Neural Computing and Applications (2023) 35:20221–20240
https://doi.org/10.1007/s00521-023-08733-4

ORIGINAL ARTICLE

Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness

Kai-Fung Chu[1] · Weisi Guo[1]

Received: 4 April 2023 / Accepted: 31 May 2023 / Published online: 20 July 2023
© The Author(s) 2023

Abstract
Multimodal transportation systems require an effective journey planner to allocate multiple passengers to transport operators. One example is mobility-as-a-service, a new mobility service that integrates various transport modes through a single platform. In such a multimodal and diverse journey planning problem, accommodating heterogeneous passengers with different and dynamic preferences can be challenging. Furthermore, passengers may behave based on experiences and expectations, in the sense that the transport experience affects their state and decision of the next transport service. Current methods of treating each journey planning optimization as a non-time varying single experience problem cannot adequately model passenger experience and memories over many journeys over time. In this paper, we model passenger experience as a Markov model where prior experiences have a transient effect on future long-term satisfaction and retention rate. As such, we formulate a multi-objective journey planning problem that considers individual passenger preferences, experiences, and memories. The proposed approach dynamically determines utility weights to obtain an optimal journey plan for individual passengers based on their status. To balance the profit received by each transport operator, we present a variant-based proportional fairness. Our experiments using real-world and synthetic datasets show that our approach enhances passenger satisfaction, compared to baseline methods. We demonstrate that the overall profit is increased by 2.3 times, resulting in a higher retention rate caused by higher satisfaction levels. Our proposed approach can facilitate the participation of transport operators and promote passenger acceptance of MaaS.

Keywords Intelligent transportation systems · Deep reinforcement learning · Passenger behavior · Optimization · Proportional fairness

1 Introduction

Mobility-as-a-service (MaaS) is a new type of mobility service that integrates various transport modes through a single platform by providing real-time information services and journey planning and booking across various transport operators. Passengers can enjoy the combined transport services offered by the MaaS scheduler across various transport operators. MaaS has the potential to reduce traffic congestion, air pollution, noise nuisance, energy consumption, and transport-related social exclusion, as well as to increase traffic safety, health, social cohesiveness, accessibility, and household expenditure [1]. To implement MaaS in intelligent transportation systems, several essential components are required: (1) a mobile application and a web-based interface that enables passengers to plan and confirm their journeys, (2) a software system that connects operators providing mobility services, (3) an intelligent scheduler that handles passenger journey queries and optimizes transport resource and utility, such as the scheduler in [2]. The quality of these components significantly impacts the successful transformation of how people travel in future intelligent transportation systems.

Among the components of MaaS, the intelligent scheduler is the only component that determines the journey offered to passengers, and thus, it influences the effectiveness of the transportation system and passenger

✉ Kai-Fung Chu
kaifung.chu@cranfield.ac.uk

Weisi Guo
weisi.guo@cranfield.ac.uk

[1] School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK

utility. Designing such a scheduler could be a challenging task and a loosely organized scheduler could diminish the utility of transport operators and passengers. Therefore, the community has put effort into the design of the scheduler. The problem of determining the best route from origin to destination can be viewed as a shortest path problem [3, 4]. Many algorithms, such as Dijkstra's algorithm [5] and A* search algorithm [6], have been proposed to determine the shortest path between two points in a transportation network. With more practical considerations such as time window, the number of transfers, and parking constraints in the transport network, various models and algorithms were proposed. In [7], the authors use a choice model to provide an optimized menu of travel options complying with seat capacity and committed time schedule constraints to passengers, which maximizes the profit of the operator. In [8], a cost allocation problem is formulated to minimize the passenger dissatisfaction in the system. We can see that all the aforementioned models in the literature are valid designs, but they are aimed at different objectives, including but not limited to the journey time, dissatisfaction, and profit.

Different journeys attract various potential passengers and their willingness to pay is not the same [9]. Especially for multimodal transportation systems such as MaaS, the weighting of each objective may have a significant effect on the choice of transport services. Therefore, a multi-objective formulation that minimizes the overall cost, time and users' discommodity in multimodal transportation networks, as in [10], would be appreciated by passengers. In particular, it presented a utility measure that takes into account different passengers' propensities. However, we argue that a static utility weight vector is still not favored by all transport operators and passengers. Moreover, the utility of transport operators and passengers can be in conflict with each other. From the transport operators' perspective, profit is the primary consideration but not the passenger's utility. They take passengers' utility into account simply because the retention rate is related to passenger satisfaction and thus affects profit. In other words, passengers may not retain to the system if transport operators provide high-profit transport services only without considering passenger satisfaction. Hence, the behavior of the passenger can be better modeled by a dynamic function (e.g., Markov chain, differential equation) instead of static utility since the transport experience influences the passenger's decision of their next transport service.

Markov decision process (MDP) is a discrete-time stochastic control process for modeling partly random outcomes under decision-making process [11]. Many transport operations can be modeled as MDP. For example, the optimal routing of a taxi searching for a new passenger was modeled as an MDP to account for long-term profit over the working period [12]. The transfer of travelers between integrated transportation was also modeled as MDP to explore the optimal transfer trip chain of different income groups [13]. The passenger purchasing behavior of air tickets was taken into account to modeled the dynamic pricing process as MDP in [14]. In the event of waiting for a taxi, the passenger may be modeled as MDP to decide whether keep waiting at the current location or move to the nearby road segment [15]. To solve an MDP problem, dynamic programming [16] is a common method to determine the optimal solution of the large problem from the values of the sub-problems, which establish the foundation of later advanced methods such as reinforcement learning (RL). RL is a learning-based method that performs action based on given observations of the system without the need for a mathematical model. The recent success in deep RL (DRL), which combines reinforcement learning and deep learning, makes it an excellent approach to solve complex problems. Therefore, DRL is a promising method for addressing the complex multi-objective journey planning problem in multimodal transport by considering dynamic passenger behavior that satisfies both transport operators' and passengers' utilities.

To the best of our knowledge, in the context of MaaS, no prior research has considered passenger satisfaction behavior as an MDP for multi-objective journey planning problems that satisfy the utilities of both transport operators and passengers. One potential way of analyzing the area of personalized behavior preferences is to assume there is a hidden customer agent with its own personal reward function and policy. In this case, inverse RL is a reasonable way of inferring these. In our case, we assume there is only a Hidden MDP common in all agents' behaviors, and that the MaaS central controller is there to infer this and benefit its own central reward function and optimize its policy. In this paper, we aim to address this research gap by leveraging DRL to obtain the utility weight vectors for both transport operators and passengers. These weight vectors are unique to each passenger for the transport operator to identify their preference of transport service. A multi-objective journey planning problem is formulated with the utility weight vector including travel time, comfortableness, price, and operating cost. Different plans can be generated from the problem with different weight vectors. The system obtains optimal utility weight vectors that enhance passenger satisfaction level and transport operator profit caused by high retention rate. A proportional fairness-based variant is presented to balance the profit received by each transport operator. Detailed experiments are conducted to evaluate the effectiveness of the proposed approach. The experimental results show that our approach is effective and efficient in enhancing the transport operator profit and passenger satisfaction. The

main contributions of this paper can be summarized as follows:

1. We propose a novel approach that considers passenger experience as a Markov model, where prior experiences have a transient effect on future long-term satisfaction and retention rate. As such, we have formulated a multi-objective journey planning problem with individual passenger preferences, experiences, and memories. This is novel compared to most multimodal transport models that consider passenger experience as a non-time varying utility function.
2. As such, we are motivated to design a DRL-based approach to determine optimal utility weights by learning the passenger's past traveling experience and transport operator profit. This is a step up from heuristic optimization of non-time varying utility functions as it can balance short-term gains with long-term success across time-varying personalized passenger experiences.
3. A variant of the DRL-based approach is designed to balance the profit received by each transport operator based on proportional fairness (PF) reward scheme. This is a novel solution to balancing competing agendas between operators and address those naturally disadvantaged.
4. We conduct experiments using both real-world and synthetic datasets, and the results show that our proposed approach enhances passenger satisfaction and transport operator profit, while maintaining PF of profit across transport operators. These results demonstrate the effectiveness of our approach in addressing the complex multi-objective journey planning problem in MaaS.

The rest of this paper is organized as follows. Section 2 reviews the related work of MaaS, transportation factors, MDP problems, and the artificial intelligence approach in transportation. Section 3 illustrates the system model including the problem formulation, the multi-objective journey planning problem, and the MDP model of the passenger. The DRL algorithm is presented in Sect. 4. Experiments and results are presented in Sect. 5. Finally, Sect. 6 concludes this paper.

2 Related work

2.1 MaaS planner

MaaS is a rapidly growing and innovative mobility concept combining different kinds of transport services. Numerous studies have explored different characteristics of MaaS, including MaaS ecosystem [17, 18], bundling services [19, 20], suppliers [21, 22], pilot projects [23, 24], and simulations [25, 26]. Since MaaS is composed of multiple transport operators, providing a high-quality route over the multimodal transport network is crucial in MaaS research. To determine a route with a minimum travel time from an origin to a destination, a dynamic shortest path problem has formulated in [27] using both historical and real-time information. The problem is solved by a hybrid Approximate Dynamic Programming algorithm with a clustering approach that combines the value function approximation and a deterministic lookahead policy. In [28], the authors developed a routing strategy in schedule-based transit networks with stochastic vehicle arrival times using an online shortest path algorithm. In [29], the authors studied the system congestion effects on the route choice model where the link capacities are a function of flow instead of link cost. To capture the structural effects that flows have on capacities and the resulting impacts on route choice utilities, the authors proposed a method to obtain unique congestible capacity shadow prices in a multimodal network to verify the capability to capture congestion effects on capacities. In [30], the authors formulated a passenger-centric vehicle routing problem to maximize the quality of service in terms of waiting and riding time. In [31], a bus transit system planner was presented to optimize the passengers' experience by limiting the probability of collisions among passengers. However, these studies focus on improving the service by a single criterion, while the issue of accommodating passengers with different preferences and longitudinal time-varying experiences and memories remains unresolved. Therefore, a problem formulation considering the successive passenger traveling experience as a series of related states and actions could better model the passenger behavior.

2.2 Studies of transportation factors

There are several studies that have examined the relationships between various factors in transportation systems. For example, Molina et al. [32] investigated the connections between loyalty and passengers' profiles, experience, and values for taxi service and private-hire driver companies. The study suggested that passenger characteristics influence the choice of transportation mode, and younger passengers tend to be less loyal, more price-sensitive, and more concerned with sustainability. In [33], the authors analyzed the public acceptance of MaaS by investigating the intention to subscribe to MaaS and the willingness to pay for extra features of the service. The study found that the service attribute characteristics such as the price and the social influence variables have a significant effect on the subscription intention, and the transportation modes included in the bundle are related to socio-demographic

profiles and individual transportation-related characteristics. Another reference [34] that investigates the effect of individual passenger preferences supports the hypothesis of the passenger preference in the MaaS. The study implies significant heterogeneity with regard to preferences, which results in different MaaS package preferences and individual characteristics in the latent class choice model. The authors also suggested that age, gender, income, education, and current travel behavior are important factors influencing an individual's propensity to purchase MaaS packages. In [35], the authors analyzed satisfaction surveys and found that the higher passenger satisfaction has led to the increase in public transport demand. From the empirical analysis of passenger satisfaction data, reference [36] concluded that the passenger perceived quality of service, passenger expectations, passenger perceived value, and passenger loyalty all have a significant correlation to the passenger satisfaction. The authors in [37] identified five factors affecting passenger satisfaction and willingness to pay: emotional, economic, social, service, and functional values. All of these studies indicate the relationship between passenger behavior and preference, as well as the causality between retention rate and satisfaction level, which can be modeled as MDP.

