

차량과 보행자를 고려한 강화학습 기반 적응형 교차로 신호제어 연구

김종민* · 김선용**

Reinforcement Learning-Based Adaptive Traffic Signal Control considering
Vehicles and Pedestrians in Intersection

Jong-Min Kim* · Sun-Yong Kim**

요 약

원활한 교통의 흐름은 현대 사회에서 매우 중요한 요소이며, 교통체증은 환경 및 경제 등 다양한 형태로 문제를 초래했다. 이러한 문제를 해결하기 위해 최근 인공지능을 활용한 지능형교통체계(Intelligent Transport System)가 주목받고 있다. 본 논문에서는 강화학습 기법을 활용하여 교차로 각 방향의 차량과 보행자를 동시에 고려하여 교통의 흐름을 원활하게 하면서 동시에 불만족도를 낮출 수 있는 알고리즘을 제안한다. 제안하는 알고리즘을 적용한 교차로 신호체계 시뮬레이션 결과, 기존의 고정형 신호체계에 비해 차량·보행자의 불만족도를 상당히 낮출 수 있으며, 교차로의 교차하는 도로의 수가 많을수록 성능의 차이가 증가함을 확인하였다.

ABSTRACT

Traffic congestion has caused issues in various forms such as the environment and economy. Recently, an intelligent transport system (ITS) using artificial intelligence (AI) has been focused so as to alleviate the traffic congestion problem. In this paper, we propose a reinforcement learning-based traffic signal control algorithm that can smooth the flow of traffic while reducing discomfort levels of drivers and pedestrians. By applying the proposed algorithm, it was confirmed that the discomfort levels of drivers and pedestrians can be significantly reduced compared to the existing fixed signal control system, and that the performance gap increases as the number of roads at the intersection increases.

키워드

Reinforcement Learning, Q-learning, Intelligent Transport System, Traffic Signal System, Discomfort Level
강화학습, Q 러닝, 지능형 교통 체계, 신호 체계, 불쾌 지수

1. 서 론

원활한 교통의 흐름은 현대 사회에서 매우 중요한

요소이다. 2010년에 텍사스 A&M 교통 연구소에서 다양한 형태의 손실을 화폐가치로 환산한 교통혼잡비용이 연간 약 1,150억 달러로 추산하였으며[1], 한국

* 동서대학교 학부생 (clay9502@naver.com)

** 교신저자 : 동서대학교 소프트웨어학과

• 접수 일 : 2023. 12. 28

• 수정완료일 : 2024. 01. 20

• 게재확정일 : 2024. 02. 17

• Received : Dec. 28, 2023, Revised : Jan. 20, 2024, Accepted : Feb. 17, 2024

• Corresponding Author : Sun-Yong Kim

Dept. of Software, Dongseo University

Email : sykim@dongseo.ac.kr

또한 한국교통연구원 보도자료에 의하면 꾸준히 증가하고 있다. 교통체증은 비용뿐만 아니라 환경에도 악영향을 끼치는데, 교통체증을 수치화한 교통체증지수가 높을수록 미세먼지가 증가하며 주요 원인으로 내연 기관 차량의 공회전이라는 연구 결과가 있다[2]. 이러한 문제는 원활하지 못한 교통의 흐름에 있는데, 국토교통부에서 집계한 자료에 의하면 대한민국의 자동차 등록 대수 또한 꾸준히 증가하고 있으며, 정부의 ‘제4차 친환경자동차 기본 계획’ 등의 정책과 함께 전기자동차 보급이 가속화될 것으로 예측되어 여전히 많은 차량으로 인해 교통체증이 유발될 것으로 추측된다[3]. 원활한 교통의 흐름을 만들기 위해 여러 해결 방법이 제안되었는데, 그 중 대표적인 방법이 신호 교차로를 회전교차로로 전환하는 것이다. 하지만 이는 2014년에 국토교통부가 발표한 ‘회전교차로 설계지침’에 의하여 교통량이 많은 교차로의 경우 회전교차로 설치가 금지되었다[4]. 또한, 신호 교차로에 비해 큰 비용이 필요로 하고, 교차로에 필요한 면적이 넓은 단점이 있으며, 운전자들의 회전교차로 우선주행 미숙지로 인해 교통 혼란을 유발하여 회전교차로의 효과를 방해하기도 한다. 최근 교통 문제를 해결하기 위한 정책 중 하나로 인공지능을 활용한 지능형교통체계(Intelligent Transport System, 이하 ITS)가 주목받고 있다. ITS란 교통수단 및 교통시설에 대하여 전자·제어 및 통신 등 첨단 교통기술과 교통정보를 개발·활용함으로써 교통체계의 운영 및 관리를 과학화·자동화하고, 교통의 효율성과 안전성을 향상시키는 교통체계를 의미한다[5]. 미국의 경우 1991년 육상교통효율화법을 시작으로 2002년 대규모 ITS 프로젝트인 ‘Vehicle Infrastructure Integration’을 진행하면서 인프라를 구축하고 있으며[6], 일본의 경우 1980년대부터 ITS 관련 부처들이 프로젝트를 진행하였다[7]. 국내에서는 1994년 고속국도 ITS 구축 시범사업(경부선)을 시작으로 4단계의 추진단계로 나누어 정부의 핵심 사업 중 하나로 추진되고 있다[8].

