ARTICLE

유턴을 고려한 교통축 적용 수요대응 자율주행 대중교통(DRDTO)의 실시간 동적 경로 생성 알고리즘

김현[1*] · 유선형[2] · 이진우[3] · 백범열[4] · 신정희[4]
[1]한국교통대학교 교통에너지융합학과 부교수, [2]뉴사우스웨일즈대학교 도시계획과 박사과정,
[3]뉴사우스웨일즈대학교 도시계획과 부교수, [4]한국교통대학교 교통ICT융합연구센터 학생연구원

Real-Time Dynamic Route Generation Algorithm for Demand-Responsive Driverless Transit Operation (DRDTO) Applied to Corridors to Consider U-Turns

KIM, Hyun[1*] · YOO, Seonhyung[2] · LEE, Jinwoo[3] · BAEK, Beomyeol[4] · SHIN, Junghee[4]

[1]Associate Professor, Department of Convergence of Mobility and Energy, Korea National University of Transportation, Chungju 27469, Korea
[2]Ph.D. Candiate, City Planning, School of Built Environment, Faculty of Built Environment, University of New South Wales, Sydney, NSW, Australia
[3]Associate Professor, City Planning, School of Built Environment, Faculty of Built Environment, University of New South Wales, Sydney, NSW, Australia
[4]Student Researcher, Center for Convergence of Vehicle, People and ICT, Korea National University of Transportation, Chungju 27469, Korea

*Corresponding author: hyunkim@ut.ac.kr

J. Korean Soc. Transp.
Vol.40, No.2, pp.260-276, April 2022
https://doi.org/10.7470/jkst.2022.40.2.260

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY
Received: 25 November 2021
Revised: 24 December 2021
Accepted: 28 March 2022

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Recently, autonomous vehicle is receiving much attention in various sectors including transportation and public transportation. This study presents an adaptive routing algorithm for the real-time demand response service of autonomous transit vehicles. The experimental setting includes the real-time demand occurring randomly over time within the study area. Each demand (service call) request the transit service from designated origin to destination. The routing algorithm is designed to make U-turn and skip stops to improve service and reduce the wait time of users. This study adapts Reinforcement Learning, one of the machine learning techniques, namely reinforcement learning, which can precede a complex calculation process in an offline process. Simulation experiments conducted on a testbed of the Chungju campus of the Korea National University of Transportation. The simulation results show that the proposed routing algorithm can improve the adaptive transit service over the fixed operation in selected performance indicators.

Keywords: autonomous vehicle, demand response, flexible route, public transport, reinforcement learning

초록

최근 자율주행 차량이 대중교통분야로 진출하여 관련 연구들이 활발히 진행되고 있다. 본 연구에서는 시공간적으로 다양하게 발생하는 실시간 호출수요에 대응하여 자율주행 차량에 동적

경로를 제시하여 효율적인 대중교통 서비스를 제공할 수 있는 알고리즘을 개발했다. 실시간 호출수요는 다수의 출발지와 다수의 목적지로 구성되어 복잡하고 다양한 경우의 수를 고려하여 최적 경로를 제시할 수 있는 방안이 필수적이다. 복잡한 계산과정은 오프라인 과정에서 선행할 수 있는 방법으로 머신러닝의 강화학습 기술을 사용하여 제시하였다. 개발된 자율주행 대중교통(DTO)를 위한 실시간 호출 수요대응 동적경로 알고리즘은 테스트베드를 선정하여 시뮬레이션 실험이 이루어 졌다. 시뮬레이션 결과는 개발된 실시간 호출 수요대응 동적경로 생성 알고리즘을 고정노선의 운행결과 대비하여 효율적인 운행결과가 나타났다. 이는 실시간 호출수요량이 적을수록 효과가 크게 나타나지만 호출 수요량이 증가하면서 그 효과 폭이 점차적으로 감소했다.

주요어: 강화 학습, 동적경로, 대중교통, 수요대응, 자율주행 대중교통

서론

자율주행자동차는 효율적인 속도제어를 통해 원할 한 교통흐름을 유지하여 전체 도로의 용량을 증대시킬 수 있으며, 위험 상황이 발생하는 경우 차량을 자동으로 제어함으로써 운전자 과실로 인한 사고 감소, 그리고 교통약자에게 차량 이용의 확장 등을 기대하고 있다(Fagnant and Kockelman, 2015; Becker and Axhausen, 2017; Faisal et al., 2019; Pettigrew et al., 2019). 자율주행 대중교통 시스템의 이점은 거동이 불편한 사람, 노인들과 여성, 저소득자 등을 포함한 모든 사회 계층을 수용할 수 있다(Raj et al., 2020). 그러나, 이런 기대효과에도 불구하고 자율주행자동차의 높은 구매 비용은 사회 소외 취약계층에게 오히려 이동성을 감소시킬 수 있다(Yigitcanlar et al., 2019). 자율주행 대중교통은 상대적으로 관심이 낮지만 많은 전문가들은 자율주행자동차가 대중교통에 도입될 경우 사회적으로 상당한 이익을 제공할 것으로 기대하고 있다(Guo et al., 2021). 예를 들어, 이동제한이 있는 사람들에게 편리한 교통 이동 수단을 제공하여 형평성을 높일 뿐만 아니라 이동의 안정성을 향상 시킬 수 있다.

한편 자율주행 대중교통은 자율주행기술에 대한 부정적 인식 때문에 오히려 대중교통수단으로 제한이 있을 수 있다. Kassens-Noor et al.(2020)의 연구에 따르면 대중교통을 이용하는 사람들의 약50%가 운전자와 같은 관리자의 부재와 기술에 대한 신뢰도 부족으로 자율주행 대중교통 이용을 주저하는 경향이 있다. 그럼에도 불구하고 세계 주요 도시에서 자율주행 대중교통은 기존 대중교통 노선 대상으로 시험운행을 활발하게 진행되고 있다(Iclodean et al., 2020; Narayanan et al., 2020; KoNUT, 2022). 이는 자율주행 대중교통 원격관제 또는 안전요원이 탑승하여 상시적 감시체계로 운행되고 있다. 최근 운전자 또는 안전요원이 동반하지 않고 운행시킬 수 있는 자율주행 차량들이 세계 몇몇 도시에서 시험 운행 중이다. 가장 대표적인 예로 Google의 자율 주행 자동차 Waymo는 미국 애리조나주 피닉스에서 실시중이다(LeBeau, 2018). 또한 인프라 협력 자율주행기술을 이용한 운전사가 없는 대중교통 운영(DTO, Driverless Transit Operation)과 DRT(Demand Responsive Transit)시스템의 융합한 서비스의 시도가 있다(KoNUT, 2022). 이는 사람 운전자가 있는 버스 노선에서 운전자가 없는 DTO로 발전시키는 과정으로 자율주행 기반의 DTO기술에 대한 대중적 신뢰성(Trust)를 얻기 위한 실증에 해당한다. Iclodean et al.(2020)에서 제시한 사례는 모두 고정 노선대상으로 DTO서비스를 일부 제공하고 있으나, 호출수요에 대응하는 동적인 대중교통 노선을 운영하는 실증은 제시되어 있지 않고 있다.

본 연구에서는 대중교통의 경영과 운영을 고정노선과 비교하여 보다 효율화 시키고 이용자에게 더 빠르고 원하는 시각에 서비스를 제공할 수 있는 실시간 기반 DRT시스템과 DTO 융합 서비스로의 진화에 초점을 두고, DRT 운영을 위한 동적 경로생성 알고리즘 개발을 목적으로 하고 있다. 따라서 호출 수요 대상으로 하는 1대의 서비스 차량에 대한 배차 알고리즘과 다르게 다수의 실시간 호출 수요 및 위치, 자율주행 대중교통 차량의 ODD(Operational Design Domain)와 정시성, 그리고 차량의 유휴시간 최소화, 호출자 이동시간 최소화 등을 고려하고자 Cordeau and Laporte(2007)의 연구에서 DARP(Dial-A- Ride Problem)에 대한 동적 경로에 중점을 두고 있다. 또한 본 논문에서는 실시간 호출수요에 대응하는 자율주행 대중교통 차량의 동적 경로를 제공할 수 있는 강화학습 알고리즘으

로 Tafreshian et al.(2021)의 접근 방식과 같이 많은 양의 복잡한 계산과정을 미리 오프라인 과정에서 거치는 지도 학습 방법을 제안하기 위해 한국교통대학 충주캠퍼스에 적용하여 동적 경로알고리즘의 유효성을 확인하고자 한다.

선행연구

출발지와 목적지 간의 이동 서비스의 배차는 n명의 사용자에 대한 차량 경로 및 일정 설계에 해당하는 것으로 목적함수는 일련의 제약 조건에서 가능한 한 많은 사용자를 수용할 수 있는 최소 비용 차량 경로를 계획하는 DARP(Dial-a-Ride Problem)이 널리 사용되었다. 가장 일반적인 예는 노인이나 장애인 대상의 Door to Door 서비스가 있다(Diana and Dessouky, 2004; Rekiek et al., 2006; Melachrinoudis et al., 2007). 또한 동적 배차에 적용하는 DARP는 Psaraftis(1980)에 의해 새로운 요청이 t_1시점에 요청되면 계획된 배차 솔루션을 사용할 수 있지만 새 요청을 포함하여 시간 t_n부터의 배차 솔루션은 다시 최적화하는 과정이 필요했다. 여기서 발생하는 계산시간의 이슈는 Gendreau et al.(2001)의 동적 구급차 재배치연구 에서 병렬 컴퓨팅을 사용하여 여러 시나리오를 미리 계산하는 방이 제안되었다.

