Optimizing Rebalance Scheme for Dock-less Bike Sharing Systems with Adaptive User Incentive

Yubin Duan and Jie Wu
Department of Computer and Information Sciences, Temple University, USA
Email: {yubin.duan, jiewu}@temple.edu

***Abstract*—Recently, the development of Bike Sharing Systems (BSSs) brings environmental and economic benefits to the public. However, BSSs frequently suffer from the imbalanced bike distribution, including dock-less BSSs. The underflow or overflow of bikes in a region may lead to a lower service level to BSSs or congestion to the city. In the paper, we consider rebalancing the dock-less BSS by providing users with monetary incentives. The long-term objective is to maximize the number of satisfied users who successfully complete their rides over a period of time. The operator of the dock-less BSS can not only encourage a user to rent bikes at the neighborhood of its source with a source incentive, but also incentivize them to return bikes at the neighborhood of its destination with a destination incentive. To learn the differentiated incentive price for rebalancing bikes across time and space, we extend a novel deep reinforcement learning framework for user incentive. The source and destination incentives are integrated in an adaptive way by adjusting the detour level at the source and/or destination by avoiding bike underflow and overflow. In the experiment, we evaluate our approach in comparison with two existing pricing schemes. The locations of sources and destinations are abstracted from a selected dataset from Mobike. The experiment results show that our adapted learning algorithm outperforms the original one that only considers source incentive as well as another state-of-the-art approach in maximizing the long-term number of satisfied users.**

***Index Terms*—dock-less bike sharing system, rebalance problem, user incentives, reinforcement learning**

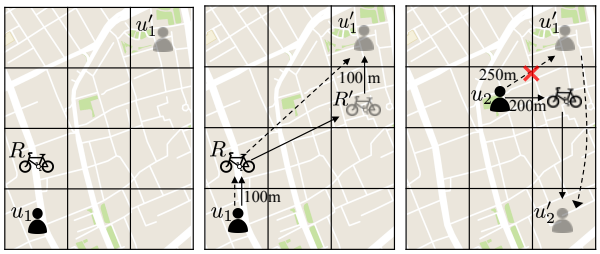
I. Introduction

Recently, the rapid development of Bike Sharing Systems (BSSs) brings environmental and economic benefits to the public [1]. The bikes of BSSs are easily accessible and affordable for users, which greatly motivates users to ride bikes for traveling a short distance. The convenience of the BSS provides residents a way of green travel. A study [2] covering 4 North American cities shows that nearly 40% of BSS users drove less after participating the system. In addition, bike sharing is an example of the sharing economy, and has potential economic benefits. Although BSSs bring attractive benefits to the public, the systems still suffer from imbalanced bike distribution. Both temporally and spatially asymmetric demands of users may cause imbalance in the distribution of bikes. For docked BSSs like *citi bike* in NYC, each station has capacity limitation and an extreme imbalance could cause underflow or overflow events. That is, a station is empty when users try to rent bikes or a station is totally full when users attempt to return bikes. For dock-less BSSs like *Mobike* in China, although there are no stations nor capacity limitations, they still face underflow and/or overflow events as shown in Fig. 1. These may lead to a lower service level to the BSS or congestion to the city. To avoid these negative impacts, it is important to rebalance the BSS in a timely and cost-efficient manner with a budget provided by BSS operators.

Fig. 1. Resolving underflow/overflow through rebalance in dock-less BSSs.

Existing user-based rebalance strategies for dock-less BSSs cannot fully exploit the advantage of user incentive. A rebalance scheme based on reinforcement learning for dock-less BSS is proposed in [3]. However, this scheme only considers encouraging users to rent bikes from nearby regions (source detour) with a source incentive and ignores the possibility that a user can also return bikes to alternative places (destination detour) with a destination incentive. We find out that the destination incentive can also help to rebalance the system. The adaptive combination of source and destination incentive can bring extra benefits to the system.

