Dynamic pricing method to maximize utilization of one-way car sharing service

Toya Kamatani
Department of
Urban Environment Systems
Chiba University
Chiba, Japan
to.kmtn@gmail.com

Yusuke Nakata
Department of
Urban Environment Systems
Chiba University
Chiba, Japan
a1019nakata@gmail.com

Sachiyo Arai
Department of
Urban Environment Systems
Chiba University
Chiba, Japan
sachiyo@faculty.chiba-u.jp

***Abstract*—A one-way car sharing service, which allows a car to be dropped off at an arbitrary station, is highly convenient for users. However, the uneven distribution of users' departures or destinations causes the situation where user cannot access to available car. This situation incurs a heavy loss for both users and the operational side of service. For this problem, we introduce a dynamic pricing scheme using reinforcement learning to set the charge for each station and propose a method to maximize the utilization rate by suppressing the uneven distribution of cars. The experimental results show that dynamic pricing improves the uneven distribution of cars compared with flat rates.**

***Index Terms*—car-sharing, dynamic pricing, reinforcement learning**

I. Introduction

Car-sharing services can be either round-trip services or one-way services, and cars are borrowed from a dedicated car sharing space (henceforth referred to as a station). The round-trip service requires the user to return the car to the station it was borrowed from, while the one-way service allows a user to drop it off at any station. Because the one-way service is more convenient, it is expected to expand in popularity shortly.

Even so, the one-way car sharing service has a major disadvantage: the cars will be unevenly distributed across stations because the departure and destination locations of the users are biased by user preference. As a result, the number of people who cannot use the service increases and the utilization rate decreases.

In this study, we propose dynamic pricing to maximize the utilization rate. This is motivated by the fact that the distribution of users in the car-sharing service changes according to the price of the service [1]. For dynamic pricing, we propose reinforcement learning. Reinforcement learning is suitable for dynamic pricing because it excels in multi-step decision making. In this study, four experiments are conducted to compare flat rates and dynamic pricing. We confirmed from experiments that proper dynamic pricing reduces the uneven distribution of cars and maximizes utilization.

II. Related research

Existing research [2] regarding redistribution involves the application of reinforcement learning to stabilize the number of vehicles based on the movement of unused cars. It also includes research on vehicle redistribution and stabilization with autonomous vehicles [3], [5]. There is also research being conducted on optimal redistribution based on the demand forecast of bicycle sharing [4]. Research on dynamic pricing includes the pricing of ride-sharing services that use private cars as taxis [6].

The technology of redistributing by automated driving technology has emerged recently [3], [5]. This research is useful because the redistribution of autonomous vehicles needs to be reduced as much as possible.

The research conducted in the current study aims not only to maximize the company's profits, but also to improve the users utilization rate.

III. Modeling of car-sharing environment

A. Definition of terms

Definitions of terms are listed in TABLEI.

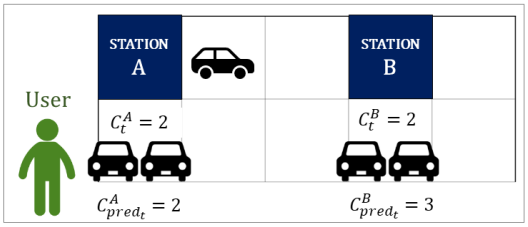
TABLE I
Definition of terms

	Details
Station N	Space to rent a car
Number of cars parked C	Parking number of each station
Predicted number of cars parked C_{pred}	C+Number of temporary destinations for cars in use
Predicted number of cars parked stability H_{pred}	Variance of C_{pred}
Charge setting F	Usage charges per minute [yen/minute]
Point of departure L	Place where users start using a car
Temporary destination D_T	Place where user temporarily goes
Actual destination D_A	Place where users would prefer to go
Preference[1] w_p	Preference for charges
Preference[2] w_d	Preference for distance

B. Car sharing service model

The car-sharing service model targeted by this research is shown in Fig.1.

