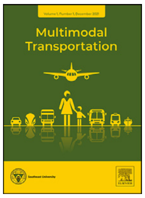
Multimodal Transportation 3 (2024) 100137

Contents lists available at ScienceDirect

Multimodal Transportation

journal homepage: www.elsevier.com/locate/multra



Full Length Article

A joint travel mode and departure time choice model in dynamic multimodal transportation networks based on deep reinforcement learning

Ziyuan Gu[a,1], Yukai Wang[b,2], Wei Ma[c,1,4], Zhiyuan Liu[a,3,*]

[a] *Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China*
[b] *Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon, South Korea*
[c] *Department of Civil and Environmental Engineering, the Hong Kong Polytechnic University, Hong Kong*

ARTICLE INFO

Keywords:
Mode choice
Departure time choice
Deep reinforcement learning
Multimodal transportation
Microscopic traffic simulation

ABSTRACT

Decision on travel choices in dynamic multimodal transportation networks is non-trivial. In this paper, we tackle this problem by proposing a new joint travel mode and departure time choice (JTMDTC) model based on deep reinforcement learning (DRL). The objective of the model is to maximize individuals travel utilities across multiple days, which is accomplished by establishing a problem-specific Markov decision process to characterize the multi-day JTMDTC, and developing a customized Deep Q-Network as the resolution scheme. To render the approach applicable to many individuals with travel decision-making requests, a clustering method is integrated with DRL to obtain representative individuals for model training, thus resulting in an elegant and computationally efficient approach. Extensive numerical experiments based on multimodal microscopic traffic simulation are conducted in a real-world network of Suzhou, China to demonstrate the effectiveness of the proposed approach. The results indicate that the proposed approach is able to make (near-)optimal JTMDTC for different individuals in complex traffic environments, that it consistently yields higher travel utilities compared with other alternatives, and that it is robust to different model parameter changes.

1. Introduction

Travel demand has been soaring for the past few decades due to the rapid growth in population and urbanization. The growth rate has far exceeded that by which transportation infrastructure is expanded. The consequence of such demand-supply imbalance is the ubiquitous traffic congestion witnessed worldwide, which explains the emergence and necessity of various travel demand management strategies (Gu et al., 2018; Qin et al., 2022). Successful implementation of these strategies, however, is heavily dependent on how we understand and model travel choices of trip-makers.

Travel choices are typically characterized as alternatives with many dimensions such as departure time, travel mode and route (Pitale et al., 2023). These decision-making processes are usually described by discrete or continuous choice models. Earlier works on

[*] Corresponding author.
E-mail addresses: ziyuangu@seu.edu.cn (Z. Gu), yukai@kaist.ac.kr (Y. Wang), wei.w.ma@polyu.edu.hk (W. Ma), zhiyuanl@seu.edu.cn (Z. Liu).
[1] Ph.D., Associate Professor
[2] Ph.D., Student
[3] Ph.D., Professor
[4] Ph.D., Assistant Professor

https://doi.org/10.1016/j.multra.2024.100137
Received 15 January 2024; Received in revised form 18 March 2024; Accepted 27 March 2024

travel choices generally focused on one dimension only, i.e., an alternative was to be chosen from a collection of mutually exclusive alternatives associated with the dimension in question. However, some dimensions of travel choices are inherently correlated and thus shall not be approached in a separate manner (Yin et al., 2014). This concern has triggered increasing attention in recent years in multi-dimensional travel decision making. Unlike traditional one-dimensional travel choice models, multi-dimensional models account for the correlation between different choices (Zimmermann et al., 2018). Two such critical and correlated choices are those associated with the mode and departure time. From a disaggregate perspective, these two choices reflect individuals' travel preferences for their trips. From an aggregate perspective, they determine the spatial-temporal travel demand for the transportation network (Souche-Le Corvec, 2023). We emphasize that the attractiveness and thus the possible selection of a transportation mode is contingent upon its level of service. This level of service can be influenced by numerous policy initiatives such as congestion pricing, public transportation priority, and various types of incentives. As a result, to evaluate such policy initiatives, a modeling framework that jointly considers the travel mode and departure time choice is necessary (Fukuda and Yai, 2010).

Most existing research on modeling the joint travel mode and departure time choice (JTMDTC) was conducted using discrete choice models (DCMs) of different types that mainly seek random utility maximization. In particular, models such as multinomial logit (MNL), nested logit (NL), cross-nested logit (CNL), and other variants were often used considering their ability to characterize correlations among different choice alternatives. Although these models are typical solutions to the travel choice problems with theoretical underpinning, their applicability particularly in complex decision-making processes may be limited by the formulation of random utility functions. Due to the lack of adaptiveness and the possibility of imperfect perception of travel information by trip-makers, travel choices informed by DCMs may not necessarily lead to the best outcomes.

Recently, (deep) reinforcement learning (RL or DRL) has emerged as one of the key machine learning methods to tackle difficult decision-making problems owing to its learning capability in complex environments. This learning capability, which DCMs do not have, can be fully utilized for travel choice modeling or recommendation. The motivation is that the time-varying information about the level of service of different transportation modes can be gained from the experience arisen from the daily travel, based on which trip-makers can learn and adjust their travel choices in a day-to-day manner. This intrinsic learning and decision-making process can be represented as a Markov decision process (MDP) and resolved by DRL. In fact, it can be perceived as a recommendation problem where the aim is to achieve as much travel efficiency as possible for individuals. DRL can be used to optimize the recommendations provided to users by taking into account both the short-term and long-term rewards associated with different travel choices. It can also absorb users attributes and/or preferences leading to more personalized recommendations.

Thus, in this paper, we harness the advantages of DRL and propose a new JTMDTC model to maximize individuals travel utilities in a dynamic multimodal transportation network. To achieve computational efficiency and enable large-scale application, a clustering method is embedded in the modeling framework. By carefully designing the learning structure and input, we show that the proposed approach is able to yield consistently better travel choices for individuals with higher utilities in complex traffic environments.

1.1. Related works

Since McFadden et al. (1973) developed the MNL model for discrete choice modeling, the big family of logit models have been extensively visited and applied to solve the JTMDTC problem. Specifically, Hendrickson and Plank (1984) modeled the JTMDTC using the MNL model and introduced a nonlinear utility function was later introduced to further account for the route choice (Weis et al., 2021). However, the MNL model requires a critical assumption that usually fails to hold in practice, namely the independence of irrelevant alternatives (IIA). Such an assumption implies that the properties not observed by the alternatives are uncorrelated. Thus, the NL model (Train and McFadden, 1978) has been proposed to overcome the IIA limitation, but the problem is that the correlations diminish among departure time alternatives that are far away from each other (Bhat, 1998). As a result, Bhat and Pulugurta (1998) used a mixed MNL structure to account for the unobserved correlations between the travel mode and departure time dimensions. A household survey in the San Francisco Bay Area was used to demonstrate that the proposed model outperforms the MNL and NL models.

In addition to the traditional Discrete Choice Models (DCMs), data-driven machine learning models have also become popular in recent years for studying travel choices. Li et al. (2000) demonstrated that artificial neural networks have the potential to establish an alternative framework replacing DCMs for travel choice modeling. Decision trees (Arentze and Timmermans, 2004; Rasouli and Timmermans, 2014; Tang et al., 2015) and support vector machines (Omrani, 2015; Semanjski et al., 2016) are some example machine learning models that have been applied. Compared with DCMs, machine learning models are more structurally flexible and thus can excel in building complex relationships between travel choices and various contributing factors. This advantage is usually obtained at the cost of reduced computational efficiency, as a data-driven method, these models are intrinsically more data demanding than DCMs.

As a machine learning paradigm, Reinforcement Learning (RL) has gained increasing attention and application in solving complex decision-making problems (Gu et al., 2023; Shi et al., 2023). This tendency is arguably due to its learning capability in the course of interacting with the environment. As a result, it is a well-suited method for tackling decision-making problems in transportation. We have already witnessed some works on the applications of RL in areas such as traffic flow management (Cruciol et al., 2013; Ning et al., 2020; Walraven et al., 2016), autonomous driving control (Aradi, 2022; Grigorescu et al., 2020; Zhu et al., 2018), and vehicle route planning (Yu and Gao, 2019). When applying RL to activity scheduling or travel choice modeling, the few existing studies generally resorted to the value-based approach. To model the activity-travel behavior of trip-makers, an RL-based choice model was developed and embedded in a learning-based transportation-oriented simulation system (Arentze and Timmermans, 2004).

Vanhulsel et al. (2009) solved the activity scheduling problem using one of the most well-known RL models called Q-learning. More recently, Idris et al. (2012) proposed a novel conceptual framework for constructing a learning-based mode shift model, where RL was integrated with random utility maximization.