본 논문에서는 신호 교차로 환경에서 ITS의 핵심 요소 중 하나인 적응형 신호체계에 주목하였으며, 적응형 신호체계에 강화학습 알고리즘을 적용하는 것을 제안한다. 성능 분석을 위해 현실과 유사한 시뮬레이션 환경을 구축하였으며, 강화학습 알고리즘으로 Q-Learning 알고리즘을 제안·적용하였다. 제안한 알고리즘은 대한민국에서 흔히 사용되는 기존의 고정형 신호체계와의 비교를 위해 시뮬레이션한 뒤, 불쾌지수를 통해 알고리즘의 성능을 비교하고 평가하였다.

II. 강화학습 기반 교차로 신호제어 기법

2.1 시스템 환경

본 논문에서는 적응형 교차로 신호제어 알고리즘으로 강화학습 기법을 활용하여 교차로 각 방향의 차량과 보행자들의 대기시간에 따른 불편함을 수치화한 불쾌지수를 최소화하도록 신호를 제어하는 방법에 대해 기재한다.

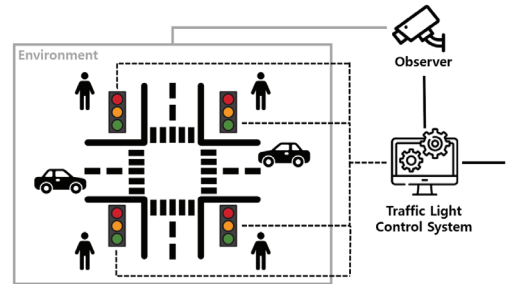


그림 1. 교차로 신호제어 시스템

Fig. 1 Traffic signal control system in intersection

그림 1은 본 논문에서 제안하는 교차로 신호제어 시스템을 보여준다. 교차로는 방향별로 통행을 위해 신호를 기다리는 차량과 보행자가 존재하며, 차량과 보행자는 시간대별로 각 방향에 따른 패턴화된 통행량을 가진다. 패턴화된 통행량 수치는 실제 도로의 구간통행시간이 정규 분포 및 쌍봉을 갖는 분포와 유사한 형태를 가지기에 가우스 분포(Gaussian Distribution) 따른다고 가정한다[9].

표 1. 패턴 예시

Table 1. Pattern example

Direction	Time 1	Time 2	...	Time m
I_1	10	9	...	5
I_2	1	2	...	2
I_3	5	2	...	5
\vdots	\vdots	\vdots	...	\vdots
I_n	t_1^n	t_2^n	...	t_m^n

표 1은 방향 I_n 에서 m 시간대의 통행량을 나타낸 것이다. 해당 수치는 방향별 추가되는 보행자와 차량의 주기로, 한 시간대가 s 초라고 할 때, 시간대 m 의 총통행량 $T(I_n, m)$ 은 수식 (1)과 같이 계산된다. 이때,

$N(t_m^n, \sigma^2)$ 는 가우스 분포를 따르는 t_m^n 을 뜻한다.

$$T(I_n, m) = \frac{s}{N(t_m^n, \sigma^2)} \quad \dots (1)$$

예를 들어, 한 시간대가 1,000초이고 방향 I_1 의 3번째 시간대의 차량 패턴 t_3^1 의 값이 5라고 하면,

$$T(I_1, 3) = \frac{1000}{N(t_3^1, \sigma^2)} \approx 200 \quad \dots (2)$$

와 같이 약 5초마다 차량 한 대가 방향 I_1 에 발생하며, 총통행량 $T(I_1, 3)$ 은 약 200대로 계산된다.

