Transformer 기반 심층 Q-네트워크를 활용한 차량-보행자 통합 교통 신호 제어 최적화

남금성, Yang Qin, 유상조

인하대학교

nks2325@inha.edu, qinyang@inha.edu, sjyoo@inha.ac.kr

Traffic Signal Optimization for Integrated Vehicle and Pedestrian Control Using Transformer-based Deep Q-Network

Keum-Seong Nam, Qin Yang, Sang-Jo Yoo

Inha Univ.

요 약

최근 Transformer 기반 딥러닝 기술이 다양한 분야에서 주목받고 있으며, 교통 신호 제어(Traffic Signal Control, TSC) 문제에서도 이를 활용하려는 연구가 활발히 이루어지고 있다. 본 논문에서는 차량과 보행자 통행을 동시에 최적화하기 위해 Transformer 구조를 활용한 심층 강화 학습(DQN) 기반 교통 신호 제어 메커니즘을 제안하며, 차량 상태와 횡단보도 상태를 효과적으로 반영함으로써, 실시간 교통 상황 변화에 복합적으로 적응하는 최적 신호 제어를 수행할 수 있도록 설계하였다. 본 논문에서 제안하는 TDDQN은 기존 DQN 기반 제어 메커니즘과 비교하여 학습 안정성과 성능 측면에서 우수함을 실험적으로 입증하였다. 이를 통해 교차로에서 차량과 보행자 통행을 동시에 고려하는 실시간 교통 신호 제어의 실현 가능성을 제시하고자 한다.

Ⅰ. 서 론

최근 교통 신호 제어(Traffic Signal Control, TSC) 문제를 해결하기 위해 강화 학습 기법을 활용한 연구가 활발히 진행되고 있다. 특히, 실시간 교통 상황에 적응하며 차량과 보행자 통행을 동시에 고려할 수 있는 제어 메커니즘에 대한 필요성이 증가하고 있다.[1]

Transformer 모델은 입력 값 간의 Attention 관계를 학습하여 복잡한 의존성을 효과적으로 처리할 수 있는 장점이 있어, 교통 신호 제어 문제에 활용 가능성이 주목받고 있다.

본 논문에서는 차량과 보행자의 통행을 동시에 최적화할 수 있는 시스템 모델에 Transformer Decoder 모델 구조를 적용한 심층 Q-네트워크(Transformer Decoder based DQN, TDDQN) 기반 교통 신호 제어 메커니즘을 제안한다. 제안하는 메커니즘은 교차로 상황과 횡단보도 신호를 모두 고려하여 실시간 교통 혼잡 완화와 안전성을 동시에 달성하는 것을 목표로 한다.

Ⅱ. 심층강화학습을 이용한 횡단보도 교차로 신호 제어 방법 제안.

2-1. 차량 및 보행자를 통행을 위한 Deep Q-Network 기반 교통 신호 제어

본 논문에서는 심층 강화 학습 기법 중 하나인 Deep Q-Network(DQN) 알고리즘[1]을 사용하여 교통 신호 제어 문제를 해결한다. 제안하는 DQN 기반 교통 신호 제어 메커니즘은 그림 1과 같이 차량과 보행자 통행을 동시에 고려하여, 교차로의 차량 상태(s_t^v), 횡단보도 상태(s_t^{CW})와 행동 히스토리(s_t^a)를 종합적으로 반영한다. Network를 통해 도출된 행동(action)을 기반으로 다음 상태와 보상을 수집하며, 이러한 데이터를 Experience Replay Buffer에 저장하여 학습 데이터로 활용한다. 저장된 데이터는 Loss Function을 통해 네트워크를 주기적으로 업데이트하여 학습 성능을 향상시킨다. DQN의 Q-value와 손실 함수는 다음과 같이 정의된다:

$$Q(s_t, a_t) = r_t + \gamma \max_{\forall a} Q(s_{t+1}, a_{t+1}) \quad (1)$$

$$Loss = \sum_{j=1}^{mb} \left[\left\{ r_t + \gamma \max_{\forall a} Q(s_{t+1}, a_{t+1}|\theta^-) \right\} - Q(s_t, a_t|\theta) \right]^2 \quad (2)$$

$Q(s_t, a_t)$는 현재 교차로에서 상태 s_t일 때, 행동(신호) a_t을 취했을 때 기대되는 누적보상이다. r_t는 현재 t시점의 상태에서 행동을 했을 때 주어지는 즉각적인 보상을 나타내고 γ는 현재 보상과 미래 보상을 조율하는 감가율이다. 손실 함수는 목표 Q값과 실제 Q값의 차이를 최소화하여 학습하며, 주기적으로 목표 네트워크(θ^-)를 업데이트하여 학습 안정성을 확보한다.