Recent advancements in DRL have been notably influential in addressing more complex decision-making scenarios within intelligent transportation systems. For instance, Shi et al. (2023) introduced an adaptive route guidance model using DRL, showcasing the potential for more responsive and efficient navigation solutions. Furthermore, the work by Zhao and Liang (2023) illustrates the application of deep inverse reinforcement learning in route choice modeling, emphasizing the significance of context-dependent rewards in understanding traveler preferences. In managing the dynamics of mobility-on-demand services, Xie et al. (2023) explored a two-sided DRL approach for optimizing operations with mixed autonomy, indicating a promising direction for future urban mobility solutions. These recent studies underscore the expanding scope of DRL applications in transportation, ranging from adaptive route guidance to sophisticated mobility management strategies.Notice that some research used DRL to study operations of intelligent transportation systems, such as dynamic pricing (Zhao and Lee, 2021) or routing (Shou et al., 2022), to improve system efficiency (Shou et al., 2022; Yang et al., 2020). The literature on DRL-based multi-dimensional travel decision-making, or recommendation, is relatively limited and needs enhancement.

1.2. *Objectives and contribution*

The objective of this study is to propose a JTMDTC model in dynamic multimodal transportation networks based on DRL. Different from most previous research on travel choice modeling that mainly relied on DCMs, the proposed framework is driven by a model-free learning-based approach where optimal travel choices are derived through extensively interacting with the environment represented by multimodal microscopic traffic simulation. The overall contribution is threefold. First, we establish a problem-specific MDP for the JTMDTC across multiple successive days, based on which a customized Deep Q-Network (DQN) is developed as the resolution scheme. Second, to deal with many individuals with travel decision-making requests, a clustering method is integrated with DRL to obtain representative individuals for model training, thus resulting in an elegant and computationally efficient approach. Third, extensive microsimulation experiments and model comparisons are conducted on a real-world network to demonstrate the effectiveness and robustness of the proposed approach.

The rest of the paper is structured as follows. Section 2 describes the detailed methodological framework for modeling the JTMDTC. Section 3 discusses the results and findings obtained from extensive microsimulation experiments. Section 4 concludes the work and provides future research directions.

2. Methodology

Application of RL to a sequential decision-making process with the Markov property requires that an MDP be first constructed, which defines the evolution of the environment considering the actions taken by the RL agent(s). Specifically, the agent constantly interacts with the environment through action exploration and exploitation according to according to the current state s_t. Once an action is taken, the environment evolves to a new state s_{t+1} and an associated reward r_t is obtained and fed back to the agent for improving its decision-making logic. Such a process iterates until the agent successfully learns a policy π (i.e., a decision maker) that is able to maximize the accumulated rewards (or return). Thus, the key to RL is to iteratively refine the policy based on the rewards.

In this study, we consider each trip-maker as an intelligent entity with learning capability, whose JTMDTC across multiple successive days is modeled as an MDP. Fig. 1 schematically illustrates the RL-based approach to modeling the Joint Travel Mode and Departure Time Choice (JTMDTC) for an individual trip-maker. The figure effectively encapsulates the iterative learning process of the agent (trip-maker) within the environment, showing how different combinations of actions are evaluated. As illustrated in Fig. 1, the actions from which an individual is able to select comprise different combinations of the travel mode and the departure time. The action a_t finally taken by the individual determines the next state s_{t+1} to which the environment evolves. This state shall reflect the individuals latest knowledge about the trip itself as well as the associated environment. The reward r_t gained by choosing this action (which is related to the cost of the trip) helps refine the decision-making logic of the individual (Idris et al., 2012). To be more precise from an RL perspective, the objective of action selection is to maximize what is called the discounted return U^t that includes rewards from multiple time steps:

$$\mathrm{U}^{\mathrm{t}} = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots + \gamma^n R_{t+n}, \tag{1}$$

where the discount parameter $\gamma \in [0, 1]$ reflects the importance attached to the long-term rewards. A larger value of γ equates to more reliance on the long-term rewards for guiding the decision making, whereas a smaller value makes individuals myopic whose actions are mainly driven by the immediate rewards.

2.1. *Construction of the Markov decision process*

As previously discussed, an MDP is a prerequisite for modeling and optimizing the JTMDTC. Here, we model the process of multi-day travel choice decisions as a sequence of states, actions, and rewards. The states represent the environment of individual trips, the

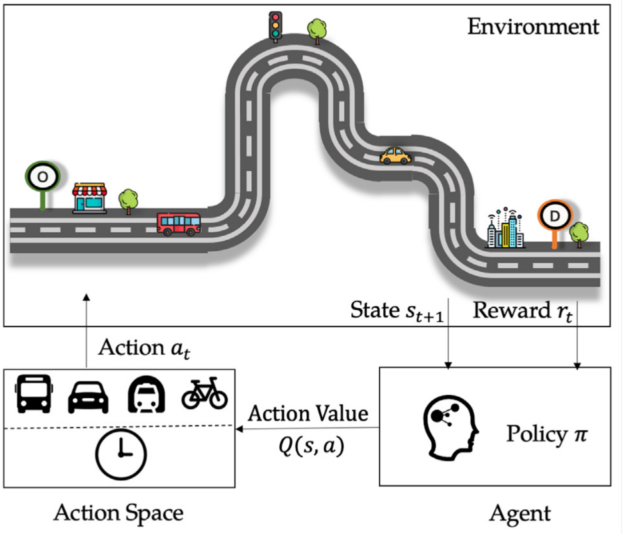


Fig. 1. Schematic illustration of the RL-based JTMDTC.

actions represent all travel options available to individuals, and the rewards represent the benefits/costs associated with the action. Mathematically speaking, it is a tuple (S, A, P, R, γ) where S represents the state space, A represents the action space, P represents the state transition probabilities, R represents the reward function, and finally γ is the discount factor. Here, an agent is a decision maker for individuals on the JTMDTC. While it is conceptually valid to treat each individual as an agent, the incurred computational complexity is a major obstacle to large-scale application, as the number of agents is equal to the number of individuals. A common decision-making logic is worthy of investigation that applies to all individuals with similar attributes, which reflects the generality of the approach. On the other hand, randomly selecting one or a few individuals for model training is insufficient and unsound. Thus, in this study, we propose to use representative individuals resulting from clustering in conjunction with DRL. In the following, we further elaborate on how the problem-specific MDP is constructed and resolved.

2.1.1. Actions

The action space defines all the actions from which the agent is able to choose given a certain state. Clearly, the action space for the JTMDTC consists of both the travel mode choice and the departure time choice. The former considers three modes of transportation including private cars, public transportation, and cycling. Public transportation is further categorized into buses and subways between which transfers are allowed. However, transfers among the three modes of transportation (e.g., park and ride) are not considered. This is part of the ongoing work requiring further investigation. To access public transportation, we assume that a trip-maker walks to the bus stop or subway station closest to the origin (Rasca et al., 2023). For the departure time choice, we assume that each trip-maker has an initial or desired departure time t_0, and that he/she is allowed to shift this departure time within a certain time window $[t_{\min}, t_{\max}]$, where $t_{\min}$ and $t_{\max}$ are the earliest and latest departure times possible, respectively. As with the literature (Zou et al., 2016), the time shift is in units of discrete intervals rather than in a continuous manner. Altogether the action space can be described by the following vector:

$$\boldsymbol{a} = \begin{bmatrix} m \\ t \end{bmatrix} \in \begin{bmatrix} \{m_1, m_2, \dots, m_N\} \\ [t_{\min}, t_{\max}] \end{bmatrix}, \tag{2}$$

where m is the mode of transportation, t is the departure time, N is the total number of available transportation modes.

2.1.2. States

The state space defines the contextual environment in which the agent chooses the actions. For the JTMDTC, the state space is designed to include not only the latest knowledge about the trip but also the earlier experience of the agent. Such a state space design resembles, to a large extent, the decision-making mechanism of rational human beings, namely learning from experience. Since the focus is not on empirical choice modeling or behavioral analysis, but on optimal choice decision or recommendation, we assume that the agent is able to fully perceive the environment and thus has perfect information about the trip. Nevertheless, this assumption can be somewhat relaxed as we will discuss later.