차량 또는 보행자는 불편함을 수치화한 불쾌지수 D_{obj} 를 가지는데, 대기시간이 길어질수록 불쾌함을 느끼도록 수식 (3)과 같이 모델링 한다[10].

$$D_{obj} = \beta(S_{obj} - C_{obj})^\eta \quad \dots (3)$$

수식 3에서 S_{obj} 은 환경에 등장한 시점, C_{obj} 은 현재 시점을 뜻하며, 예를 들어 $\beta=1$, $\eta=2$ 인 경우, 대기시간이 길어질수록 2차 함수 형태의 불쾌지수 D_{obj} 을 가진다. 차량과 보행자는 서로 다른 패턴의 통행량과 함께, 통행을 위한 신호 또한 다르게 가진다. 이는 대한민국의 교통신호체계에서 흔히 사용되는 신호체계를 기준으로 하였으며, 보행신호는 차량 신호와 연동되어 평행한 방향의 우측 신호가 함께 동작하도록 하였다. 이러한 환경에서 본 논문에서 제안하는 알고리즘은 교차로의 신호를 적재적소에 보내어 차량과 보행자의 불쾌지수를 모두 최소화하는 것을 목표로 한다. 따라서 주어진 환경에서 차량 전체의

불쾌지수인 $\sum_{j=1}^J D_j^{veh}$ 와 보행자 전체의 불쾌지수인

$\sum_{k=1}^K D_k^{ped}$ 을 동시에 최소화하는 R (Reward)를 찾는 문제가 되며, 이를 수식 (4)로 나타낼 수 있다[10]. 이때, $0 \leq w \leq 1$ 는 차량과 보행자 간에 가중치를 조절하는 매개변수이다.

$$R = w \sum_{j=1}^J D_j^{veh} + (1-w) \sum_{k=1}^K D_k^{ped} \quad \dots (4)$$

학습을 위한 R 을 차량과 보행자 전체의 통행량이 아닌 불쾌지수를 사용하는 이유는 전체의 통행량으로 R 을 계산할 시, 통행량이 현저히 적은 방향의 경우 신호를 전혀 받지 못할 경우가 발생할 수 있다. 따라서 대기시간을 기반한 불쾌지수를 사용하여 이러한 문제를 해결하였다.

2.2. 제안하는 적응형 교차로 신호제어 알고리즘

본 논문에서는 차량과 보행자를 고려한 강화학습 기반 적응형 교차로 신호제어 알고리즘으로 Q-Learning 알고리즘을 사용할 것을 제안한다. Q-Learning 알고리즘은 모델프리(model-free) 강화학습 중 하나로, 유한 마르코프 결정 과정(Markov Decision Process, MDP)에서 Agent가 최적의 정책(Policy)을 찾는 것을 목표로 한다[11]. Q-Learning 알고리즘은 어떠한 상태(State) s 에서 Agent가 어떠한 행동(Action) a 을 함으로써 가져다줄 수 있는 기댓값을 예측하는 함수 $Q(s, a)$ 를 학습함으로써 Q-Value를 구한다[12]. 이는 수식 (5)로 나타낼 수 있다. 이때 α 는 Learning Rate를, γ 는 Discount Factor를 뜻한다.

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha[R + \gamma \min_{a'} Q(s', a')] \quad \dots (5)$$

Q-Learning 알고리즘에서는 미래를 알고 있다고 가정한다. 즉, 수식 (5)에서의 $\min_{a'} Q(a', s')$ 은 미래의 상태 s' 에서 행동 a' 을 했을 때의 Q-Value의 최솟값을 말한다.