2.3 MDP in transportation

MDP is widely used in modeling stochastic processes, especially the transport system that are full of uncertainty. Lautenbacher et al. [38] solved the single-leg airline revenue management problem by formulating it as a discrete-time MDP formulation. Hong et al. [39] applied the MDP formulation to the ordering and delivery problems with different transportation modes, costs, and inventory issues. In [40], the authors modeled the process of the passenger seeking a taxi as an MDP that finds the best move for a vacant taxi to maximize the total revenue. To understand driver behavior and driving decisions, MDP was used to analyze basic safety message data from vehicles [41], which discovered that the driver prefers to accelerate in order to escape the crowdedness around them. For motion planning for connected and automated vehicles, MDP can be used to incorporate network-level data and make decisions on platoon membership, lane changing, and route choice [42]. Traffic light control problem can be modeled as MDP to enhance the junction flow rate. Khamis et al. presented a multi-agent-based multi-objective traffic signal control that simulates the driver's behavior continuously in space and time dimensions based on Bayesian interpretation of probability [43] and RL [44].

2.4 Artificial intelligence in transportation

The success of artificial intelligence and deep learning in recent years accelerated research in various fields including transportation [45]. For example, deep neural networks can be used to accurately predict short-term traffic flow [46, 47]. Besides traffic flow, travel demand [48] and origin–destination pairs [49] can also be predicted by a deep learning model, called Multi-Scale Convolutional LSTM Network, which considers temporal and spatial correlations, and high-level prediction results of the historical traffic data. Lane detection is another example that uses deep neural networks [50]. Among various artificial intelligence technologies, DRL is a sub-field that focuses on complex control problems [51], especially for those with MDP characteristics. For example, the capacitated vehicle routing problem was solved by the RL-based method in [52], which outperformed existing heuristics and optimization tools on medium-sized instances. In [53], routing strategies determined by a trained DRL agent were treated as the initial solution for a local search method to refine the solution quality. In autonomous driving, the RL-based strategy used the shared information to improve travel efficiency, fuel consumption, and safety at traffic intersections [54]. RL would also be helpful in situations where multiple autonomous vehicles coordinate the driving maneuvers with each other [55]. In the traffic signal control problem, DRL can intake images from road surveillance cameras to efficiently control the signal duration in an end-to-end fashion [56]. Using multi-agent DRL, the problem size can be extended to city-scale [57]. For the lane change problem, although rule-based models may perform well in known scenarios, controlling in an unforeseen scenario could be prone to failure. A RL-based controller can perform lane changes under complex and unforeseen scenarios [58, 59]. By considering the planning decision as the action in DRL, multi-modal journey planning can be solved by DRL [60, 61]. Although these works have not incorporated passenger behavior, they serve as a great pilot study for solving a multi-modal journey planning problem with DRL. All the above literature indicates that DRL has excellent potential to deal with the highly dynamic transportation systems filled with uncertainty.

2.5 Research gap

From the existing works, we learn that multimodal transport, including MaaS, is an actively studied mobility concept for future intelligent transportation systems. However, current journey planning optimization methods treat passengers with the same origin and destination as identical passengers and non-time varying single experience

problem, which is incapable of modeling passenger preferences, experience, and memories over many journeys over time. Based on this observed property, MDP appears to be an ideal model for the aforementioned heterogeneous and dynamic process. Among the state-of-the-art technologies, deep reinforcement learning is widely used to solve MDP problems and could be integrated into the journey planning optimization methods to address these challenges.

3 System model

In this section, we first introduce the model and problem formulation of multimodal transport and then discuss the MDP behavior of the passenger and the effect on the model. The notations used in this section are summarized in Table 1.[1]

3.1 Problem formulation

MaaS consists of multiple transport operators and the transport operators may operate different transport services for the same route. Indirect or detours may be taken if the journey fits the passenger's expectation in terms of the utility. For instance, passengers may purchase transit flight tickets to obtain a cheaper price, even if it results in a longer flight duration. The transport network can be modeled by a directed graph $G(\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ and $\mathcal{A}$ is the set of nodes and links in the network, respectively. For example, a node represents the staring or ending point of a mobility service, while a link represents the journey of a mobility service. Let $\mathcal{F}$ be the set of transport operators and each operator $f \in \mathcal{F}$ manages their own sub-network $\mathcal{A}_f \in \mathcal{A}$. Each link (i, j) that provides transport service from i to j is associated with a time cost[2] β_{ij}^f, discomfort cost δ_{ij}^f, price ρ_{ij}^f, and operation cost μ_{ij}^f, as managed by the operator f. When a passenger requests transport service from origin $O(k)$ to destination $D(k)$, an intelligent scheduler determines a planned journey that satisfies the transport requirements. The journey includes services from different transport operators f that are offered to the passenger.

We first define two decision variables x_{ij}^k and y_{ij}^f to facilitate the formulation of the problem. Binary variables

Table 1 Notation summary

Notations	Meaning
$G(\mathcal{N}, \mathcal{A})$	Graph of the transport network
$\mathcal{F}$	Set of transport operators
f	Transport operator $f \in \mathcal{F}$
$\mathcal{N}$	Set of nodes in the network
$\mathcal{A}$	Set of links in the network
$\mathcal{A}_f$	Subset of $\mathcal{A}$ operated by $f \in \mathcal{F}$
$\mathcal{N}^+(i)$	Set of incoming locations of i
$\mathcal{N}^-(i)$	Set of outgoing locations of i
$\mathcal{K}$	Set of passengers
k	Passenger k
$O(k)$	Origin node of passenger $k \in \mathcal{K}$
$D(k)$	Destination node of passenger $k \in \mathcal{K}$
U_{ij}^k	Utility vector of $k \in \mathcal{K}$ for link $(i,j) \in \tilde{\mathcal{A}}_k$
U_{ij}^f	Utility vector of $f \in \mathcal{F}$ for link $(i,j) \in \mathcal{A}_f$
β_{ij}^f	Travel time cost of link $(i,j) \in \mathcal{A}_f$
δ_{ij}^f	Discomfort measures of link $(i,j) \in \mathcal{A}_f$
ρ_{ij}^f	Price of link $(i,j) \in \mathcal{A}_f$
μ_{ij}^f	Operating cost of link $(i,j) \in \mathcal{A}_f$
W^k	Utility weight vector of passenger k
w_β^k	Weight of β_{ij}^f
w_δ^k	Weight of δ_{ij}^f
w_ρ^k	Weight of ρ_{ij}^f
w_μ^f	Weight of μ_{ij}^f
C_{ij}^f	Capacity of link $(i,j) \in \mathcal{A}_f$
x_{ij}^k	Binary variable for link $(i,j) \in \tilde{\mathcal{A}}_s$ that recommends to passenger $k \in \mathcal{K}$
y_{ij}^f	Binary variable for an operator to serve a link $(i,j) \in \mathcal{A}_f$

x_{ij}^k are used to indicate which transport service will be offered to the passenger:

$$x_{ij}^k = \begin{cases} 1 & \text{if link } (i,j) \text{ offers to } k, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Binary variables y_{ij}^f are defined for the transport link operation:

$$y_{ij}^f = \begin{cases} 1 & \text{if link } (i,j) \text{ is operated by } f, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Let U_{ij}^k and U_{ij}^f be the utility vector of passengers and transport operators, respectively. The objective function is to minimize the total utility cost of passengers and transport operators as

[1] Many of these parameters vary with time or discrete steps, but this is not written for simplicity.

[2] In reality, travel time may include waiting, in-vehicle, and transfer time. We use a single parameter to represent all the time cost for simplicity.

$$\sum_{(i,j)\in\mathcal{A}_f,k\in\mathcal{K}} U_{ij}^k x_{ij}^k + \sum_{(i,j)\in\mathcal{A}_f f\in\mathcal{F}} U_{ij}^f y_{ij}^f. \tag{3}$$

In this paper, we consider four utility costs formulated as

$$U_{ij}^k = [w_\beta^k \beta_{ij}^f; w_\delta^k \delta_{ij}^f; w_\rho^k \rho_{ij}^f], \tag{4}$$

$$U_{ij}^f = [w_\mu^f \mu_{ij}^f], \tag{5}$$

where β_{ij}^f, δ_{ij}^f, ρ_{ij}^f, and μ_{ij}^f are the travel time, discomfort, price, and operating cost of link $(i,j) \in \mathcal{A}_f$, respectively. $W^k = [w_\beta^k; w_\delta^k; w_\rho^k; w_\mu^f]$ are the weighting of the corresponding utility cost terms. Without loss of generality, more utility costs such as carbon emissions can be formulated and this paper considers four utility cost only for simplicity.

Let $\mathcal{N}^+(i)$ and $\mathcal{N}^-(i)$ be the sets of incoming and outgoing locations of i, respectively, i.e., $\mathcal{N}^+(i) = \{j \in \mathcal{N} | (j,i) \in \mathcal{A}_f\}$ and $\mathcal{N}^-(i) = \{j \in \mathcal{N} | (i,j) \in \mathcal{A}_f\}$. The following equation ensures the feasibility of flow in the network:

$$\sum_{j\in\mathcal{N}^-(i)} x_{ij}^k - \sum_{j\in\mathcal{N}^+(i)} x_{ji}^k = \begin{cases} 1 & \text{if } i = O(k), \\ -1 & \text{if } i = D(k), \forall i \in \mathcal{N}, k \in \mathcal{K} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

We have to ensure the transport capacity constraint is not violated in the problem:

$$\sum_{k\in\mathcal{K}} x_{ij}^k \le C_{ij}^f y_{ij}^f, \quad \forall (i,j) \in \mathcal{A}_f, f \in \mathcal{F} \tag{7}$$

where C_{ij}^f is the capacity of transport link $(i,j) \in \mathcal{A}_f$ operated by $f \in \mathcal{F}$.

Therefore, the multi-objective journey planning problem is formulated as follows:

Problem 1 (Multi-objective Journey Planning Problem)

minimize (3)
subject to (6)-(7)

The problem is an integer linear program. To solve this problem, the scheduler must coordinate with the transport operators and obtain information required in the problem such as transport link details. However, the utility weight vector W^k is an abstract term that cannot be easily quantified. For example, a passenger may not explicitly state that they prefer a journey with a combination of 40% travel time, 20% discomfort, and 40% price. Additionally, passenger behavior and expectation are influenced by their past experience which can be modeled as MDP. Too address these challenges, we propose using DRL (to be discussed in Sect. 4), which has the ability to learn from the passenger status and infer their expectation. Once the W^k vector is determined, the scheduler can solve Problem 1 using a conventional integer programming solver or heuristic.

