Transformer 기반 PPO를 활용한 cycle 기반 차량-보행자 통합 교통 신호 제어 최적화 연구

*남금성, 양친, 유상조
*인하대학교, 인하대학교, 인하대학교

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

A Study on Cycle-Based Integrated Vehicle-Pedestrian Traffic Signal Control Optimization with Transformer-Based Proximal Policy Optimization

*Keum seong Nam, Qin Yang, Sang-Jo Yoo
*Inha Univ., Inha Univ., Inha Univ.

요 약

본 논문은 V2X 통신 기반 연결형 ITS 환경에서 교차로 신호제어가 차량, 보행자 정보를 실시간으로 활용해 cycle 기반 신호 운영을 최적화하는 문제를 다룬다. 기존 CNN 중심의 단기 스냅샷 기반 학습은 이전 신호 결정의 지연 효과와 사이클 단위의 누적 변화를 충분히 반영하기 어렵다는 점에 착안하여, Transformer 기반 시계열 정책과 PPO 를 결합한 적응형 교통신호제어 방법을 제안한다.

Ⅰ. 서 론

자율주행 기술 확산과 함께 지능형교통시스템(ITS)은 차량과 인프라가 실시간으로 정보를 주고받는 연결형(connected) ITS 방향으로 빠르게 전환되고 있다. V2V 및 V2I 를 포함한 V2X 통신이 보편화되면서, 차량의 주행 상태와 교차로 접근 정보, 대기 및 정체 수준과 같은 교통 상태를 수집, 공유할 수 있게 되었고, 이는 신호 운영의 정밀도를 높이는 기반이 된다. 교통신호제어(traffic signal control, TSC)는 도심 혼잡, 통행시간, 에너지 및 배출, 안전과 같은 핵심 문제와 직결되는 영역으로, 통신 기반 ITS 환경에서 활용 가치가 큰 대표적 응용 분야이다.

실제 교통 운영에서는 운전자 예측 가능성과 안전, 그리고 방향별 형평성을 고려하여 cycle 기반 신호 운영이 널리 사용된다.[1] cycle 기반 운영은 신호 순서가 반복되는 구조를 갖고, 한 사이클 동안의 선택과 그 누적이 다음 사이클의 상태를 좌우한다. 또한 도심 교차로에서는 보행자 안전과 편의가 필수 요소이므로, 차량 흐름만이 아니라 보행자 신호 및 횡단 상태까지 함께 고려하는 통합 운영이 요구된다. 결과적으로 cycle 기반 차량-보행자 통합 신호제어는 실용성과 필요성이 크지만, 그 구조적 특성상 시간적 문맥을 충분히 반영하는 학습 방식이 필요하다. 교통흐름 또한 시간에 따른 연속성이 크며 출퇴근, 요일 효과, 이벤트/사고 등에 의해 반복적인 패턴이 나타난다. 이러한 특성은 단일 교차로에서 직전 구간의 유입, 유출 흐름과 신호 운영 결과가 이후 혼잡 형성에 영향을 미친다.

그럼에도 기존 교통신호제어 연구의 상당수는 큐 길이, 대기시간 등 단기 관측치 중심으로 상태를 구성하고 CNN 기반 인코딩을 통해 정책을 학습하는 방식에 머물러 왔다. 이 경우 단기 반응성은 확보할 수 있으나, 이전 신호 결정의 영향이 시간이 지난 뒤 나타나는 특성이나 사이클 단위로 전개되는 흐름 변화를 충분히 반영하기 어렵다.

본 논문은 이러한 한계를 보완하기 위해 Transformer 기반 시계열 정책과 PPO 를 결합한 cycle 기반 교통신호제어 방법을 제안한다.

Ⅱ. 본론

본 연구는 V2X 통신 기반 연결형 교통 환경에서, 교차로 신호제어기가 도로 이용자(차량·보행자)와 양방향 정보 교환을 수행하며 신호 운영을 최적화하는 구조를 가정한다. 교차로에는 RSU 가 설치되어 차량으로부터 주행 상태(위치, 속도 등)를 수신하고, 보행자 신호 요청 및 횡단 상태는 인프라 센서 또는 보행자 검지기로부터 취득한다. 제어기는 수집된 정보를 기반으로 현재 교통 상황을 추정하여 신호 제어 결정을 수행하며, 결정된 신호 계획은 SPaT(Signal Phase and Timing) 등 V2X 메시지를 통해 주변 차량에 브로드캐스트함으로써 폐루프(closed-loop) 상호작용을 형성한다.