An example in Fig. 2 illustrates our observation. The setting is shown in the figure where the map is divided into 4×3 square regions. The performance of rebalancing is quantified by the service level, i.e. the number of users who successfully finish their trip. In the example, there is one bike located at R. User u_1 arrives first with destination u_1', and user u_2 with destination u_2' arrives after u_1 finishes his trip. The traces of u_1 and u_2 under different incentive schemes are plotted in Fig. 2. The dashed line represents the movement of following the source-incentive-only scheme. The solid line shows a better way which combines source and destination incentives, which can further improve the service level. Assume the maximum walk distance of users is 200m, which includes the source detour and destination detour. Under source-incentive-only scheme, user u_2 cannot successfully rent a bike after u_1 returns the bike at u_1', since the distance between u_2's source and u_1's destination exceeds 200m. The service level is 1. In contrast, if a user is allowed to both rent and return a bike at neighbor



(a) User u_1 arrives with destination u_1' (b) Give source and dest. incentive to u_1 (c) User u_2 arrives with destination u_2'

Fig. 2. A motivation example.

regions, the service level can be improved to 2. As shown by the solid line, after u_1 arrives, he rents the bike at R and returns the bike at R' by receiving a mandatory incentive. When u_2 arrives, he can rent the bike at R' and finish his trip to u_2'. Fig. 2 shows that the service level can be improved by allowing users to return bikes to neighbor regions.

Motivated by this observation, we propose the *Dock-less BSSs Rebalancing* (DBR) problem. The setting up scenario is that the operator of the dock-less BSS offers both source and destination incentives for users and encourages them to rent or return bikes at specific locations with a limited budget. If the source or destination incentive is larger than user's detour cost, which contains an initial fee plus the fair related with detour distance, then the user will accept the offer. We aims to design an adaptive incentive scheme that maximizes the total number of satisfied users over a day. Designing the adaptive incentive scheme is not trivial. The asymmetric user demands in both temporal and spatial domains bring the challenge of deciding differentiated pricing for users temporally and spatially. The rebalance scheme needs to adaptively adjust the ratio of the source and destination incentive in a timely and cost-efficient manner, which brings another challenge to the problem.

In this paper, we extend the novel deep reinforcement learning framework proposed by [3] to rebalance the dock-less BSS with user incentive. We propose to build a hybrid incentive scheme and take the benefits brought by destination incentive into account instead of just considering the source incentive. For simplicity, the city map is divided into square regions and the temporal domain is discretized into time-slots. The architecture is shown in Fig. 3, and the system can adaptively adjust the ratio of the source and destination incentives. Specifically, when training the reinforcement learning network, the rebalancing scheme takes the bike trace information of BSS and the budget provided by the operator as the state and takes the source and destination incentive prices for each region as an action. The environment feeds back the number of satisfied users as the reward and updates the state.

The contributions of our paper are summarized as follows:

- We propose the Dock-less BSS Rebalancing (DBR) problem, where the BSS is balanced by incentivizing users. Both source and destination incentives are considered.
- We illustrate the benefits brought by the destination incentive and adapt the deep reinforcement learning architecture designed for source incentive to optimize the incentive scheme for considering destination incentive.

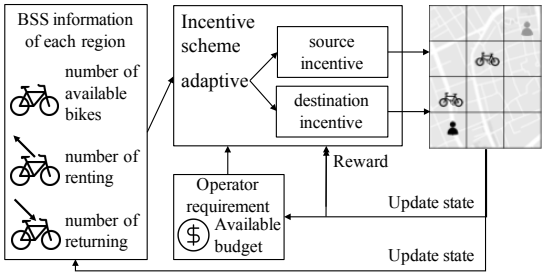


Fig. 3. An overview of the architecture for rebalancing dock-less BSSs.

- We further consider two ways to adaptively combine source and destination incentives under the same budget constraint, and we set up experiments on a real-world dataset to examine the performance of our approaches.

II. Problem Statement

A. Overview

In our model, we propose an adaptive approach for rebalancing dock-less BSS. Given a limited budget, which is not sufficient enough to totally balance the BSS, our approach adaptively allocates it to incentivize users to conduct a detour at source and/or destination based on the underflow/overflow distribution across time and space. The objective is to maximize the overall service level of the system over a day. The service level is quantified by the number of satisfied users.

Specifically, the incentive that is used to encourage users to rent bikes at neighbor regions of their sources is denoted as source incentive, while it is called destination incentive on the other side. For source incentive, the BSS operator provides locations of available bikes to each user along with incentive prices of bikes in neighbor regions. For destination incentive, the operator suggests users return bikes to neighbor regions of the user's destination. The price of the source and destination incentives is determined by the incentive scheme. A reinforcement learning based price scheme for source incentive has been studied in [3]. We propose to jointly consider the source and destination incentive inspired by the benefits of destination incentive we observed. Users' choice of accepting incentives or not is simulated by the environment model. The performance of the rebalance is evaluated via the service level which is the number of satisfied users. The architecture of our hybrid incentive architecture is shown in Fig. 3.