Trip information first includes the travel distance L and the memory travel time $\bar{T}$ associated with each mode of transportation. While the former is the travel distance by mode m, the latter is the average experienced travel time by mode m. The reason is twofold to take advantage of the experience and to maintain robustness in the presence of traffic stochasticity. Two other variables as part of the trip information are the initial departure time t_0 and the departure time difference or shift Δt relative to t_0. Putting everything together gives the following state vector specific to trip information:

$$s_{\text{trip}}^{m} = \begin{bmatrix} L_m \\ \bar{T}_m \\ t_0 \\ \Delta t \end{bmatrix} = \begin{bmatrix} \text{travel distance} \\ \text{memory travel time} \\ \text{initial departure time} \\ \text{departure time difference} \end{bmatrix}. \tag{3}$$

Environment information basically comprises factors that contribute to the travel costs of different modes of transportation. For public transportation, two such factors are considered, namely accessibility and fares (Gu et al., 2022). Here, we define and measure accessibility p as the total walking distance required to complete the first and last legs of the trip:

$$p = d_{\text{origin}} + d_{\text{destination}}, \tag{4}$$

where d_{origin} and $d_{\text{destination}}$ are the walking distances from the closest bus stop or subway station to the origin and the destination, respectively. Public transportation fares are the monetary cost one must pay to access the service. For private cars, we consider the fuel price as one influencing factor and put it in the state. Thus, the state vector specific to the environment information is as follows:

$$s_{\text{environment}} = \begin{bmatrix} p \\ f \\ o \end{bmatrix} = \begin{bmatrix} \text{accessibility} \\ \text{fare} \\ \text{fuel price} \end{bmatrix}. \tag{5}$$

Notice that to consider the impact of the value of time (VOT) on travel decision making, this individual-specific attribute is incorporated into the state representing individuals preferences and perceptions of travel time vs. monetary cost. Combining the VOT with Eqs. (3) and (5) leads to the complete state vector or space under the assumption of perfect information. We can slightly relax this assumption by reducing the full state vector to Eq. (6) corresponding to the case of partial information. From a human behavioral perspective, the reduced state space might be more relevant, but the ability to inform better action selections weakens. Evidence in this regard will be provided in the results.

$$s_{\text{reduced}} = \begin{bmatrix} \bar{T} \\ t_0 \\ \Delta t \end{bmatrix} = \begin{bmatrix} \text{memory travel time} \\ \text{initial departure time} \\ \text{departure time difference} \end{bmatrix}. \tag{6}$$

2.1.3. Reward function

When the agent selects and performs an action, it causes the environment to change from the current state to a new one. Along with this change is a feedback or reward that is intended to improve the agents decision-making logic. The reward can be positive meaning that the action is sensible, or negative meaning the opposite. Here, we define the reward as the travel utility of the trip, which is mainly comprised of various monetary costs. Specifically, the reward obtained at step i is calculated as follows:

$$r_i = \frac{E_1 - C_m^i}{E_2}, \tag{7}$$

where C_m^i is the total travel cost of transportation mode m, and E_1 and E_2 are two constants for mapping and scaling the cost to the reward. The total travel cost is further decomposed into three parts, namely the total travel time T_m^i, the schedule delay $\delta\left(t^i\right)$, and the other travel-related cost F_m^i.

$$C_m^i = \alpha T_m^i + \delta\left(t^i\right) + F_m^i, \tag{8}$$

where α is the value of time.

While the total travel time for private cars and cycling is essentially the travel time spent on the road, public transportation requires that three distinct types of travel time be considered including the in-vehicle travel time t_{inv}, the waiting time t_{wt}, and the transfer time t_{tr}. Thus, mode- specific total travel time is expressed as follows:

$$T_m^i = \begin{cases} t_{\text{car}} & \text{if} \quad m = \text{ car }, \\ t_{\text{inv}} + t_{\text{wt}} + t_{\text{tr}} & \text{if} \quad m = \text{ public transportation }, \\ t_{\text{cycling}} & \text{if} \quad m = \text{ cycling }. \end{cases} \tag{9}$$

The problem with only considering the total travel time is the neglect of the actual arrival time. That is, the arrival time might be far away from the desired one despite the fact that the total travel time is rather low. Thus, the schedule delay is introduced, a concept that can be traced back to Small (1982). It is assumed that each individual has a desired arrival time, and that both early and late arrivals would incur a so-called schedule delay cost. As the actual arrival time deviates from the desired one, the schedule delay cost grows in a linear fashion. Mathematically speaking, it is expressed as follows:

$$\delta(t^i) = \begin{cases} \beta\left(t + T_m - t_{\mathrm{d}}\right) & \text{if} \quad t + T_m - t_{\mathrm{d}} < 0, \\ 0 & \text{if} \quad t + T_m - t_{\mathrm{d}} = 0, \\ \gamma\left(t_{\mathrm{d}} - t - T_m\right) & \text{if} \quad t + T_m - t_{\mathrm{d}} > 0, \end{cases} \tag{10}$$

where β and γ are the unit costs of the schedule delay for early and late arrivals, respectively, and t_{d} is the desired arrival time. Finally, the other travel-related cost mainly refers to the total fuel expenses for private cars that are linearly dependent on the travel distance, or the public transportation fares associated with the trip:

$$F_m^i = \begin{cases} L_{\text{car}} \cdot o & \text{if} \quad m = \text{ car }, \\ f & \text{if} \quad m = \text{ public transportation }, \\ 0 & \text{if} \quad m = \text{ cycling}, \end{cases} \tag{11}$$

where f is further expressed as follows:

$$f = \mathbb{I}(\text{ bus }) \cdot f_{\text{bus}} + \mathbb{I}(\text{ subway }) \cdot f_{\text{subway}} , \tag{12}$$

where $\mathbb{I}(\cdot)$ is an indicator function that returns 1 when the travel mode is chosen and 0 otherwise. Bus fares f_{bus} are constant while subway fares f_{subway} increase with the distance traveled.

2.2. *Deep reinforcement learning with representative individuals*

In this section, we elaborate on the solution to the JTMDTC problem based on the constructed MDP. We first present a customized DQN as the resolution scheme. We then describe how representative individuals are determined through clustering and utilized for model training, so that the JTMDTC problem for different individuals can be solved in a computationally efficient manner.

2.2.1. *Deep Q-network as the resolution scheme*

To solve the JTMDTC problem, we resort to model-free value-based RL as the solution algorithm. It learns the state-action values associated with the MDP, based on which the optimal policy is implicitly derived by always choosing the action that leads to the maximum state-action value. By definition, the state-action value resulting from policy π is calculated as the expected discounted return conditional on the state-action pair (s_t, a_t):

$$Q_\pi\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right) = E\left[U_t \mid S_t = \boldsymbol{s}_t, A_t = \boldsymbol{a}_t\right]. \tag{13}$$

It characterizes how promising action a_t is at state s_t after which policy π is followed. By maximizing the above state-action value function with respect to the policy, the optimal Q-value $Q^*(s_t, a_t)$ is obtained. The optimal action a_t^* is then the one that leads to the maximum Q-value:

$$\boldsymbol{a}_t^* = \underset{\boldsymbol{a} \in A}{\operatorname{argmax}} Q^*\left(\boldsymbol{s}_t, \boldsymbol{a}\right). \tag{14}$$

The DQN algorithm is at the forefront of model-free value-based RL, which uses a neural network to approximate the optimal state-action value function, thereby extending the application of traditional Q-learning to high-dimensional and/or continuous-space problems:

$$Q\left(\boldsymbol{s}_t, \boldsymbol{a}_t; \mathbf{w}\right) \approx Q^*\left(\boldsymbol{s}_t, \boldsymbol{a}_t\right), \tag{15}$$

where $\mathbf{w}$ is the vector of weights of the neural network.

The success of DQN, or RL in general, lies in the action exploration and exploitation in the course of interactions with the environment. Exploration is basically random selection of actions to fully explore the environment, which is the opposite to exploitation that seeks to select the action that maximizes the optimal Q-value. There is a clear trade-off between exploration and exploitation, which is achieved via the ϵ-greedy strategy. That is, the agent randomly explores actions with probability ϵ and with probability $1 - \epsilon$, it selects and performs the current best action:

$$\boldsymbol{a}_t = \begin{cases} \underset{\boldsymbol{a} \in A}{\operatorname{argmax}} Q\left(\boldsymbol{s}_t, \boldsymbol{a}\right) & \text{with probabilty } 1 - \varepsilon, \\ \operatorname{rand}(\boldsymbol{a}) & \text{with probabilty } \varepsilon. \end{cases} \tag{16}$$

Note that ϵ is not fixed but gradually decaying as the agent explores the environment and accumulates the knowledge. This is sensible because a relatively large value of ϵ helps the agent to quickly and sufficiently explore the environment at the early stage of learning. Toward the later stage, however, a smaller value is of more help so that the agent can fully exploit the potentially optimal actions.

To obtain a successfully trained agent, the Q-value function must be iteratively updated via temporal difference (TD) until convergence or stabilization:

$$Q\big(\boldsymbol{s}_t, \boldsymbol{a}_t; \mathbf{w}\big) \approx r_\mathrm{t} + \gamma \max_{\boldsymbol{a}\in A} Q\big(\boldsymbol{s}_{t+1}, \boldsymbol{a}; \mathbf{w}\big). \tag{17}$$

Thus, the weights of the neural network approximating the Q-value function are optimized to minimize the TD error in the form of the following loss function:

$$L(\mathbf{w}_t) = \left(Q(\boldsymbol{s}_t, \boldsymbol{a}_t) - \left(r_t + \gamma \max_{\boldsymbol{a}'\in A} Q\big(\boldsymbol{s}_{t+1}, \boldsymbol{a}_{t+1}; \mathbf{w}\big)\right)\right)^2. \tag{18}$$

In the presence of mini-batch sampling, the loss function is actually the mean squared error.