978-1-7281-4026-1/19/$31.00 ©2019 IEEE



1. The number of parked are and the estimated number of parked cars are observed and the rate is presented
2. See **the rate table for departure** decide whether to use or not
3. See **the rate table for arrival** and decide the temporary destination
4. Update the number of cars parked and the estimated number of parked cars

Fig. 1. Car sharing service station model

The station observes the number of parked cars C and the predicted number of parked cars C_{pred}. Based on these observations, a charge is presented to the user. The user decides which start station to use based on the usage rates corresponding to the stations' charges for departure. Utilization rates, which are defined according to the existing research [1], to calculate the number of users per day from the number of reservations received for each fee offer. Additionally, according to the utilization rate, the actual destination D_A station is determined. Next, the user determines a temporary destination D_T, which is a station that is on the way to their destination. The temporary destination is determined from the preference $\mathbf{w} = (w_p, w_d)$ of each user and the charge for arrival of each station. D_A and D_T are indicated by two-dimensional vectors. The degree of preference indicates what each user places weight on. In this research, two weights, w_p "charge" or w_d "distance between the actual destination and another station", are used to determine the temporary destination. Moreover, w_p and w_d are given by Eq.(1).

$$w_p + w_d = 1 \quad (w_p, w_d \geq 0) \tag{1}$$

In this case, the selection of stations according to the selection level is defined as Eq.(2). The user chooses the station with the highest preference (their preferred station) as the temporary destination and starts using it. Finally, the number of parked cars at each station and the estimated number of cars parked are updated. As described above, the model determines the charge F of the station to be used for each unit time and fluctuates with the utilization rate of the user.

$$\underset{D_T \in \mathcal{N}}{\arg\min} \quad \frac{||D_A - D_T||_2}{x_{\max}} \times w_d + \frac{F_{D_T}}{F_{\max}} \times w_p \tag{2}$$

First, the distance between the actual destination D_A and the other stations N is calculated. Normalization is then performed by multiplying it by the user preference w_d and dividing the result by the station group maximum distance $x_{\max}$. Next, the user's preference w_p is multiplied by the station charge F. Similarly, the result is divided by the maximum charge $F_{\max}$ to perform normalization. The station with the least value for a given preference is the tentative destination D_T.

C. User behavior

Algorithm 1 User behavior algorithm

Determine the departure location L according to the charge
Determine actual destination D_A
Determine temporary destination D_T according to preference
—Lending—
if $C_L > 0$ **then** ▷ There is a car available: USE
else$C_L = 0$ ▷ There is no car available: NOT USE
end if
—Return—
if $C_{D_T} < C_{\max}$ **then** ▷ There is space to return: RETURN
else$C_{D_T} = C_{\max}$ ▷ There is no space to return: NON-RETURN
Re-determine temporary destination D_T according to preference
end if

The user's behavior is outlined by Algorithm 1 above. It generates users with different preferences. At this time, the origin L and the actual destination D_A are determined according to the distribution of user occurrence. The temporary destination D_T to be adopted is also determined from the preferences described above. If there is an available car at the selected departure location L, it can be used. Finally, the user returns the car. At this time, it is determined whether parking at the temporary destination D_T is possible. If parking is not possible, a different temporary destination D_T is determined according to preferences.

D. User outbreak

One user comes to a start station per step. At this time, there are no users at a station where the number of cars is zero. In the existing research [1], a one-way car-sharing service was simulated based on different charges. Simulations spanning 90 days were run 100 times, and the number of reservations accepted and tickets not taken were calculated. The incidence rate according to the charge in this research was calculated from the existing analysis [1]. The number of users per day was derived from the number of reservations received for each charge setting in the current study. The number of users for each charge is shown in TABLEII.

TABLE II
NUMBER OF USERS ACCORDING TO CHARGE

Charge setting[yen/minutes]	Daily number of users[People/minutes]
15	2.32
20	1.20

It is assumed that the prices are 15 yen and 20 yen. The users per minutes are 2.32 and 1.20, respectively, and the ratio of the number of users is 2:1. In this study, the ratio of the number of users is used as the generation ratio according to the charge.

