최소녹색시간 기반 다중 교차로 강화학습 신호 제어 기법

김건민, 최민준, 김경백†
전남대학교 인공지능융합학과
{geonminkim, larrychoi, kyungbaekkim}@jnu.ac.kr
†교신저자(Corresponding Author) : 김경백

Minimum Green Time-Based Reinforcement Learning for Multi-Intersection Traffic Signal Control

Geonmin Kim, Minjun Choi, Kyungbaek Kim
Chonnam National University, Dept. of Artificial Intelligence Convergence

요 약

강화학습 기반 신호제어에서는 최소녹색시간으로 인해 요청한 신호 전환이 실행되지 않을 수 있다. 이때 요청 행동을 학습에 사용하면 실제 실행 행동과 학습 데이터의 행동 라벨이 달라진다. 본 논문은 최소녹색시간을 기준으로 수요를 관측하는 HED(Horizon-effective Demand)와 실행 가능 행동을 학습에 반영하는 LCB(Lock-consistent Backup)를 결합한다. SUMO의 네 시나리오에서 LCB는 HED-CoLight의 진입 대기 포함 정지 지표를 2.5~7.3% 감소시켰다.

1. 서론

다중 교차로 강화학습 신호제어에서는 교차로 간 협력[1], 압력[2], 접근 수요[4]를 고려하는 방법이 연구되었다. 그러나 최소녹색시간으로 요청한 전환이 실행되지 않으면 학습 행동과 실제 행동이 달라질 수 있다. 강화학습에서는 실행 가능한 행동으로 학습 목표를 계산하거나[5,6] 대체된 행동으로 갱신하는 방법[7]이 제안되었다. 본 논문은 이를 최소녹색시간 제약에 적용하고, 동일한 시간을 기준으로 수요를 관측하는 HED-CoLight+LCB를 제안한다.

2. 최소녹색시간에 따른 전환 제한

결정 주기 5초, 황색시간 3초, 최소녹색시간 10초인 환경에서는 신호 전환 후 두 번의 전환 요청이 무시된다. asym에서 학습 완료된 HED-CoLight 12회 평가 결과, 회당 교차로별 결정 11,520건 중 21.8~41.6%가 전환 제한 상태, 9.6~15.9%는 전환 요청이 실제로 유지된 경우였다.

3. 제안 방법:HED-CoLight+LCB

3.1 HED 기반 수요 관측

HED는 최소녹색시간 이내에 대기열에 도달할 것으로 예상되는 차량을 관측한다. 차량 v의 대기열 꼬리까지 예상 도달 시간은 다음과 같다.

$$\tau_v = \max(0, L - x_v - 7.5 \cdot h_l) / \max(v_v, 0.5) \quad (1)$$

L은 차로 길이, x_v와 V_v는 차량 위치와 속도, h_l은 해당 차로의 정지 차량 수이다. 관측에는 위상별 정지 차량 수, 도달 시간이 0~5초와 5초 초과 ~ 10초인 이동 차량수, 현재 위상 및 정규화된 녹색 경과 시간을 사용한다.

3.2 LCB 행동 학습

LCB는 요청한 위상 대신 실제 실행된 위상을 전이에 저장하고, 다음 상태에서 실행 가능한 위상으로 학습 목표를 계산한다.

$$y_i = r_i + \gamma(1-d) \cdot \max_{a \in F_i(s')} \overline{Q_i}(s', a) \quad (2)$$

r_i는 보상, γ는 할인율, d는 종료 여부, $\overline{Q_i}$는 타깃 네트워크의 Q값이다. $F_i(s')$는 녹색 경과시간이

최소녹색시간보다 짧으면 현재 위상만, 그렇지 않으면 전체 위상을 포함한다.