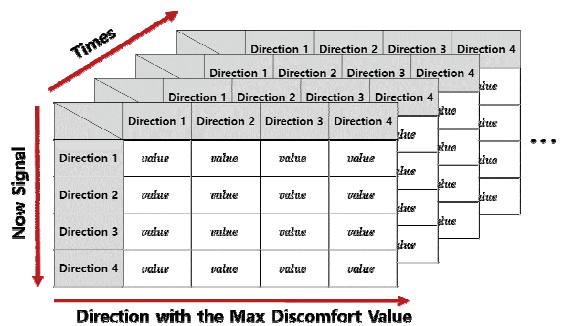


그림 2. 제안하는 Q-Table 예시 구성도
Fig. 2 Propose Q-Table example configuration chart

Q-Value는 Table 형태로 관리가 되는데, 이를 Q-Table이라고 한다. 그림 2는 4개의 방향을 가지는 사거리 교차로 경우에서의 Q-Table을 구성도로 나타낸 것이다. 제안하는 적응형 교차로 신호제어 알고리즘에서는 Q-Table이 3차원 형태의 Table을 가지며, 시간대 $m \in 1, \dots, M$ 과 현재 교차로에서 신호가 주어진 방향 $n \in 1, \dots, N$, 불쾌지수가 가장 높은 방향 $n^* \in 1, \dots, N$, 총 $M \times N \times N$ 개의 Q-Value를 가진다. Agent는 시간대 m 에서 교차로 신호가 주어진 방향이 n 일 때(State), Q-Table에 근거하여 신호를 제어하고(Action), 교차로 내의 차량과 보행자의 불쾌지수 총합인 R 을 구해, 이를 최소화하도록 학습하고 반복한다. 이때, Agent는 교차로 환경에서 같은 시간대에 오직 하나의 신호만 주도록 제어하고 Q-Table에서 수식 (5)에 의하여 시간대 m , 교차로에서의 현재 신호 n 에서의 $n^* \in 1, \dots, N$ 의 최솟값을 $Q(a', s')$ 로 가진다. 이런 일련의 과정을 의사코드로 표현하면 그림 3과 같이 표현할 수 있다. 이때, 탐색(Exploration)과 활용(Exploitation)을 하기 위해 ϵ -Greedy 정책을 사용한다.

Algorithm Q-Learning-based Proposed Algorithm

```

1: // Initialization
2:  $N \leftarrow \#$  of Directions in intersection,
3:  $M \leftarrow \#$  of Times
4:  $\epsilon \leftarrow$  Explore weight
5: DECLARE  $V_{M,N,N^*}$ : Q-Table
6: // Main Loop
7: for  $i = 1$  to # of Learning iterations do
8:   for Loop by Time interval do
9:      $L_n \leftarrow$  List the discomfort value in each directions
10:     $n^* \leftarrow \arg \max L_n$ 
11:     $R \leftarrow \sum_{n=1}^N L_n$ 
12:    // Update  $V_{M,N,N^*}$ 
13:     $V_{m,n,n^*} \leftarrow (1 - \alpha) \cdot V_{m,n,n^*} +$ 
14:       $\alpha \cdot (R + (\gamma \cdot V_{m,n,n^*}'))$ 
15:     $x \leftarrow$  Random from  $0 \leq x \leq 1$ 
16:    if  $x < \epsilon$  then
17:       $n \leftarrow$  Random from  $(1, \dots, N)$ 
18:    else
19:       $n \leftarrow \arg \min_{n^*} V_{m,n,n^*}$ 
20:    Now signal  $\leftarrow D_n$ 

```

그림 3. 제안하는 강화학습 기반 적응형 교차로 신호제어 Q-Learning 알고리즘

Fig. 3 Propose reinforcement learning-based adaptive traffic signal control Q-Learning algorithm

III. 성능평가

3.1 시뮬레이션 환경

본 논문에서 제안하는 강화학습 기반 적응형 교차로 신호제어 Q-Learning 알고리즘을 성능 평가하기 위해 시뮬레이션 환경을 표 2와 같이 가정하고 세팅하였다.

표 2. 시뮬레이션을 위한 Hyper-Parameter 세팅 값
Table 2. Hyper-Parameter settings for simulation

Name	Values
Learning iteration	$i = 100000$ (sec)
Time interval	1000 sec
Weight between the vehicle and the pedestrian	$w = 0.5$
Intersection Type (Direction)	$N = 4$
Learning rate	$\alpha = 0.2$
Discount factor	$\gamma = 0.9$
ϵ -greedy parameter	$\epsilon = 0.1$

시뮬레이션에 사용될 패턴은 10개의 시간대를 가진다고 가정하고 1,000초마다 시간대가 변경된다. 각 방향 I_n 은 가우스 분포를 따른 $N(t_m^n, 1)$ 초마다 차량과 보행자가 생성된다. 생성되는 즉시, 각 차량과 보행자의 불쾌지수 D_{obj} 는 수식 (3)에 의거하여 실시간으로 갱신(Update)되며, 차량의 경우 교통체증을 표현하기 위해 자료구조 Queue의 형태로 처리한다.