현재 운행되고 있는 대부분의 대중교통 서비스는 고정된 노선과 정해진 스케줄로 운영되고 있어 실제 실시간 수요에 능동적으로 대응하는 데 한계가 있다. 반면, 택시와 같은 교통수단은 실시간 수요 대응으로 서비스를 제공할 수 있지만 일대일 수요대응 방식 서비스를 제공하기에 많은 비용을 필요로 한다. 최근 우버(Uber)나 디디(Didi), 그리고 카카오(Kakao)와 같은 공유 택시 플랫폼의 등장으로 실시간 호출 서비스가 가능하게 되었다. 실시간 호출수요와 서비스 제공을 적절히 매치시킴으로 차량의 비어 있는 시간을 최소화 하였지만, 여전히 단일 서비스 제공이라는 제한적인 효율성 향상이 이루어졌다. 즉 동일한 차량으로 다수의 호출 수요자 대상으로 우회 시간을 최소화 하며 다수의 목적지를 가진 승객들에 수요대응 서비스를 제공하기 위한 연구들이 최근 진행되었다(Sultana et al., 2018; Lee et al., 2021).

Schilde et al.(2011)의 연구는 외래환자들의 실시간 수요에 대응하여 집에서 병원까지 왕복 교통 서비스를 제공하기 위한 메타 휴리스틱 알고리즘을 구현했다. 이 시스템은 과거 서비스 사용이력을 바탕으로 확률적으로 실시간 수요를 예측하고 돌아오는 귀가 호출 시기에 맞는 서비스 제공을 위해 설계되었다.

Ritzinger et al.(2016)의 연구는 확률적으로 수요를 예측하여 차량 경로 서비스를 유동적으로 제시 하려는 연구들을 전반적으로 검토했으며, 이와 같은 수요대응 서비스 방법으로 고정적인 차량경로 운행에 비해 최대 60%까지의 운행 비용을 절감할 수 있었다. Lee et al.(2019) 연구에서는 미래 자율주행차 운행을 위한 모의 어닐링 알고리즘(simulated annealing algorithm)을 이용하여 다중 역과 열차에 대한 버스 재배치를 고려한 최적의 First and Last Mile 제공을 위한 버스 경로 알고리즘을 개발하였다. 이 알고리즘은 특정 역에서 사용할 수 있는 버스를 사용하여 최적의 버스 라우팅이 가능하지 않을 때 버스 재배치를 성공적으로 처리했다. 또한 총 차량 운영 비용과 총 승객 차량 내 이동 시간 비용을 포함한 총 비용을 최소화했다.

실시간 사전 예약 수요 대응형 차량의 배차 관련 연구(Huang et al., 2019)는 기존 실시간 수요대응형 차량 배치 알고리즘과 다르게 실시간 및 사전 예약 수요에 대한 차량 배차 최적화 알고리즘을 개발하여 실제 자료에 기반하여 알고리즘을 평가하였다. 한편 Li et al.(2019)은 차량과 승객을 직접 연결하여 배차를 하는 것이 아닌 유사한 승객의 수요를 이용하여 배차하는 방식 연구하였으며 이를 위해 cluster기반 승객의 유사성을 묶는 알고리즘을 개발하여 총 통행거리와 차량의 수 감소에 기여하는 것을 확인하였다.

최근 Tafreshian et al.(2021)은 대규모의 실시간 수요에 대응하여 다수의 버스들을 적절히 배차시킬 수 있는 방법을 고안했다. 오프라인 단계에서 복잡하고 시간이 많이 걸리는 부분을 처리 할 수 있는 방법을 고안하여 실시간 단계에서 신속하고 효과적인 솔루션을 제안했다. 버스 운영에 대한 동적 경로 서비스 이점은 우회 시간 비용을 고려하면서 수요량을 최대화시킬 수 있다. 이 문제를 효과적으로 해결하기 위해 Lee et al.(2021)은 차량 경로선택 문제와 승객 대 차량 할당 문제로 분리하여 승객수요와 우회 시간 신뢰성을 최적화시키기 위해 Gradient와 Greedy Search

솔루션을 결합하는 연구를 수행하였다. 이는 OD정보와 요청당 승객 수를 바탕으로 수요량과 우회 시간을 확률 분포로 추론하여 서비스 지역을 구역으로 세분화 한 다음 유동적인 실시간 수요대응 서비스 제공방안에 해당한다.

대부분의 수요대응서비스에 관한 연구들은 운행비용과 소비자 측면의 비용을 최소화 할 수 있는 방법들을 고안하는데 중점을 두었다. 서비스 운행비용은 기존 고정된 서비스와 대비하여 계산되었고, 서비스를 이용하는 수요자 입장에서는 기존 고정 노선과 비교해 서비스 대기시간이나 우회시간으로 향상된 서비스 질을 평가하였다. 지역내 기존 고정노선이 없는 경우 자가용 대비 우회시간으로 수요자의 측면의 비용을 계산했다. 자율주행 관련 기술의 발전으로 승객의 서비스 질을 보다 향상시키기 위해 기존 DARP(Cordeau and Laporte, 2007)의 문제를 수리적으로 최적화하는 방식에서 인공지능을 활용하는 방식으로 방법론은 진화하고 있다.

본 연구에서는 자율주행 대중교통이 고정노선에 제한되어 운행되어 지는 것이 아니라 실시간 대중교통 수요를 바탕으로 최적의 동적 경로를 생성할 수 있는 AI 알고리즘 개발을 목표로 하고 있다. 선행연구들은 대중교통 버스에 동적 경로 안내의 대부분이 출발지 혹은 목적지가 한 곳으로 지정되어 있는 교통축 서비스에 중점을 두어 진행되었다. 본 연구는 실시간 수요가 다수의 출발지로 부터 일어나며 다수의 도착지에 서비스를 제공하는 것을 목적으로 실용적인 해결방안을 제시하고자 Tafreshian et al.(2021)의 접근 방식과 같이 많은 양의 복잡한 계산과정을 미리 오프라인 과정에서학습하는 방법을 포함하고 있다.

모델 모형화 및 알고리즘 개발

1. 연구 흐름도

본 연구의 수행 절차는 Figure 1에서와 같이 4단계로 나누어져 있다. 먼저 문헌조사와 관련 실시간 수요대응 서비스 문제를 상세히 파악하여 동적경로 알고리즘을 개발한다. 개발된 알고리즘은 테스트 베드를 선정하여 가상의 시뮬레이션을 실시하고, 그 결과에 근거하여 적정한 평가 기준을 마련하여 본 연구에서 개발된 알고리즘의 효용성을 검증한다. 그리고 종합적인 연구과정과 실험결과에 근거하여 결론과 향후 발전방향에 대해 토론한다.

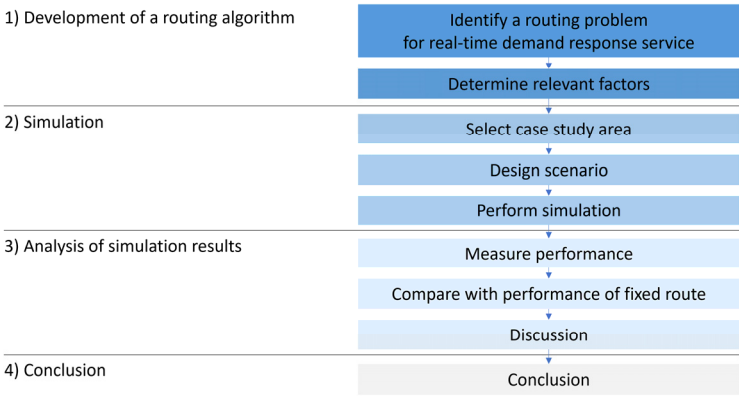

Figure 1. Research flowchart

2. 동적 경로 알고리즘 개발

본 연구는 실시간 호출수요에 대응하여 DTO에 유동적인 경로를 제시하기 위하여 강화학습을 도입한다. 강화학습은 특정 환경 내에서 에이전트가 반복적인 학습을 통해 경험을 축적한 다음 특정한 상황에서 학습에 기반하여 최대의 보상을 성취하도록 개발된 머신러닝 기술이다(Sutton and Barto, 1998). 강화학습에는 크게 다섯 가지의 주요소로 구성되어 주변 상태(state)에 따라 어떤 행동(action)을 할지 판단을 내리는 주체인 에이전트(agent)가 있다.

에이전트가 속한 환경(environment)에서 에이전트가 행동을 하면 그에 대응하여 상태가 변하며, 변화된 상태에 따라 보상(reward)을 받을 수도 있다. 강화학습내의 주 다섯 가지 요소의 관계도는 Figure 2로 표현하였다.

