TD3 기반 자율주행차량의 병목지점 통과 정책 연구

엄찬인, 이동수, 권민혜
숭실대학교

{eci0623, movementwater}@soongsil.ac.kr, minhae@ssu.ac.kr

TD3-based Autonomous Driving Strategy for Bottleneck Traffic

Chanin Eom, Dongsu Lee, Minhae Kwon
Soongsil University

요 약

본 연구는 심층 강화학습을 통해 차량 정체 현상이 빈번하게 발생하는 병목구간을 통과할 수 있는 자율주행차량의 학습을 목적으로 한다. 문제 해결을 위해 Partially Observable Markov Decision Process (POMDP) 모델을 제안하며 대표적인 심층 강화학습 알고리즘인 Twin Delayed DDPG (TD3)를 이용하여 자율주행차량을 학습시킨다. 그 결과 학습된 자율주행차량은 병목구간에서 일반차량 대비 7.7% 빠른 평균 속력으로 주행하는 것을 확인하였으며 인접 도로 환경을 유지한 채 효과적으로 병목구간을 통과할 수 있음을 확인하였다.

Ⅰ. 서 론

최근 자율주행차량과 일반차량이 혼재된 도로 환경에서 강화학습을 활용한 자율주행 연구가 활발히 진행되고 있다[1]. 일반 운전자와 공존하는 도로에서 자율주행 차량의 잘못된 주행전략은 심각한 교통 혼잡 및 사고로 이어질 수 있다. 특히, 병목지점과 같은 도로 환경에서는 차선 감소로 인해 도로 정체 현상이 극대화될 수 있기 때문에 주변 환경을 고려한 자율주행 정책이 적용되어야 한다. 이에 본 논문에서는 자율주행차량의 병목구간 통과를 위한 POMDP 모델을 제안한다. 제안된 POMDP 는 심층 강화학습 알고리즘 TD3[2]를 통해 인접 차량의 주행에 미치는 영향을 최소화하며 병목구간을 통과하는 자율주행 개체를 학습한다.

Ⅱ. 병목 구간 통과를 위한 강화학습 문제정의

도로 주행과 같이 개체가 제한된 관측 정보를 통해 의사 결정을 수행해야 하는 강화학습 문제는 POMDP 를 통해 모델링 할 수 있다. POMDP 는 튜플 $\langle S, A, O, R, \gamma \rangle$로서 정의되며, 에이전트는 시간 t 에서의 모든 유한한 상태 $s_t \in S$를 관측한 정보 $o_t \in O$를 통해 행동 $a_t \in A$를 결정한다. 에이전트는 상태 s_t에서의 행동 a_t에 대한 결과로 보상 R_t를 획득하며, 시간에 따른 감가율 γ가 적용된 누적 보상을 최대화하는 행동 정책을 학습한다.

Ⅱ.1. 도로 환경 정의

본 연구에서는 N 대의 차량 $C = \{c_1, \cdots, c_N\}$와 B개의 병합지점(merge point) $M = \{m_1, \cdots, m_B\}$이 존재하는 도로 상황을 가정한다. 전체 차량 집합 C는 1대의 자율주행차량과 $N-1$ 대의 일반차량을 포함한다. 이때, $c_i\ (i \neq N)$는 일반차량, c_N은 자율주행차량을 의미한다. 이와 같은 환경에서 상태정보 s_t는 다음과 같이 정의한다($s_t \in \mathbb{R}^{4N}$).

$$s_t = [\boldsymbol{v}_t^T, \boldsymbol{p}_t^T, \boldsymbol{k}_t^T, \boldsymbol{n}_t^T]^T$$

이때, $\boldsymbol{v_t} = [v_{t,1}, \cdots, v_{t,N}]^T$는 모든 차량의 속력, $\boldsymbol{p_t} = [p_{t,1}, \cdots, p_{t,N}]^T$는 차량의 위치, $\boldsymbol{k_t} = [k_{t,1}, \cdots, k_{t,N}]^T$는 각 차량이 위치한 차선번호를 의미한다. 여기서 임의의 차량 c_i가 위치한 차선번호 $k_{t,i}$는 가장 바깥차선을 0 으로 정의하고, 안쪽으로 이동할수록 1 씩 증가한다. 마지막으로, $\boldsymbol{n_t} = [n_{t,1}, \cdots, n_{t,N}]^T$는 각 차량이 위치한 도로의 총 차선 개수이다.