To achieve better training performance, the DQN algorithm features two key techniques, one being the experience replay and the other the target network. The former refers to the construction of a memory pool where experience is stored in real time. As will be shortly introduced, each memory pool in our approach consists of the experience of the selected representative from each cluster. Compared with directly using the current experience for training, experience relay not only allows historical experience to be repeatedly used, but also eliminates the undesirable correlation between successive experience samples. The latter technique refers to the preparation and use of two neural networks, one being the online network and the other the target network. Both networks have the same initial weights, but the target network only copies the updated parameters of the online network every fixed number of steps. This makes the DQN algorithm more stable compared with the standard online Q-learning. Putting everything together, we arrive at Algorithm 1 for solving the JTMDTC problem.

Algorithm 1: Joint travel mode and departure time choice modeling and training

```
initialize the replay memory $D$, the policy network parameters $\mathbf{w}$ and $\mathbf{v}$
for $episode = 1$ to $M$ do
    observe initial state $s_0$ from the environment
    for $t = 1$ to $T$ do
        determine the JTMDTC using Eq. (16) for time step $t$
        calculate the corresponding total travel cost for the trip as per Subsection 2.1.3,
          and obtain the reward $r_t$ using Eq. (7)
        record the trip and environment information, and establish the next state $s_{t+1}$
        store the tuple $(\boldsymbol{s}_t, \boldsymbol{a}_t, r_t, \boldsymbol{s}_{t+1})$ into $D$
        sample random mini-batch from $D$
        $Q \leftarrow r_t + \gamma \operatorname*{argmax}_{\boldsymbol{a}\in A} \hat{q}(\boldsymbol{s}_{t+1}, \boldsymbol{a}, \mathbf{v})$
        $\hat{Q} \leftarrow \hat{q}(\boldsymbol{s}_t, \boldsymbol{a}, \mathbf{w}_t)$
        $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \alpha\nabla\frac{1}{2}(Q - \hat{Q})^2$
        set $\mathbf{v} \leftarrow \mathbf{w}_t$ every fixed number $c$ of time steps
    end
end
```

2.2.2. Obtaining representative individuals for efficient model training

In the previous section, we elaborate on the DQN algorithm as the basic resolution scheme for solving the JTMDTC problem. However, how to generalize the algorithm for solving the same problem with a large number of individuals remains an open question. We have previously discussed that storing all the experience of individuals for training is computationally unwise and inefficient, and that randomly selecting one or a few individuals is insufficient and unsound. Thus, we propose an elegant yet effective method to obtain representative individuals for efficient model training, which is to cluster individuals based on their travel characteristics. For individuals within the same cluster, we consider their travel characteristics similar. Thus, each of them can be treated as a representative of the cluster, whose experience can be utilized to train the DQN on behalf of the rest. In this way, we not only avoid deploying as many agents as the number of individuals, but also effectively and efficiently utilize the experience of representative individuals for sufficient model training. In fact, the JTMDTC problem with many individuals can be efficiently solved by the proposed method without sacrificing too much optimality in the decision making. Supporting evidence will be provided in the results.

To obtain representative individuals, we resort to a widely used clustering algorithm called density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996). It is a non-parametric method without the need to assume or specify the distribution of data. The core of DBSCAN is to first identify high-density samples and then gradually connect those similar samples into bigger clusters. Two travel characteristics are considered as input to clustering, namely the travel distance and the accessibility of public transportation (see the methodology). They jointly provide a general picture of the decision-making environment faced by individuals.

The overall workflow to cluster individuals and obtain representatives is presented in Algorithm 2. These representatives are simultaneously simulated in the same environment rather than in a separate manner, whose experience is stored into their respective memory pools for model training. Note that one representative is selected from each cluster of individuals with equal probability, and that different combinations would not lead to significant performance variations. This will be demonstrated in the results.

Algorithm 2: Clustering individuals and obtaining representatives

```
input : travel characteristics of all individuals $D = \{\boldsymbol{x}_1 = [L_1, p_1]^T, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\}$, distance
        threshold $eps$, minimum number of points required to form a cluster $MinPts$
output: clusters of individuals for obtaining representatives
data normalization: $x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
for each non-clustered individual $\boldsymbol{x} \in D_{normalized}$ do
    label individual $\boldsymbol{x}$ as clustered into a group
    $N \leftarrow GetNeighbors(\boldsymbol{x}, eps)$
    if $|N| < MinPts$ then
        label individual $\boldsymbol{x}$ as noise
    else
        set the new cluster $C \leftarrow \boldsymbol{x}$
        for each individual $\boldsymbol{x}' \in N$ do
            $N \leftarrow N \backslash \boldsymbol{x}'$
            if the individual $\boldsymbol{x}'$ is non-clustered then
                label individual $\boldsymbol{x}'$ as clustered
                $N' \leftarrow GetNeighbors(\boldsymbol{x}', eps)$
                if $|N'| \geq MinPts$ then
                    $N \leftarrow N \cup N'$
                end
            end
            if $\boldsymbol{x}'$ is noise then
                $C \leftarrow C \cup \{\boldsymbol{x}'\}$
            end
        end
    end
end
```

Combing the customized DQN with the process of clustering individuals and obtaining representatives results in the final integrated algorithm for solving the JTMDTC problem with many individuals. The number of agents to be trained is equal to the number of representatives or clusters. These agents are trained simultaneously with their respective memory pools. Once sufficiently trained, they are jointly utilized to make travel choice decisions for different individuals without the need to redo clustering. That is, the action performed by the agent that yields the highest reward is chosen to be implemented.

3. Numerical experiments and results

3.1. Simulation setup

We use Simulation of Urban MObility (SUMO) in this study as our traffic simulation engine. The chosen road network is part of the urban area of Suzhou, China of approximately 20 km^2 (see Fig. 2a). To build and simulate this multimodal transportation network, we use OpenStreetMap (OSM) to acquire the network geometry and configuration (2,423 nodes and 4,970 edges), and feed

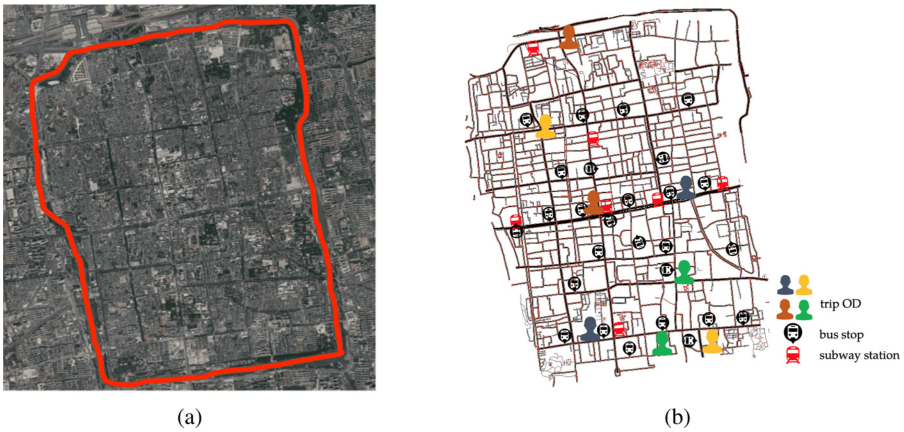


Fig. 2. (a) Map of the study area in Suzhou, China; (b) multimodal traffic simulation network constructed in SUMO.

Table 1
Simulation parameters used in the numerical experiments.

Parameter	Description	Value
$t_{\min}$	Earliest departure time	07:00
$t_{\max}$	Latest departure time	09:00
t_{unit}	Unit interval for departure time choice (min)	30
E_1	Constant for mapping cost to reward	100
E_2	Constant for mapping cost to reward	0.1
α	Value of time (CNY/min)	0.5
β	Unit costs of the schedule delay for early arrivals (CNY/min)	0.05
γ	Unit costs of the schedule delay for late arrivals (CNY/min)	0.3
o	Fuel price (CNY/km)	0.56
\	Bus fare (CNY)	2
\	Subway base fare (CNY)	1
\	Subway fare per km increase (CNY/km)	0.2
\	Peak frequency of bus (veh/h)	10
\	Peak frequency of subway (veh/h)	14
\	Off-peak frequency of bus (veh/h)	6
\	Off-peak frequency of subway (veh/h)	8
\	Dwell time of bus (s)	40
\	Dwell time of subway (s)	30

Table 2
Travel demand configuration for the morning peak period across five consecutive workdays.