Optimization: 본논문에서는 차량($\mathcal{U}_V$)과 보행자($\mathcal{U}_p$)의 최적화에 목적이 있다.

$$\mathcal{U}_V = \mathbb{E}[w_1 WT_V + w_2 QL_V + w_3 VCount_{maxWT}] \quad (1)$$

차량 함수는 대기시간(WT_V)와 대기열(QL_V) 그리고 최대대기시간을 초과한 차량 수($VCount_{maxWT}$)의 가중합으로 구성된다.

$$\mathcal{U}_P = \mathbb{E}[w_4 VT_{maxCWSI}]$$

또한 보행자 함수는 각 횡단보도가 기다릴 수 있는 최대대기시간을 넘긴 시간으로 정의된다.

이러한 차량과 보행자 함수가 최소가 될 수 있도록 최적화하는 것을 목적으로 한다.

$$\mathbf{P}: \min_{a_t} \mathcal{U}(t) = \min_{a_t} [\lambda_V \mathcal{U}_V(t) + \lambda_P \mathcal{U}_P(t)] \quad (2)$$

System Model: 본 연구는 V2X 기반 연결형 교차로 환경에서, 심층강화학습 기반의 적응형 교통신호제어기를 설계한다. 제어기(agent)는 의사결정 시점 t에서 수집된 상태를 입력으로 받아, 현재 위상의 녹색 지속시간을 결정하며, 위상 전이는 사전에 정의된 cycle 순서를 따른다. 정책 학습에는 안정적인 정책 업데이트 특성을

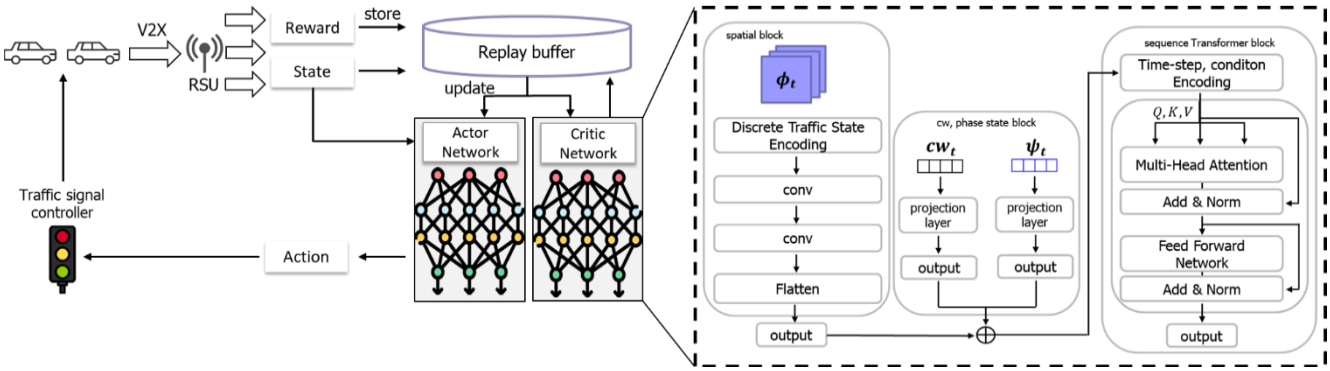

그림 1. V2X 통신 기반 Transformer-PPO 교통신호제어기 구조

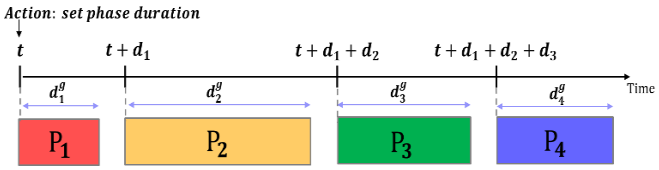

그림 2. Cycle 기반 녹색 위상 지속시간 액션 정의

갖는 PPO 를 사용한다. 정책망은 단기 스냅샷 상태뿐 아니라 최근 여러 구간에 걸친 교통 흐름 변화와 신호 결정의 누적 효과를 반영하기 위해 Transformer 기반 구조로 구성한다. Transformer 의 self-attention 은 긴 시계열에서 중요한 시점과 상호작용을 선택적으로 강조할 수 있어, CNN 중심의 단기 특징 추출 대비 시간 문맥을 보다 효과적으로 모델링할 수 있다.[2]

State: 의사결정 시점 t에서 상태s_t는 차량 상태, 보행자(횡단보도) 상태, 신호 상태로 구성한다.

$$s_t = [\phi_t, cw_t, \psi_t] \quad (3)$$

차량 상태 ϕ_t는 V2X 로 수집되는 차량 $n \in V_t$의 위치($p_n(t)$), 속도($v_n(t)$), 대기시간($w_n(t)$)을 사용한다.

$$x_n(t) = [p_n(t), v_n(t), w_n(t)],\ \phi_t = \{xn(t) \mid n \in Vt\} \quad (4)$$

보행자 상태 cw_t는 횡단보도 $j \in \{1, \dots, J\}$에 대해 허용 최대대기시간 T_j^{max}대비 남은 대기시간($I_j(t)$)을 사용한다.

$$I_j(t) = max(0, T_j^{max} - T_j(t)),\ cw_t = [I_1(t), \dots, I_J(t)] \quad (5)$$

여기서 $T_j(t)$ 는 시점 t 에서 횡단보도 j 의 누적 대기시간이다. 마지막으로 신호 상태 ψ_t는 현재 활성화된 신호의 인덱스(또는 one-hot)이다.