3.2 Markov decision process

The transport journey and retention of a passenger are modeled as a 4-tuple Markov decision process $\langle \mathcal{S}, \mathcal{A}, P, R \rangle$, where $\mathcal{S}$ and $\mathcal{A}$ are the set of states and actions, respectively. $P(s_{t+1}|s_t, a_t)$ represents the state transition probability from States $s_t \in \mathcal{S}$ to $s_{t+1} \in \mathcal{S}$ for performing Action $a_t \in \mathcal{A}$. $R(s_t, a_t, s_{t+1})$ is the received reward due to the transition from s_t to s_{t+1} after performing a_t. In particular, the passenger satisfaction level in the State is a N-level integer value indicating the retention rate. In general, passengers tend to retain to the system if they are satisfied with it. Therefore, satisfaction is proportional to the retention rate. A sample relationship is shown in Table 2.

The satisfaction level of passenger k, H^k, can be changed after each journey offered to the passenger. For example, if the offered journey matches the passenger expectation, passenger satisfaction increases by n level. If the service is far below what the passenger is expecting, say more expensive than usual, the satisfaction decreases and thus reducing the retention rate of the journey. The expectation difference for passenger k depends on the expected and actual utility, which is defined as

$$E^k = \tilde{w}_\beta^k(\tilde{\beta}_{od}^k - \beta_{od}^k) + \tilde{w}_\delta^k(\tilde{\delta}_{od}^k - \delta_{od}^k) + \tilde{w}_\rho^k(\tilde{\rho}_{od}^k - \rho_{od}^k), \tag{8}$$

where $\tilde{W}^k = [\tilde{w}_\beta^k; \tilde{w}_\delta^k; \tilde{w}_\rho^k]$ is the actual weighting of passenger, $\tilde{\beta}_{od}^k, \tilde{\delta}_{od}^k, \tilde{\rho}_{od}^k$ are the utility expected by the passenger, o and d are the origin and destination, respectively. Expected utility is the utility of the ideal best journey the passenger can get. It can be determined by solving the planning problem as if the passenger is alone in the system. The actual utility is the utility of the journey planned by the scheduler, which may be deviated from the expected utility

Table 2 Sample relationship between satisfaction level and retention rate

Satisfaction level, H^k	Retention rate, $P(H^k)$
5	1
4	0.75
3	0.5
2	0.25
1	0

with an incompetent planner and limited capacity. Therefore, an incompetent planner may have a negatively large expectation difference on average.

The satisfaction function modeling the change in satisfaction can be expressed as

$$H^k := \begin{cases} H^k + n & \text{if } E^k \geq \overline{E}^k, \\ H^k - n & \text{if } E^k \leq \underline{E}^k, \\ H^k & \text{otherwise}, \end{cases} \tag{9}$$

where $\overline{E}^k$ and $\underline{E}^k$ are the upper and lower expectation threshold, and n is the step change of the satisfaction level. As a whole, a sample state diagram of $n = 1$ is summarized in Fig. 1. Similar customer satisfaction models can be found in supply chain [62] and product management [63].

4 Deep reinforcement learning with proportional fairness

In this section, we present the algorithm of the DRL agent and the corresponding components interacting with the agent. A typical interaction framework between the DRL agent and the environment is adopted as shown in Fig. 2. When the MaaS scheduler receives a journey query, the trained DRL agent determines the output actions for the journey planner. Then, the journey planner solves problem 1 to determine the optimal journey. The journey is offered to the passenger who may take the suggestion based on the retention rate. The process repeats for another iteration when the passenger submits a new journey query.

4.1 Environment

The environment represents the set of passengers and the transportation system. At each time t, the state s_t of a set of passengers is transmitted to the agent. Based on the action a_t received from the agent, an optimal journey planning problem is solved for the passengers as presented in Sect. 3.1. Note that the time t is not representing a fixed time interval as in other commonly seen RL problems. The time interval in our model depends on when the consumer chooses to query for their next journey, and the agent performs journey-to-journey sequential actions. The transport service taken by the passenger affect the next state s_{t+1}, which can be denoted by transition probability $P(s_{t+1}|s_t, a_t)$.

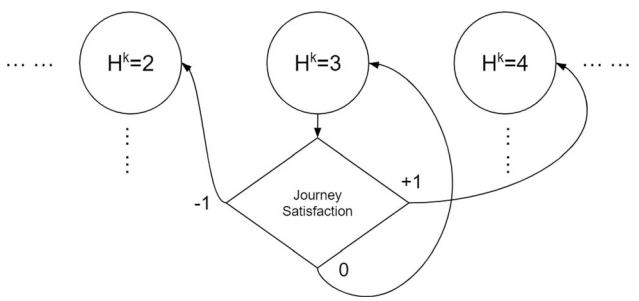


Fig. 1 Satisfaction transition representation

4.1.1 State

The state s_t of a passenger is the satisfaction to the system together with the passenger's characteristics such as income and age. Satisfaction to the system is dynamic which follows the MDP as discussed in Sect. 3.2, while the characteristics are static for each passenger. In this paper, we assume the characteristics of the passengers are the factors affecting their travel expectations for simplicity. Nevertheless, other characteristics can be included in the state without loss of generality.

4.1.2 Reward

The profit is set to be the reward of the transport operators, which is equal to the price minus the operation cost of all passengers, given by Eq. (10).

$$\sum_{k,f} (\rho^{kf} - \mu^{kf}). \tag{10}$$

The agent's objective is to maximize this reward function by selecting an optimal action. Here, we introduce the concept of proportional fairness into the decision-making process of the agent. PF is a well-known principle in resource allocation that states that the gain of one party should be greater than the sum of the losses of all other parties together [64, 65]. To ensure PF across the transport operators, according to [66], a PF variation of the reward function can be formulated as

$$\sum_{f} w_{PF}^{f} \log \sum_{k} (\rho^{kf} - \mu^{kf}), \tag{11}$$

where w_{PF}^{f} is the proportional weighting factor associated with each transport operator, which can be simplified as a common constant. This PF variation reward function is expected to ensure proportional fairness of profit among the transport operators. A naive approach to increase profit would be to increase the price and decrease the cost in the utility weight vector. However, simply increasing the price and lowering the cost may offend the passenger and lower their satisfaction. According to the MDP, passengers with a lower satisfaction have a lower chance to return to the system. Thus, the agent should learn to increase the passenger satisfaction by selecting a reasonable utility weight vector in order to increase the long-term profit.

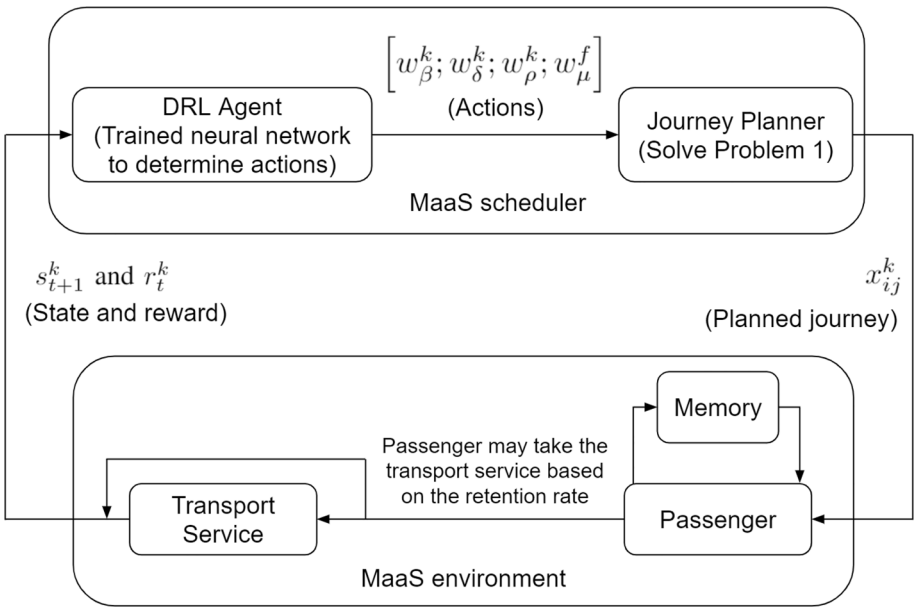


Fig. 2 Interaction between the environment and the scheduler

4.2 DRL agent

The DRL agent refers to the component of the intelligent scheduler in the system, which determines the utility weight vector. Based on the given state s_t from the passenger, the agent computes an optimal utility weight vector that matches the passenger expectation of the journey in order to increase passenger satisfaction and the retention rate. For example, an expensive first-class itinerary versus an economy trip; a comfortable detour journey versus a crowded direct route. The route is determined by solving the multi-objective journey planning problem in Sect. 3.1 together with the given action (utility weight vector) produced from the DRL agent.

4.2.1 Action

The action is the utility weight vector with a dimension equal to the number of objective function terms. For example, the problem formulation presented in Sect. 3.1 contains four objectives: time, discomfort, price, and operation cost, and thus, the action is a four dimension utility weight vector indicating the weighting of each objective to be solved. Without loss of generality, the aforementioned four dimension utility weight vector is used in the experiments for simplicity. Ones can add more utilities to the action and problem formulation such as carbon emission.

4.2.2 Neural network

Neural networks are used to learn and perform the transition functions of states, actions, and rewards. To preserve the generality of the proposed approach, we use deep fully connected neural networks as the neural network structure in this paper. The reason is that the input and output of the neural network are numeric features and values where simple fully connected neural networks might be able to model the relationship already, unlike other data types such as images and time series data that usually use convolutional neural networks and recurrent neural network, respectively. We are fully aware that many state-of-the-art neural networks structure can replace the fully connected neural network, but the neural network structure is out of the scope of this paper. Nonetheless, the deep neural networks have to be designed to match the problem nature. For example, the action is an array ranging from 0 to 1, and thus, we use the *sigmoid* as the activation function.

4.3 Deep reinforcement learning algorithm

Our DRL algorithm is modified based on the deep deterministic policy gradient (DDPG) [67] which is a model-free off-policy algorithm for continuous control problems. The algorithm is an actor-critic approach where an actor function $\pi(s|\theta^\pi)$ is responsible for deterministically performing an action based on a given state, and a critic function $Q(s, a)$ learns the Q-value of the state and action pair following the Bellman equation.

A replay buffer R is used to store the transition tuple (s_t, a_t, r_t, s_{t+1}) for sampling minibatch to update the actor and critic. The motivation for using a replay buffer is that the training of neural networks usually assumes the samples are independently and identically distributed, which cannot be achieved if the updates are based on sequential transitions in reinforcement learning.

In the classic DDPG, Ornstein-Uhlenbeck (OU) noise is added to the action for exploration. However, in our scenario, the action space is a vector with values ranging from 0 to 1. Adding the OU noise to the action does not guarantee the range and thus we use the epsilon-greedy action selection method instead of OU noise for exploration. A uniformly random action vector is generated under the probability defined by ϵ.