Demand[veh/h]	7:00-7:30	7:30-8:00	8:00-8:30	8:30-9:00
Monday	$D_1 \sim \mathcal{N}(2700, 108^2)$	$D_1 \sim \mathcal{N}(5000, 200^2)$	$D_1 \sim \mathcal{N}(3800, 152^2)$	$D_1 \sim \mathcal{N}(2500, 100^2)$
Tuesday	$D_2 \sim \mathcal{N}(3300, 132^2)$	$D_2 \sim \mathcal{N}(4400, 176^2)$	$D_2 \sim \mathcal{N}(4000, 160^2)$	$D_2 \sim \mathcal{N}(3300, 132^2)$
Wednesday	$D_3 \sim \mathcal{N}(3000, 120^2)$	$D_3 \sim \mathcal{N}(4800, 192^2)$	$D_3 \sim \mathcal{N}(3200, 128^2)$	$D_3 \sim \mathcal{N}(3000, 120^2)$
Thursday	$D_4 \sim \mathcal{N}(2800, 112^2)$	$D_4 \sim \mathcal{N}(4200, 168^2)$	$D_4 \sim \mathcal{N}(3500, 140^2)$	$D_4 \sim \mathcal{N}(3500, 140^2)$
Friday	$D_5 \sim \mathcal{N}(1500, 60^2)$	$D_5 \sim \mathcal{N}(4000, 160^2)$	$D_5 \sim \mathcal{N}(4900, 196^2)$	$D_5 \sim \mathcal{N}(3600, 144^2)$

these information into SUMO to establish the simulation environment.Fig. 2a illustrates the urban layout of Suzhou within which the simulation takes place, clearly marked by a red outline to provide geographical context. As a multimodal network, public transportation including buses and subways must be configured. This is achieved by first extracting the information of public transportation operations from a map service application and then performing map matching. The information extracted includes line IDs, stop or station IDs, and their geographical locations. These locations are illustrated in Fig. 2b.

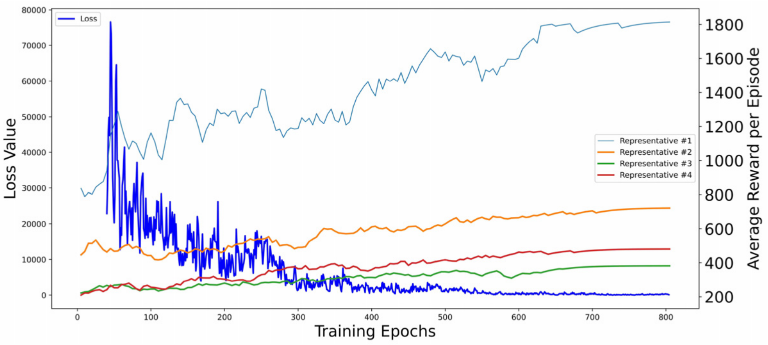
The operational characteristics as well as necessary simulation parameters are summarized in Table 1. Our approach is to use the existing knowledge of the utility function parameters (Tian et al., 2009; Västberg et al., 2020), which are usually obtained through survey data and observed choices, to inform our RL model and make it more accurate and effective.

The JTMDTC during a typical morning peak period between 7 a.m. and 9 a.m. is considered in the simulation. Such choices across five consecutive workdays consist of a single episode in the training. Although synthetic travel demand is used, it does not affect the validity of the proposed approach. The 2-hr simulation horizon is divided into four 30-min time intervals. For each workday, we have a typical demand pattern for the morning peak period, and for each time interval, the travel demand is treated as a random variable following a Gaussian distribution (see Table 2). Thus, each day of one episode during the simulation and training is associated with a slightly different travel demand. This stochasticity is in line with the daily traffic fluctuations around a certain pattern observed in reality (i.e., recurrent congestion). However, in the case of non-recurrent congestion or unexpected events, the proposed approach might no longer be applicable because the effects of such events on the travel choice are not experienced and learned. Further investigation in this regard is worthwhile.

3.2. Training results

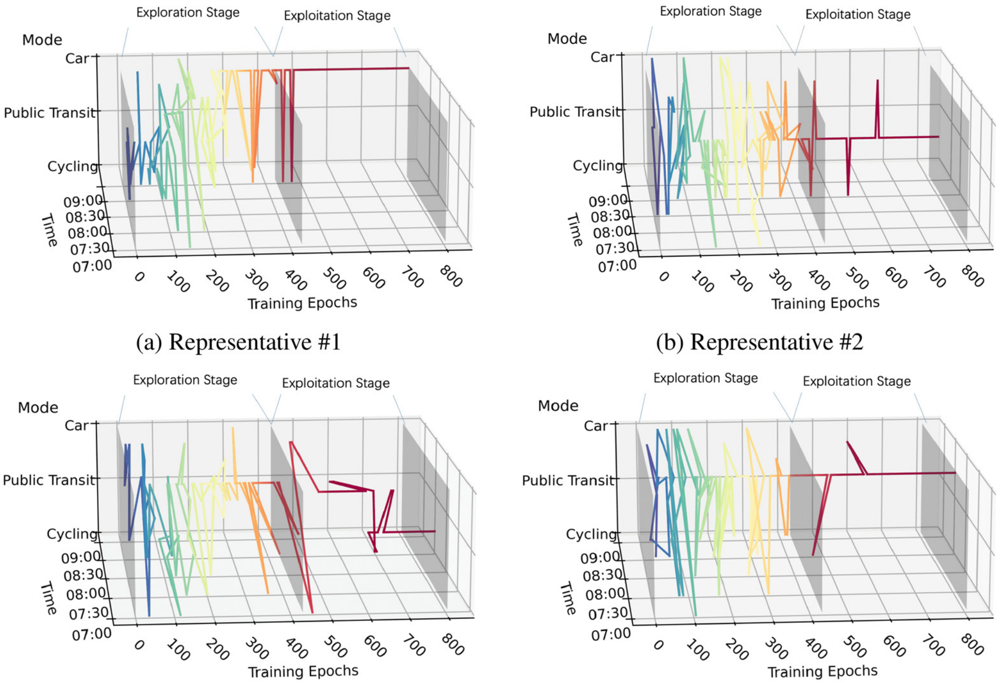
We choose 60 time-dependent origin-destination (OD) trips of individuals in the network and input their respective travel characteristics into the clustering method. By setting the minimum number of points to 10 and the minimum distance threshold to 0.05, we arrive at four clusters. From each cluster, a representative individual is selected whose experience is put into the common memory pool for training the DQN. The locations of these four ODs in the network are illustrated in Fig.. All numerical experiments are conducted on a standard computer with Intel Core (TM) i5-9400F 2.90 GHz CPU and 8 GB RAM.

Fig. 3 shows the convergence pattern of the loss function resulting from 800 episodes of training. The overall decreasing trend is obvious and desired. At the early stage, notable fluctuations in the loss function value exist due to the action exploration as well as the fact that the agent knows nothing about the environment at the very beginning. Such fluctuations, however, persist mainly



(a)

Fig. 3. Convergence pattern of the loss function and variations of the reward for the four representative individuals.



(a) Representative #1

(b) Representative #2

(c) Representative #3

(d) Representative #4

Fig. 4. Action selections of the four representative individuals.

within the first 300 episodes and do not last long. In fact, starting from about 500 episodes, the loss function value shows minimal variations and almost lands on a straight line. This observation clearly indicates the convergence of the algorithm.

We now examine the variations of the reward each representative individual receives by following the DRL-suggested actions in the course of training. The results are shown in Fig. 3. Given that the trip of each representative has a different time-dependent OD pair, the associated reward is of different levels of magnitude. In fact, one can easily tell that representative #1 is likely to have the longest trip while representatives #3 and #4 might have much shorter trips. Nevertheless, regardless of the absolute values of the reward, all the representatives exhibit an increasing trend in the curve meaning that they are constantly improving their travel choices by interacting with and learning from the environment. Within approximately 700 episodes, the reward for each representative is basically stabilized without changing too much thereafter, an observation that is consistent with Fig. 3.

In Fig. 4, we graphically present the DRL-suggested actions followed by the four representative individuals in the course of training. One can immediately recognize the action exploration stage at the beginning of the training, where representatives actions frequently change. But once entering the action exploitation stage, we no longer see such volatility, and each representative seems to have found the optimal solution to its own JTMDTC problem. The occasional action changes during exploitation are likely the result of random action selection triggered by the ϵ-greedy strategy (where ϵ has decreased to a very small value).