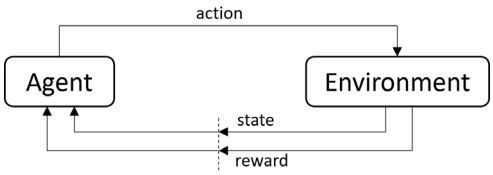

Figure 2. The agent-environment interaction in reinforcement learning

상태(State)는 현재 시점에서 상황이 어떤지 나타내는 값의 집합으로 정의할 수 있다. 가능한 모든 상태의 집합은 S로 표현한다. 여기서 특정시간에서의 상태 값은 S_t로 표현한다. 행동(Action)은 본 연구에서 자율주행 대중교통 시슬템에 해당하는 에이전트가 어느 상태에서 결정하는 선택으로 정의한다. 따라서 가능한 모든 행동의 집합은 A로 표현한다. 특정 시각 t의 상태 S_t에서 선택한 행동은 a로 표현한다.

강화학습 내에서 에이전트가 일련의 행동을 통한 환경과의 상호작용은 강화학습내의 환경이 Markov property를 가진다고 볼 수 있다. Markov 의사결정 프로세스(decision process)는 행동으로 주어지는 즉각적인 보상과 차후의 상태들 및 보상들에 대한 의사결정 모델링이 가능하게 한다. 이와 같은 강화학습 문제를 수학적으로 표현할 수 있게 하며, 상태 S_t에서 상태 S_{t+1}로 가는 행동 a가 선택하는 확률은 Equation 1과 같이 표현된다.

$$p(s'|s,a) = \Pr\{S_{t+1} = s' \mid S_t = s,\ A_t = a\} \tag{1}$$

where, S : A set of possible states interacted with the environment;
A : Aset of possible actions of the agent;
$s', s \in S$ and $a \in A$.

에이전트(Agent)가 어떤 행동을 했을 때 주어지는 이득은 보상(Reward)으로 표현되며 상태 S_t에서 행동 a을 선택해서 상태가 S_{t+1}로 변화되면서 주어지는 보상을 r이라고 하고 이 함수는 Equation 2로 표현된다.

$$r(s,a,s') = E\{R_t | S_t = s,\ A_t = a,\ S_{t+1} = s'\} \tag{2}$$

where, R : A set of rewards; $r \in R$.

강화학습 알고리즘의 최종 목표는 에이전트가 주어진 특정 환경과 효과적으로 상호작용하여 최적의 행동들을 취하여 사전 정의된 평가 척도에 기반하여 최적 결과를 도출하는 데 목표가 있다. 그러므로 에이전트가 적절한 행동을 취하기 위한 인지되는 환경의 상태는 Equation 3과 같이 합리적으로 구조화되는 것이 필수적이다. 주어진 환경에서 취 할 수 있는 여러 행동들 중에 에이전트가 의사결정한 다음 하나의 행동을 판단하는 방식은 정책(Policy)에 의해 결정된다. 환경과 정책에서 에이전트의 현재 상황은 가치함수로 평가될 수 있으며, 대부분의 강화학습 알고리즘은 Equation 4의 가치함수를 추정하여 주어진 상태에서의 수행될 행동들의 가치가 업데이트 된다.

$$v_\pi(s) \ = \ E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s\right\} \tag{3}$$

$$q_\pi(s,a)=E_\pi\left\{\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} \mid S_t=s,\ A_t=a\right\} \tag{4}$$

where, π : Policy;

$v_\pi(s)$: The value of state a under policy π;

$q_\pi(s,a)$: The value of action a in state under policy π;

γ^t : A discount rate at time step $t\,(0 \le \gamma \le 1)$.

가치함수는 강화학습내에서 에이전트가 장기적인 결과를 기다리지 않고 현재 상황을 판단할 수 있는 중요한 역할을 한다. 장기적으로 최적의 보상을 얻기 위해 기대되는 가치 함수는 Equations 5, 6으로 표현할 수 있다. 최적 상태 가치 함수는 일련의 일어날 상황들을 기다리지 않고 즉시 장기적으로 기대되는 보상을 정량화 할 수 있다. 그러므로 장기적으로 최적의 행동들을 즉각적으로 예견할 수 있다.

$$v_*(s)=\max_{a\in A(s)} q_{\pi^*}(s,a)=\max_{a\in A(s)}\sum_{s',r}p(s',r|s,a)\left[r+\gamma v_*(s')\right] \tag{5}$$

$$q_*(s,a)=\sum_{s',r}p(s',r|s,a)\left[r+\gamma\max_{a'}q_*(s',r')\right] \tag{6}$$

본 연구는 자율주행 대중교통 차량에 배정된 실시간 수요들의 정보를 바탕으로 최적의 동적 경로를 제공하기 위한 알고리즘 개발을 목적으로 하였다. 기존의 자율주행 대중교통 차량 운행에 관한 연구들은 고정된 노선에 운행되며, 자율주행 대중교통 차량이 여러 다양한 상황 속에서도 운행자 간섭 없이 운행될 수 있도록 주변 환경인식 능력 향상에 초점을 두었다. 하지만 본 연구는 자율주행차량이 운행자가 상시적인 간섭이 필요 없는 레벨4 또는 4+ 단계에서 운행된다는 가정 하에 고정 노선으로 왕복 운행하는 것이 아니라, 실시간 수요에 대응하여 다양한 현재 상태를 고려한 최적의 동적 경로를 DTO 차량에게 제시하기 위한 것이다.

Figure 3에서는 실시간 호출수요 대응 DTO의 동적경로 알고리즘을 위한 강화학습내의 구성요소는 다음과 같이 분류되어 정의 될 수 있다. 우선, 환경(environment)는 자율주행 차량이 운행될 자세한 도로정보와 승하차가 가능한 정류장들 위치정보를 기반으로 설계되어 진다. 이렇게 설계된 환경에서 운행될 자율주행 차량은 강화학습 내에서

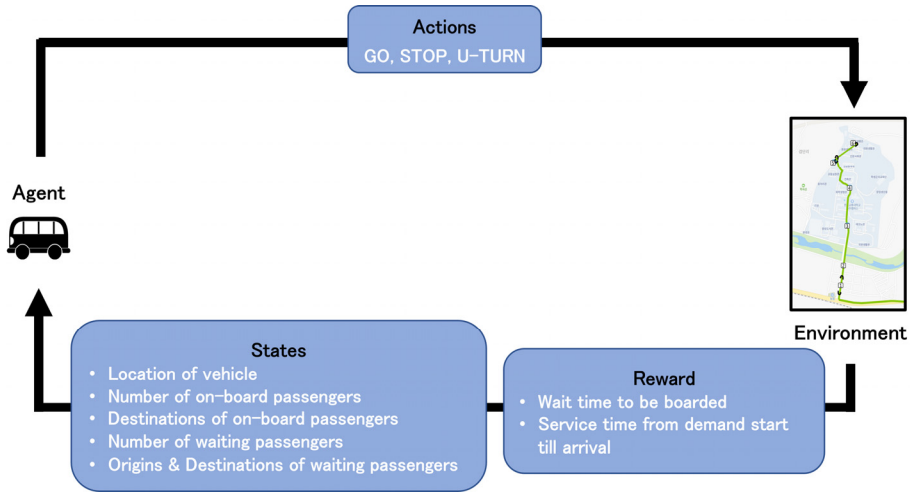

Figure 3. Interaction of components in reinforcement learning for real-time demand response service

에이전트(agent)로 정의 되며, 환경 위에서 에이전트가 취할 수 있는 행동(action)은 전진, 정차 및 유턴이다. 유턴의 행동은 모든 곳에서 이루어지는 것이 아니라 특정한 위치에서만 가능하도록 환경이 설계되어 질 것이다. 즉 본 연구에서는 교통축을 대상으로 First and Last Mile 서비스를 지향하는 도로망으로 동적 경로를 제공하기 위해 유턴 지점을 3개소에 설계하였다. DTO 차량이 승객을 버스에 탑승시키거나 하차 시킬 때 적정한 보상(reward)이 이루어진다. 모든 호출 수요를 탑승할 때 주어지는 보상은 일정하며, 최종 목적지에 도착하여 하차시 주어지는 보상은 OD 거리에 대비하여 결정되어 진다. 또한 매 행동마다 추가적으로 패널티 보상을 고려하여 최단의 효율적인 동적 경로로 학습을 유도한다. 강화학습내의 상태는 DTO 차량(agent)의 위치, DTO 차량 내의 탑승객 수 및 각 탑승객의 최종 목적지, 아직 탑승하지 않은 수요들의 탑승 위치 및 최종 목적지의 정보들을 고려하여 결정되어 진다.

본 연구에 적용되는 강화학습 알고리즘은 강화학습의 연구가 시작되고 초기에 개발된 Q-learning을 적용한다. Q-learning 알고리즘은 현재 상태에서 에이전트가 어떠한 행동이 최선의 결정인지를 알려주는 정책을 학습시키는 목적으로 한다(Sutton and Barto, 1998). 그 구체적인 방법으로는 일련의 행동들로 주어지는 보상들의 합이 최대가 될 수 있는 정책이 되도록 학습된다. 미래의 모든 가능한 보상의 합은 매회 계산하는 것이 아니라 Q-learning 현재 행동으로 주어지는 즉각적인 보상과 현재 행동으로 변경되는 상태에서의 보상들의 조합을 이용하여 행동의 가치를 추정하는 시간차 알고리즘으로 알려져 있다. 이와 같이 추정된 값은 Q-value로 정의되어 상태-동작의 쌍의 항목으로 이루어진 테이블에 저장된다. Q-value가 추정되는 수학적 표현과 계산된 Q-value가 이전에 테이블에 저장된 값에 업데이트되는 방법은 Equation 7로 표현된다.

$$Q_{t+1}(s_t,a_t)=Q_t(s_t,a_t)+\alpha\left[r_{t+1}+\gamma\max_a Q_t(s_{t+1},a)-Q_t(s_t,a_t)\right] \tag{7}$$

where, α : A learning rate coefficient that keeps avoidingpremature convergence;

r_{t+1} : The current reward;

$\gamma\max_a Q_t(s_{t+1},a)$: The current estimated value of the sum of discounted future rewards.

