

다중 교차로에서 협동적 신호제어를 위한 보상함수 설계

배요한 · 장진현* · 송문혁
경기과학고등학교

Designing Reward Function for Cooperative Traffic Signal Control at Multi-intersection

Yo-han Bae · Jin-heon Jang* · Moon-hyuk Song
Gyeonggi Science High School

E-mail : gs21052@gs.hs.kr / gs21094@gs.hs.kr / gs21062@gs.hs.kr

요 약

신호를 제어하는 방식은 기존의 전통적인 수학적 방식을 이용한 최적화를 넘어 이제 인공지능이 본격적으로 활용되기 시작하는 단계까지 발전하였다. 이에 따라 인공지능을 적용하는 방안에 대해 다양한 연구들이 진행되고 있는데, 현행 연구에서는 주로 좋은 교통 상황에 대한 마땅한 고려 없이 간단히 지체도만을 고려하여 보상함수를 설정하는 방식을 주로 채택하고 있다. 그러나 이 경우 현실성이 떨어지는 신호 제어 방식을 인공지능이 학습할 가능성이 존재한다는 문제점을 지닐 뿐더러, 보상 함수에서 좋다고 평가하는 것이 실질적인 서비스 수준의 정의에 부합하지 않음을 확인할 수 있다. 따라서 본 연구에서는 기존의 보상함수 설정 사례를 분석하고, 개선 방향을 제시하고자 한다.

ABSTRACT

Nowadays, breaking through the conventional traffic signal control method based on mathematical optimization, artificial intelligence began to be used in the area. In response to this trend, many studies are ongoing to figure out how to utilize AI technology properly for traffic signal optimization. They just simply focus on which method will work well besides lots of machine learning techniques and abandon the reward function engineering. In many cases, the reward function consists of the average delay of the vehicles in the intersection. However, this may lead to AI's misunderstanding about the traffic signal control: what AI regards as a good situation may not be realistic. Even the reward function itself may not meet the service level. Therefore, this study analyzes the problems of previous reward functions and will suggest how to reward function can be enhanced.

키워드

Reward Function Engineering, Traffic Signal Control, Reinforcement Learning, Urban Traffic Modeling,
Traffic Simulation

I. 서 론

현재 우리나라 수도권에서 사용되고 있는 신호

체계는 COSMOS 신호체계로, 도로에 설치된 차량 검지기를 통해 수집한 자료로 실시간으로 신호제어를 하는 시스템이다. 이때, 시간이 지남에 따라 교통수요는 지속적으로 증가하고 있으나 신호체계는 기존의 것이 유지되면서 자연스럽게 COSMOS

* speaker

가 감당할 수 있는 범위를 넘게 되었다. 예를 들어, 포화로 인해 교통량이 감소한 것을 실제로 도로 위에 차량이 적게 존재한다는 것으로 인식하기도 한다.

이에 따라 많은 연구에서 고전적인 방식을 탈피하여 인공지능을 도입하였지만 이들 대다수는 보상함수를 임의로 설정하고, 그에 대한 고찰은 깊이 있게 진행하지 않았다. 따라서 본 연구에서는 보상함수의 타당성을 검증하고, 추후 보상함수를 제안할 것이다.

II. 이론적 배경

교통류 이론 교통류(traffic flow)란 한방향으로 주행하는 연속적인 차량의 흐름을 의미한다. 교통 신호가 있어 주행 중 부득이하게 정지를 해야 하는 교통류를 단속교통류라고 한다. 교통류를 나타내는 변수는 속도, 통행시간, 교통량, 교통류율, 차두시간, 밀도, 점유율, 차두거리, 지체시간 등이 있다.