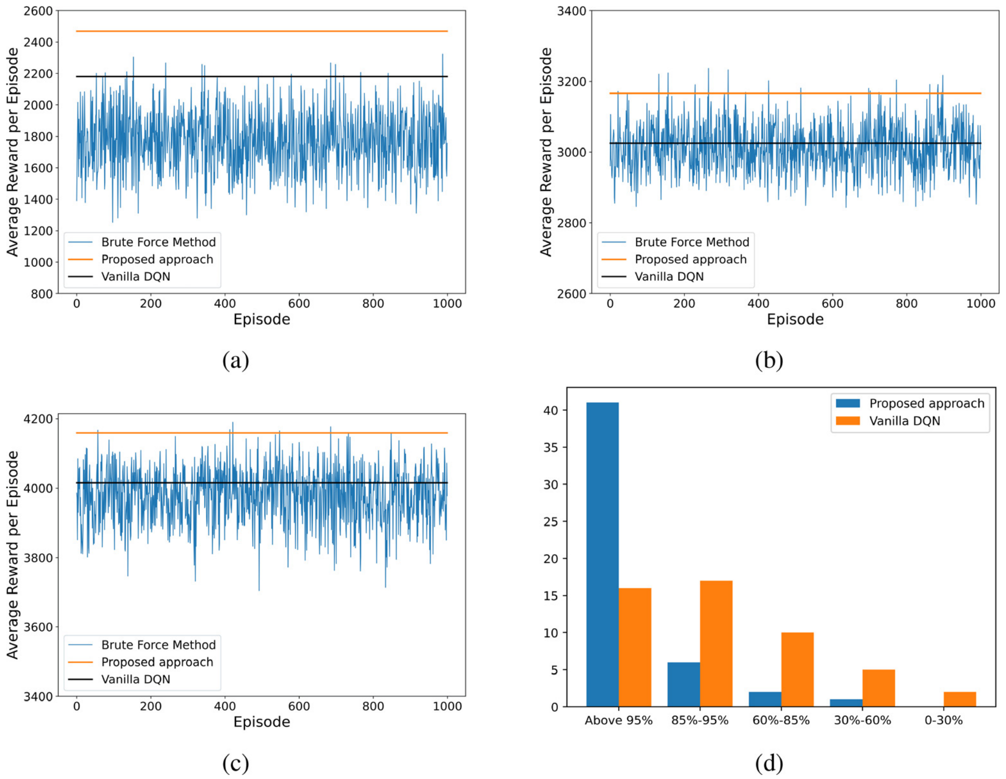


Fig. 5. (a), (b), and (c) show the performance comparison of different approaches for three test individuals; (d) is a global picture of the performance difference for all the 50 test individuals.

All the results obtained so far demonstrate that the proposed approach is effective to obtain a good solution to the JTMDTC problem. But how good or optimal the solution is has not been answered. Another open question relates to the applicability or transferability of the trained agent to other individuals who are not part of the training. Both questions will be answered in the following.

3.3. Performance comparison of different approaches

To examine the goodness or optimality of the solution obtained from the proposed approach, we perform a comparative analysis where the proposed approach is compared with the vanilla DQN and a brute force method that randomly selects and tries all the possible actions. 50 new time-dependent OD trips of test individuals in the network, who are not considered in the training, are chosen for this comparative analysis. To perform the comparison, we let each of the test individuals to undertake actions according to one of these compared decision makers, respectively, and collect the resulting reward. 1,000 episodes are performed for the brute force method in order to fully explore the action space.

The comparative results for three selected test individuals are shown in Fig. 5a, b, and c. Significant fluctuations of the reward can be observed for the brute force method, as expected, due to the random action selection that does not always guarantee a good result. For both the proposed approach and the vanilla DQN, the reward obtained is a single value represented as a straight line and associated with the optimal solution to the JTMDTC problem. Clearly, for all the three test individuals, the solution given by the proposed approach not only outperforms that resulting from the vanilla DQN, but also dominates most of the solutions given by the brute force method. In fact, without clustering individuals and utilizing representatives, about half of the solutions given by the brute force method can beat that resulting from the vanilla DQN (see Fig. 5a and b).

To see a bigger picture of the comparison for all the 50 test individuals, we find the maximum reward obtained by the brute force method for each of them, which is treated as the (near-)optimal solution to the JTMDTC problem. By comparing the solutions given by the proposed approach and the vanilla DQN with this reference value, we can observe the performance difference from a global perspective. As shown in Fig. 5d, the majority of the solutions (over 40 out of 50) given by the proposed approach are over 95 percent of the reference value, meaning that these solutions are close to optimality. This is clearly not the case for the vanilla DQN, because only about 30 percent of the solutions are over 95 percent of the same reference value, not to mention that quite a few solutions are even lower than 60 percent of the reference. Thus, the comparative results demonstrate the effectiveness of the proposed approach when applied to solve the JTMDTC problem with many individuals, as well as the important role that representatives play in fulfilling this task. Since the test individuals are not part of the training, the results indicate good transferability of the proposed approach.

Table 3
Performance evaluation of different alternative models.

Model	NLL	Avg Reward	Accuracy
First-order MC model	19.23	1,354	0.24
Discrete choice model	24.17	1,217	0.23
Proposed approach with partial information	17.31	1,449	0.29
Proposed approach with perfect information(baseline)	\	1,735	\

We now try to examine the performance of the proposed approach in the case of partial in- formation. As previously discussed, partial information is more relevant from a human behavioral perspective, and thus is expected to yield worse action selections. Here, we also consider two other models for comparison purposes. The first one is a first-order Markov chain (MC) model that only uses the current travel distance and departure time difference to decide on the next choice. Its states, transition and initial state probabilities, and the reward function are all derived from the DRL model. The main difference between the two is the decision-making process. The MC model uses the transition and initial state probabilities to simulate behavior over time, while the DRL model uses an iterative trial-and-error process.

The second one is the traditional MNL model that uses the reward function, Eq. (7). We use the travel choices of individuals resulting from the proposed approach with perfect information as the baseline, based on which the performance of the other models are compared. Three metrics for performance evaluation and comparison are considered. Apart from the reward, the other two are the negative log-loss (NLL) and the Jaccard index. Both metrics measure the closeness or similarity of the actions yielded by the other models to those obtained by the baseline. Thus, they can reflect the level of optimality of the other models relative to the baseline. Note, however, that a lower value for the NLL is desired whereas for the Jaccard index, the higher the better.

Table 3 summarizes the three performance metrics for comparison. As expected, the proposed approach with perfect information provides the best performance in achieving as much reward as possible. All the other models exhibit worse performance, as can be readily seen by comparing the average rewards that are all lower than that of the baseline. This trend is also true for either the NLL or the Jaccard index. Nevertheless, the proposed approach with partial information still exhibits slightly better performance compared with the first-order MC model and the MNL model, which shows the effectiveness of the proposed approach even in the presence of partial information.

3.4. Sensitivity analysis on the model parameters

We now perform two sensitivity analyses to investigate the performance changes of the proposed approach in response to changes in the model parameters. The first parameter to be examined is the number of representatives, and the second is the set of the training individuals. To see the effects of the former, we perform further experiments with 1, 10, 20, and 40 representatives, respectively. We keep the same experiment setup where the 60 time-dependent OD trips of individuals are clustered into the above numbers for selecting training representatives, while the other 50 test individuals are used for evaluation and comparison. Again, the brute force method is used as a reference.

The comparative results are summarized in Table 4. The experiment with 40 representatives cannot be accomplished on the same machine due to the memory overflow, and thus the result is not reported. With more and more representatives, the required training or computational time increases, as expected. However, the increased number of representatives does lead to a better reward. The greatest improvement in the reward is achieved by turning one representative into four. Further increasing this number to 10 or 20 does not improve the reward much. This result indicates that increasing the number of representatives is not necessarily cost effective. In fact, a small number of representatives can already produce a rather good result with reasonable computational time. Using the maximum reward obtained by the brute force method as the reference value, we compare the number of test individuals whose rewards given by the proposed approach are over 95 percent of the reference. As expected, a similar pattern is observed where this number remains large with little change for 4, 10, and 20 representatives. Fig. 6 further shows the comparison of the results for four selected test individuals with different numbers of representatives. One representative only is clearly insufficient to beat the brute force method, while four or more representatives yield promising results.

To show that the performance of the proposed approach does not change much due to different sets of the training individuals, we perform another group of experiments with four representatives by selecting different individuals from the clusters for training the DQN. All the other experiment setup remains the same. Table 5 summarizes the results for four such experiments. Since a different

Table 4
Performance changes of the proposed approach in response to different numbers of representatives.