강화학습(Q-learning) 알고리즘 내에서 상태에 따른 Q-value들이 업데이트 되어 정책이 학습되어지는 일련의 과정은 Table 1에 표현했다. 이는 학습이 시작되기 전에, 모든 상태-동작 쌍의 항목으로 이루어진 Q-value들은 0의 값으로 초기화된다. 에이전트는 행동을 취하면서 취한 행동에 대한 보상, 이전 상태, 그리고 취한 행동에 따른 새로운 상태 정보를 바탕으로 Q-value를 업데이트한다. 이와 같은 과정을 반복하여 다수의 에피소드가 수행됨에 따라 상태-동작 쌍의 Q-value들이 최적의 값으로 수렴한다. 최종적으로 학습된 정책은 에이전트가 최선의 행동들을 취할 수 있도록 하는 것이다.

Table 1. Procedures of reinforcement learning (Q-learning) algorithm

1:	Set $t=0$ and initialise $Q_0(S_0,a)$for all $s_0\in S, a\in A$
2:	Observe the initial state s_0
3:	Repeat (for each episode):
4:	Initialise $S=s_0$
5:	Repeat (for each step of the episode):
6:	Choose from using policy derived from Q (e.g. e-greedy)
7:	Take action a, observe, r_{t+1}, S_{t+1}
8:	$Q_{t+1}(s_t,a_t)=Q_t(s_t,a_t)+\alpha\left[r_{t+1}+\gamma\max_a Q_t(s_{t+1},a)-Q_t(s_t,a_t)\right]$
9:	Set $s_t=s_{t+1}, t=t+1$
10:	Until S is terminal
11:	End repeat

강화학습 알고리즘 내에서 에이전트는 두 가지 방식으로 행동을 결정한다. 첫 번째 방식은 무작위로 행동하는 것이다. 이와 같은 무작위 행동은 탐색(exploration)이라고 정의되며, 환경을 탐색하며 그에 따른 행동의 가치(보상값) 정보를 얻기 위함이다. 에이전트의 탐색을 통한 행동방식은 강화학습 초기에 새로운 상태-행동의 대한 정보를 얻기 위해 필수적이다. 두 번째 행동방식은 이미 학습된 경험을 참조하여 에이전트가 특정 상태에 가능한 행동들 중 최고의 Q-value를 기준으로 행동을 선택하여 취하는 방식이다. 이와 같이 이미 경험한 정보를 바탕으로 행동을 취하는 방식을 탐사(exploitation)라고 정의한다. 방대한 탐색이 없으면 에이전트는 로컬 최적 값에 빠질 수 있기에 탐사와 탐색의 적절한 균형이 매우 중요하다. 본 연구에서는 가장 대표적으로 쓰이는 탐사와 탐색간을 결정하는 ε(epsilon)-greedy 정책을 사용할 것이다. ε-greedy 정책에서 에이전트의 행동 선택은 Equation 8로 표현된다. 에이전트는 ε의 확률로 무작위 행동의 탐색을 하고, (1- ε)의 확률로 학습에 결과에 기초한 최적의 탐사행동을 취한다. 학습과정에서 점차적으로 감소하는 ε값을 채택하여 초기단계에서 여러 다양한 환경을 많이 탐색하며, 학습 최종 단계에서 너무 무작위적인 행동 없이 탐사를 통한 최적의 Q-value로 수렴하도록 할 것이다.

$$a_t = \begin{cases} \underset{a \in A}{\arg max}\, Q_t(s_t, a); & with\ \ probabilty(1-\epsilon) \\ \underset{a \in A}{rand}(a); & with\ \ probabilty(\epsilon) \end{cases} \tag{8}$$

일정한 시간 간격내에 발생하는 실시간 수요량을 점차적으로 늘려가며 에이전트가 최선의 행동들을 결정하도록 학습시킬 것이다. 일정한 에피소드를 기점으로 현재까지 학습된 에이전트가 100% 탐사의 행동을 취하도록 했을 때, 일정한 대중교통 서비스 기준을 충족시킨다면, 현재의 실시간 수요량의 학습이 완료되었다고 가정하고 수요량을 늘린다. 반면, 서비스기준을 충족시키지 못한다면 같은 실시간 수요량으로 학습과정을 반복한다.

주행 시뮬레이션 실험

1. 주행 시뮬레이션 환경 구축

본 연구에서 개발하는 실시간 수요대응 동적 경로 알고리즘은 교통대학교 캠퍼스와 캠퍼스 타운앞 교차로를 대상으로, DTO 차량의 주행 시뮬레이션 환경을 구축한다. DTO 차량이 정차하여 승객이 승하차 할 수 있는 총 13개의

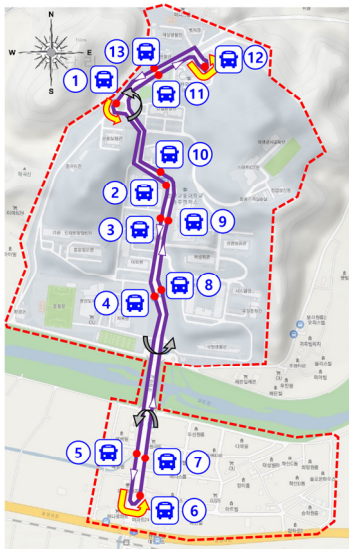

Figure 4. Environment of demand-response service for autonomous transit vehicle

정류장을 사전에 정하여 교통대 충주캠퍼스 내외에서 실시간 수요가 발생시 가장 가까운 버스 정류장으로 안내한다. DTO차량이 캠퍼스내외에서 운행할 도로망은 Figure 4에 제시하고 있다. 도로망의 6번과 12번 정류장에서는 회차 가능한 환경이 조성되었다. 또한 중간에 2개 유턴 장소와 1개의 회전교차로가 있어서 실시간 대중교통 호출수요 위치 정보에 따라 총 12개의 동적 경로가 생성될 수 있도록 시뮬레이션 환경을 구축하였다.

정해진 13개의 정류장은 남/북행 정류장으로 구분되어진다. 북행 정류장은1, 6, 7, 8, 9, 10, 11, 12번 정류장들을 포함하며, 남행 정류장은 1, 2, 3, 4, 5, 6, 12, 13번 정류장들을 포함한다. 양끝의 6, 12번 정류장과 회전교차로에 위치한 1번 정류장은 남북행 정류장에 모두 포함시킨다. 출발 정류장과 목적 정류장은 모두 같은 행 정류장에 항상 속해 있어야 된다는 가설에 본 연구는 남/북행 정류장으로 구분하는 것이다. 예를 들어, 3번 정류장에서 출발해서 9번 정류장으로 도착하는 여행은 반대 방향 정류장에 속해 있기에 요청되지 않는다. 실시간 수요가 발생했을 때, 남행과 북행으로 분류되어 정해진 여행 방향에 정류장들 중 가장 가까운 정류장으로 출발지가 배정되어 안내되어진다. 가능한 출발지-목적지 정류장의 쌍은 Table 2에 정류장간 거리와 함께 나타내어 졌다.

Table 2. Origin-destination pairs with corresponding distances

Distance (m)		Origins												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Destinations	1						1,095	1,015	595	405	270	290	165	
	2	270											560	435
	3	370	100										660	535
	4	570	300	200									860	735
	5	970	700	600	400								1,260	1,135
	6	1,180	910	810	610	210							1,470	1,345
	7						80							
	8						500	420						
	9						690	610	190					
	10						825	745	325	135				
	11	90					1,185	1,105	685	495	360			
	12	200					1,295	1,215	795	605	470	110		
	13												125	

현재 전국에서 시험운행중인 자율주행 셔틀 버스는 최대 25km/h의 속도로 운행되고 있으며, 본 연구에서는 실시되는 주행시뮬레이션에서는 여러 정차 및 유턴 환경을 고려하여 평균 20km/h로 설정했다. 조성된 시뮬레이션 환경의 최소 격자 사이즈는 10×10m이며 DTO 차량은 20km/h로 운행하는 조건에서 에이전트로 매회의 행동마다 2초의 시간이 흐른다. 이는 2초마다 새로운 호출수요가 발생한다면 이에 대응하여 DTO 차량에 새로운 동적 경로에 해당되는 행동들을 제공하게 된다. 실제적으로 교통대 충주캠퍼스 대상으로 운영예정인 DTO 차량의 승차용량(최대 15인)과 동일한 에이전트 역할이 되어 강화학습 동적 경로 알고리즘 내에서 학습이 진행된다.