단속교통류 시설에서 차량들은 교통신호로 인해 전체 신호 주기 중 일부만 동안만, 즉 유효 녹색시간 동안만 주행하게 된다. 따라서 최대 통과가능 교통량, 용량은 포화교통량과 유효녹색시간 비를 곱한 것과 같다. 식으로 나타내면 $c_i = s_i \times (\frac{g}{C})_i$ 이다. 또한 교통량 대 용량의 비는 포화도라고 한다.

교통류의 3변수인 속도, 밀도, 교통량 간의 상관관계를 교통류 모형이라 부르며 이들간의 관계는 $q = \mu k$ 와 같다. 현장관측에 의한 속도-밀도 관계를 수식적으로 설명하기 위해 Greenshields, Greenberg 등의 모형이 사용된다.

교통류를 유체와 같은 것으로 보고 이에 대한 수리역학적인 원리를 적용시킨 것이 교통류의 충격파이다. 교통량과 밀도가 변하는 지점이 이동하는 것을 충격파의 이동으로 해석하여, 충격파 발생 지점을 통과하는 차량의 수를 기준으로 식을 세워 충격파의 속도를 구하면 다음과 같다.

$$u_w(1-2) = \frac{q_2 - q_1}{k_2 - k_1} \dots\dots\dots (1)$$

교통감응식 신호 교차로에 유입하는 교통량에 따라 녹색신호 시간을 자율적으로 조절하는 방식을 교통감응식 신호라고 한다. 교차로에 설치된 검지기를 통해 얻은 교통 정보로 산출된다. 유입교통량, 인접 교차로와의 연동 등 다양한 알고리즘이 연구되고 있다.

III. 연구과정 및 방법

3.1. 선행 연구 분석

선행 연구들의 평가함수 사례를 분석해보는 작

업을 진행하였다.[1][2][3] 이들의 주된 공통점은 신호제어를 교통 소통의 측면에서 얼마나 원활한가를 기준으로 평가하고 있다는 점이다. 특히, 이 중에서도 전체 차량의 지체도 평균을 평가하는 방안이 지배적인 것으로 관찰된다. 다만, 이는 위의 연구들에서도 언급되었듯이, 교차로에서 주교통류와 부교통류 간 차이가 커질 경우, 부교통류가 겪게 될 지체가 일정 수준 이상으로 커질 수 있다는 점이다.

그래서 다른 연구들에서는 이 문제점을 해결하기 위해서 균등화 전략을 취하기도 하였다. 균등화 전략은 각 방향의 차들이 겪게 될 지체도를 균등하게 맞춰주는 것을 의미한다. 이때, 더 높은 정확도를 위하여 진입 방향을 넘어 교통류별로 지체를 추정하는 경우도 있었다.

3.2. 기존 보상함수의 문제점 분석

먼저, 위에서 언급했던 두 전략을 수식화하고 실제로 해당 문제점이 발생하는 지를 증명해보겠다. 편의를 위해 교통 상황이 Adolf D. May의 정주기 신호 대기행렬 모형을 따른다고 가정하고, 도로는 포화상태에 이르지 않는다고 가정해보자.

표 1. 정주기 신호 모형에서 변수 정의

Symbol	Definition
D_{tot}, D_{avg}	차량 지체도(총, 평균) (h*대, h)
Δt	녹색 현시 점등 후 대기 행렬 길이 소진시까지 걸리는 시간 (h)
T_R	하나의 교통류에서 한 주기에 유효하게 적색 현시가 켜진 시간 (h)
T	한 신호주기의 시간 (h)
p	하나의 교통류에서 주기에 대한 녹색 현시의 비율: $p := 1 - \frac{T_R}{T}$
q	단위시간당 진입하는 차량 수 = 도착 교통류율 (대/h)
s	단위시간당 빠져나가는 최대차량 수 = 출발 교통류율 (대/h)
χ	포화도: $\chi := \frac{q}{s}$
π	신호주기 설정 전략, 교통 상황을 입력받아 최적 p 값을 반환하는 함수