There are four function approximators, namely actor local, actor target, critic local, and critic target network. The update rules of the four function approximators are given as follows as presented in [67]. Critic local θ^Q is updated based on the loss function:

$$L = \frac{1}{B}\sum_i (r_i + \gamma Q'(s_{i+1}, \pi'(s_{i+1}|\theta^{\pi'})|\theta^{Q'}) - Q(s_i, a_i|\theta^Q))^2 \tag{12}$$

where B is the minibatch size, i is the index of sample in the minibatch, and γ is the discount factor. Actor local θ^π is updated using policy gradient:

$$\nabla_{\theta\pi} J \approx \frac{1}{B}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\pi(s_i)} \nabla_{\theta\pi}\pi(s|\theta^\pi)|_{s_i} \tag{13}$$

The use of local and target networks is a technique to stabilize the training. In each iteration, the parameters of local networks are updated using the aforementioned rules. For the target networks, they are updated by copying the parameters from the local networks softly with a scale of τ. For critic target $\theta^{Q'}$:

$$\theta^{Q'} := \tau\theta^Q + (1-\tau)\theta^{Q'}. \tag{14}$$

Similarly, for actor target $\theta^{\pi'}$:

$$\theta^{\pi'} := \tau\theta^\pi + (1-\tau)\theta^{\pi'}. \tag{15}$$

Detailed algorithm is shown in algorithm 1.

Algorithm 1 DRL algorithm for utility weights

Initialize actor local $\pi(s|\theta^\pi)$ and critic local networks $Q(s, a|\theta^Q)$ with parameters θ^π and θ^Q
Initialize parameters of actor target $\pi'(s|\theta^{\pi'})$ and critic target networks $Q'(s, a|\theta^{Q'})$ with parameters $\theta^{\pi'} \leftarrow \theta^\pi$ and $\theta^{Q'} \leftarrow \theta^Q$
for episode $m = 1$ to M **do**
  Initialize passengers' parameters
  for iteration $t = 1$ to T **do**
    for passenger $k = 1$ to $|\mathcal{K}^m|$ **do**
      $j^k \leftarrow$ uniform random number between 0 and 1
      if $j^k < \epsilon$ **then**
        $a_t^k \leftarrow$ uniform random vector between 0 to 1
      else
        $a_t^k \leftarrow \pi(s_t^k|\theta^\pi)$
      end if
    end for
    Execute action $a_t = [a_t^1, \ldots, a_t^{|\mathcal{K}|}]$ to obtain new state $s_{t+1} = [s_{t+1}^1, \ldots, s_{t+1}^{|\mathcal{K}|}]$ and reward r_t
    Store transition $(s_t^k, a_t^k, r_t^k, s_{t+1}^k)$ for all k to replay buffer R
    if transitions in $R \geq$ minibatch size B **then**
      Sample a random minibatch of B transitions $(s_i^k, a_i^k, r_i^k, s_{i+1}^k)$ from R
      Update critic local θ^Q using loss function: $L = \frac{1}{B}\sum_i (r_i^k + \gamma Q'(s_{i+1}^k, \pi'(s_{i+1}^k|\theta^{\pi'})|\theta^{Q'}) - Q(s_i^k, a_i^k|\theta^Q))^2$
      Update actor local θ^π using policy gradient: $\nabla_{\theta\pi} J \approx \frac{1}{B}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i^k, a=\pi(s_i^k)} \nabla_{\theta\pi}\pi(s|\theta^\pi)|_{s_i^k}$
      Update critic target $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
      Update actor target $\theta^{\pi'} \leftarrow \tau\theta^\pi + (1-\tau)\theta^{\pi'}$
    end if
  end for
  $\epsilon_{m+1} = \epsilon_m\bar{\epsilon}$
end for
return actor parameters θ^π

5 Experiments

5.1 Experiment setup

We conduct experiments in two scenarios: New York City (NYC) and a synthetic scenario to simulate the multimodal transport system and passenger behaviors.

5.1.1 New York City scenario

To evaluate our proposed approach, we use a real-world transportation network and datasets of NYC. The transportation network is extracted from the Manhattan region of NYC based on the taxi zone maps.[3] Each zone represents a node in the network, and we assume there is an edge between two nodes if they are connected on the map. Isolated regions without connection are ignored. The graph contains 63 nodes in an irregular shape. We assume 3 transport operators on each connection, resulting in 3 edges between the connected nodes, composing a total of 963 edges. Each edge is associated with time, discomfort, price, and operation cost. The values are randomly generated between 0 and 1, except that the price must be larger than the operation cost for each edge.

A set of passengers $\mathcal{K}$ is constructed by open datasets of NYC. We sample the traffic query and passengers' characteristics from the NYC Taxi and Limousine Commission Trip Record Data[4] and Citywide Mobility Survey,[5] respectively. The expected utility weight vector is calculated based only on the characteristics for Eq. (8), which is unknown to the intelligent scheduler throughout the experiments. All initial passenger satisfaction levels are at level 3. The multi-objective journey planning problem with a given utility weight vector from the DRL agent is solved by a standard optimizer in CVXPY [68].

5.1.2 Synthetic scenario

We also evaluate our proposed approach with a synthetic system. We use a square grid with 36 nodes to represent the transportation system, where each node is a stop or transit to other transport, and each edge is a transport sub-journey. Between any two neighboring nodes, 6 edges, each representing a transport operator, are set, resulting in a total of 360 edges for the 36 nodes. The network costs are randomly generated, as in the NYC scenario.

A set of passengers $\mathcal{K}$ is generated with the random origin, destination, and characteristics. The calculation of the expected utility weight vector, initial passenger satisfaction levels, and multi-objective journey planning problem-solving process is the same as the NYC scenario. Table 3 summarized the parameters used in the experiments.

5.1.3 Benchmarks

The main purpose of the DRL approach is to determine an utility weight vector for each passenger. Hence, two benchmark approaches for determining the utility weight vector are used to compare with the DRL approach, namely "fixed" and "random" policy. The utility weight vector of the former benchmark is fixed to ones and that of the latter benchmark is uniformly random vector between 0 and 1.

5.2 Profit

We test the transport operator profit received using the proposed and benchmark approaches in both NYC and synthetic scenarios. As discussed in Sect. 4.1.2, the rewards received from the environment are equivalent to the profit as in Eq (10), or Eq (11) for PF.

5.2.1 NYC scenario

The average rewards of 2000 episodes for the NYC scenario are shown in Table 4. Among the compared approaches, the DRL approach performs the best in terms of obtaining the highest rewards (profits), with an average of 336.48 units per episode. The DRL with PF is the second highest, with an average of 287.47 units per episode. The random policy is the lowest, resulting in 171.20 units per episode. The fixed utility weight vector policy is between random and DRL agents, obtaining an average reward of 223.55 units per episode. We noted that there is a drop in reward for DRL with PF. This is expected since the PF tried to ensure that the sum of proportional changes among transport operators is negative but not the highest reward in total, which can be considered an additional constraint to the maximization problem. The fairness of transport operators will be discussed in Sect. 5.5.

We are also interested in the variation of the approaches throughout the episode. To mitigate the fluctuation due to the randomness and see the trends clearly, we calculated the moving average of reward with a time window equal to 20 episodes and plotted the approaches in Fig. 3. In general, the DRL approach performs the best in all episodes. The DRL with PF is slightly lower than the DRL approach. Similarly, the random policy is the worst, and the fixed utility weight vector policy is between random and the

[3] https://data.cityofnewyork.us/Transportation/NYC-Taxi-Zones/d3c5-ddgc.

[4] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

[5] https://www1.nyc.gov/html/dot/html/about/citywide-mobility-survey.shtml.

Table 3 Parameter settings

Parameter	Definition	Value
NYC transportation network and datasets		
$\|\mathcal{N}\|$	Number of nodes	63
$\|\mathcal{A}\|$	Number of links	963
$\|\mathcal{F}\|$	Number of transport operators	3
Synthetic transportation network and datasets		
$\|\mathcal{N}\|$	Number of nodes	36
$\|\mathcal{A}\|$	Number of links	360
$\|\mathcal{F}\|$	Number of transport operators	6
$\|\mathcal{K}^m\|$	Number of passengers per episode	10
C_{ij}^f	Capacity	3
$\|\mathcal{R}\|$	Replay buffer size	10^6
B	Minibatch size	128
γ	Discount factor	0.99
τ	Target network soft update rate	0.001
–	Actor learning rate	0.0001
–	Critic learning rate	0.0003
–	Neural network optimizer	Adam
ϵ_0	Initial random explore rate	1
$\bar{\epsilon}$	Explore rate decay per episode	0.995
T	Number of iteration per episode	100
M	Number of episodes	2000
–	Number of neural network layers	3
–	Number of neurons of each layer	256
–	Range of satisfaction level	1 to 5
$\overline{E}^k$	Upper expectation threshold	0.0
$\underline{E}^k$	Lower expectation threshold	−0.1

Table 4 Average reward of approaches in the NYC scenario

Approach	Average reward (profit)
DRL with PF	287.47
DRL agent	336.48
Fixed utility weight vector	223.55
Random utility weight vector	171.20

DRL approaches for all episodes. A clear increasing trend is observed for the DRL agents. It starts from a low reward at the beginning of the episode and then increases with the training episode until it converges as expected. The agent behaves like this for two reasons. First, the ϵ value is high at the beginning, which means random policy dominates the policy of the DRL agent, and thus, the performance is similar to random policy at the beginning. Second, the initial random policy provides a replay experience for the DRL agent to explore and learn. In later episodes when the ϵ value gradually decays along the episode, the actions are mainly performed by the agent, which is helpful for further exploitation.

5.2.2 Synthetic scenario

The average rewards of 2000 episodes for the synthetic scenario are shown in Table 5. Similar to the NYC scenario, among the compared approaches, the DRL agent performs the best in terms of obtaining the highest rewards on an average of 376.17 units per episode. The DRL with PF is the second-highest, with an average of 363.41 units per episode. The random policy is the lowest, resulting in 111.58 units per episode. The fixed utility weight vector policy is between random and DRL approaches, obtaining an average reward of 164.32 units per episode.

The moving average of the reward with a time window of 20 episodes is plotted in Fig. 4. Similar to the results in the NYC scenario, the DRL approaches outperform all other approaches in all episodes. The random policy performs the worst, and the fixed utility weight vector policy is between the random policy and the DRL agent for all episodes. For both DRL agents, a clear increasing trend is observed, as in the NYC scenario.

We can observe similar results in both scenarios. Therefore, the proposed approach can successfully maximize the profit of the transport operators when compared to the benchmarks.

5.3 Passenger satisfaction

We study the passenger satisfaction level in both scenarios to understand the performance of each approach from the passengers' perspective. For the simulated 2000 episodes and the 100 iterations in each episode, we plotted the average satisfaction level of each episode, each iteration, the total number of the satisfaction levels of each method, and the average satisfaction level against the number of nodes in Figs. 5, 6, 7 and 8, respectively, for NYC scenario. The same plots for the synthetic scenario are shown in Figs. 9, 10, 11 and 12. The color in each figure corresponds to the satisfaction level, as indicated by the color bar on the right side of the plots. Recall that a higher the satisfaction level H^k indicates better service for the passenger.