IV. Evaluation Method

In one-way car sharing, users are unevenly distributed. At this time, if the charge is such that the number of cars becomes uniform, cars will be unevenly distributed by utilization. Therefore, if the difference between the user occurrence distribution and the distribution of the predicted number of parked cars at each station is closer, the charge can be set appropriately. Therefore, the Kullback-Leibler information

amount is calculated from the occurrence distribution and the ratio of the predicted number of parked cars at each station. The Kullback-Leibler information amount measures the difference between two probability distributions in the actual probability distribution P and in other probability distributions P'. When P and P' are discrete distributions, the Kullack-Leibler information amount for P' in P is indicated by Eq.(3).

$$D_{\mathrm{KL}}(P||P') = \sum_i P(i) \log \frac{P(i)}{P'(i)} \tag{3}$$

$P(i)$ and $P'(i)$ are probabilities that the value selected by the probability distributions P and P' is i. It indicates that the probability distributions of P and P' are close when the value of the Kullback-Leibler information amount is low. In this study, as the user occurrence distribution P and the distribution P' of the predicted number of cars parked approach 0, the bias can be reduced, and the need for redistribution will be reduced in the future.

V. Proposed method

This study is formulated by the Markov decision process $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma \rangle$. $\mathcal{S}$ is the state set, $\mathcal{A}$ is the action set, $\mathcal{R}$ is the reward vector, $\mathcal{P}$ is the transition probability set, and γ is the discount rate. We express the discount rate and apply Q-learning to achieve dynamic pricing. Q-Learning is a typical reinforcement learning method that learns the optimal policy in the environment of the Markov decision process [7]. In Q-Learning, an agent chooses its action a_t in current state s_t according to $Q(s_t, a_t)$. Then, by executing a_t and transitioning to the next state s_{t+1}, the agent is updated from $Q(s_t, a_t)$ with the obtained reward r_t to $Q(s_{t+1}, a_{t+1})$, which is the maximum Q-value in s_{t+1} according to Eq.(4). Here, α is the learning rate.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big(r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\Big) \tag{4}$$

The status input in this study is the predicted number of parked cars at each station C^n_{pred}. The procedure is to observe the state $s \in \mathcal{S}$ of each step, change the charge F with the action $a \in \mathcal{A}$, and observe s' in the next state.

In this study, we seek to minimize D_{KL}. However, it is difficult to minimize D_{KL} directly. So the problem is solved by keeping the expected number of parked cars within an acceptable range. The predicted number of parked cars stability H_{pred} is defined in Eq.(5). Here, N is the number of stations, C_{pred} is the predicted number of parked cars for each station, $\overline{C}_{\mathrm{pred}}$ is the average of the predicted number of parked cars for each station.

$$H_{\mathrm{pred}} = \frac{1}{N} \sum_{n=1}^{N} (C^t_{\mathrm{pred_n}} - \overline{C^t}_{\mathrm{pred}})^2 \tag{5}$$

We use H, given by Eq.(5), for the reward calculation instead of minimizing D_{KL}. Thus, we introduce the metric defined in Eq.(6). The reward will be applied to the predicted number of parked cars stability H_{pred}. The allowable predicted number of parked cars stability $H^{\mathrm{tolerance}}_{\mathrm{pred}}$ is the value of the stability of the permitted number of parked cars. A negative reward is applied when this value exceeds the forecasted parking stability $H^{\mathrm{tolerance}}_{\mathrm{pred}}$.

$$r_t = C \quad \text{if} \quad H_{\mathrm{pred}} > H^{\mathrm{tolerance}}_{\mathrm{pred}} (C \text{ is negative reward}) \tag{6}$$

VI. Computer experiments

In this section, we confirm the applicability of the proposed method using four experiments.

A. Experimental settings

The experimental environment indicates that the maximum number of parked cars is 6, the number of stations is N = 6, and the total number of cars is 18. The initial number of parked cars at each station is 3, and the maximum predicted number of parked cars is $\max(\mathrm{C^n_{pred}})$=18. The locations of the stations are shown in Fig.2. The environment is a 5 x 5 grid world, and the six stations are located at (0, 0), (0, 1), (0, 4), (1, 4), (4, 3), and (4, 4). The red frame shows one station, the blue frame shows a parking space, and the green structure shows a pair of stations. Moreover, 1 cycle represents 1 minute, 10 hours are simulated, and the average of 1000 episodes is evaluated for each experiment.