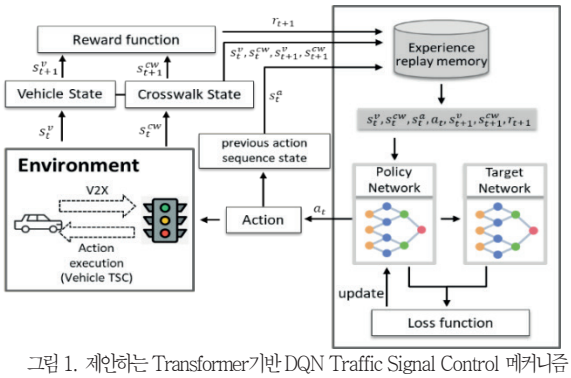

그림 1. 제안하는 Transformer기반 DQN Traffic Signal Control 메커니즘

2-2. Transformer 기반 모델 구조

본 연구에서는 교차로 교통 신호 제어 문제를 해결하기 위해 Transformer Decoder기반 심층 Q-네트워크(TDDQN)를 사용한다.

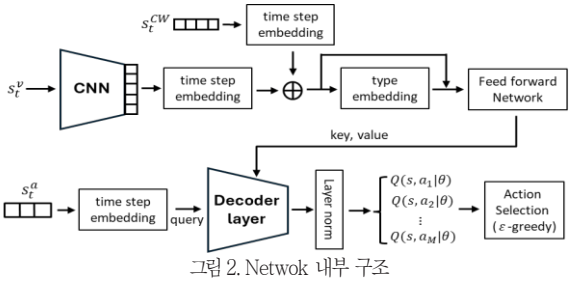

그림 2. Netwok 내부 구조

그림 2는 에이전트내부 네트워크에서 교차로의 상태 정보를 활용하는 전체 구조를 나타낸다. 교차로에서 수집되는 차량 상태와 보행자 상태는 각각 CNN과 MLP를 통해 인코딩되어 Decoder Layer의 Key/Value로 입력된다. 또한, 행동 히스토리는 MLP를 거쳐 인코딩되어 Decoder Layer의 Query로 입력된다. 이와 같이 차량 상태, 보행자 상태, 행동 히스토리를 모두 활용하여 DQN 에이전트가 최적의 행동을 예측하도록 설계하였다.

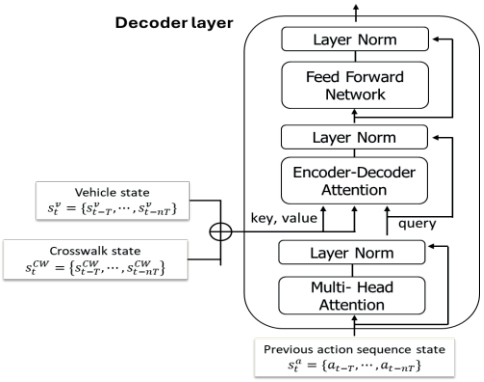

그림 3. Decoder layer 내부 구조

그림 3은 Transformer 기반 Decoder Layer의 내부 구조를 나타낸다.[2] Decoder Layer는 Multi-Head Attention을 통해 과거 행동 히스토리의 순차적 관계를 학습하며, Encoder-Decoder Attention을 통해 과거 행동 히스토리와 인코딩된 차량 상태, 횡단보도 상태 간의 관계를 학습한다. 이때, 과거 행동 히스토리는 Query로 사용되며, 과거부터 현재까지의 차량 상태와 횡단보도 상태는 Key와 Value로 사용하여 Attention을 수행한다. 이러한 구조를 통해 모델은 행동 히스토리와 상태 정보 간의 관계를 효과적으로 학습하여, 교차로 상황에서 신호 제어가 상태 변화에 미치는 영향을 반영할 수 있다. 이를 통해 차량 흐름과 보행자 흐름을 종합적으로 고려한 실시간 최적 교통 신호 제어를 학습할 수 있다.

