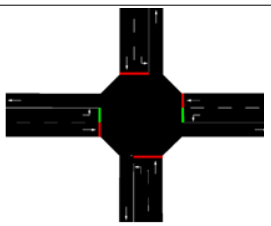
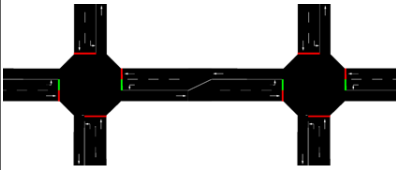
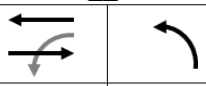
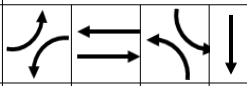
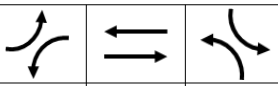



Figure 4. Update process of DQN based signal model

Table 1. Training scenario

| | Case A | | Case B | | | | Case C | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Network structure |  | |  | | | |  | | | |
| Signal sequence |  | |  | | | |  | | | |
| Initial green time (second) | 50 | 50 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| Min green time(second) | 30 | 10 | 10 | 24 | 10 | 24 | 10 | 24 | 10 | 24 |
| Traffic (veh/hour) | Random traffic between (0, 300) generated for each direction | | | | | | | | | |
| $\triangle t$ (second) | 1 | | | | | | | | | |
| $T$ (second) | 100,000 | | | | | | | | | |
| Episode | 100 | | | | | | | | | |

오이다. 시나리오는 3지 교차로 형태의 케이스 A, 4지 교차로 형태의 케이스 B, 2개의 연속된 4지 교차로에서 학습을 진행한 케이스 C로 구성된다. 케이스 C는 각 교차로 당 하나의 신호 모형이 있는, 즉, 2개의 독립적인 신경망을 학습시키는 시나리오이다.

케이스 C는 다른 케이스들과 달리 주방향에 비보호 좌회전이 포함되어 있다. 학습을 통해 결정되는 것은 각 신호별 녹색시간이고, 신호별 초기 녹색시간(initial green time)은 100초를 신호의 개수만큼 균일하게 나누어 적용하였다. 학습에 적용한 교통량은 1시간 단위로 랜덤하게 생성하였고, 방향별로 차량이 최소 0대에서 최대 300대까지 나타나게 하였다.

최소 녹색시간은 일반적인 교차로의 경우 녹색신호 7초와 황색신호 3초를 포함하여 총 10초가 부여된다. 횡단보도가 존재할 경우, 도로의 폭과 보행속도를 고려한 횡단시간이 최소 녹색시간에 포함된다. 본 연구에서는 관련 연구 및 교통 매뉴얼을 참고하여 교차로 형태에 따른 최소 녹색시간을 적용하였다(Kang and Oh, 2004).

모든 케이스에 대하여 학습 주기는 1초이고, 1회 시뮬레이션(episode)은 100,000초까지 수행하였다. 학습 주기가 1초인 이유는 실시간으로 교통량을 수집하고, 상황에 맞는 대응 역시 실시간으로 수행하게끔 신호 모형을 학습시키기 위함이다. 각 케이스별 시뮬레이션, 즉, 에피소드는 100회씩 반복하였다.

### 4.2 학습 결과

우선, 모형이 충분히 학습되어 수렴된 해를 도출하는지 판단하기 위해, Fig. 5와 같이 모든 케이스에 대하여 각 에피소드별 누적 대기시간을 살펴보았다. 대체로 모든 케이스에서 10번째 에피소드 이전에 수렴된 결과가 나타났다. 이는 한 회 시뮬레이션 시간이 100,000초이기 때문에, 일부 에피소드만 진행하여도 모형이 충분히 학습되는 것으로 판단된다.

학습된 신호 모형의 평가를 위해 전통적인 교통이론 기반 최적화 프로그램인 PASSER Ⅱ와 비교를 수행하였다. 신호 모형의 학습이 잘 수행되었다면, 유동적인 교통상황에 반응하여 고정형 신호 모형보다 더 원활한 통행을 기대할 수 있다. Table 2는 전통 모형과 본 모형의 성능 비교에 활용된 실험 교통량과 각 모형을 통해 산출된 신호패턴을 나타낸다. PASSER Ⅱ에 주어진 교차로 구조와 1시간 교통량을 입력하면, Table 2와 같은 신호패턴을 얻을 수 있다. 본 모형의 결과는 SUMO를 통해 시뮬레이션 환경(교차로, 교통량)을 구축하고, 학습된 신경망으로 1시간 동안 신호를 제어한 결과이다.