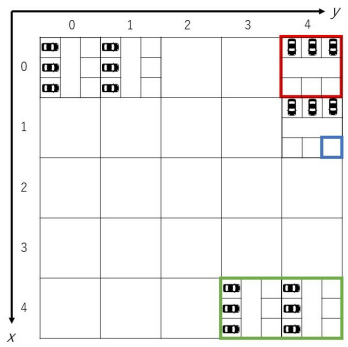


Fig. 2. Schematic of the experimental environment

Reinforcement learning is applied to each environment's dynamic pricing. State $s \in \mathcal{S}$ splits $0 \leq C^n_{pred} \leq \max(\mathrm{C^n_{pred}})$ into discrete values and observes this state at all stations. Specifically, the state can be divided into three states, $C^n_{pred} = 0$, $C^n_{pred} = 1, 2$, and $C^n_{pred} \leq 3$, which can be distinguished at each station. The action is selected from $\mathcal{A}$ = {15,20}. The reward is defined as $r_t = -5$, if $H_{\mathrm{pred}} > H^{\mathrm{tolerance}}_{\mathrm{pred}}$.

In this study, four experiments are conducted to compare flat prices, and dynamic pricing obtained by reinforcement learning. For evaluation, the difference D_{KL} of the distribution of the predicted number of parked cars at each station is used.

Experiment 1 and Experiment 2 verify user-generated uneven distribution. Experiment 1 is an environment where the occurrences at all stations are uniform, and Experiment 2 is an environment where the events at the stations are unevenly distributed. Additionally, in all experiments, the choice of station for departure and the actual destination depend on the utilization rate based on the charge. Experiment 3 and Experiment 4 determine the temporary destination according to user preference. In Experiment 3, the preference considers only the "charge", and in Experiment 4, the preference is

determined from “charge” and “the distance between the actual destination and another station.” Additionally, in Experiment 3 and Experiment 4, the occurrence of users is an uneven distribution.

B. Experimental results and considerations

TABLE III
EVALUATION OF EXPERIMENTS 1, 2, 3, 4 (RL=REINFORCEMENT LEARNING)

	Experiment 1		Experiment 2		Experiment 3		Experiment 4	
	Flat	RL	Flat	RL	Flat	RL	Flat	RL
D_{KL}	0.20	0.16	0.54	0.43	0.53	0.34	0.44	0.34

The evaluation value considers the difference D_{KL} of the distribution of the predicted number of parked cars at each station, as presented in TABLEIII. The reinforcement learning is denoted as RL.

1) Applicability of dynamic pricing: As a result of comparing the flat charge of 20 yen and the evaluation value of dynamic pricing in Experiment 1, the variance H_{pred} of the predicted number of parked cars was 4.12 for the flat charge, and 2.49 for the charge based on reinforcement learning. In a stable outbreak environment, it is desirable that the predicted number of cars parked at each station be uniform, so dynamic pricing with a small variance of the predicted number of parked cars is useful. Thus, we compare the flat charge and dynamic pricing with the difference between the distribution of outbreaks and the distribution of the predicted number of parked cars at each station. As a result, a significant difference was found at the 1% level of significance in reinforcement learning. Therefore, it can be stated that providing a one-way car-sharing service at a flat charge is difficult to implement without requiring redistribution.

2) An environment in which the emergence of users is biased: Experiment 2 is an environment where the occurrence of users is unevenly distributed, Experiment 3 is an environment where the user’s preference is only “ charge ”, and Experiment 4 is an environment where the user’s preference is “ charge ” and “ Distance ”. For each experiment, we compare the flat charge and dynamic pricing evaluation values. The difference between the evaluation values was verified using the T-test for the flat charge and reinforcement learning charge setting schemes, which takes two samples with equal population variance from each scheme for comparison. As a result, in each comparison, the difference D_{KL} between the distribution of user occurrence and the distribution of the predicted number of parked cars at each station was significant (significance level of 1%). The result is displayed as a box plot in Fig.3. From these results, in an environment that takes preference into consideration, using reinforcement learning makes it possible to set charges that match the user occurrence distribution.