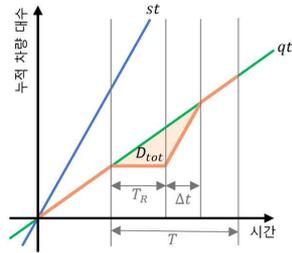


그림 1. 정주기 신호모형의 그래프

먼저, Δt 를 계산해보자. 비포화 상태를 가정했으므로, 식 (2)와 같이 한 주기에 교차로 방향 i 에 진입한 차량 수는 빠져나간 차량 수와 동일하다 (2). 그리고 식 (3)과 같이 총 차량 지체도는 그림 1에 표시된 면적과 동일하다 (3). 이를 토대로 $D_{avg,i}$ 를 계산하면 식 (4)와 같다.

$$s\Delta t_i = q(\Delta t_i + T) \rightarrow \Delta t = \frac{q_i}{s - q_i} T_{R,i} \dots\dots\dots (2)$$

$$D_{tot,i} = \frac{1}{2} T_{R,i} \times s\Delta t_i = \frac{1}{2} T^2(1 - p_i)^2 \left(\frac{q_i}{1 - \chi_i} \right) \dots\dots\dots (3)$$

$$D_{avg,i} = \frac{1}{q_i T} D_{tot,i} = \frac{1}{2} T(1 - p_i)^2 \left(\frac{1}{1 - \chi_i} \right) \dots\dots (4)$$

이제 한 교차로에서 서로 상충되는 교통류 2개를 잡아 이를 1과 2로 해보자(e.g., 동서방향, 남북방향). 그렇다면, 교차로 1에서 q_1 의 도착교통류를 갖고, $p_1 = p$ 의 녹색 현시 비율을 가지면, 나머지 하나의 교통류는 자동으로 $p_2 = 1 - p$ 의 녹색 현시비율을 최대 가질 수 있다. 이를 토대로 교차로 전체에 대한 평균 지체도(D_{avg})를 구하면, (4)와 같다.

$$D_{avg} = \frac{T}{2(\chi_1 + \chi_2)} \left\{ \left(\frac{\chi_1}{1 - \chi_1} + \frac{\chi_2}{1 - \chi_2} \right) p^2 - \left(\frac{2\chi_1}{1 - \chi_1} \right) p + \frac{\chi_1}{1 - \chi_1} \right\} \dots\dots\dots (5)$$

따라서 전체 지체도 최소화 전략(π_{SO})이 도출해낼 p 값은 (5)를 따른다.

$$\pi_{SO} = \operatorname{argmin}_p (D_{avg}) = \frac{\chi_1 - \chi_1\chi_2}{\chi_1 + \chi_2 - 2\chi_1\chi_2} \dots\dots (6)$$

만약 $q_1 = 0.5s$, $q_2 = 0.2s$ 정도라면, $\pi_{SO} = 0.8$ 가 되어, $D_{avg,2} = 10D_{avg,1}$ 이 된다. 즉, 한쪽이 교통량이 2.5배이지만, 겪는 지체는 평균 10배인 것이다. 따라서 이 방식은 교통량 차이가 큰 상황에서 문

제를 발생시킴을 알 수 있다.

균등화 전략(π_{UE})의 경우, $D_{avg,2} = D_{avg,1}$ 가 목표라고 할 수 있다. 이를 토대로 전략을 도출하면 (6)과 같다.

$$\pi_{UE} = \left(1 + \sqrt{\frac{1 - \chi_1}{1 - \chi_2}} \right)^{-1} \dots\dots\dots (7)$$

이를 토대로 위의 예시에서 $\pi_{UE} \approx 0.56$ 이 된다. 그러나 이를 (4)에 대입하여 평균 지체도를 구해보면, 36% 가량 증가했다는 사실을 알 수 있다. 따라서 교통 소통 측면에서는 상당히 안 좋아진 것이다. 다만, 교통 상황에서 실제로 운전자들 각각이 체감하는 지체시간 자체가 일정 수준을 넘김에 따라 사람들이 느낄 불만이 급격히 증가할 것이므로, 균등화 측면도 강조되어야 한다. 따라서 둘 사이에 적절한 조화가 필요하다.