전통 모형은 고정형 패턴을 산출하기 때문에 일정한 신호주기가 반복된다. 신호주기란 각 방향별 신호 순서(현시순서)를 한번 완료하는 데 필요한 시간을 의미한다. 전통 모형은 모든 케이스에 대하여 신호주기를 110초로 결정하였고, 교통량이 많은 방향에 녹색시간의 비

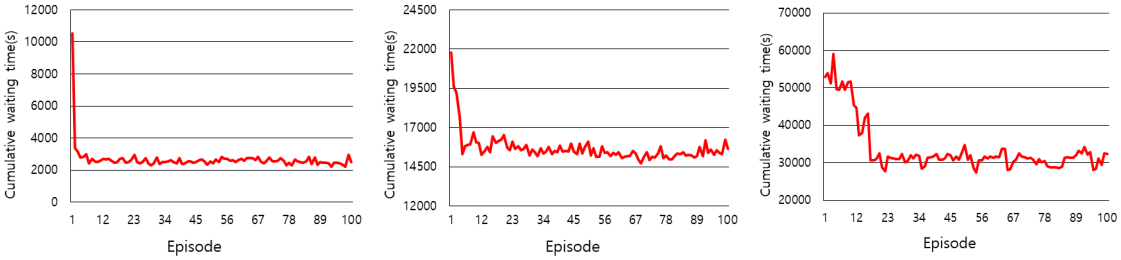
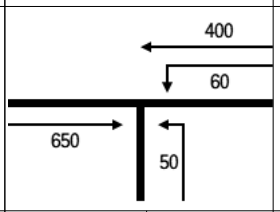
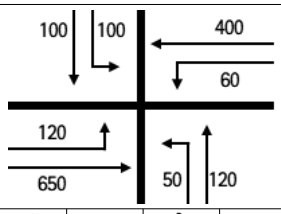
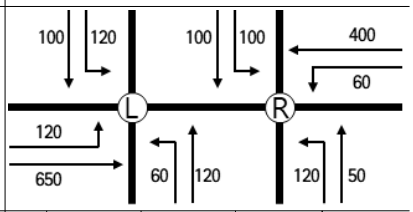
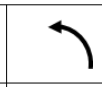



Figure 5. Cumulative waiting times(s) each episode; (a) Case A, (b) Case B, (c) Case C

Table 2. Test data and optimal signal patterns

| 구분 | Case A | | Case B | | | | Case C | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Test traffic (veh/hour) |  | |  | | | |  | | | | |
| Signal sequence |  |  |  |  |  |  | L |  |  |  |  |
| | | | | | | | R | | | | |
| Duration(s) [PASSER Ⅱ] | 97 | 13 | 14 | 56 | 13 | 27 | L | 14 | 56 | 13 | 27 |
| | | | | | | | R | | | | |
| Duration(s) [RL model1] (mean/st.dev) | 54.1/4.4 | 10.0/0.2 | 10.0 /0.1 | 34.9 /4.6 | 10.3 /0.2 | 24.0 /0.1 | L | 10.0 /0.1 | 53.7 /11.9 | 10.1 /0.9 | 24.0 /0.1 |
| | | | | | | | R | 10.0 /0.1 | 53.2 /11.5 | 10.0 /0.1 | 24.1 /0.1 |

율을 높게 산정하였다. 또한 케이스 B와 케이스 C에 대해서는 동일한 패턴을 산출하였고, 케이스 C의 왼쪽과 오른쪽 교차로도 동일한 패턴을 적용하였다.

본 신호 모형은 적응형이기 때문에, 신호주기의 길이와 각 신호별 녹색시간이 일정하지 않다. 이에 Table 2의 신호별 녹색시간은 평균과 표준편차를 갖는다. Fig. 6는 이를 보다 직관적으로 나타낸다. Fig. 6는 한 시간 동안 관측된 신호주기별 녹색시간 비율을 나타낸 것이다. x축은 신호주기, y축은 해당 신호주기에서의 각 신호별 녹색시간 비율이다. 녹색시간 비율의 변화가 크다는 것은 실시간 교통상황에 동적으로 대응하는 신호패턴을 산출했다는 의미로 해석할 수 있다.