상태는 현재 DTO 차량의 위치, 만차 여부, 현재 탑승한 승객들의 목적지 정류장 정보, 탑승을 기다리는 승객들의 출발지/목적지 정보가 포함된다. 현재 테스트 베드를 대상으로 한 DTO 차량은 행동을 취하게 된다. 강화학습 내에서 에이전트는 최대의 보상 값을 찾는 일련의 행동들로 학습되어지기 때문에 적절한 보상 값을 정하는 것이 매우 중요하다. 본 연구에서는 승객의 탑승이 이루어졌을 경우와 목적지에서 하차가 이루어 졌을 때로 구분하여 크게 2가지 경우에 주어진다. 모든 호출수요에 대해 동일한 보상 값이 주어진다면, 에이전트는 원거리 수요는 등한시하고 단거리수요에 대해서만 중요하게 인식하여 편중된 서비스를 제공하려고 한다. 따라서 탑승 승객을 목적지 정류장에 하차시켰을 때 주어지는 보상 값은 Table 2와 같이 거리에 대비하여 전반적인 수요에 균형된 서비스를 제공하도록 유도한다. 반면 탑승에 대한 보상 값은 너무 높게 책정한다면, DTO 차량이 승객을 탑승시키는 데에만 중점을 두고 탑승 이후 목적지까지 서비스제공이 적절히 이루어지지 않을 수 있다. 그러므로 전체적인 하차 보상 값은 Table 2에 대한 평균값 500이 승객을 하차시켰을 때 주어질 것이다. 또한 탑승과 하차의 행동이 아닌 경우 패널티 보상이 10을

부여하여 최단 거리 경로가 학습되어 지도록 유도한다. 추가적으로 hyper parameter에 해당하는 discount factor는 0.95를 적용하여 최단거리 동적 경로의 패널티 보상 값과 함께 유도한다.

강화학습 내에서 에이전트는 한개 에피소드 내에서 각 시점의 상태에 따라 이루어지는 일련의 행동들과 그에 따른 보상 값들의 이력으로 학습이 이루어진다. 본 연구와 같이 실시간 호출수요의 OD 위치와 발생시점이 일정하지 않는 경우 에피소드의 길이와 학습을 완료하는 적절한 조건을 정하는 것이 매우 중요하다. 학습 중 한 개의 에피소드가 시작될 경우 발생되는 총 수요량이 정해진다. 이때 정해진 모든 수요들이 목적지까지 서비스 제공이 완료되면 에피소드가 끝나는 조건으로 설정한다. 반면, 호출수요가 발생한 이후 너무 많은 차내시간(in vehicle time)을 소비하여 목적지까지의 경로는 가치가 없는 학습에 해당한다. 따라서 이런 학습조건은 지양하기 위해 호출수요 기준 차내시간(in vehicle time)이 10분을 초과하는 경우 에피소드가 정지하도록 한다.

상태 정보는 DTO 차량의 현재 위치, 호출 수요 OD정보 및 탑승한 승객의 최종목적지를 모두 포함하기에 충분한 에피소드 길이에 다양한 시점과 다양한 OD종류의 수요분포가 되어 DTO 차량의 동적 경로 학습이 이루어져야 한다. 이는 초기 호출수요들이 발생하는 기간, 초기 호출수요들을 탑승시킨 이후 호출수요들이 발생할 기간, 초기 초기 호출수요 하차 또는 후기 호출수요의 탑승 기간, 후기 호출수요의 하차 기간 등을 모두 고려할 경우 약 4번의 왕복운행을 고려하여 40분은 최대 에피소드 길이로 설정한다. 한 에피소드의 40분은 처음 30분내에 실시간 수요가 발생되어 호출 수요가 발생한 이후 DTO 차량이 목적지까지 서비스를 제공할 수 있는 최소 10분의 시간을 고려하여 설정한 값이다.

한 에피소드 내에서 발생하는 호출수요량은 1명의 승객부터 시작하여 점차적으로 증가하면서 최적 동적 경로 학습을 진행한다. 여기서 10,000개의 에피소드 대상으로 DTO 차량이 강화학습 알고리즘 내에서 동일한 수요량으로 학습이 이루어진 이후 중간 평가를 수행하여 호출 수요량을 증가시킬지를 결정한다. 10,000개의 에피소드 중 마지막 100개 에피소드의 모든 행동은 탐사(exploitation)로 실시되도록 설계하여 100개 에피소드의 결과가 모든 실시간 수요에 서비스를 제공하고, 전체 수요의 85% 이상의 출발지에서 서비스 대기시간이 10분 이내에 달성했을 때 현재 수요량을 충분히 학습하였다고 가정하고 수요량을 증가시킨다. 반대로 모든 실시간 수요에 서비스를 제공하지 못하거나, 85%미만의 수요가 10분 이내 서비스 대기시간을 충족시킬 시에는 현재의 수요량을 유지하고 또 다른 10,000개의 에피소드 학습이 이루어지게 된다.

10,000개의 에피소드에서 이루어지는 행동의 탐색과 탐사의 확률은 앞에서 언급된 ε-greedy 정책으로 정해진다. 첫 번째 에피소드의 ε값은 1로 정해져, 모든 행동이 무작위 탐색으로 진행한다. 매회 에피소드 마다 0.99999의 값이 이전 에피소드의 ε값에 곱해져서 점차적으로 ε값이 감소하여 9,000번째 에피소드에는 ε값이 0.9048이 된다. Q-value의 점차적인 수렴을 보기 위해서 9,001번째 에피소드부터는 매 에피소드마다 0.9995의 값이 이전 에피소드의 ε값에 곱해진다. 9,501번째 에피소드부터는 ε값을 0으로 하여 현재까지 학습된 자율주행 차량의 Q-value들을 수렴시킨다.

학습이전에 총 에피소드의 숫자를 정해 두고 학습이 이루어지는 것이 아니라 일정한 에피소드를 기점으로 현재까지 학습된 에이전트가 100% 탐사의 행동을 취하도록 설계하였다. 그러므로 learning rate는 항상 1로 hyper parameter를 설정해 두었다. 이는 이미 학습된 경로를 다시 경험하려는 overshooting 문제가 발생 할 수 있지만, 학습모델에서 해결할 최적경로문제가 여러 복잡한 요소들이 관련되어 global optimum이 아닌 local optimum에 빠지는 것을 방지 하는 것을 중요하게 다루기 위해서 이와 같은 hyper parameter를 적용하였다.

2. 알고리즘 평가 척도

DTO 차량이 동적경로 생성 알고리즘에서 제시한 동적 노선경로를 따라 운행시켰을 경우 DTO 차량의 서비스 성능을 파악하기 위해 고정노선(Fixed route)은 6번 정류장에서부터 12번 정류장까지 왕복 운행하는 시나리오 노선과 비교 분석할 수 있도록 시뮬레이션에서 목적지까지 서비스 완료 건수에 대해 서비스 대기시간과 서비스 시간

(통행시간), 동적 경로생성 빈도, 10분 이내 대기시간 및 통행시간 달성 분포율 등으로 구분한다.

서비스 대기 시간은 실시간 호출수요가 발생한 시점부터 DTO 차량이 출발지에 도착하여 탑승이 이루어질 때까지 걸리는 시간으로 정의한다. 서비스 제공시간(in vehicle time)의 정의는 실시간 호출 수요가 발생한 시점부터 탑승하여 도착지 정류장에서 하차가 이루어지는 시점까지의 총 걸린 시간으로 정의한다. 또한 얼마나 많은 수요가 에피소드 내(40분)에 출발 정류장에서의 승차 서비스와 도착 정류장에서의 하차 서비스를 제공받을 수 있었는지 각각 분석한다. 특히 일반적으로 사람들이 불편함 없이 기다릴 수 있는 대기시간은 Arhin et al.(2019)이 제안한 10분 이내로 설정하여 얼마나 많은 호출승객의 수요에 대해서 DTO 차량을 배차 할 수 있었는지에 대한 달성도로 설정한다.

알고리즘의 성능평가

Figure 4의 호출 수요는 이전의 주행 기록이나 수요예측을 근거로 진행된 것이 아니므로 모두 시공간적으로 유니폼 몬테카를로 방법을 적용하여 호출수요를 발생시켰다. 이는 모든 시공간적으로 실시간 호출수요의 발생할 확률이 모두 동일하다. 다만, 에피소드내 발생될 실시간 호출수요량을 제어하기 위해 최소 1명부터 최대 23명까지 한 개의 에피소드내에 발생하도록 설정하였다. 이는 23명 초과의 수요량에서는 고정노선과 동적노선의 운행결과가 거의 동일하게 수렴하기 때문에 최대 실시간 수요량은 23명으로 설정했다. 각 수요량별로 실시된 100개의 평가 에피소드는 실시간 대비 50시간의 주행시뮬레이션으로 간주되어 주행하였다. 실시된 실험의 결과는 수요량별로 정리하여 고정노선과 실시간 동적경로 생성알고리즘 노선으로 운행하는 것에 대해 비교한다.

1. 호출수요 승객에 대한 목적지까지 서비스율

발생된 실시간 호출수요는 출발지에서 탑승 시킨 후 목적지까지 서비스를 제공한 비율을 비교했다. 고정노선과 동적경로 생성알고리즘 노선은 발생한 모든 호출수요에 대해 출발정류장에서 탑승시킬 수 있었다. 탑승시킨 모든 호출 승객들은 동적경로 생성알고리즘 노선으로 목적지까지 서비스 제공이 가능하였지만 고정노선은 일부 승객들에 대해 시뮬레이션이 끝나기 이전에 하차가 완료되지 못하는 결과에 해당 한다(Figure 5 참조).