2-3. 강화학습 상태, 행동, 보상

이 섹션에서는 교통 신호 제어 메커니즘에서 정의된 상태, 행동, 보상을 소개한다.

상태: 위 논문에서는 discrete traffic state encoding(DTSE)방법[1]을 사용하여 차도를 여러 개의 고정된 길이의 셀로 나누어 교통 상태에 대한 정보를 관리한다. 교통 상태 정보는 각 셀에서 차량의 위치, 속도, 대기시간 3가지 정보를 반영하여 3개의 행렬로 표현한다.

$$s_t^{cw} = \begin{bmatrix} I_1^i(t) & \cdots & I_1^i(t+(J-1)T) \\ \vdots & \ddots & \vdots \\ I_j^i(t) & \cdots & I_j^i(t+(J-1)T) \end{bmatrix} \quad (3)$$

$$I_j^i(t) = \begin{cases} 1, & \text{if } CW_j^i > maxCWSI_j^i \text{ at time } t \\ 0, & \text{else} \end{cases} \quad (4)$$

본 논문에서는 교통 상황과 함께 횡단보도 신호의 주기를 계산하여 상태정보로 함께 사용한다. 수식(4)와 같이 각 방향(j)의 횡단보도의 대기시간(CW_j^i)이 미리 설정된 최대대기시간($maxCWSI_j^i$)을 넘기는 시점t을 계산하여 1로 나타낸다. 이를 통해 실시간 교통상황과 횡단보도 상황을 직관적이고 정확하게 활용할 수 있다.

행동: 이 연구에서 행동은 교통신호제어기의 신호이다. 가능한 교차로 신호는 총 6가지의 신호가 있고 각 신호는 횡단보도 신호와 연결된다. 보행자가 안전하게 통행할 수 있도록 우회전의 경우 지나가는 횡단보도가 초록색일 경우 정지하도록 한다.

보상: 위 연구는 교차로에서 보행자와 차량의 통행을 동시에 최적화하는 목적이다. 이를 위해 보상은 횡단보도 신호가 최대대기시간($maxCWSI_j^i$)을 넘기지 않고 차량의 통행이 최적화될 수 있도록 고려하여 설계한다. $R_W(t)$는 차량의 대기시간($waiting\ time$)을 의미하고 $R_V(t)$는 최대대기시간을 넘긴 횡단보도의 수($crosswalk\ violation$)를 의미하며 보행자의 횡단보도가 너무 오랫동안 대기하지 않도록 보상을 통해 학습하도록 하는 인자이다. 보상 값r_t는 다음과 같이 정의한다:

$$r_t = -(w_W R_W(t) + w_V R_V(t))$$

w_W, w_V는 가중치이고 가중치의 합은 1이다.

Ⅲ. 모의 실험 결과

Parameter	Value	Parameter	Value
Number of Vehicle	1300	$maxCWSI_j^i$	N=75s, E, S=60s, W=45
Learning rate (α)	10^{-5}	Number of Head	2
Discount factor (γ)	0.9	Number of Decoder layer	2
Batch size	128	Optimizer	Adam

표 1. 시뮬레이션 파라미터

본 논문에서 사용한 모의 실험은 교통 시뮬레이션 도구인 SUMO(Simulation of Urban MObility)환경에서 구축하였다. 실험은 4방향 교차로를 기반으로 하며, 총 1,500대의 차량을 실제 교통 상황을 모방하여 생성하였다. 차량 생성은 규칙성과 랜덤성을 모두 반영하여 교차로 환경의 불확실성을 고려하였다. 표 1은 시뮬레이션에 사용된 주요 파라미터를 제시하며, 이를 통해 다양한 교통 상황을 고려한 실험을 수행하였다. 실험 결과로 얻어진 보상 값은 그림 4를 통해 시각적으로 비교하여 각 모델의 성능을 분석하였다.