본 모형은 모든 케이스에 대하여, 교통량이 적은 좌회전 차선에는 최소 녹색시간만 부여하였다. 이는 최소 녹색시간만으로 좌회전 차선의 교통량을 감당할 수 있고, 교통량이 많은 직진 방향에 녹색신호를 주는 것이 보상을 증가시키는데 더 유리하다고 모형이 학습했기 때문이다.

Fig. 7은 3가지 모형을 SUMO에 적용하여 1시간 동안 측정한 모든 차량의 누적 대기시간과 누적 정지횟수를 나타낸 것이다. 차량의 대기시간과 정지횟수는 신호모형의 효과를 평가하는데 매우 핵심적인 척도이다 (Roess et al., 2004). 3가지 모형은 PASSER II, 본 신호 모형(RL model1), 강화학습 기반 비교 모형(RL model2)으로 구성하였다. 강화학습 기반 비교 모형은 본 모형과 동일한 DQN 알고리즘을 적용하였지만, 신호 순서 유지 조건과 모든 신호가 한 주기에 반드시 한 번은 녹색시간이 유지되는 조건을 제거하고 학습된 모형이다.

케이스 A에서 C 순서로 누적 대기시간은, 본 모형이 PASSER II에 비해, 54%, 8%, 27% 감소하였다. 이와 같은 결과가 나타난 이유는 해당 차선에 차량이 없더라도 녹색신호가 적용되어, 다른 차선에 대기행렬을 발생시키는 고정형 모형과 달리, 강화학습 기반 신호 모형은 주어진 상황에 맞는 대응이 가능하기 때문에 누적 대기시간을 현저히 감소시킨 것으로 판단하였다.

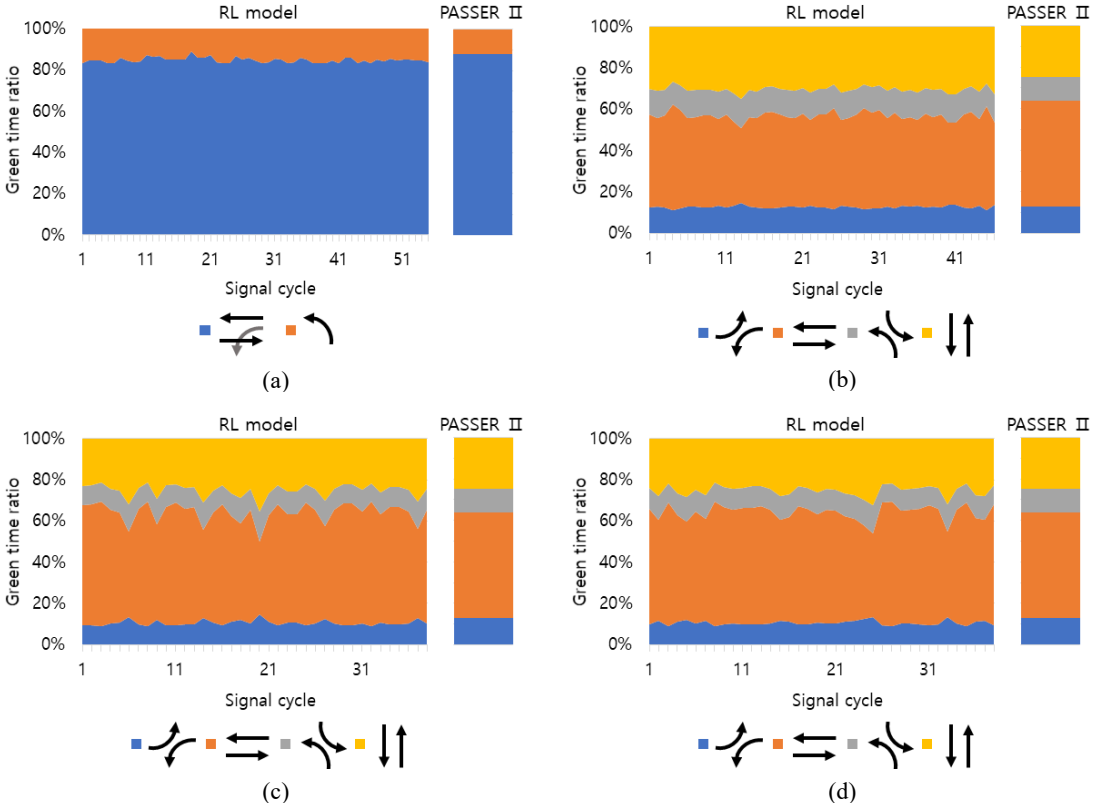



Figure 6. Green time ratio by case; (a) Case A, (b) Case B, (c) Case C: left intersection, (d) Case C: right intersection

반면, 비교 모형에 비해 본 모형의 대기시간은 31%, 68%, 33% 크게 나타났다. 이는 신호의 순서를 유지하고 최소 녹색시간으로 인해 대기차량이 많은 차선에 즉각적으로 녹색 신호를 부여할 수 없기 때문이다.