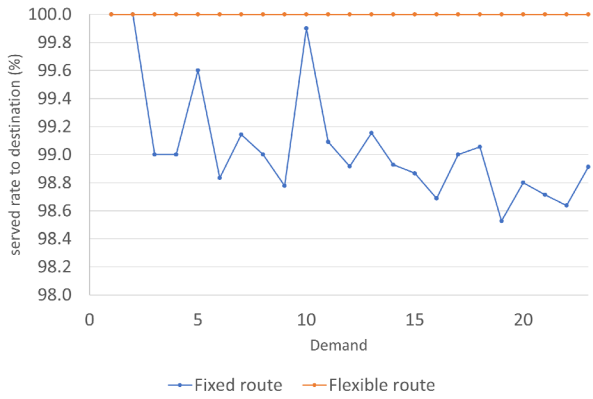

Figure 5. Rates of passengers served to destinations

2. 서비스 평균 대기시간과 평균 통행시간 (in vehicle time)

동적 경로생성 알고리즘 성능은 좀 더 자세하게 검토하기 위해 호출수요에 대응한 평균 서비스 대기시간과 평균 통행시간 지표를 수요량별로 계산하여 고정노선과 비교하여 Table 3과 Figures 6, 7에 제시하고 있다. 여기서 고정노선 서비스가 목적지까지 제공받지 않은 호출승객들은 대기시간과 통행시간 계산에서 제외하였다. 호출 수요량에

관계없이 고정노선의 평균 서비스 대기시간과 평균 통행시간은 각각 350초, 500초로 나타났다. 이는 실시간 호출수요가 공간적으로 균등한 확률로 랜덤하게 발생하기 때문에 서비스를 제공받은 승객들은 전체적인 평균시간이 거의 일정하게 나타나게 된다. 동적경로 생성알고리즘의 경로 지시받은 DTO 차량은 고정노선과 비교하여 보다 짧은 서비스 대기시간과 보다 빠른 통행속도로 목적지까지 도착 할 수 있었다.

Table 3. Averages of service wait times and in vehicle time

Demand	No. of call demand	1	2	3	4	5	6	7	8
Fixed route	Wait times (sec)	325.1	362.8	348.5	349.2	347.1	352.6	351.3	344.7
Flexible route		189.9	197.2	191.3	213.1	215.3	215.1	226.5	238.3
Fixed route	In vehicle time (sec)	478.5	520.9	496.6	495.4	490.4	498.3	500.3	490.6
Flexible route		343.4	348.2	329.3	346.7	361.5	354.2	377.4	381.7
Demand	No. of call demand	9	10	11	12	13	14	15	16
Fixed route	Wait times (sec)	348.6	339.9	347.9	340.8	343.6	346.8	347.5	349.8
Flexible route		243.5	237.6	244.8	254.4	260.4	253.9	265.8	269.6
Fixed route	In vehicle time (sec)	501.8	494.8	495.2	493.6	497.3	498.5	493.2	502.2
Flexible route		391.4	380.7	392.1	405.9	415.7	407.3	422.4	430.9
Demand	No. of call demand	17	18	19	20	21	22	23	Average
Fixed route	Wait times (sec)	347.5	344.5	346.1	349.9	349.6	348.8	348.9	347.0
Flexible route		264.6	272.8	270.0	270.2	280.3	275.0	283.7	244.9
Fixed route	In vehicle time (sec)	474.0	494.7	462.9	472.8	497.5	499.0	494.1	493.2
Flexible route		423.4	427.3	431.4	430.8	443.6	433.9	445.1	396.7

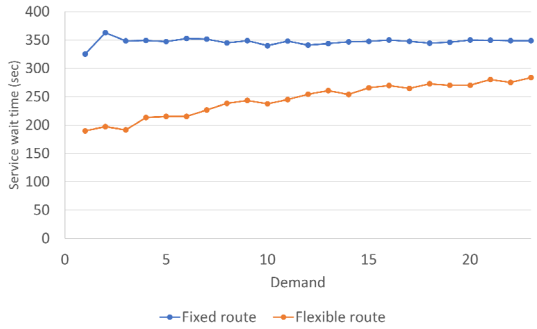

Figure 6. Average of service wait times

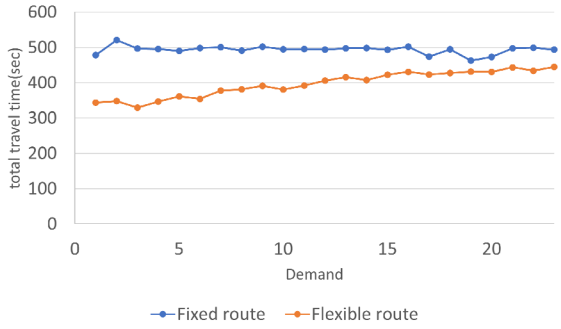

Figure 7. Average of in vehicle time

동적 경로생성 알고리즘은 Figures 6, 7에서 보는 바와 같이 실시간 호출수요량이 적을수록 더 큰 효과가 나타났으며 동시 호출 수요량이 증가하면서 고정노선 서비스의 대기시간과 통행시간으로 수렴해 가는 것을 확인할 수 있다.

3. 동적 경로 생성 빈도

실시간 호출 수요에 따른 동적 경로생성은 이번 실험 환경조건에서는 동적 경로생성 알고리즘에서 발생한 유턴의 빈도수로도 측정할 수 있다(Figure 8 참조). 여기서 호출 수요량이 증가함에 따라 유턴의 빈도수도 증가하는 것을 볼 수 있지만, 호출 수요량이 10이상인 경우에서는 동적 경로생성의 빈도가 감소하는 로그성장(logarithmic growth)이 나타난다. 이와 같은 현상은 에피소드 초기의 적은 수의 실시간 호출 수요량에서는 동적경로 노선이 보다 큰 효과를 볼 수 있지만, 동시 호출 수요량이 많은 경우 일정 배차 간격으로 고정한 운행이 보다 효율적인 것으로 해석된다. 이는 일반적으로 대중교통 수요가 많은 경우는 현재와 같은 일정 배차 간격으로 운행하고 있는 도시 및 광역철도와 같은 고정 스케줄링 운행이 효율적이라는 의미로 해석할 수 있다. 한국교통대학교 데스트베드 구간은 학기 중의 경우 노선버스가 1일 24회 운영되고 있지만, 방학기간과 휴일과 같은 경우 노선버스가 1일 4회 운영되고 있다(KoNUT,

2022). DTO 차량의 동적 경로 생성 알고리즘은 호출 수요대응 모빌리티가 수요가 낮은 기간(예를 들어, 휴일과 방학기간 등)의 경우 이용자와 운영자 모두에게 효과적임을 알 수 있다.

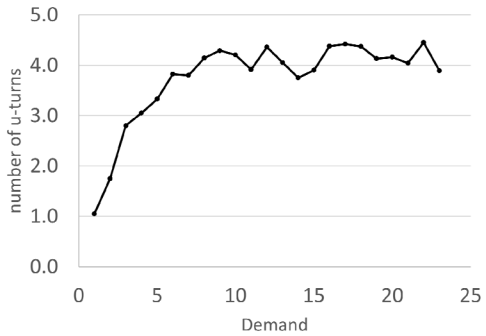

Figure 8. The average number of U-turns per episode

Table 4에서는 동적 경로생성 알고리즘의 지시받은 결과 유턴을 통해 호출 승객의 서비스가 제공되는 시점이 지연되는 정도를 간접적으로 알아보기 위해서 호출 수요량별 평가 에피소드내 유턴횟수의 최소값, 최대값, 평균값, 4분위수와 편차의 통계량을 제시하고 있다. 최소값 0은 유턴을 전혀 지시 받지 않은 에피소드를 모든 호출 수요량에 걸쳐 포함하고 있고, 최대로는 10 이상의 많은 유턴의 동적 경로 제시 받은 경우도 나타난다. 에피소드별 유턴횟수의 전체적인 편차는 2.3으로 평균값 3.7 기준으로 크게 벗어나지 않는 것으로 알 수 있었다. 이와 같은 작은 편차는 사분위수로도 확인할 수 있었다. 첫 번째 사분위 수인 2와 세 번째 사분위 수인 5는 50%의 평가 에피소드들의 유턴횟수가 2 이상 혹은 5 이하에 분포되는 것을 알 수 있다. 이 낮은 평균값과 편차는 유턴으로 인해 서비스가 지연되는 정도가 본 연구의 주행 시뮬레이션에서는 크지 않았다는 것을 알 수 있다.