5.3.1 NYC scenario

Figures 5 and 6 illustrate the variation of satisfaction levels across iterations and episodes, respectively. In Fig. 5, the

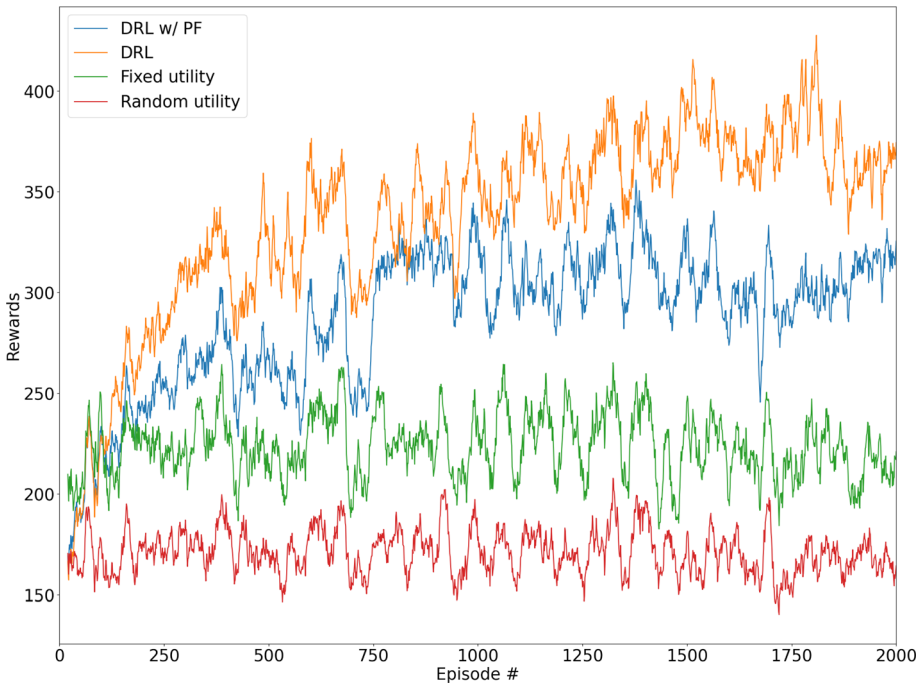


Fig. 3 Moving average reward of NYC scenario. The time window of moving average is equal to 20

Table 5 Average reward of approaches in the synthetic scenario

Approach	Average reward (profit)
DRL with PF	363.41
DRL agent	376.17
Fixed utility weight vector	164.32
Random utility weight vector	111.58

average satisfaction level of the DRL agent is the highest across the iterations compared to other approaches. DRL with PF is slightly lower than DRL. Since the satisfaction level shown is the average of 2000 episodes, the fluctuation is mitigated in the figure and shows a clear distinction between the approaches. This suggests that the methods converge to certain satisfaction levels across iterations. Figure 6 shows the average satisfaction level across

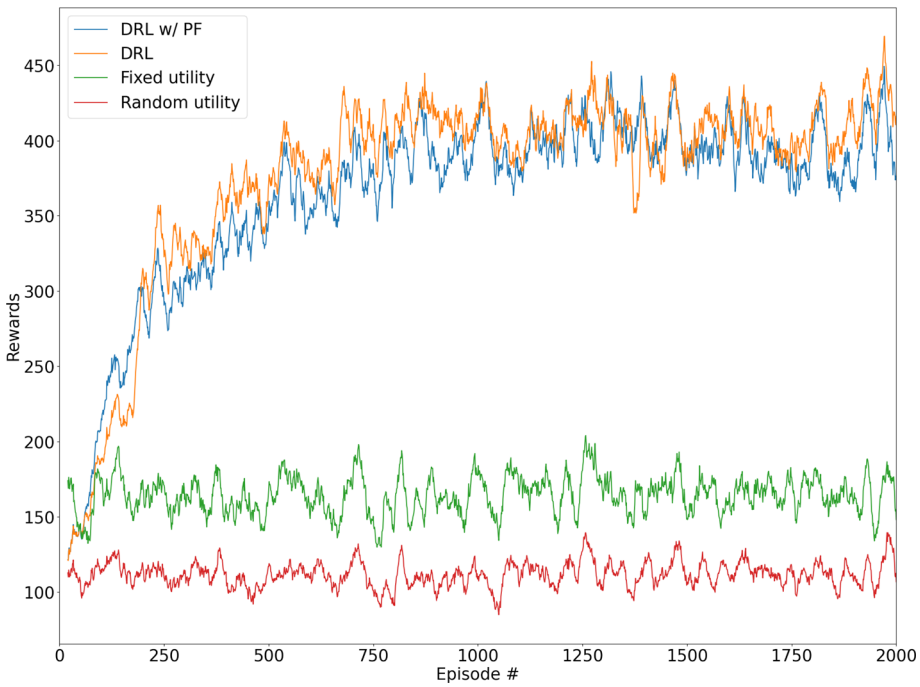


Fig. 4 Moving average reward of synthetic scenario. The time window of moving average is equal to 20

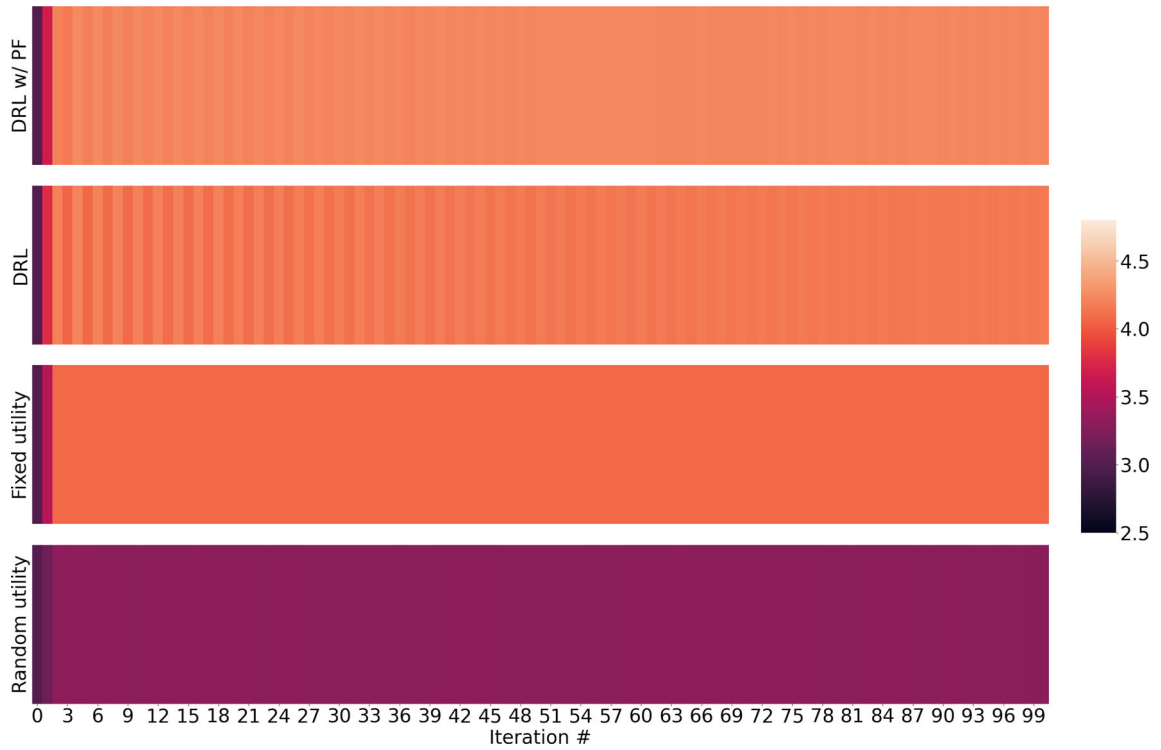


Fig. 5 Average satisfaction level of each iteration of NYC scenario

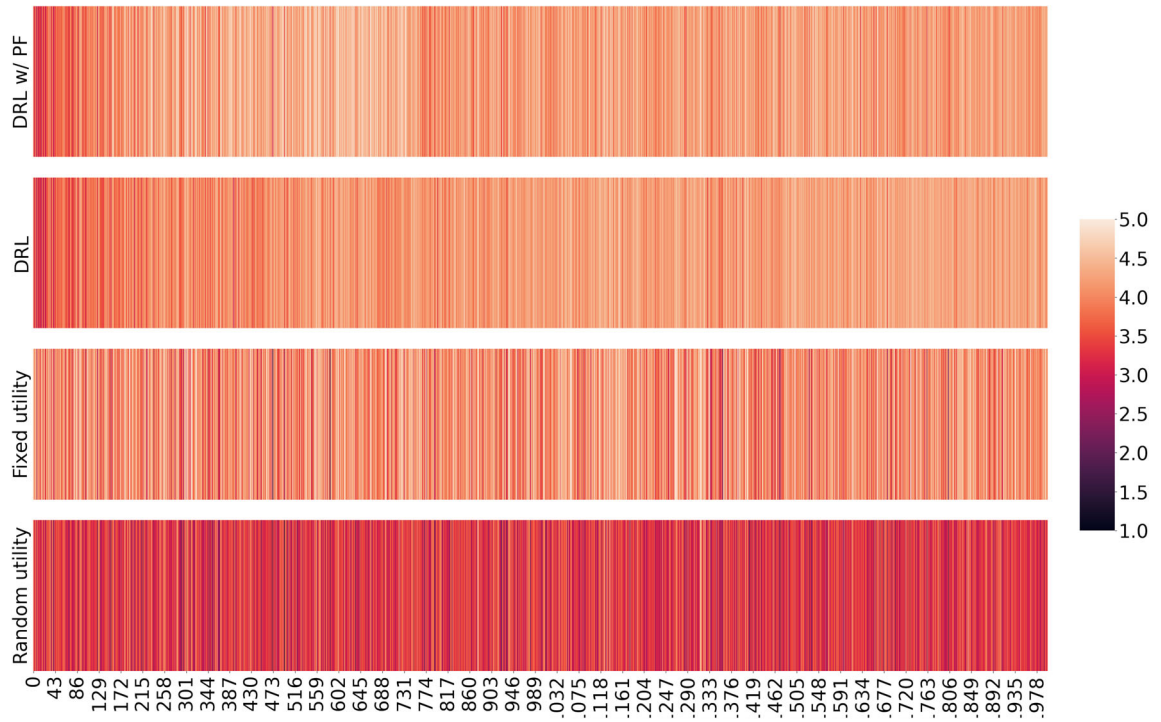


Fig. 6 Average satisfaction level of each episode of NYC scenario

episodes, which fluctuates more compared to Fig. 5. Nonetheless, we can still observe that the satisfaction level of DRL and DRL with PF is higher than those of the fixed and random policies. The fluctuation may be caused by the random origin, destination, and characteristics of the passengers across episodes.