케이스 A와 B에서 전통 모형의 주기는 110초이고, 본 모형은 약 65초, 80초이다. 이는 강화학습 모형을 적용할 경우, 신호가 더 자주 바뀐다는 것을 의미한다. 이는 교차로에 정지한 차량의 수만큼 음의 보상을 받기 때문에, 정체현상(대기시간)을 줄이는 방향으로 학습을 수행하였지만, 자주 신호를 바꿈으로써 오히려 차량들의 정지 빈도는 증가했다고 볼 수 있다.

케이스 C에서 PASSER II에 비해 비교 모형의 대기시간은 절반가량으로 감소했지만 정지횟수는 유사하게 나타났다. 즉, 신호의 순서가 없고 짧은 신호시간으로

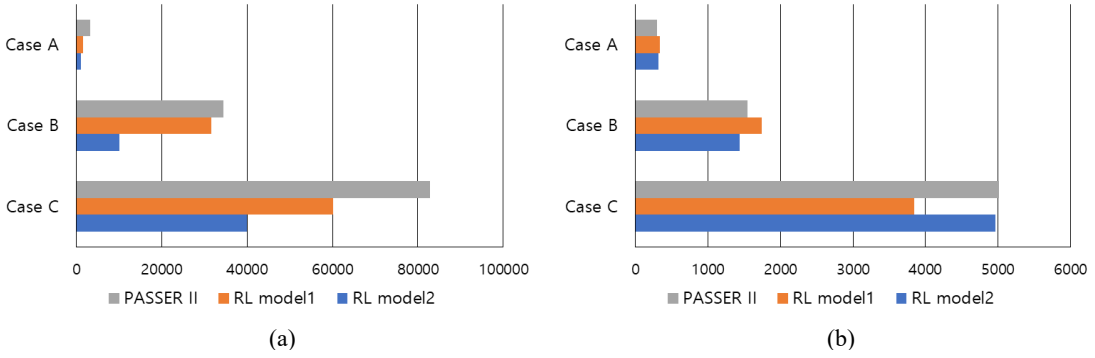



Figure 7. Measure of effectiveness; (a) cumulative waiting times(s), (b) cumulative number of stops

인해 나타난 것으로 이와같은 신호 패턴은 차량간의 사고를 유발할 수 있기에 지양되어야 한다.

반면, 케이스 C에서는 본 모형을 적용했을 때 가장 이상적인 결과가 나타났다. 차량들의 대기시간과 정지횟수 모두 감소하였기 때문이다. 본 모형을 통해 산출된 신호주기는 약 100초로, 기존 모형과 유사하다. 이는 신호를 자주 바꾸지 않으면서, 실시간 교통상황에 적절한 대응을 했다고 판단할 수 있다.

## 5. 결 론

본 연구는 강화학습 알고리즘을 이용하여 차량의 정체를 최소화하는 교차로 신호 모형을 제안하였다. 기존 머신러닝 기반의 신호 모형 연구들은 수학적인 최적화에만 초점을 맞추어 현실에 적용하는데 한계가 있다. 예를 들어, 총 대기시간 최소화를 위해, 교통량이 적은 도로에는 녹색신호를 거의 주지 않는 케이스, 일정한 순서 없이 교통량에만 반응하는 녹색신호를 주어 운전자들에게 혼란을 야기할 수 있는 케이스 등이 있다. 본 연구에서는 이러한 현상을 방지하기 위해 일반적인 교차로 신호 순서를 지키면서, 각 신호는 최소 녹색시간을 갖도록 가정하고, 모형의 학습을 수행하였다.

일반적인 3가지 형태의 교차로 환경에서, 마이크로 교통 시뮬레이터를 이용하여 기존 전통 모형과 본 모형의 평가를 수행하였다. 누적 대기시간은 모든 케이스에서 전통 모형보다 효과적인 것으로 나타났고, 누적 정지횟수는 일부 케이스에서 증가된 결과를 보였다. 신호모형은 교통량에 반응하여 신호패턴을 결정하기 때문에, 전통 모형보다 신호주기가 짧게 나타날 수 있다. 이는 신호가 자주 바뀌는 것을 의미하고, 이로 인해 차량의 정지횟수는 증가하는 결과가 나타났다. 기존 전통 모형과 신호주기가 유사하게 결정된 케이스에서는, 대기시간과 정지횟수 모두 감소되는 가장 이상적인 결과를 보였다.