Table 4. The statistics: the number of U-turns per episode at each demand

Demand	1	2	3	4	5	6	7	8	9	10	11	12
Minimum value	0	0	0	0	0	0	0	0	0	0	0	0
First quartile (25%)	0	1	1	2	2	2	2	2	3	2	2	3
Second quartile (50%)	1	2	2	3	3	3	4	4	4	4	4	4
Third quartile (75%)	2	2	4	4	4	5	5	5	6	6	5	5
Maximum value	2	7	9	10	12	11	10	13	8	10	13	12
Average	1.1	1.8	2.8	3.1	3.3	3.8	3.8	4.1	4.3	4.2	3.9	4.4
Standard deviation	0.9	1.3	1.9	2.1	2.1	2.1	2.1	2.4	2.1	2.3	2.3	2.5
Demand	13	14	15	16	17	18	19	20	21	22	23	Total
Minimum value	0	0	0	0	0	0	0	0	0	0	0	**0**
First quartile (25%)	2	2	3	3	3	3	3	2	3	3	2	**2**
Second quartile (50%)	4	3	4	4	4	4	3	4	4	4	4	**3**
Third quartile (75%)	5	5	5	6	6	5	5	6	5	6	5	**5**
Maximum value	11	10	9	12	10	11	11	11	11	11	10	**13**
Average	4.1	3.8	3.9	4.4	4.4	4.4	4.1	4.2	4.0	4.5	3.9	**3.7**
Standard deviation	2.3	2.1	2.0	2.4	2.2	2.3	2.3	2.3	2.1	2.2	2.3	**2.3**

4. DTO 차량의 목표 서비스 점유율

실시간 동적경로 생성알고리즘 서비스 목표는 호출 이후 서비스 개시시간을 10분 이내로 승차이후 목적지까지 통행시간을 10분 이내로 설정하였다. 따라서 Table 5에서는 DTO 차량의 목표 서비스가 고정노선과 비교한 결과, 호출 수요량과 관계없이 고정노선을 이용한 승객들의 약 93.5%가 10분이 내에 탑승할 수 있었다. 또한 약 67.5%의

승객이 10분 이내에 원하는 도착정류장에서 하차할 수 있었다. 실시간 호출 수요대응 동적 경로생성 알고리즘의 지시를 받은 DTO 차량은 호출수요 승객 대상으로 대기시간과 통행시간을 10분 이내 기준으로 95% 이상이 서비스를 받았다. 그러나 호출 수요량이 증가하면서 동적경로 생성알고리즘이 지시하는 동적 경로 중 약 5% 이내에서 10분을 초과하는 사례가 발생하고 있다(Figures 9, 10 참조). 그러므로 동적 경로 알고리즘의 높은 효과는 실시간 호출 수요량과 서비스 네트워크의 규모를 고려하여 적정한 자율주행 대중교통 차량이 공급되어야 한다는 점을 확인할 수 있었다.

Table 5. Served rates at origin stops and arrival rates at destination stops in 10 minutes

	Demand	1	2	3	4	5	6	7	8
Served rates at origin stops in 10 minutes (%)	Fixed route	93.8	93.4	93.7	93.5	93.5	93.4	93.2	93.1
	Flexible route	100.0	99.5	100.0	98.3	99.0	98.7	98.9	97.9
Arrival rates at destination stops in 10 minutes (%)	Fixed route	69.0	61.5	69.3	69.5	71.0	68.3	67.3	68.9
	Flexible route	94.0	92.0	95.0	92.3	90.2	90.8	87.4	87.6
	Demand	9	10	11	12	13	14	15	16
Served rates at origin stops in 10 minutes (%)	Fixed route	93.3	93.1	93.2	93.1	93.1	93.9	93.3	93.8
	Flexible route	97.1	97.1	97.9	96.4	96.3	97.7	96.0	95.7
Arrival rates at destination stops in 10 minutes (%)	Fixed route	66.4	68.8	67.5	67.9	67.5	67.2	68.5	66.5
	Flexible route	85.4	86.6	86.2	83.1	81.5	83.3	80.4	78.8
	Demand	17	18	19	20	21	22	23	
Served rates at origin stops in 10 minutes (%)	Fixed route	93.4	93.8	93.6	93.1	93.5	93.5	93.3	
	Flexible route	96.2	95.8	96.1	96.1	95.8	95.2	95.3	
Arrival rates at destination stops in 10 minutes (%)	Fixed route	68.0	67.7	68.2	67.8	67.4	67.2	67.7	
	Flexible route	80.4	79.8	78.1	79.5	77.0	78.1	75.8	

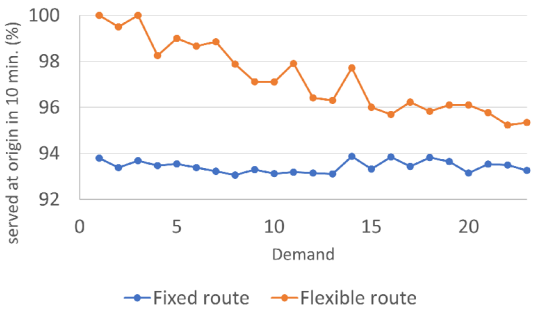

Figure 9. Served rates at origin stops in 10 min

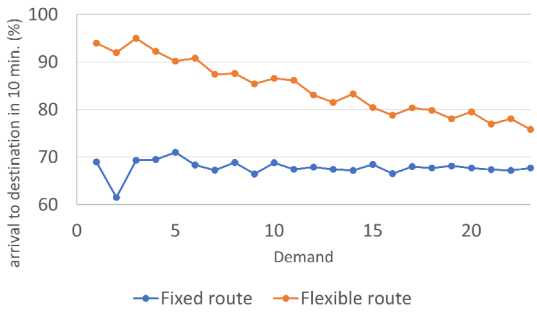

Figure 10. Arrival rates to destination stops in 10 min

결론 및 향후 연구과제

본 논문에서는 버스 이용자에게 대기시간과 목적지까지의 통행시간을 최적으로 관리하기 위해 고정 노선에서 동적 노선으로 전환하여 원하는 시각에 목적지에 도착할 수 있는 DRT시스템과 무인운영이 가능한 DTO 차량의 융합 서비스에 초점을 두고 강화학습 기반으로 동적 경로생성 알고리즘을 개발하였다. 동적 경로생성 알고리즘은 실시간 호출수요의 출발지 혹은 도착지가 한 곳으로 고정되어 있는 것이 아니라 출발지와 도착지가 다수로 발생할 수 있는 DARP 문제를 해결 할 수 있도록 하였다.

강화학습 요소는 상태, 행동, 에이전트, 환경, 보상으로 구성하였다. 에이전트는 호출승객의 출발지와 목적지, DTO 차량의 위치 등의 주변 상태에 따라 어떤 행동의 판단(정지, 주행, 유턴 등)을 내리는 주체로 정의하였다. 해당 에이전트가 속한 환경에서 행동하고, 특정 상태에 따른 행동의 결과에 대해 보상하는 과정을 통하여 DTO 차량이 최적의 동적 경로로 행동할 수 있도록 학습시켰다. 상태와 행동의 항목이 Q-value로 평가된 학습데이터가 만들어 졌다.

동적 경로생성 알고리즘의 성능은 고정 노선 시나리오 대상 기준으로 4개 지표(호출수요 승객 대상으로 목적지까지의 서비스 완료률, 평균 대기시간과 통행시간, 유턴빈도, 10분 이내 서비스 달성율)를 동적 노선 결과 비교하여 검증한 결과는 다음과 같다. 첫째, 호출수요 승객을 대상으로 하는 서비스 완성은 동적 경로의 경우 100% 달성하였으나 고정 노선 운영의 경우 약 3%가 시뮬레이션하는 동안 달성하지 못했다(Figure 5 참조). 둘째, 고정 노선 보다 동적 노선이 평균 서비스 대기시간과 평균 통행시간 지표에서 각각 약 1.67분이 단축되었다. 이 결과는 동적 노선의 운영이 호출수요가 낮은 경우 대기시간과 통행시간이 감소하지만 호출수요가 일정 수준으로 증가하면서 고정 노선 운영과 동일한 상태로 수렴하는 경향이 나타났다. 셋째, 동적 경로생성 알고리즘의 구현은 교통축 대상으로 호출 수요량이 증가함에 따라 유턴의 빈도수도 증가하는 것을 볼 수 있지만, 동시 호출 수요량이 10인 이상인 경우에서는 동적 경로의 생성 빈도가 감소하는 로그성장이 나타난다. 이는 일반적으로 대중교통 수요가 많은 경우는 현재와 같은 일정 배차 간격으로 운행하고 있는 도시 및 광역철도와 같은 고정 노선 운행이 효율적이라는 의미로 해석할 수 있다. 넷째, 동적 경로생성 알고리즘의 최적화 목표는 대기시간과 통행시간이 각각 10분이내 목표로 설정하여 학습하였다. 그 결과 동적 노선은 호출수요가 낮은 수준에서 모든 승객에게 최적화 목표가 달성되었지만 호출수요가 증가하면서 96% 달성으로 수렴하였다. 반면 고정 노선은 69-93% 범위로 그 차이가 크게 나타나고 있다.

이상의 결과에 근거하면 동시 호출수요가 적을수록 동적 노선이 고정 노선 대비 운영 효율성이 높다는 점을 확인할 수 있었지만 자율주행 대중교통의 동적 노선의 효과는 호출 수요량이 증가에 따라 점차적으로 감소하는 경향이 나타났다. 이는 동시 다발적으로 많은 호출수요에서는 고정 노선의 운행이 동적 노선의 운영과 수렴하게 된다. 따라서 본 연구에서 수행한 동적 경로생성 알고리즘의 경우 다음과 같은 보상함수 관련하여 다음과 같은 시사점을 확인할 수 있었다.