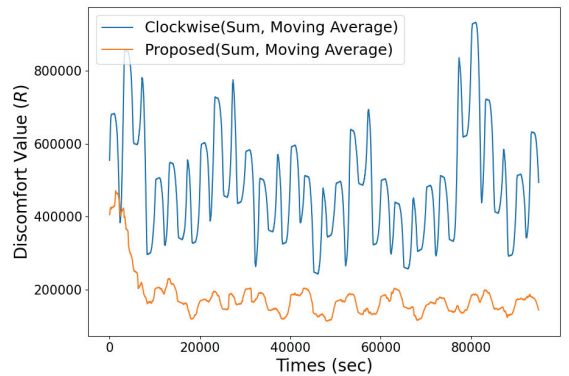


그림 4. 시계방향과 제안한 알고리즘 간 불쾌지수 시간별 변화 비교

Fig. 4 Discomfort value between clockwise pattern and proposed algorithm pattern

3.2 성능평가 결과

그림 4는 대한민국에서 흔하게 사용되는 시계방향으로만 신호를 전달, 반복하는 고정형 신호체계 Clockwise 패턴과 본 논문에서 제안하는 알고리즘을 적용한 적응형 신호체계와의 전체 불쾌지수 R 을 초단위로 측정하여 비교한 것이다.

모델의 학습이 진행 중인 초기에는 R 값이 Clockwise 패턴과 비슷한 값을 가지나, 점차 학습이 진행됨으로써 Clockwise 패턴보다 낮은 불쾌지수를 가지도록 수렴하는 것을 확인할 수 있다. 이를 통해 제안한 Q-Learning 알고리즘을 통해 차량과 보행자를 함께 고려하여 개선된 신호체계를 구축할 수 있음을 확인하였다.

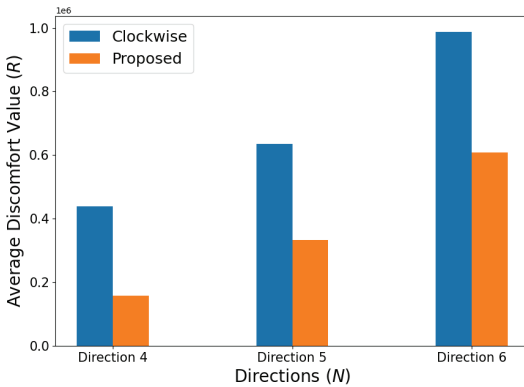


그림 5. 교차로 유형별 알고리즘 간 평균 불쾌지수 비교

Fig. 5 Comparison of average discomfort value according to the number of intersection directions

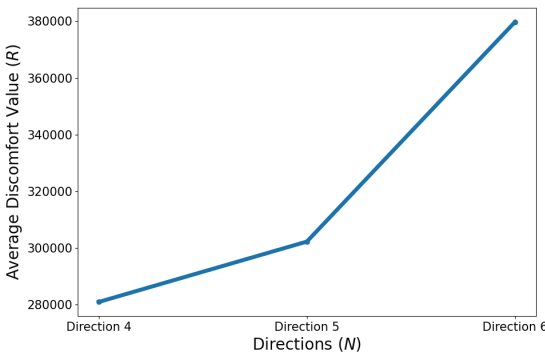


그림 6. 교차로 유형별 알고리즘 간 평균 불쾌지수 차이

Fig. 6 Average discomfort value difference between algorithms by the number of intersection directions

그림 5는 그림 5에서 제안한 사거리($N=4$)를 비롯하여 교차로가 오거리($N=5$), 육거리($N=6$)일 때 Clockwise와의 평균 불쾌지수 수치를 비교한 것이다. 제안한 모델의 평균 불쾌지수는 학습이 되지 않은 경우 R 값이 높은 경향이 있으므로, 충분히 학습된 시뮬레이션 기준 마지막 시간대에서 측정하여 비교하였다. Clockwise와 제안한 모델의 차는 그림 6과 같이 나타나는데, 더 많은 방향 N 을 가질 때, 성능이 더 좋아지는 것을 확인할 수 있다.