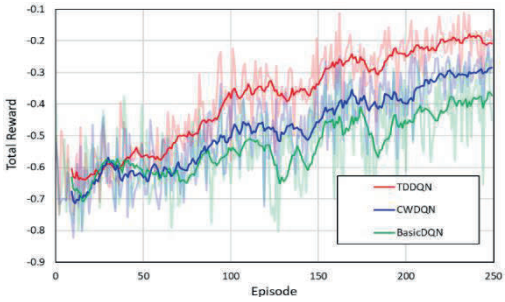

그림 4. 보상 학습 결과

	Waiting time		Crosswalk violation	
Method	Episode 100	Episode 200	Episode 100	Episode 200
BasicDQN	58101.3	29383.3	189	44
CWDQN	51537.0	27066.1	111	22
TDDQN	45064.7	19462.4	59	12

표2. 보상 항목 기반 성능 비교

BasicDQN은 기존 연구[3]에서 사용된 기법으로, 횡단보도 상태를 고려하지 않고 학습된 DQN 기반 모델이다. 반면, CWDQN은 횡단보도 상태와 과거 행동 히스토리 정보를 다층 퍼셉트론(MLP) 구조를 통해 함께 학습함으로써 통합 제어 성능을 개선하고자 한 모델이다. 본 연구에서 제안하는 TDDQN은 Transformer 구조를 기반으로 차량 상태, 횡단보도 상태, 그리고 행동 히스토리 간의 복합적인 관계를 효과적으로 학습하도록 설계되었다.

그림 4는 각 모델의 학습 보상 수렴 과정을 비교한 결과이며, 표 2는 학습 초반(Episode 0)과 학습 후반(Episode 200) 시점에서의 보상 구성 요소별 결과, 즉 평균 대기 시간과 신호 위반 횟수를 나타낸다. 이를 통해 세 모델 간 성능 차이를 정량적으로 비교할 수 있다.

BasicDQN은 횡단보도 상태를 고려하지 않기 때문에, 학습 과정에서 최대 허용 대기 시간을 초과하는 보행자가 다수 발생하며, 전체적인 보상 수치도 낮고 수렴 또한 불안정한 경향을 보인다. 반면, CWDQN과 TDDQN은 보행자와 차량의 상태를 종합적으로 고려하여 평균 대기 시간을 효과적으로 단축하고, 보상 수치도 안정적으로 증가하는 양상을 나타낸다. 특히, 제안하는 TDDQN은 Transformer 구조의 장점을 활용하여 복잡한 교차로 상황에서도 각 상태 간 상호작용을 효과적으로 파악하고, 이를 기반으로 한 정책을 학습함으로써 기존 모델 대비 학습 효율성과 제어 성능 측면 모두에서 우수한 결과를 도출함을 확인할 수 있다.

Ⅳ. 결론

본 논문에서는 Transformer Decoder 기반 심층강화학습(DQN)을 적용하여 횡단보도 대기시간을 고려한 교통신호 제어 시스템(TDDQN)을 제안하였다. 제안된 모델은 행동 히스토리와 차량, 횡단보도 상태 정보를 attention 메커니즘으로 통합하여 최적의 행동을 학습한다.

TDDQN은 교통신호제어 목적에 맞게 설계되어 교통신호에 효과적으로 최적화되며, 횡단보도 대기시간이 최대대기시간을 넘기는 횟수와 시간이 이전 모델에 비해 안정적으로 감소하고 있다. 이 메커니즘은 횡단보도가 있는 교차로에서 차량과 보행자의 통행을 통합하여 최적화하는 모델임을 보여주며, 특히 행동 히스토리와 환경 상태 간의 관계를 학습함으로써 효과적인 정책 학습과 안정적인 신호 제어 성능을 달성한다는 점에서 의의가 있다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임 (RS-2021-II212052)

참 고 문 헌

[1] F. Rasheed, K. -L. A. Yau, R. M. Noor, C. Wu and Y. -C. Low, "Deep Reinforcement Learning for Traffic Signal Control: A Review," IEEE Access, vol. 8, pp. 208016-208044, 2020.

[2] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., Long Beach, CA, pp. 6000-6010, Dec. 2017.

[3] F. Qi, R. He, L. Yan, J. Yao, P. Wang and X. Zhao, "Traffic Signal Control with Deep Q-Learning Network (DQN) Algorithm at Isolated Intersection." 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, pp. 616-621, 2022.