첫째, 보상함수는 탑승한 승객을 목적지 정류장에 하차 시켰을 때 Table 2에서와 같이 거리기준으로 반영하여 호출 수요에 균형적인 서비스를 고려해야 한다. 여기서 탑승 보상 값이 너무 높게 설정되는 경우 DTO 차량이 승객을 탑승시키는 데에만 중점을 두어 탑승 이후 목적지까지 서비스제공이 적절히 이루어지지 않을 수 있다는 점을 반영할 수 있도록 전체적인 하차 보상 값이 평균값에 가까운 500의 값으로 설정하여 승객을 탑승시키도록 유도할 수 있었다. 둘째, 탑승 또는 하차의 행동이 아닌 경우는 패널티 값 10을 부여하여 최단거리 경로가 학습될 수 있도록 유도할 수 있었다. 특히 최단거리의 동적 경로의 패널티 보상 값은 hyper parameter와 함께 유도할 수 있다.

본 연구에서 사용된 Q-learning 강화학습 알고리즘은 모든 경우의 상태와 그에 따른 행동에 대한 경험학습을 전제로 높은 신뢰도가 보장된다. 반면 실제상황에서 경험하지 않은 상태에 행동을 취해야 할 경우 돌발적인 결정을 에이전트가 내릴 수 있다는 단점이 있다. 또한 테스트 환경이 커지거나 복잡한 네트워크 환경을 경험하지 않은 상태에 한계가 있기 때문에 이를 극복할 수 있는 DQL(Deep Q-Network)과 같은 강화학습 알고리즘으로 확장이 필요하였다. 본 연구에서 실시한 주행 시뮬레이션 환경은 사전에 예측한 호출수요를 고려하지 않는 설계에 해당한다. 따라서 시공간적 수요발생 확률이 유니폼하게 가정한 몬테카를로 시뮬레이션 방법을 적용하고 있다. 향후 연구에서는 실제 실적 호출수요 기반으로 몬테카를로 시뮬레이션을 실시하여 강화학습내 에이전트를 학습시킨다면 보다 다양한 상황에서 DTO 차량의 배차 및 동적 경로 지시의 성능이 개선될 것으로 기대된다.

Funding

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21AMDP-C161756-01).

ORCID

KIM, Hyun http://orcid.org/0000-0002-8350-0670
YOO, Seonhyung http://orcid.org/0000-0002-6952-5054
LEE, Jinwoo http://orcid.org/0000-0002-4148-4115
BAEK, Beomyeol http://orcid.org/0000-0002-2887-0695
SHIN, Junghee http://orcid.org/0000-0002-8652-9372

References

Arhin S. A., Gatiba A., Anderson M., Manandhar B., Ribbisso M. (2019), Acceptable Wait Time Models at Transit Bus Stops, Engineering, Technology & Applied Science Research, 9(4), 4574-4580.

Becker F., Axhausen K. W. (2017), Literature Review on Surveys Investigating the Acceptance of Automated Vehicles, Transportation, 44(6), 1293-1306. doi:10.1007/s11116-017-9808-9

Cordeau J. F., Laporte G. (2007), The Dial-a-ride Problem: Models and Algorithms, Annals of Operations Research, 153, 29-46.

Diana M., Dessouky M. M. (2004), A New Regret Insertion Heuristic for Solving Large-scale Dial-a-rideproblems with Time Windows, Transportation Research Part B, 38, 539-557.

Fagnant D. J., Kockelman K. (2015), Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations, Transportation Research Part A: Policy and Practice, 77, 167-181. doi: https://doi.org/10.1016/j.tra.2015.04.003

Faisal A., Kamruzzaman M., Yigitcanlar T., Currie G. (2019), Understanding Autonomous Vehicles: A Systematic Literature Review on Capability, Impact, Planning and Policy, Journal of Transport and Land Use, 12(1), 45-72. Retrieved from https://www.jstor.org/stable/26911258

Gendreau M., Laport G., Semet F.(2001), A Dynamic Model and Parallel Tabu Search Heuristic for Real-time Ambulance Relocation, Parallel Computing, 27(12), 1641-1653. doi: https://doi.org/10.1016/S0167-8191(01)00103-X

Guo J., Susilo Y., Antoniou C., Pernestål A. (2021), When and Why Do People Choose Automated Buses over Conventional Buses? Results of a Context-dependent Stated Choice Experiment, Sustainable Cities and Society, 69(102842), 1-9. doi: https://doi.org/10.1016/j.scs.2021.102842

Huang T., Fang B., Bei X., Fang F. (2019), Dynamic Trip-Vehicle Dispatch with Scheduled and On-Demand Requests, arXiv.

Iclodean C., Cordos N., Varga B. O. (2020), Autonomous Shuttle Bus for Public Transportation: A Review, Energies, 13(11, 2917), 1-45. Retrieved from https://www.mdpi.com/1996-1073/13/11/2917

Kassens-Noor E., Kotval-Karamchandani Z., Cai M. (2020), Willingness to Ride and Perceptions of Autonomous Public Transit, Transportation Research Part A: Policy and Practice, 138, 92-104. doi: https://doi.org/10.1016/j.tra.2020.05.010

Korea National University for Transportation (2022), Smart Campus Challenge Project: A Demonstration of Multi-purpose Mobility Service Using OHMIO Shuttle.

LeBeau P. (2018), Waymo Starts Commercial Ride-share Service, Retrieved from https://www.cnbc.com/2018/12/05/waymo-starts-commercial-ride-share-service.html

Lee E., Cen X., Lo H. K. (2021), Zonal-based Flexible Bus Service under Elastic Stochastic Demand, Transportation Research Part E: Logistics and Transportation Review, 152, 1-25. doi: https://doi.org/10.1016/j.tre.2021.102367

Lee Y.-J., Meskar M., Nickkar A., Sahebi S. (2019), Development of an Algorithm for Optimal Demand Responsive Relocatable Feeder Transit Networks Serving Multiple Trains and Stations, Urban Rail Transit, 5(3), 186-201.

Li M., Zheng N., Wu X., Huo X. (2019), An Efficient Matching Method for Dispatching Autonomous Vehicles, 2019 IEEE Intelligent Transportation Systems Conference (ITSC) Auckland, NZ, October 27-30.

Melachrinoudis E., Ilhan A. B., Min H. (2007), A Dial-a-ride Problem for Client Transportation in a Healthcare Organization, Computers & Operations Research, 34, 742-759.

Narayanan S., Chaniotakis E., Antoniou C. (2020), Shared Autonomous Vehicle Services: A Comprehensive Review, Transportation Research Part C: Emerging Technologies, 111, 255-293. doi: https://doi.org/10.1016/j.trc.2019.12.008

Pettigrew S., Cronin S. L., Norman R. (2019), Brief Report: The Unrealized Potential of Autonomous Vehicles for an Aging Population, Journal of Aging & Social Policy, 31(5), 486-496. doi: https://doi.org/10.1080/08959420.2018.1500860

Psaraftis H. N.(1980), A Dynamic Programming Solution to the Single Vehicle Many-to-Many Immediate Request Dial-a-Ride Problem, Transportation Science 14(2), 130-154. doi: https://doi.org/10.1287/trsc.14.2.130

Raj A., Kumar J. A., Bansal P. (2020), A Multicriteria Decision Making Approach to Study Barriers to the Adoption of Autonomous Vehicles, Transportation Research Part A: Policy and Practice, 133, 122-137. doi: https://doi.org/10.1016/j.tra.2020.01.013

Rekiek B., Delchambre A., Saleh H. A. (2006), Handicapped Person Transportation: An Application of the Grouping Genetic Algorithm, Engineering Application of Artificial Intelligence, 19, 511-520.

Ritzinger U., Puchinger J., Hartl R. F. (2016), A Survey on Dynamic and Stochastic Vehicle Routing Problems, International Journal of Production Research, 54(1), 215-231. doi: https://doi.org/10.1080/00207543.2015.1043403

Schilde M., Doerner K. F., Hartl R. F. (2011), Metaheuristics for the Dynamic Stochastic Dial-a-ride Problem with Expected Return Transports, Computers & Operations Research, 38(12), 1719-1730. doi: https://doi.org/10.1016/j.cor.2011.02.006

Sultana Z., Mishra S., Cherry C. R., Golias M. M., Tabrizizadeh Jeffers S. (2018), Modeling Frequency of Rural Demand Response Transit Trips, Transportation Research Part A: Policy and Practice, 118, 494-505. doi: https://doi.org/10.1016/j.tra.2018.10.006

Sutton R. S., Barto A. G. (1998), Introduction to Reinforcement Learning (Vol.2): MIT Press Cambridge.

Tafreshian A., Abdolmaleki M., Masoud N., Wang H. (2021), Proactive Shuttle Dispatching in Large-scale Dynamic Dial-a-ride Systems, Transportation Research Part B: Methodological, 150, 227-259. doi: https://doi.org/10.1016/j.trb.2021.06.002

Yigitcanlar T., Mohamed A., Kamruzzaman M., Piracha A. (2019), Understanding Transport-Related Social Exclusion: A Multidimensional Approach, Urban Policy and Research, 37(1), 97-110. doi: https://doi.org/10.1080/08111146.2018.1533461