3.3 실험 설정

SUMO의 10×10 격자에서 균일 수요(med), 간선 플래툰(asym), 중심부 수요 급증(burst), 중심 간선 저속(incident)을 평가하였다. 에피소드는 600초이며, 학습 기반 방법은 45개 에피소드로 학습하고 6회 재시작하였다. 평가는 고정된 수요·경로에서 운전자 무작위성이 달라지는 SUMO 시드 4~13으로 수행하였다. 주 지표는 관측 차로의 중복 제거 정지 차량과 진입 대기를 매초 합산한 값이며, 괄호 안은 기존 정지 지표이다.

4. 실험 결과

제안 방법은 진입 대기 포함 지표에서 네 시나리오 모두 가장 낮은 평균값을 기록하였다. HED-CoLight에 LCB를 적용하면 해당 지표가 2.5~7.3% 감소하였다(Holm 보정 $p \le 0.022$). asym의 기존 지표에서는 CoLight(대기열+위상)가 제안 방법보다 낮았으나, 진입 대기는 각각 26.3×10^3과 3.1×10^3 veh·s였다. LCB의 효과는 관측에 따라 달랐으며, 대기열+위상 관측의 asym에서는 기존 지표가 49.1% 증가하였다. 이동 시간과 진입 대기를 포함한 총 시간은 최고 기준선과 −0.9~+0.3% 차이를 보였다.

표 1. 진입 대기 포함 정지 지표와 기존 정지 지표(괄호 안), $\times10^3$ veh·s

방법	med	asym	burst	incident
MaxPressure[3]	12.2 (46.9)	41.0 (132.5)	25.9 (101.7)	14.1 (52.4)
ATS-MP(10초)	6.1 (23.1)	30.9 (113.1)	19.8 (77.7)	8.4 (30.6)
CoLight (대기열+위상)	12.1 (46.1)	51.9 (102.2)	25.6 (100.1)	14.2 (52.6)
CoLight (접근 차량)	7.4 (27.9)	37.5 (138.6)	21.5 (84.2)	10.5 (38.4)
HED-MP	6.0 (22.7)	31.0 (113.0)	19.4 (76.0)	8.5 (30.5)
HED-CoLight	5.8 (21.7)	31.6 (112.9)	19.6 (76.7)	8.2 (29.7)
HED-CoLight + LCB	**5.3 (20.0)**	**29.6 (105.8)**	**19.1 (74.7)**	**7.9 (28.5)**

5. 결론

본 논문은 최소녹색시간을 기준으로 수요를 관측하고 실행 가능 행동을 학습하는 HED-CoLight+LCB를 제안하였다.네 시나리오에서 진입 대기 포함 정지 지표가 가장 낮았다. 다만 LCB의 효과는 관측에 따라 달랐다. 향후 새로운 수요와 실제 도로망에서 검증할 계획이다.

Acknowledgement

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2026-RS-2022-00156287)(34%).본 연구는 2026년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 (S3471392)(33%).본 연구성과는 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(RS-2025-25398164)(33%)

참고 문헌

[1] H. Wei et al., “CoLight: Learning network-level cooperation for traffic signal control,” Proc. ACM CIKM, pp. 1913–1922, 2019.

[2] H. Wei et al., “PressLight: Learning max pressure control to coordinate traffic signals in arterial network,” Proc. ACM KDD, pp. 1290–1298, 2019.

[3] P. Varaiya, “Max pressure control of a network of signalized intersections,” Transportation Research Part C, Vol. 36, pp. 177–195, 2013.

[4] L. Zhang et al., “Expression might be enough: Representing pressure and demand for reinforcement learning based traffic signal control,” Proc. ICML, pp. 26645–26654, 2022.

[5] C. Boutilier et al., “Planning and learning with stochastic action sets,” Proc. IJCAI, pp. 4674–4682, 2018.

[6] T. Zahavy et al., “Learn what not to learn: Action elimination with deep reinforcement learning,” Proc. NeurIPS, pp. 3562–3573, 2018.

[7] M. Alshiekh et al., “Safe reinforcement learning via shielding,” Proc. AAAI, pp. 2669–2678, 2018.