또한, 균등화 전략 적용 시에 지체도 대신에 포화도를 기준으로 하는 것도 하나의 방안인데, 지체도만을 고려 시에 포화 상태를 넘는 경우 지체도를 낮은 방향으로 갈도록 하는 것이 아니라 상향평준화가 되도록 작동할 수 있기 때문이다. 다만, 포화도를 이용하면 이를 방지할 수 있다.

추가적으로, 현재 보상함수는 교통소통의 측면 즉, 빠른 시간 내로 도로를 통과하는 것에만 초점이 맞춰져 있다. 때문에 인공지능이 신호를 바꾸는 행위 자체를 좋다고 판단하는 비현실적인 사례들도 발견된다[4]. 따라서 이를 직접적으로 억제해줄 수 있는 변수의 도입(e.g., 정지횟수) 등의 도입으로 방지해줄 필요가 있다. 또한, 단순히 빠른 시간 내 통과가 중요한 것이 아니라, 안전한 통행 역시 중요하므로, 서비스 수준의 본질적인 의미에 입각하여 사고율의 반영도 필요하다.

마지막으로, 다중 교차로에서 단순히 하나의 교차로에서의 지체보다는 전체 도로망 통과 시간을 이용하면 자연스럽게 협력적으로 신호제어가 일어나게 유도할 수 있다.

IV. 결 론

본 연구에서는 먼저 현행 연구에서 어떤 방식으로 보상함수를 설계하고 있는지를 분석하였다. 그리고 그 과정에서 정주기 신호주기 모형을 활용하여 정량적으로 문제점을 분석해보았다. 그 결과 기존의 변수들이 어떠한 측면에서 교통 상황을 평가하고, 서로 어떤 영향을 주고받는지를 알 수 있었다. 그리고 이를 바탕으로 어떻게 보상함수를 개선할 수 있을 지를 알아보았다. 그 결과, 지체도 자체를 최소화하는 것도 중요하나 그 과정에서 증가하는 지체도의 차이가 상당히 커지므로 이 측면의 고려도 중요함을 정량적으로 보였다. 그리고 기존의 보상함수에 더해 추가적으로 고려되어야 하는

변수로 사고율, 정지횟수, 등도 제시하였다.

앞으로 본 연구는 인공지능, 특히 강화학습을 이용하여 교통신호제어 문제를 해결하려고 함에 있어 어떤 식으로 보상을 설계하는 것이 좋은지 길라잡이 역할을 할 수 있을 것이다. 그리고 추후에 각 가중치나 변수의 조합에 관한 연구를 진행하여 직접 최적의 보상함수도 만들고 이를 적용한 학습을 진행하여 그 효용도 검증해볼 것이다.

Acknowledgement

이 논문은 경기과학기술대학교의 지원에 의함.

References

- [1] J. Y. Lee, M. S. Chang, and J. T. Kim, “Development of Neural Network Based Cycle Length Design Model Minimizing Delay for Traffic Responsive Control”, *Korean Society of Transportation*, 22, 3, 145-157, Mar. 2005
- [2] S. M. Park, “Development of Traffic Signal Control Model Using AI Technique”, Ajou University, Feb. 2020
- [3] M. G. Pi, H. S. Lee, and M. Y. Chung, “Reinforcement Learning-based Traffic Signal Control under Real-World Constraints”, *Journal of KIISE*. 48, 4, Aug. 2021
- [4] S. H. Ryu, H. S. Lim, “교통공학과 기계학습의 연구 추이와 한계점 - 신호교차로 운영방법론을 중심으로”, *Monthly KOTI Magazine on Transport*, 259, Sept. 2019