3) Practicality of reinforcement learning: We simulated a one-way car-sharing service with tolls that incorporates reinforcement learning. Reinforcement learning makes it possible to respond to changes in the environment, the maximum number of parked stations, and changes in the number of cars as automatic charge setting parameters, so it can be stated that dynamic pricing based on reinforcement learning is practical.

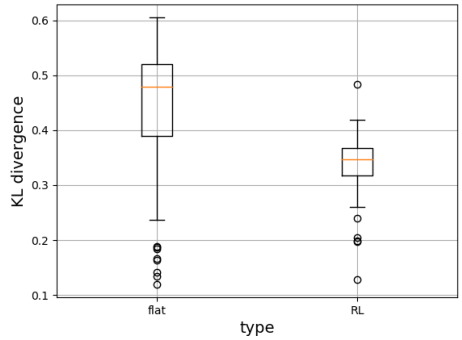


Fig. 3. Difference between the distribution of user occurrence and the distribution of the predicted number of parked cars at each station (Experiment 4). The lower part of the square is the first quartile, the orange line is the median, the upper part of the square is the third quartile, and the circles are outliers.

VII. SUMMARY AND FUTURE ISSUES

In this study, applying dynamic pricing using reinforcement learning reduces the number of unavailable stations. This is achieved by stabilizing the planned number of parked vehicles within an acceptable range.

On the other hand, there is still room for improvement in terms of the reward design. The current reward design is given when the variance of the predicted number of parked cars increases, but is not considered according to the bias of the user occurrence distribution. Therefore, a possible improvement would be to consider the reward design in consideration of the bias of user occurrence distribution.

Furthermore, we should consider the number of stations in real one-way car sharing services. At this time, the state input is limited by the currently used Q-learning, and it is not effective in an environment where the number of stations may change. Concerning the number of stations N, the state $\mathcal{S}$ and operation $\mathcal{A}$ can be shown to represent 3^N and 2^N computations, respectively. Therefore, methods such as deep reinforcement learning [8] that can handle more complex environments are attracting attention. We will consider the introduction of a method that can handle more complex environments as a future task.

REFERENCES

[1] Takeshi Mizokami, Yuta Nakamura, Tatsuya Hashimoto.“SIMULATION MODEL FOR INTRODUCTION OF ONE-WAY MICRO ELECTRIC CAR SHARING SCHEME”.2015.

[2] Jian Wen, Jinhua Zhao, Patrick Jaillet.“Rebalancing Shared Mobility-on-Demand Systems: a Reinforcement Learning Approach”.IEEE, 2017.

[3] M.Pavone, S.L.Smith, E.Frazzoli, and D.Rus.“Robotic load balancing for mobility-on-demand systems”.The International Journal of Robotics Research.vol.31,no.7,pp.839-854, 2012.

[4] Supriyo Ghosh, Pradeep Varakantham, Yossiri Adulyasak, Patrick Jaillet.“Dynamic Repositioning to Reduce Lost Demand in Bike Sharing Systems”.2017.

[5] K.Spieser, S.Samaranayake, W.Gruel, and E.Frazolli.“Shared-car mobility-on-demand systems: a fleet operators guide to rebalancing empty cars”.in Transportation Research Board 96th Annual Meeting, 2016.

[6] M.K.Chen and M.Sheldon.“Dynamic pricing in a labor market: Surge pricing and flexible work on the uber-platform”.in EC.2016, p.455.

[7] Watkins, Christopher JCH and Dayan, Peter.“Q-learning,”Machine learning”.vol.8,pp.279-292,1992.

[8] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, Demis Hassabis.“Human-level control through deep reinforcement learning”. Nature interational journal of science, 2015.