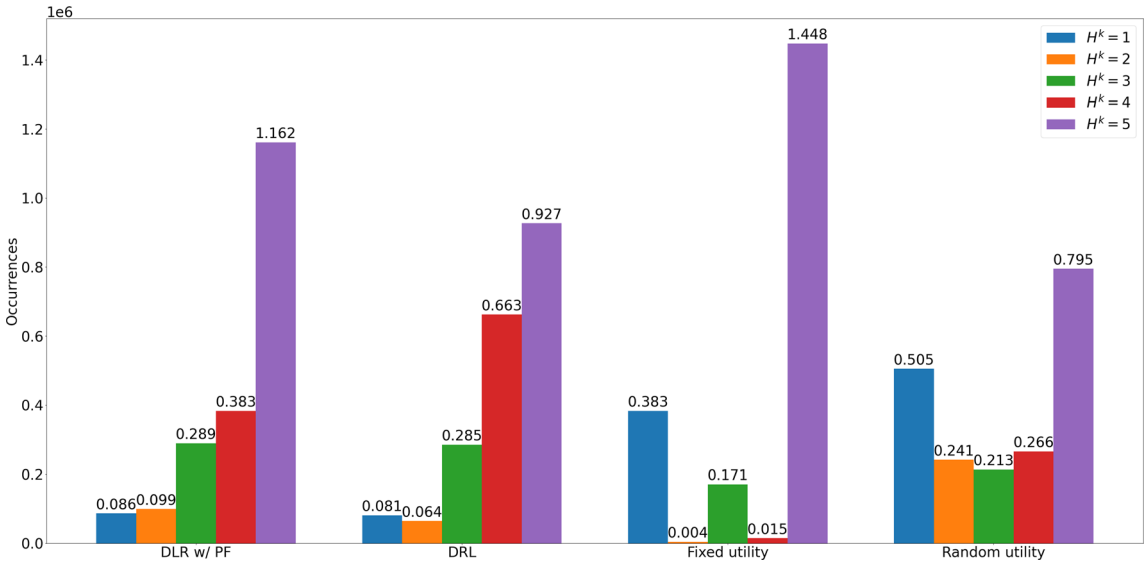


Fig. 7 Total number of satisfaction level of each method of NYC scenario

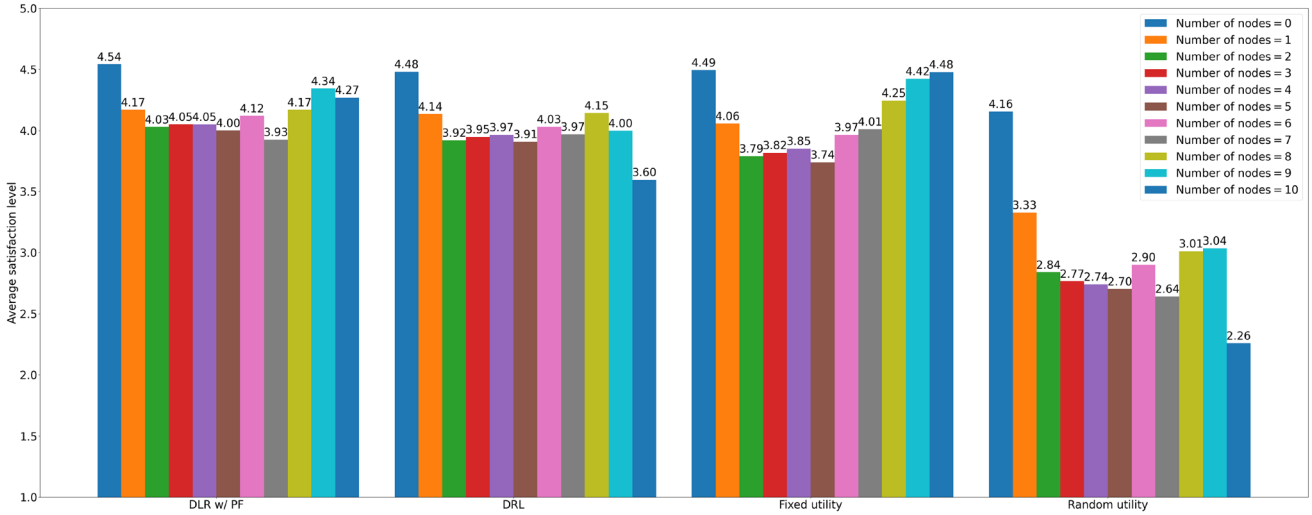


Fig. 8 Average satisfaction level by number of nodes of NYC scenario

5.3.2 Synthetic scenario

The observation of synthetic scenario results is similar to those of the NYC scenario. In Fig. 9, the average satisfaction level of both DRL approaches is the highest across the iterations compared to fixed and random policies. Figure 10 shows the average satisfaction level across episodes, which fluctuates more compared to Fig. 9. Nevertheless, we can still observe that the satisfaction level of both DRL methods is higher than those of the fixed and random policies.

5.4 Interpreting the satisfaction level

5.4.1 NYC scenario

In Fig. 7, we can observe that most of the satisfaction levels are high in both DRL methods compared to fixed and random, which means the agent can offer journeys that meet passengers' preferences. An notable observation about the fixed policy is that the satisfaction level vary widely in levels 1 and 5. In other words, passengers' satisfaction tends to converge toward either the lowest or the highest level when using the fixed utility policy. This is because the fixed utility policy is static and cannot adapt to the preferences of different passengers, as is the case with a static policy that has