B. Incentive Scheme Model

We first describe our incentive scheme. Both temporal and spatial domains are discretized. The BSS operator provides different source and/or destination incentive prices for each region at each time-slot. Specifically, each day is separated into m time-slots, denoted by $T = \{t_1, t_2, \ldots, t_m\}$. A city H is divided into n square regions, i.e., $H = \{h_1, h_2, \ldots, h_n\}$. The neighbors of a region h_i are defined as the four regions directly adjacent to h_i, and the set of neighbor regions for h_i is denoted as $N(h_i)$. Users to the BSS system are denoted by $U = \{u_1, u_2, \ldots u_o\}$. Although the actual user demands vary in temporal and spatial domains, the patterns on their demands in both domains provide basis for our incentive scheme. Our

statistic on traces data from *Mobike* shows the existence of rush hour and demand hot spots. The number of users' rent events and return events at region h_i during time-slot t is modeled as random variables $D_i(t)$ and $\Lambda_i(t)$ respectively. The number of bikes in h_i at the beginning of time-slot t is denoted as $\varphi_i(t)$.

To deal with the imbalance of the BSS, we assume that the provider can provide a budget B for user incentive, including a source incentive budget B^+ and a destination incentive budget B^-. Our incentive scheme is used to decide the different price of source incentive $p_i^+(t)$ and destination incentive $p_i^-(t)$ for each region h_i at each time-slot t. If a user rents bikes at a neighborhood region h_i of his/her source region during time-slot t, he/she can obtain an incentive $p_i^+(t)$. Each neighbor region may contain more than one bike, and the bikes in the same region have the same incentive price. Similarly, destination incentive $p_i^-(t)$ is given to users who return bikes to h_i that is adjacent to users' destination region during time-slot t. Different from the source incentive, we assume that each region only contains one potential return location which is the center of the region. This simplification can reduce the complexity of the model.

C. Environment Model

The environment mainly models user dynamics and provides feedback to the incentive scheme. Based on the source and destination incentive price vectors generated by the scheme, the environment simulates each user's choice of accepting the incentive or not.

We assume users know the source and destination incentive prices of all regions, and have costs when walking from their sources to rent locations (source detour) and walking from return locations to their destinations (destination detour). The user would accept the source incentive if the source incentive price is larger than the source detour cost. It is symmetric for the destination incentive. Both source and destination detour costs share the same model, which is built based on the model in [3, 4]. In our model, a user u_k has an initial cost C for either source or destination detour. Besides, the cost is also relevant to the detour distance δ. Specifically, let $c_k(h_i, h_j, \delta)$ and $c'_k(h_i, h_j, \delta')$ denote the source and destination detour cost respectively. h_i and h_j represent regions where u_k rent and return a bike respectively. δ and δ' are the corresponding source and destination detour distance. If the user u_k rents (or returns) a bike at a region which is the neighbor of his/her source (or destination), his/her source detour cost $c_k(h_i, h_j, \delta) = C + \eta\delta^2$ (or destination detour cost $c'_k(h_i, h_j, \delta') = C + \eta\delta'^2$), where η is a constant coefficient. We assume users are not willing to rent or return bikes at regions further than neighbor regions, and the cost of renting or returning bikes in these regions is infinity. If the user u_k rent (or return) bikes in the same region as his/her source (or destination), there is no cost. Note that if a user detours at both source and destination, he/she will receive two Cs as an incentive to conduct source and destination detour, which helps to resolve overflow and underflow problem of the BSS, in one trip.

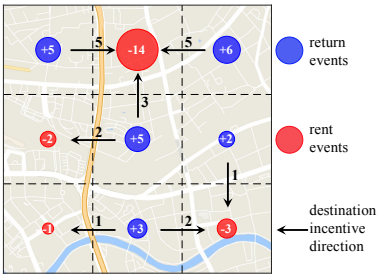


Fig. 4. An illustration of the destination incentive.