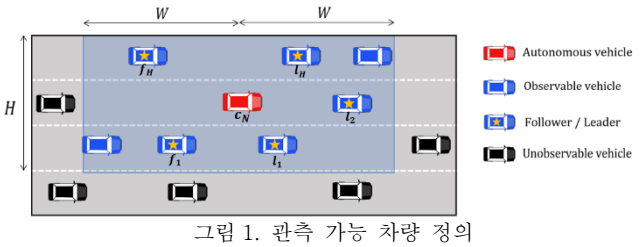

그림 1. 관측 가능 차량 정의

Ⅱ.2. Partially Observable Markov Decision Process 정의

에이전트는 상태정보 s_t에 대한 관측정보 o_t를 통해 행동 a_t를 결정한다. 여기서 관측정보 o_t는 관측 가능 거리 $2W$와 관측 가능 차선 개수 H 내의 상태정보로 제한된다 (그림 1). 관측 가능 차량은 후방차량 $C_f = \{c_i | -W \leq \Delta p_{t,i} < 0\}$와 전방차량 $C_l = \{c_i | 0 \leq \Delta p_{t,i} \leq W\}$로 정의하며, $\Delta p_{t,i}$는 일반차량과의 상대거리($p_{t,i} - p_{t,N}$)를 의미한다. 이때, C_f, C_l에는 각각 부분집합 $F \subset C_f, L \subset C_l$이 존재한다. $F = \{f_1, \cdots, f_H\}$는 후방 차량 중 차선 별 상대거리의 절댓값이 가장 작은 차량들로 follower 집합을 의미하며, $L = \{l_1, \cdots, l_H\}$은 leader 집합을 의미한다. 전체 POMDP 는 위의 관측 집합을 통해 표현할 수 있으며 아래와 같이 정의한다($o_t \in \mathbb{R}^{5H+5}$).

$$o_t = [\Delta \boldsymbol{v}_{t,f}^T, \Delta \boldsymbol{v}_{t,l}^T, \Delta \boldsymbol{p}_{t,f}^T, \Delta \boldsymbol{p}_{t,l}^T, \boldsymbol{\rho}_t^T, n_{t,N}, n_{t,m}, \Delta p_{t,m}, v_{t,N}, k_{t,N}]^T$$

여기서 $\Delta \boldsymbol{v}_{t,f} = [(v_{t,f_1} - v_{t,N}), \cdots, (v_{t,f_H} - v_{t,N})]^T$은 follower 집합 F 내의 각 차량과 자율주행차량 사이의 상대속도이며, $\Delta \boldsymbol{v}_{t,l} \in \mathbb{R}^H$는 leader 집합 L 내 차량과의 상대속도를 의미한다. $\Delta \boldsymbol{p}_{t,f} = [(p_{t,f_1} - p_{t,N}), \cdots, (p_{t,f_H} - p_{t,N})]^T$는 F와의 상대거리를 의미하며, $\Delta \boldsymbol{p}_{t,l} \in \mathbb{R}^H$는 L과의 상대거리를 의미한다. $\boldsymbol{\rho}_t \in \mathbb{R}^H$는 자율주행차량의 전방

차선 별 차량밀도로 관측 가능 거리 W 대비 C_l 차량이 차지하고 있는 차선 별 도로 비율을 의미한다. $n_{t,m}$, $\Delta p_{t,m}$은 자율주행차량과 가장 가까운 병합지점 $m_j \in M$에 대한 관측으로, m_j가 관측 범위내에 있을 때 즉, $\Delta p_{t,j} \leq W$일 때, 병합 이후 도로의 차선 개수와 병합 시작점까지의 거리를 의미한다.