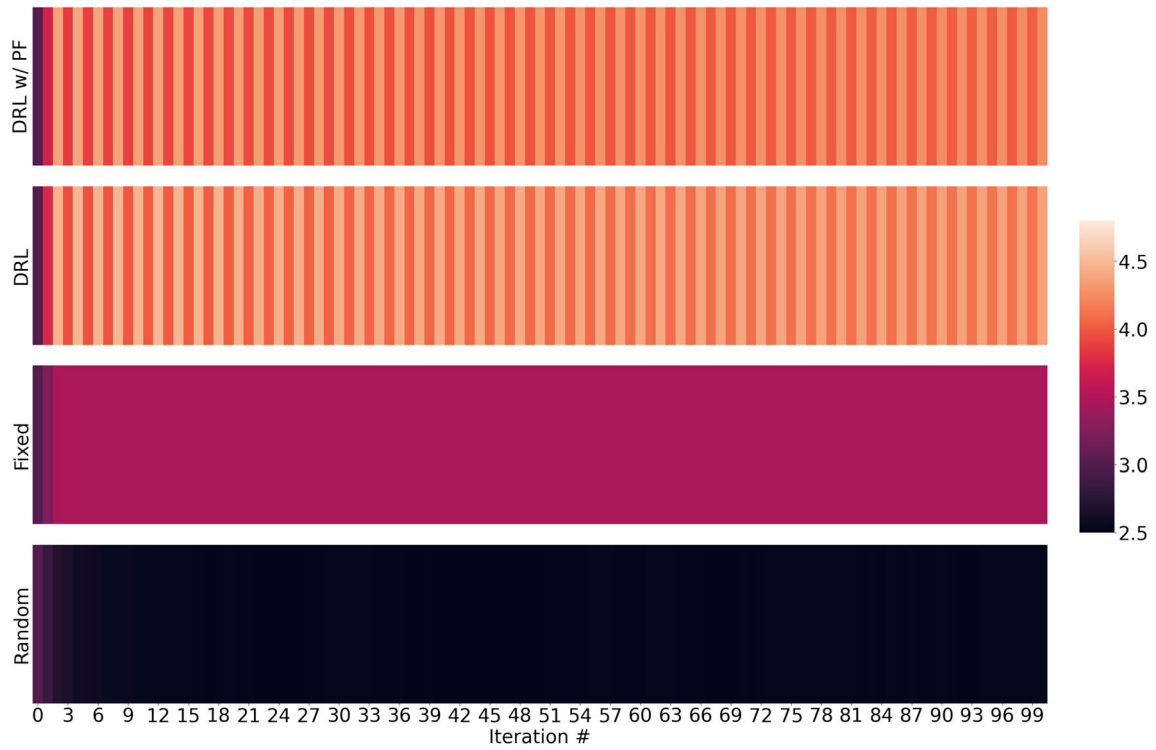


Fig. 9 Average satisfaction level of each iteration of synthetic scenario

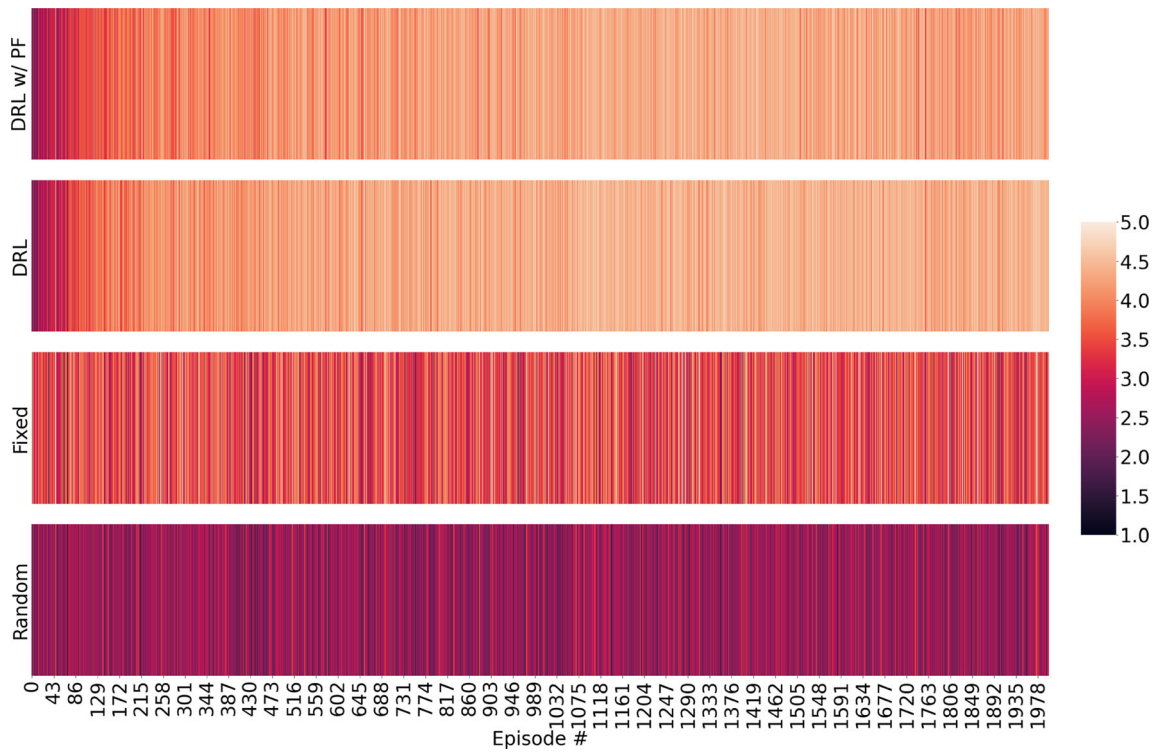


Fig. 10 Average satisfaction level of each episode of synthetic scenario

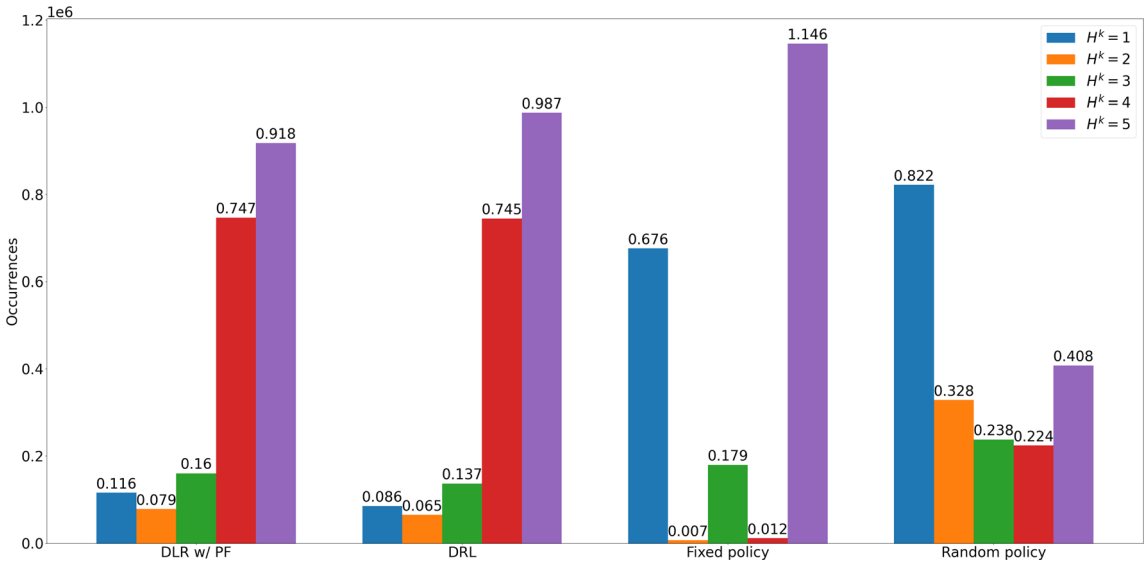


Fig. 11 Total number of satisfaction level of each method of synthetic scenario

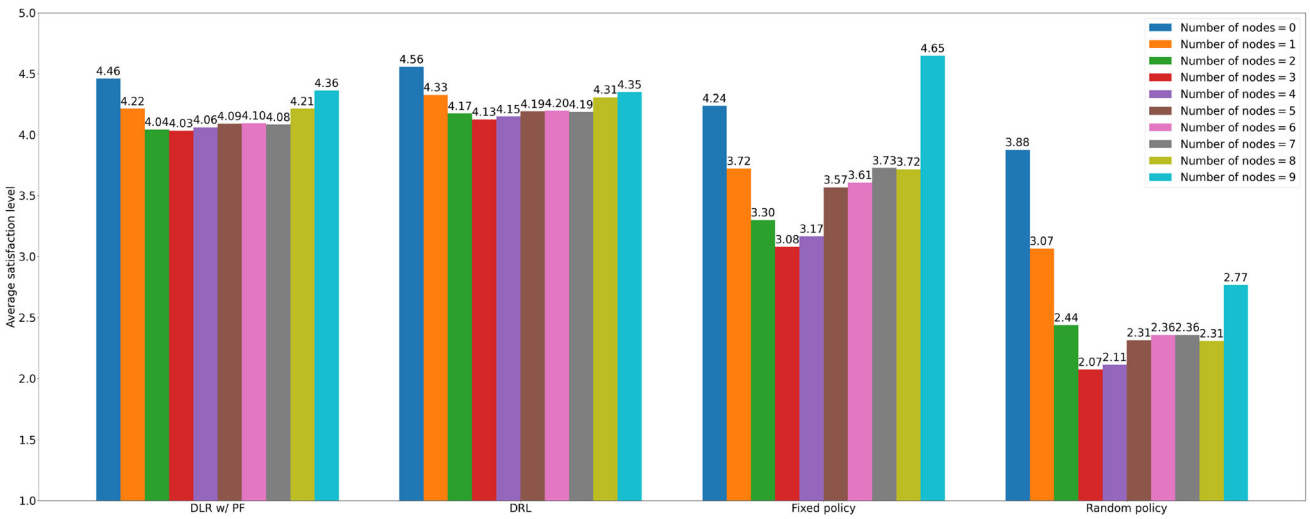


Fig. 12 Average satisfaction level by number of nodes of synthetic scenario

no intelligent design. For the random policy, level 1 occurs most of the time, indicating its instability. To interpret this observation, we plotted the average satisfaction level against the number of nodes between the origin and destination in Fig. 8. We observe a valley in the satisfaction levels of the fixed utility policy from the third to the fifth bar. This could be related to the complexity of the journey search space and the number of the origin–destination combination. First, for journeys with a smaller number of nodes in between, the search space is smaller, so there are fewer choices for the scheduler, leading to a sharp decrease trend in satisfaction from the first to the third bar. Second, there are fewer origin–destination combinations for journeys with more nodes in between in the network. Table 6 lists the origin–destination combinations of the networks. Origin–destination with three nodes in between has the largest number of combinations for the NYC scenario. The decreasing combination from 3 to 13 implies less variation in the problem and the expectation difference is mitigated. Hence, satisfaction rises from the fourth bar when the two effects are combined. In Fig. 11, level 1 occurs most of the time for the random utility policy, indicating the instability of this policy.

5.4.2 Synthetic scenario

In Fig. 11, we see that both DRL approaches yield high satisfaction levels compared to fixed and random policies, indicating that the agent can offer journeys that meet

Table 6 The shortest number of nodes in between origin and destination against the number of the origin–destination combination

Shortest number of nodes between two nodes	Combination of NYC	Combination of synthetic
0	63	120
1	321	196
2	556	232
3	673	232
4	617	200
5	519	140
6	398	80
7	286	40
8	216	16
9	154	4
10	86	–
11	48	–
12	26	–
13	6	–

passengers' preferences. Similar to the NYC scenario, the satisfaction level of fixed utility policy is quite diverse, with satisfaction levels varying widely between levels 1 and 5. We plotted the average satisfaction level against the number of nodes between the origin and destination in Fig. 12 to interpret the observation. We observe a clear valley in the satisfaction levels of the fixed utility policy against the number of nodes, where the journeys with the number of nodes equal to three have the lowest satisfaction. This could be related to the same reasons for complexity as in the NYC scenario. Table 6 lists the origin–destination combination of the networks. Origin–destination with two and three nodes in between has the largest combination. The decreasing combination from 3 to 9 implies less variation in the problem and the expectation difference is mitigated. Hence, satisfaction rises from the fourth bar when the two effects are combined. In Fig. 11, level 1 occurs most of the time for the random utility policy, which indicates the instability of this policy.

Therefore, based on the results of both NYC and synthetic scenarios, we can conclude that the DRL approaches can offer journeys that meet passengers' preferences, and passengers are generally satisfied with the service.

5.5 Fairness among transport operators

We analyze the reward of transport operators for the DRL approach with and without PF. Table 7 summarizes the results of both scenarios.

5.5.1 NYC scenario

For the three transport operators in the NYC scenario, their normalized rewards are slightly different using the DRL with and without PF. Although the reward of all three transport operators is close to the mean, we can observe the fairness by comparing their variance and mean absolute deviation around the mean as shown in Table 7. A smaller value of variance and mean absolute deviation means the reward difference between transport operators is smaller and the passengers are more evenly distributed to the transport operators. We can see that the reward is more balanced for DRL with PF.

5.5.2 Synthetic scenario

For the six transport operators in the synthetic scenario, their normalized rewards are slightly different using the DRL with and without PF. Similar to the observation in the NYC scenario, the reward differences among transport operators are smaller as indicated by the variance and mean absolute deviation around the mean.

Therefore, from the results in both scenarios, we can conclude that the DRL with PF may balance the profit and passengers across the transport operators.

Table 7 Analysis of reward of transport operators

Metric	DRL with PF	DRL
NYC scenario		
Normalized reward	0.3411, 0.3327, 0.3262	0.3260, 0.3434, 0.3306
Variance	3.7421×10^{-5}	5.4281×10^{-5}
Mean absolute deviation	0.005201	0.006713
Synthetic scenario		
Normalized reward	0.2140, 0.0791, 0.1641, 0.1240, 0.1746, 0.2442	0.2284, 0.0872, 0.1558, 0.1205, 0.1571, 0.2509
Variance	0.002968	0.003261
Mean absolute deviation	0.04427	0.04867

6 Conclusion

MaaS is a new concept that unites various modes of transportation into a single platform for future intelligent transportation systems. However, passengers in multimodal transportation systems have diverse demands and preferences for their transport service when facing multimodal transport choices. It is even more challenging to consider the consequences of disappointing the passengers as they may not retain. To address this challenge, we proposed a DRL approach to handle the unknown utility weight issue in the multi-objective journey planning problem. The agent can determine appropriate weights for each passenger based on their characteristic. Given the weights, the scheduler can determine the optimal journey that matches the passenger's expectations instead of the non-dominated Pareto solution. A PF-based variant of the DRL approach is presented to balance the profit across transport operators during the journey planning. Our experiments with real-world and synthetic datasets show that the proposed approach effectively increases the passenger satisfaction level and can lead to a 2.3 times increase in profit. This research can serve as a useful reference for implementing practical intelligent transportation systems dealing with journey planning problem between multiple passengers and transport operators.

Furthermore, there are possible future research directions in addition to this research work. First, although the problem 1 defined a necessary set of objectives and constraints for demonstration and experimentation purposes, there are many variants focusing on different subtle applications that require further study. Second, as a pioneering work, we used classic fully connected neural networks as function approximators, which produced promising results. However, exploring different function approximator structures could enhance the performance further. Third, we will aim to also integrate the research with EV-charging grid management for a low-carbon MaaS ecosystem [69].

Funding This work was supported by EPSRC MACRO - Mobility as a service: MAnaging Cybersecurity Risks across Consumers, Organisations and Sectors (EP/V039164/1).