# of representatives	1	4	10	20	40
Training time (h)	5	22	68	140	\
Average reward	2,684	3,203	3,351	3,394	\
Above 95(out of 50)%	16	41	45	46	\

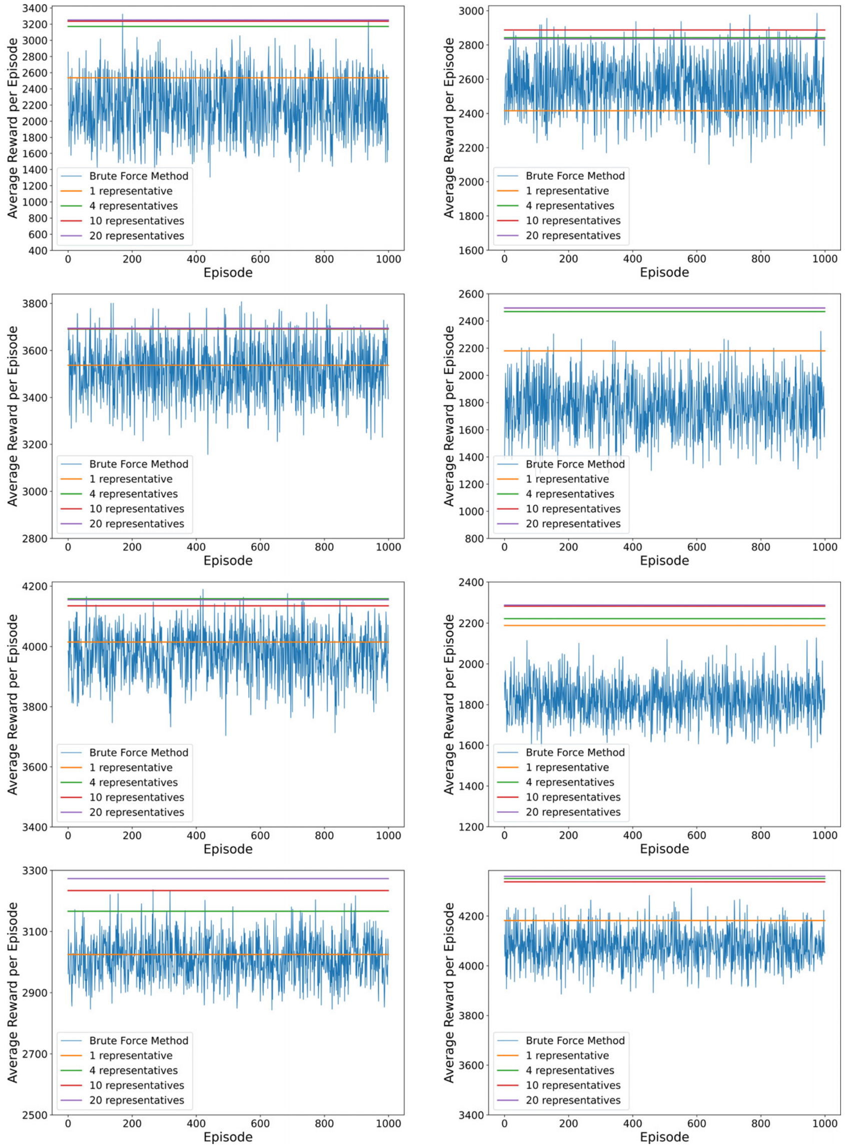


Fig. 6. The results of eight selected test individuals with different numbers of representatives.

Table 5
Performance changes of the proposed approach when using different sets of the training individuals.

	Set 1	Set 2	Set 3	Set 4
Average reward	3,203	3,179	3,280	3,248
Above 95% (out of 50)	41	39	43	43

set of the training individuals does not change the computational time (which is 22 hr for four representatives), this metric is no longer reported. From the results, it is clear that the performance of the proposed approach is stable without exhibiting significant variations as different representative individuals are used for training. Similar to Fig. 6. Fig. 7 shows the comparison of the results for four selected test individuals when different sets of the training individuals are used, which manifests the robustness of the proposed approach to the selection of representative individuals.

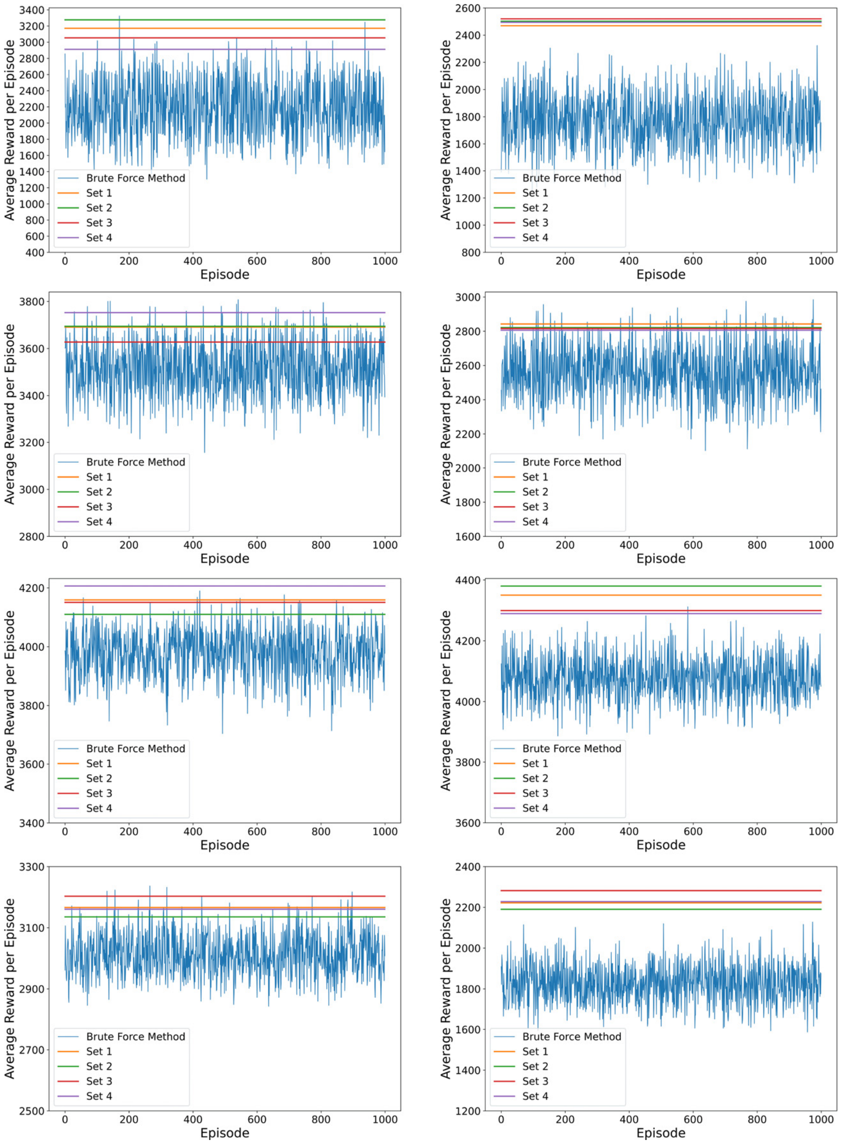


Fig. 7. The results of eight selected test individuals with different sets of the training individuals.

4. Conclusion

In this paper, we propose a new JTMDTC model based on DRL to maximize individuals travel utilities in a dynamic multimodal transportation network. Unlike traditional DCMs that mainly reply on the random utility theory for characterizing travel behavior, the proposed approach is driven by a learning mechanism whereby the agent keeps improving the decision-making logic via constantly interacting with the complex traffic environment across multiple days. This process is treated as a sequential decision-making problem and the solution to the problem is the (near-)optimal JTMDTC that helps individuals achieve as high travel utilities as possible. Such a decision, however, is not necessarily from a behavioral perspective. Rather, it is more like a guidance or recommendation as in mobility-as-a-service, where individuals make travel requests to the system and the system processes all the requests to yield suggestions on individuals travel choices.

To effectively apply DRL in this context, a problem-specific MDP is constructed to characterize the multi-day JTMDTC. A customized DQN is then developed as the resolution scheme that is well suited for high-dimensional and/or continuous-space problems. To render the approach applicable to dealing with many individuals with travel decision-making requests, a clustering method is integrated into the modeling framework so that representative individuals are obtained for training the agent, thus resulting in an elegant and computationally efficient approach. We perform extensive numerical experiments based on multimodal microsimulation in a real-world network of Suzhou, China to demonstrate the effectiveness of the proposed approach. By comparing it with several other models, we show that the proposed approach is able to make (near-)optimal decisions on the JTMDTC with consistently higher travel utilities for different individuals in complex traffic environments, and that the approach is robust to different model parameter changes.

The overall modeling framework proposed in this study is actually inspired by the decision-making mechanism of rational human beings, namely learning from experience. This is the foundation of the constructed MDP and the DRL model. Although the reward function is still driven by the random utility theory, it can be designed differently to account for possible human behavioral characteristics such as travel inertia. In other words, the proposed approach has a more flexible modeling structure which can be utilized to develop models with heterogeneous trip-making objectives. The ultimate goal is perhaps to provide individuals with some personalized guidance or recommendation on the JTMDTC.