강화학습 기반 교차로 신호 모형은 행동에 대한 선택을 분류하는 가치함수 기반의 알고리즘으로, 향후 교차로 간의 신호시간 등을 조절하는 모형에서는 Policy Gradient 등과 같은 정책기반의 강화학습 알고리즘을 적용할 필요가 있다. 또한 은닉층의 구성, 활성화 함수, 상태 및 보상 변수 등 다양한 방향의 모형 튜닝을 고려하지 않았다. 향후 연구를 통해, 모형의 성능을 최적화할 수 있는 파라미터 튜닝, 대규모 네트워크 및 실제 도로 데이터 반영 등이 이루어진다면, 보다 우수한 성능의 모형을 도출할 수 있을 것으로 판단한다.

## 감사의 글

본 연구는 국토교통부 국토공간정보연구사업의 연구비지원(20NSIP-B135746-04)에 의해 수행되었습니다.

## References

1. Al Islam, S. B. and Hajbabaie, A., 2017, Distributed coordinated signal timing optimization in connected transportation networks, Transportation Research Part C: Emerging Technologies, Vol. 100, No. 80, pp. 272-285.
2. Arel, I., Liu, C., Urbanik, T. and Kohls, A. G., 2010, Reinforcement learning-based multi-agent system for network traffic signal control, IET Intelligent Transport Systems, Vol. 4, No. 2, pp. 128-135.
3. Cai, C., Wong, C. K. and Heydecker, B. G., 2009, Adaptive traffic signal control using approximate dynamic programming, Transportation Research Part C: Emerging Technologies, Vol. 17, No. 5, pp. 465-474.
4. Chang, E. C. P., Messer, C. J. and Marsden, B. G., 1985, Reduced-delay optimization and other enhancements in the PASSER II-84 program, Transportation Research Board, No. 105, pp. 80-89.
5. Kang, D. M., and Oh, Y. T., 2004, The method of the phase split adjustment considering the minimum green time in COSMOS, Journal of Korean Society of Transportation, Vol. 22, No. 7, pp. 147-154.
6. Kim, D. and Jeong, O., 2020, Cooperative traffic signal control with traffic flow prediction in multi-intersection. Sensors, Vol. 20, No. 1, pp. 137.
7. Kim, J. and Kim, Y., 2019, Development of network-wide traffic signal control strategy for preventing bockage at intersection, Journal of Korean Society of Transportation, Vol. 37, No. 2, pp. 178-192.
8. Liang, X., Du, X., Wang, G. and Han, Z., 2018, Deep reinforcement learning for traffic light control in vehicular networks, Machine Learning, Vol. 68, No. 2, pp. 1-11.
9. Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D. and Mirchandani, P., 2003, ACS-Lite algorithmic architecture: applying adaptive control system technology to closed-loop traffic signal control systems, Transportation Research Record, Vol. 1856, No. 1, pp. 175-184.

10. Mannion, P., Duggan, J. and Howley, E., 2016, An experimental review of reinforcement learning algorithms for adaptive traffic signal control, Autonomic Road Transport Support Systems, No. 4, pp. 47-66.
11. Mousavi, S. S., Schukat, M. and Howley, E., 2017, Traffic light control using deep policy-gradient and value-function-based reinforcement learning, IET Intelligent Transport Systems, Vol. 11, No. 7, pp. 417-423.
12. Pandit, K., Ghosal, D., Zhang, H. M. and Chuah, C. N., 2013, Adaptive traffic signal control with vehicular ad hoc networks, IEEE Transactions on Vehicular Technology, Vol. 62, No. 4, pp. 145-147.
13. Rasheed, F., Yau, K. L. A. and Low, Y. C., 2020, Deep reinforcement learning for traffic signal control under disturbances: A case study on Sunway city, Malaysia, Future Generation Computer Systems, Vol. 109, pp. 431-445.
14. Roess, R. P., Prassas, E. S. and McShane, W. R., 2004, Traffic engineering, Pearson/Prentice Hall, pp. 102-114.
15. Sutton, R. S. and Barto, A. G., 1998, Introduction to reinforcement learning, Cambridge: MIT press, pp. 58-64.
16. 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.