IV. 결론 및 향후 연구

본 논문에서는 강화학습 기반 적응형 교차로 알고리즘을 제안하여 차량과 보행자를 함께 고려하고, 이를 대한민국에서 흔하게 사용되는 고정형 신호체계 Clockwise 패턴과 비교·분석하였다. 제안하는 알고리즘은 Clockwise 패턴으로는 원활한 교통흐름을 만들 수 없는 교차로에서 원활한 교통흐름을 만들 수 있음을 확인하였고, 이는 더 많은 방향을 가지는 교차로에서 더 높은 성능을 확인할 수 있었다.

향후 연구로써, 현재 사용되는 복잡한 신호체계를 함께 고려하여 시뮬레이션을 정교화하고, 응급차량 및 사고와 같은 변수 등을 고려하여 현실에 가깝게 환경을 구축하여 심층강화학습 기법을 적용하는 연구를 진행할 계획이다[13]. 또한, 실제 측정되는 교통량 데이터를 활용하여 성능평가를 진행할 계획이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1G1A1011513)

References

- [1] T. Lomax, D. Shrank, S. Turner, L. Geng, Y. Li, and N. Koncz, "Real-timing the 2010 urban mobility report," *Texas A&M Transportation Institute*, no. UTCM 10-65-55, 2011.

- [2] J. Jin and J. Jin, "A Study on the Effect of Traffic Congestion on Particulate Matter Concentration in Seoul : Big Data Approach," *J. of the Korea Planning Association*, vol. 56, no. 1, 2021, pp. 121-136.
- [3] H. Han and I. Yun, "Prediction of the Electric Vehicles Supply and Electricity Demand Using Growth Models," *J. of the Korea Institute of Intelligent Transportation Systems*, vol. 22, no. 4, 2023, pp. 132-144.
- [4] Ministry of Land, Infrastructure and Transport, "Guidelines for roundabout," *Report*, 2014.
- [5] A. Gohar and G. Nencioni, "The role of 5G technologies in a smart city: the case for intelligent transportation system," *Sustainability*, vol. 13, no. 9, 2021, pp. 5188.
- [6] M. Sharafsaleh, J. VanderWerf, A. Misener, and E. Shladover, "Implementing vehicle-infrastructure integration: real-world challenges," *Transportation Research Record*, vol. 2086, no. 1, 2008, pp. 124-132.
- [7] M. Kuwahara, "Intelligent transport system-overview and Japanese systems," *Prometheus*, vol. 86, 1998, pp. 94-103.
- [8] S. An, B. Lee, and D. Shin, "A survey of intelligent transportation systems," In *proceedings of International Conference on Computational Intelligence, Communication System an Networks*, Bali, Indonesia, 2011.
- [9] Y. Lee and J. Lee, "A study on link travel time estimating methodology for traffic information service," *J. of the Korean Society of Transportation*, vol. 20, no. 3, 2002, pp. 55-67.
- [10] J. Kim and S. Kim, "Reinforcement learning-based illuminance control method for building lighting system," *J. of the Institute of Korean Electrical and Electronics Engineers*, vol. 26, no. 1, 2022, pp. 56-61.
- [11] D. Kim, Y. Choi, B. Roh, and J. Choi "Reinforce learning based cooperative sensing for cognitive radio networks," *J. of the Korea Institute of Electronic Communication Science*, vol. 13, no. 5, 2018, pp. 1043-1050.
- [12] O. Khin and S. Lee, "Performance analysis of deep reinforcement learning for crop yield prediction," *J. of the Korea Institute of Electronic Communication Science*, vol. 18, no. 1, 2023, pp. 99-106.
- [13] S. Lee, "A routing algorithm based on deep reinforcement learning in SDN," *J. of the Korea Institute of Electronic Communication Science*, vol. 16, no. 6, 2021, pp. 1153-1160.

저자 소개

김종민(Jong-Min Kim)



2017~현재 동서대학교 소프트웨어학과 재학

※ 관심분야 : 게임 인공지능, 강화학습

김선용(Sun-Yong Kim)



2014년 한동대학교 전산전자공학부 졸업(공학사)

2019년 University of Maryland, Baltimore County 방문연구원

2020년 GIST 전기전자컴퓨터공학부 졸업(공학박사)

2020년 GIST AI연구소 연구원

2020년~현재 동서대학교 소프트웨어학과 조교수

※ 관심분야 : 스마트그리드, 강화학습