The current approach to travel recommendation is flexible with a reward function that can be adjusted for different users or situations, but is limited to single travelers and requires one day of demand data. Future research could improve the real-time nature of the system and incorporate behavioral and sociodemographic characteristics, inter-modal transfers, and dynamic interactions among individuals to better utilize transportation network resources and improve system efficiency.

Declaration of competing interest

Authors declare no conflict of interest.

CRediT authorship contribution statement

Ziyuan Gu: Conceptualization, Writing – original draft, Writing – review & editing. **Yukai Wang:** Formal analysis, Writing – original draft, Writing – review & editing. **Wei Ma:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Zhiyuan Liu:** Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

This study is supported by the Key Project (52131203) of the National Natural Science Foundation of China, the High-Level Personnel Project of Jiangsu Province (JSSCBS20220099), the Fundamental Research Funds for the Central Universities (2242022R40025), and the Start-up Research Fund of Southeast University (RF1028623057).

References

Aradi, S., 2022. Survey of deep reinforcement learning for motion planning of autonomous vehicles. IEEE Trans. Intell. Transp. Syst. 23 (2), 740–759.

Arentze, T.A., Timmermans, H.J.P., 2004. A learning-based transportation oriented simulation system. Transp. Res. Part B: Methodol. 38 (7), 613–633.

Bhat, C.R., 1998. Analysis of travel mode and departure time choice for urban shopping trips. Trans. Res. Part B: Methodol. 32 (6), 361–371.

Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transp. Res. Part B: Methodol. 32 (1), 61–75.

Cruciol, L.L., de Arruda Jr, A.C., Weigang, L., Li, L., Crespo, A.M.F., 2013. Reward functions for learning to control in air traffic flow management. Transp. Res. Part C: Emerg. Technol. 35, 141–155.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231.

Fukuda, D., Yai, T., 2010. Semiparametric specification of the utility function in a travel mode choice model. Transportation 37 (2), 221–238.

Grigorescu, S., Trasnea, B., Cocias, T., Macesanu, G., 2020. A survey of deep learning techniques for autonomous driving. J. Field Robot. 37 (3), 362–386.

Gu, Y., Chen, A., Kitthamkesorn, S., 2022. Accessibility-based vulnerability analysis of multi-modal transportation networks with weibit choice models. Multimodal Transp. 1 (3), 100029.

Gu, Z., Liu, Z., Cheng, Q., Saberi, M., 2018. Congestion pricing practices and public acceptance: a review of evidence. Case Stud. Transport Policy 6 (1), 94–101.

Gu, Z., Yang, X., Zhang, Q., Yu, W., Liu, Z., 2023. Terl: two-stage ensemble reinforcement learning paradigm for large-scale decentralized decision making in transportation simulation. IEEE Trans. Knowl. Data Eng..

Hendrickson, C., Plank, E., 1984. The flexibility of departure times for work trips. Transp. Res. Part A: General 18 (1), 25–36.

Idris, A., Shalaby, A., Habib, K., 2012. Towards a learning-based mode shift model: a conceptual framework. Transp. Lett.: Int. J. Transp. Res. 4 (1), 15–27.

Li, K.K., Lai, L.L., David, A.K., 2000. Application of artificial neural network in fault location technique. In: DRPT2000. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Proceedings (Cat. No. 00EX382). IEEE, pp. 226–231.

McFadden, D., et al., 1973. Conditional logit analysis of qualitative choice behavior. In: Frontiers in Econometrics., pp. 105–142.

Ning, Z., Zhang, K., Wang, X., Obaidat, M.S., Guo, L., Hu, X., Hu, B., Guo, Y., Sadoun, B., Kwok, R.Y.K., 2020. Joint computing and caching in 5g-envisioned internet of vehicles: a deep reinforcement learning-based traffic control system. IEEE Trans. Intell. Transp. Syst. 22 (8), 5201–5212.

Omrani, H., 2015. Predicting travel mode of individuals by machine learning. Transp. Res. Procedia 10, 840–849.

Pitale, A.M., Parida, M., Sadhukhan, S., 2023. Factors influencing choice riders for using park-and-ride facilities: a case of delhi. Multimodal Transp. 2 (1), 100065.

Qin, X., Ke, J., Wang, X., Tang, Y., Yang, H., 2022. Demand management for smart transportation: a review. Multimodal Transp. 1 (4), 100038.

Rasca, S.I., Markvica, K., Biesinger, B., 2023. Persona design methodology for work-commute travel behaviour using latent class cluster analysis. Multimodal Transp. 2 (4), 100095.

Rasouli, S., Timmermans, H., 2014. Using ensembles of decision trees to predict transport mode choice decisions: effects on predictive success and uncertainty estimates. Eur. J. Transport Infrastruct. Res. 14 (4), 412–424.

Semanjski, I., Lopez, A., Gautama, S., 2016. Forecasting transport mode use with support vector machines based approach. Trans. Maritime Sci. 5 (02), 111–120.

Shi, Y., Gu, Z., Yang, X., Li, Y., Chu, Z., 2023. An adaptive route guidance model considering the effect of traffic signals based on deep reinforcement learning. IEEE Intell. Transp. Syst. Mag..

Shou, Z., Chen, X., Fu, Y., Di, X., 2022. Multi-agent reinforcement learning for markov routing games: a new modeling paradigm for dynamic traffic assignment. Transp. Res. Part C: Emerg. Technol. 137, 103560.

Small, K.A., 1982. The scheduling of consumer activities: work trips. Am. Econ. Rev. 72 (3), 467–479.

Souche-Le Corvec, S., 2023. Which transport modes do people use to travel to coworking spaces (CWSs)? Multimodal Transp. 2 (2), 100078.

Tang, L., Xiong, C., Zhang, L., 2015. Decision tree method for modeling travel mode switching in a dynamic behavioral process. Transp. Plan. Technol. 38 (8), 833–850.

Tian, Q., Lam, W.H.K., Huang, H., Mou, H., 2009. Modeling time-dependent travel choice problems in a mixed-mode network with park-and-ride facilities. In: 2009 International Joint Conference on Computational Sciences and Optimization, Vol. 2. IEEE, pp. 119–123.

Train, K., McFadden, D., 1978. The goods/leisure tradeoff and disaggregate work trip mode choice models. Transp. Res. 12 (5), 349–353.

Vanhulsel, M., Janssens, D., Wets, G., Vanhoof, K., 2009. Simulation of sequential data: an enhanced reinforcement learning approach. Expert Syst. Appl. 36 (4), 8032–8039.

Västberg, O.B., Karlström, A., Jonsson, D., Sundberg, M., 2020. A dynamic discrete choice activity-based travel demand model. Transp. Sci. 54 (1), 21–41.

Walraven, E., Spaan, M.T.J., Bakker, B., 2016. Traffic flow optimization: a reinforcement learning approach. Eng. Appl. Artif. Intell. 52, 203–212.

Weis, C., Kowald, M., Danalet, A., Schmid, B., Vrtic, M., Axhausen, K.W., Mathys, N., 2021. Surveying and analysing mode and route choices in switzerland 2010–2015. Travel Behav. Soc. 22, 10–21.

Xie, J., Liu, Y., Chen, N., 2023. Two-sided deep reinforcement learning for dynamic mobility-on-demand management with mixed autonomy. Transp. Sci. 57 (4), 1019–1046.

Yang, J., Zhang, J., Wang, H., 2020. Urban traffic control in software defined internet of things via a multi-agent deep reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. 22 (6), 3742–3754.

Yin, W., Murray-Tuite, P., Ukkusuri, S.V., Gladwin, H., 2014. An agent-based modeling system for travel demand simulation for hurricane evacuation. Transp. Res. Part C: Emerg. Technol. 42, 44–59.

Yu, X., Gao, S., 2019. Learning routing policies in a disrupted, congestible network with real-time information: An experimental approach. Transp. Res. Part C: Emerg. Technol. 106, 205–219.

Zhao, Z., Lee, C.K.M., 2021. Dynamic pricing for EV charging stations: a deep reinforcement learning approach. IEEE Trans. Transp. Electrificat. 8 (2), 2456–2468.

Zhao, Z., Liang, Y., 2023. A deep inverse reinforcement learning approach to route choice modeling with context-dependent rewards. Transp. Res. Part C: Emerg. Technol. 149, 104079.

Zhu, M., Wang, X., Wang, Y., 2018. Human-like autonomous car-following model with deep reinforcement learning. Transp. Res. Part C: Emerg. Technol. 97, 348–368.

Zimmermann, M., Västberg, O.B., Frejinger, E., Karlström, A., 2018. Capturing correlation with a mixed recursive logit model for activity-travel scheduling. Transpo. Res. Part C: Emerg. Technol. 93, 273–291.

Zou, M., Li, M., Lin, X., Xiong, C., Mao, C., Wan, C., Zhang, K., Yu, J., 2016. An agent-based choice model for travel mode and departure time and its case study in beijing. Transp. Res. Part C: Emerg. Technol. 64, 133–147.