Availability of data and materials The datasets generated during and/or analyzed during the current study are available in the NYC Taxi and Limousine Commission Trip Record Data and Citywide Mobility Survey, https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page and https://www1.nyc.gov/html/dot/html/about/citywide-mobility-survey.shtml, respectively.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Nikitas A, Michalakopoulou K, Njoya ET, Karampatzakis D (2020) Artificial intelligence, transport and the smart city: definitions and dimensions of a new mobility era. Sustainability 12(7):2789
2. Chu KF, Lam AY, Li VO (2021) Joint rebalancing and vehicle-to-grid coordination for autonomous vehicle public transportation system. IEEE Trans Intell Transp Syst
3. Dreyfus SE (1969) An appraisal of some shortest-path algorithms. Oper Res 17(3):395–412
4. Chu K-F, Lam AY, Li VO (2019) Dynamic lane reversal routing and scheduling for connected and autonomous vehicles: formulation and distributed algorithm. IEEE Trans Intell Transp Syst 21(6):2557–2570
5. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271
6. Hart PE, Nilsson NJ, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern 4(2):100–107
7. Atasoy B, Ikeda T, Song X, Ben-Akiva ME (2015) The concept and impact analysis of a flexible mobility on demand system. Transp Res Part C Emerg Technol 56:373–392
8. Lu W, Quadrifoglio L (2019) Fair cost allocation for ridesharing services-modeling, mathematical programming and an algorithm to find the nucleolus. Transp Res Part B Methodol 121:41–55
9. Ho CQ, Hensher DA, Mulley C, Wong YZ (2018) Potential uptake and willingness-to-pay for mobility as a service (maas): a stated choice study. Transp Res Part A Policy Pract 117:302–318
10. Modesti P, Sciomachen A (1998) A utility measure for finding multiobjective shortest paths in urban multimodal transportation networks. Eur J Oper Res 111(3):495–508
11. Bellman R (1957) A markovian decision process. J Math Mech:679–684
12. Yu X, Gao S, Hu X, Park H (2019) A markov decision process approach to vacant taxi routing with e-hailing. Transp Res Part B Methodol 121:114–134
13. Yue M, Ma SH, Zhou W, Chen XF (2022) Estimation markov decision process of multimodal trip chain between integrated transportation hubs in urban agglomeration based on generalized cost. J Adv Transp 2022
14. Lu J, Feng T, Timmermans H, Yang Z (2017) An integrated markov decision process and nested logit consumer response

model of air ticket pricing. Transportmetrica A Transp Sci 13(6):544–567
15. Rong H, Zhang X, Li Z, Ai Z (2020) Waiting or moving? A crossroad network-based markov decision process approach to catch vacant taxis. IEEE Access 8:10528–10542
16. Bellman R (1966) Dynamic programming. Science 153(3731):34–37
17. Kamargianni M, Matyas M (2017) The business ecosystem of mobility-as-a-service. transportation research board. Transp Res Board 96
18. Pangbourne K, Mladenović MN, Stead D, Milakis D (2020) Questioning mobility as a service: unanticipated implications for society and governance. Transp Res Part A Policy Pract 131:35–49
19. Ho CQ, Mulley C, Hensher DA (2020) Public preferences for mobility as a service: insights from stated preference surveys. Transp Res Part A Policy Pract 131:70–90
20. Guidon S, Wicki M, Bernauer T, Axhausen K (2020) Transportation service bundling-for whose benefit? Consumer valuation of pure bundling in the passenger transportation market. Transp Res Part A Policy Pract 131:91–106
21. Polydoropoulou A, Pagoni I, Tsirimpa A, Roumboutsos A, Kamargianni M, Tsouros I (2020) Prototype business models for mobility-as-a-service. Transp Res Part A Policy Pract 131:149–162
22. Smith G, Sochor J, Karlsson IM (2020) Intermediary maas integrators: a case study on hopes and fears. Transp Res Part A Policy Pract 131:163–177
23. Meurs H, Sharmeen F, Marchau V, Heijden R (2020) Organizing integrated services in mobility-as-a-service systems: principles of alliance formation applied to a maas-pilot in the netherlands. Transp Res Part A Policy Pract 131:178–195
24. Singh M (2020) India's shift from mass transit to maas transit: insights from kochi. Transp Res Part A Policy Pract 131:219–227
25. Becker H, Balac M, Ciari F, Axhausen KW (2020) Assessing the welfare impacts of shared mobility and mobility as a service (maas). Transp Res Part A Policy Pract 131:228–243
26. López D, Lozano A (2020) Shortest hyperpaths in a multimodal hypergraph with real-time information on some transit lines. Transp Res Part A Policy Pract 137:541–559
27. Sever D, Zhao L, Dellaert N, Demir E, Van Woensel T, De Kok T (2018) The dynamic shortest path problem with time-dependent stochastic disruptions. Transp Res Part C Emerg Technol 92:42–57
28. Khani A (2019) An online shortest path algorithm for reliable routing in schedule-based transit networks considering transfer failure probability. Transp Res Part B Methodol 126:549–564
29. Xu SJ, Chow JY (2021) Online route choice modeling for mobility-as-a-service networks with non-separable, congestible link capacity effects. IEEE Trans Intell Transp Syst
30. Ning F, Jiang G, Lam S-K, Ou C, He P, Sun Y (2021) Passenger-centric vehicle routing for first-mile transportation considering request uncertainty. Inf Sci 570:241–261
31. Shi R, Steenkiste P, Veloso MM (2020) Improving the on-vehicle experience of passengers through sc-m*: a scalable multi-passenger multi-criteria mobility planner. IEEE Trans Intell Transp Syst 22(2):1026–1040
32. Molina MDM, Molina BDM, Pérez DC, Campos VS (2021) Connecting passenger loyalty to preferences in the urban passenger transport: trends from an empirical study of taxi vs. vtc services in spain. Res Transp Bus Manag 41:100661
33. Caiati V, Rasouli S, Timmermans H (2020) Bundling, pricing schemes and extra features preferences for mobility as a service: sequential portfolio choice experiment. Transp Res Part A Policy Pract 131:123–148
34. Matyas M, Kamargianni M (2021) Investigating heterogeneity in preferences for mobility-as-a-service plans through a latent class choice model. Travel Behav Soc 23:143–156
35. Efthymiou D, Antoniou C (2017) Understanding the effects of economic crisis on public transport users' satisfaction and demand. Transp Policy 53:89–97
36. Zhang C, Liu Y, Lu W, Xiao G (2019) Evaluating passenger satisfaction index based on pls-sem model: evidence from chinese public transport service. Transp Res Part A Policy Pract 120:149–164
37. Yang H, Xia L (2022) Leading the sharing economy: an exploration on how perceived value affecting customers' satisfaction and willingness to pay by using didi. J Glob Scholars Market Sci 32(1):54–76
38. Lautenbacher CJ, Stidham S Jr (1999) The underlying markov decision process in the single-leg airline yield-management problem. Transp Sci 33(2):136–146
39. Zhao Q, Chen S, Leung SC, Lai K (2010) Integration of inventory and transportation decisions in a logistics system. Transp Res Part E Logist Transp Review 46(6):913–925
40. Rong H, Zhou X, Yang C, Shafiq Z, Liu A (2016) The rich and the poor: a markov decision process approach to optimizing taxi driver revenue efficiency. In: Proceedings of the 25th ACM international on conference on information and knowledge management, pp 2329–2334
41. Kamrani M, Srinivasan AR, Chakraborty S, Khattak AJ (2020) Applying markov decision process to understand driving decisions using basic safety messages data. Transp Res Part C Emerg Technol 115:102642
42. Liu X, Masoud N, Zhu Q, Khojandi A (2022) A markov decision process framework to incorporate network-level data in motion planning for connected and automated vehicles. Transp Res Part C Emerg Technol 136:103550
43. Khamis MA, Gomaa W, El-Shishiny H (2012) Multi-objective traffic light control system based on bayesian probability interpretation. In: 2012 15th international ieee conference on intelligent transportation systems. IEEE, pp 995–1000
44. Khamis MA, Gomaa W (2014) Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework. Eng Appl Artif Intell 29:134–151
45. Haydari A, Yilmaz Y (2020) Deep reinforcement learning for intelligent transportation systems: a survey. IEEE Trans Intell Transp Syst
46. Huang W, Song G, Hong H, Xie K (2014) Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Trans Intell Transp Syst 15(5):2191–2201
47. Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2014) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16(2):865–873
48. Chu KF, Lam AY, Li VO (2018) Travel demand prediction using deep multi-scale convolutional lstm network. In: Proceedings of the 21st international conference on intelligent transportation systems (ITSC). IEEE, pp 1402–1407
49. Chu K-F, Lam AY, Li VO (2019) Deep multi-scale convolutional lstm network for travel demand and origin-destination predictions. IEEE Trans Intell Transp Syst 21(8):3219–3232
50. Li J, Mei X, Prokhorov D, Tao D (2016) Deep neural network for structural prediction and lane detection in traffic scene. IEEE Trans Neural Netw Learn Syst 28(3):690–703
51. Chu K-F, Lam AY, Fan C, Li VO (2020) Disturbance-aware neuro-optimal system control using generative adversarial control networks. IEEE Trans Neural Netw Learn Syst 32(10):4565–4576

52. Nazari M, Oroojlooy A, Snyder L, Takác M (2018) Reinforcement learning for solving the vehicle routing problem. Adv Neural Inf Process Syst 31
53. Zhao J, Mao M, Zhao X, Zou J (2020) A hybrid of deep reinforcement learning and local search for the vehicle routing problems. IEEE Trans Intell Transp Syst 22(11):7208–7218
54. Zhou M, Yu Y, Qu X (2020) Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach. IEEE Trans Intell Transp Syst 21(1):433–443
55. Yu C, Wang X, Xu X, Zhang M, Ge H, Ren J, Sun L, Chen B, Tan G (2020) Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs. IEEE Trans Intell Transp Syst 21(2):735–748
56. Chu KF, Lam AY, Li VO (2021) Traffic signal control using end-to-end off-policy deep reinforcement learning. IEEE Trans Intell Transp Syst
57. Chu T, Wang J, Codecà L, Li Z (2019) Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Trans Intell Transp Syst 21(3):1086–1095
58. Wang P, Chan CY, La Fortelle A (2018) A reinforcement learning based approach for automated lane change maneuvers. In: Proceedings of the IEEE intelligent vehicles symposium (IV). IEEE, pp 1379–1384
59. Shi T, Wang P, Cheng X, Chan CY, Huang D (2019) Driving decision and control for automated lane change behavior based on deep reinforcement learning. In: Proceedings of the IEEE intelligent transportation systems conference (ITSC). IEEE, pp 2895–2900
60. Codeca L, Cahill V (2021) Using deep reinforcement learning to coordinate multi-modal journey planning with limited transportation capacity. In: SUMO conference proceedings, vol 2, pp 13–32
61. Farahani A, Genga L, Dijkman R (2021) Online multimodal transportation planning using deep reinforcement learning. In: 2021 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 1691–1698
62. Lam C, Ip W (2011) A customer satisfaction inventory model for supply chain integration. Expert Syst Appl 38(1):875–883
63. Martín-Consuegra D, Molina A, Esteban Á (2007) An integrated model of price, satisfaction and loyalty: an empirical analysis in the service sector. J Prod Brand Manag
64. Kelly FP, Maulloo AK, Tan DKH (1998) Rate control for communication networks: shadow prices, proportional fairness and stability. J Oper Res Soc 49(3):237–252
65. Srikant R, Başar T (2004) The mathematics of internet congestion control. Springer
66. Petrou K, Liu MZ, Procopiou AT, Ochoa LF, Theunissen J, Harding J (2020) Operating envelopes for prosumers in lv networks: a weighted proportional fairness approach. In: Proceedings of the IEEE PES innovative smart grid technologies Europe (ISGT-Europe). IEEE, pp 579–583
67. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D (2016) Continuous control with deep reinforcement learning. In: Proceedings of the international conference on learning representations
68. Diamond S, Boyd S (2016) CVXPY: a Python-embedded modeling language for convex optimization. J Mach Learn Res 17(83):1–5
69. Zhang X, Xia X, Liu S, Cao Y, Li J, Guo W (2022) An integrated framework on autonomous-ev charging and autonomous valet parking (avp) management system. IEEE Trans Transp Electrif

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



https://dspace.org/

Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness

Chu, Kai-Fung; Guo, Weisi

2023-09-01
Springer
https://doi.org/10.1007/s00521-023-08733-4
Chu K-F, Guo W. (2023) Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness. Neural Computing and Applications, Volume 35, September 2023, pp. 20221-20240

Downloaded from a DSpace repository. https://dspace.org/