에이전트는 행동 $a_t = \{a_{t,acc}, a_{t,lc}\}$의 선택을 통해 가속도 조절 $a_{t,acc}$과 차선 변경 $a_{t,lc}$을 수행한다. 이때, $a_{t,acc}$는 제한된 범위 내의 연속된 값을 갖는다($a_{t,acc} \in [acc_{min}, acc_{max}]$). 차선 변경 $a_{t,lc}$의 경우에는 $[-1, 0, 1]$중 하나의 값을 가지며, $a_{t,lc} = 0$일 경우 차선 유지, $a_{t,lc} = 1$일 때 안쪽 차선으로 이동, $a_{t,lc} = -1$일 때 바깥 차선으로 이동한다. t 시점에서의 보상 R_t는 아래와 같이 정의한다.

$$R_t = R_{t,rew} + R_{t,pen} \quad (1)$$

$R_{t,rew}$는 t 시점 행동에 대한 보상 항, $R_{t,pen}$은 처벌 항을 의미하며, 보상 항 $R_{t,rew}$은 아래와 같이 정의한다.

$$R_{t,rew} = \eta_1 \left(1 - \left|\frac{\bar{v}_{t+1} - v^*}{v^*}\right|\right) \quad (2)$$

보상 항에서 v^*는 도로 내 모든 차량의 목표속도로, 환경 내 차량이 도달하고자 하는 속력을 의미한다. $\bar{v}_{t+1}$는 자율주행차량과 관측 가능 거리 $2W$ 내에 있는 전, 후방 차량의 평균 속력을 의미하며 아래와 같이 정의한다.

$$\bar{v}_{t+1} = \frac{v_{t+1,N} + \sum_d^D v_{t+1,d}}{D+1} \quad (3)$$

여기서 $v_{t+1,d}$는 $t+1$ 시점에서 관측 가능한 임의의 차량 $c_{t+1,d}$의 속력이며, 이때 $c_{t+1,d} \in C_f \cup C_l$을 만족한다. D는 관측 가능한 차량의 총개수를 의미한다. 다음으로, 처벌 항은 다음과 같이 정의한다.

$$R_{t,pen} = \eta_2 \left(min\left[0, 1 - \left(\frac{s^*}{\Delta p_{t+1,f}}\right)\right]\right) + \eta_3 \left(|a_{t,lc}| min[0, \Delta p_{t+1,l} - \Delta p_{t,l}]\right) \quad (4)$$

여기서 $\left(min\left[0, 1 - \left(\frac{s^*}{\Delta p_{t+1,f}}\right)\right]\right)$은 무리한 차선 변경 및 급정거 등의 위험 행동을 약화하기 위한 처벌 항으로 t 시점의 행동으로 인해 동일 차선 후방차량의 안전거리 s^*를 침범한 경우 페널티를 부여한다. 두 번째 처벌 항 $\left(|a_{t,lc}| min[0, \Delta p_{t+1,l} - \Delta p_{t,l}]\right)$은 의미 없는 차선 변경을 처벌한다. 즉, 차선 변경을 수행($|a_{t,lc}| = 1$)한 뒤의 동일 차선 선두 차량의 상대 거리($\Delta p_{t+1,l}$)가 행동 수행 전 동일 차선 선두 차량의 상대 거리($\Delta p_{t,l}$) 보다 작은 경우 ($\Delta p_{t+1,l} < \Delta p_{t,l}$), 의미 없는 차선 변경으로 간주하여 페널티를 부여한다.

Ⅲ. 실험 및 성능평가

본 연구에서는 자율주행차량이 존재/부재 하는 총 2 가지 환경에 대해 실험을 진행하며, 도로 환경은 병목지점이 포함된 타원형 도로를 고려한다. 또한, 도로 구간별 평균 속력 그래프를 통해 개체의 행동을 분석하고 병목 구간 평균 속력을 통한 정량적 분석을 진행한다. 자율주행차량 학습에는 심층 강화학습 알고리즘인 TD3 를 사용한다.