Action: 실제 교통 운영에서는 운전자 예측 가능성과 안전, 그리고 방향별 형평성을 고려하여 cycle 기반 신호 교차로 제어에서 행동은 의사결정 시점 t에서 활성화된 신호의 녹색시간(d_t^g)을 결정하는 문제로 정의한다.

$$d_t^g \in [D_{min}, D_{max}] \quad (6)$$

이때 D_{min}과 $D_{\max}$는 각각 최소/최대 녹색시간 제약을 의미한다.

Reward: 본 논문에서는 차량 효율과 보행자 서비스 수준을 동시에 고려하기 위해 다목적 보상함수를 구성한다. 차량 흐름과 보행자 대기 제약을 각각 반영한 항들을 정의하고, 이를 가중합으로 결합한다.

차량들의 대기시간은 차량의 대기시간의 총합으로, 대기열은 정체 차량($v_n<v_{th}$)의 수로 정의할 수 있으며 다음과 같이 정의한다.

$$R_W(t) = \sum w(t),\ R_Q(t) = \sum 1(v(t) < v_{th}) \quad (7)$$

최대 허용 대기시간($maxWT$)을 초과한 차량 수는 다음과 같이 정의한다.

$$R_{VMW}(t) = \sum 1(w(t) > maxWT) \quad (8)$$

횡단보도 $j \in \{1, \dots, J\}$ 에 대해 최대대기시간 초과 정도(초과 대기)는 다음과 같이 정의한다.

$$R_V(t) = \sum (T_j(t) - T_j^{\max}) \quad (9)$$

각 항을 최소화하는 것이 목적이므로, 최종 보상(r_t)은 비용의 음수 가중합으로 정의한다.

$$r_t = -(w_W R_W(t+T) + w_Q R_Q(t+T) + w_{VMW} R_{VMW}(t+T) + w_V R_V(t+T)) \quad (10)$$

Ⅲ. 시뮬레이션 결과

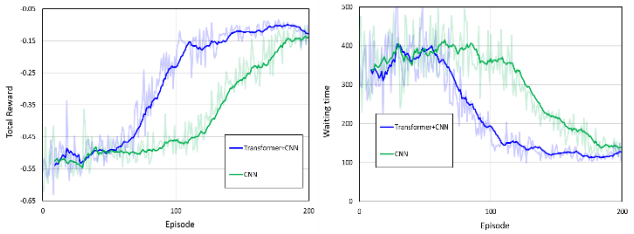

그림 1. 시뮬레이션 결과

시뮬레이션 기반 학습 결과는 그림 3 과 같다. 제안한 Transformer 기반 모델은 CNN 베이스라인 대비 에피소드 진행 초반부터 총 보상이 더 빠르게 증가하며, 동일한 보상 수준에 도달하는 데 필요한 에피소드 수가 유의하게 감소한다.

표 1 모델의 파라미터 규모 및 추론/학습 시간

Model (A+C)	Params (M)	Forward ms/step (A/C)	Train ms/step (A/C)
Transformer+Cnn	**87.81**	**3.18 / 3.07**	**19.8 / 20.0**
CNN baseline	**45.32**	**0.91 / 2.14**	**3.21 / 16.5**

표 1 에서 확인되듯이 Transformer 기반 모델은 파라미터 수 및 학습 단계(latency) 관점에서 더 높은 계산 비용을 요구한다. 그럼에도 불구하고 학습 수렴 속도가 개선되어, 동일 또는 더 높은 성능 수준에 약 100 에피소드 더 이르게 도달함으로써 전체 학습 효율(성능 대비 학습 시간/샘플 효율) 측면에서 우수한 특성을 보인다.

ACKNOWLEDGMENT

This work was supported by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT) (RS-2021-II212052). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-21212972).

참 고 문 헌

[1] M. Wang, Y. Xu, X. Xiong, Y. Kan, C. Xu, and M.-O. Pun, "ADLight: A Universal Approach of Traffic Signal Control with Augmented Data Using Reinforcement Learning." arXiv, Mar. 18, 2023.

[2] M. Wang, X. Xiong, Y. Kan, C. Xu and M. -O. Pun, "UniTSA: A Universal Reinforcement Learning Framework for V2X Traffic Signal Control," in IEEE Transactions on Vehicular Technology, vol. 73, no. 10, pp. 14354-14369, Oct. 2024