Ⅲ.1. 실험 설정

본 연구에서는 교통제어 시뮬레이터인 FLOW [3]를 사용한다. 실험 환경 내 전체 차량 수 N은 32 대로 일반차량 31 대와 자율주행차량 1 대로 구성된다. 일반차량은 도로가 넓어지는 구간을 제외하고는 차선 변경이 불가능하며 Intelligence Driving Model (IDM) [4]에 의한 속력 조절만이 가능하다. 이때 도로 내 모든 차량의 목표속도 v^*는 12.5m/s 이다. 도로의 총길이는 465m 로 설정하였으며, 병합 구간의 수 B는 2 로 설정하였다. 자율주행차량은 최대 5 개의 차선을 관측할 수 있으며, 관측 가능 거리 W는 30m 로 설정하였다.

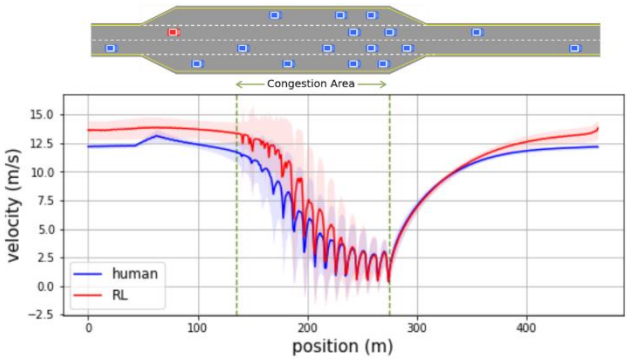

그림 2. 도로 구간 별 속력 비교

표 1. 실험환경 및 차량종류 별 병목구간 내 속력 비교

실험환경	차량종류	평균속력
RL 차량 부재	일반	3.10 m/s
RL 차량 존재	일반	3.05 m/s
	자율주행	3.34 m/s

Ⅲ.2. 병목구간 통과 정책 성능평가

그림 2 는 도로 구간별 일반차량(human)과 자율주행차량(RL)의 속력 평균을 나타낸 그래프이다. 그래프에서 혼잡지역(congestion area)은 병목지점(285m)으로 인해 차량 밀도가 높은 구간을 의미한다. 그래프를 통해 자율주행차량이 병목구간 내에서 일반차량보다 높은 속력으로 주행하는 것을 확인할 수 있다. 구체적인 수치는 표 1 을 통해 확인할 수 있다. 자율주행차량의 병목구간 평균 속력은 3.34m/s 로, 자율주행차량이 부재한 환경 내 일반차량의 평균 속력에 비해 0.24m/s 빠르게 주행하는 것을 확인할 수 있다. 이에 비해, 각 환경의 일반차량 평균 속력은 자율주행차량이 존재하는 환경에서 0.05m/s 라는 상대적으로 적은 감소율을 보였다. 이는 학습된 개체가 병목구간에서 인접한 주행 차량에 대한 영향은 최소화하면서, 성공적으로 정체 구간을 통과하고 있음을 의미한다.

Ⅳ. 결론

본 논문에서는 POMDP 설계를 통해 병목구간을 성공적으로 통과하는 개체를 학습하였다. 학습된 개체는 자율주행차량이 부재한 환경 내 일반차량과 비교했을 때, 병목구간에서 7.7% 높은 속력을 유지하는 것을 확인하였다. 또한, 자율주행차량은 인접 도로 환경을 유지하면서 차량 밀집 지역을 통과하는 것을 확인하였다.

사 사

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단(NRF-2020R1F1A1069182) 및 정보통신기획평가원(2021-0-00739, 분산/협력 AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] D.S Lee, M.H Kwon, "ADAS-RL: Safety Learning Approach for Stable Autonomous Driving," ICT Express, vol.8, no.3, pp. 479-483, 2022.

[2] S. Fujimoto, H. van Hoof, et al., "Addressing Function Approximation Error in Actor-Critic Methods," ICML, 2018.

[3] C. Wu, A. Kreidieh, et al., "Flow: Architecture and Benchmarking for Reinforcement Learning in Traffic Control," arXiv preprint arXiv:1710.05465, 2017.

[4] M. Treiber, A. Hennecke, et al., "Congested Traffic States in Empirical Observations and Microscopic Simulations," Physical Review E, vol.62, no.2, pp. 1805-1824, 2000.