D. Problem Formulation

Based on the system and environment models, the Dockless BSS Rebalancing (DBR) problem is proposed. In DBR, we aim to maximize the service level of a BSS in a one-day service circle. In each service circle, the BSS operator provides budget B^+ and B^- for source and destination incentives, and $B^+ + B^- = B$. Formally, our problem can be expressed as:

$$\max \quad \sum_{t=1}^{m}\sum_{i,j=1}^{n}\tau_{ij}(t) \tag{1}$$

$$\text{s.t.} \quad \sum_{t=1}^{m}\sum_{i=1}^{n}p_i^+(t) < B - B^- \tag{2}$$

$$\sum_{t=1}^{m}\sum_{i=1}^{n}p_i^-(t) \le B^- \tag{3}$$

$$\sum_{j=1}^{n}\tau_{ji}(t) - \sum_{j=1}^{n}\tau_{ij}(t) \le \varphi_i(t), \forall i, t \tag{4}$$

$$\varphi_i(t+1) = \varphi_i(t) + \sum_{j=1}^{n}(\tau_{ji}(t) - \tau_{ij}(t)) \forall i, t. \tag{5}$$

Note that the difference with the existing price scheme can be found in Eq. (6) and (7), where we consider two kinds of incentives. The overall budget of source and destination incentive remains as B. The difference is that some part of the budget B^- is assigned for destination incentive.

III. Hybrid Incentive Scheme

A. An Existing Pricing Scheme for Source Incentive

A pricing algorithm for source incentive is proposed by Pan et. al [3]. Their pricing scheme is based on a Markov Decision Process (MDP) and is optimized by using a reinforcement learning approach inspired by the hierarchical reinforcement learning [5–7] and Deep Deterministic Policy Gradient algorithm [8]. The pricing algorithm is briefly stated in the section and our adaptive incentive scheme is built upon it.

The MDP is used to model the interaction between the pricing scheme and the environment. Specifically, the MDP is a 5-tuple (S, A, P, r, γ), where S is the set of states $\{s_t\}$, A is the set of actions $\{a_t\}$, P describes the transition possibility between states under an action, r denotes the immediate reward and γ is the discount factor. The weights of future rewards and the present reward are determined by the discount factor $\gamma \in [0, 1]$. $\gamma = 1$ represents that future rewards share the same importance as the present reward, i.e. the overall reward is the additive sum of the reward from each time-slot. The pricing scheme treats source incentive prices for all regions as an action and the number of satisfied users as a reward. The MDP ends when the budget B is used up. The pricing scheme finds a policy π_θ, which maps states to actions, through optimizing the MDP based on reinforcement learning. The number of bikes rented from h_i and returned to h_j during time slot t is denoted by $\tau_{ij}(t)$.

Algorithm 1 The source (or destination) incentive schema.

Input: The source (or destination) of user u_k

Output: Alternative bike b to rent (or location b') to return

h_i (or h_j) $\leftarrow$ index of the region of location u_k (or u'_k)
$N(h_i)$ (or $N(h_j)$) $\leftarrow$ neighboring regions of h_i (or h_j)
Incentives set $I = (p_i^+(t), p_i^-(t)) \leftarrow$ the pricing scheme learned by the actor-critic network
for all bikes in $N(h_i)$ (or return locations in $N(h_j)$) **do**
    u_k calculates the net profit of source detour (or destination detour), i.e., incentive price minus detour cost
u_k chooses the bike b (or return location b') with maximum net profit which is denoted as p_{max}
if $p_{max} < 0$ **then**
    u_k refuses source incentive and leave the system, (or u_k returns bike to the original destination u'_k)
else
    return b (or b') as the target

B. A Hybrid Incentive Scheme

Either source incentive or destination incentive has its shortage of certain user dynamics. Therefore, besides only considering the source or destination incentive, we propose to combine these two kinds of incentives and build a hybrid incentive scheme. The hybrid incentive scheme could adaptively adjust the proportion between the source and destination incentive based on different imbalance situations.

In the hybrid incentive scheme, the system shows the source (or destination) incentive price for each nearby bike when users try to rent (or return) a bike. We assume users' decisions are made based on the pricing model. The state and action spaces in the MDP are enlarged because of the destination detour budget B^- and incentive price p^-. Specifically, a state vector s_t is constructed by $\sum_{h_i} \varphi_i(t)$, $\sum_{h_i} D_i(t-1)$, $\sum_{h_i} \Lambda_i(t-1)$, $B^+ - \sum_{h_i,t} p_i^+(t)$, $B^- - \sum_{h_i,t} p_i^-(t)$ and out-of-service events in previous time-slots. The first term represents the number of unused bikes over the city at the beginning of t. The total amount of bikes over the city is constant, but the number of unused bikes may vary over time because of the fluctuated usage of users. The $\sum_{h_i} D_i(t-1)$ and $\sum_{h_i} \Lambda_i(t-1)$ represent the total number of rent and return events over the city, which captures the temporal bike usage information to the MDP. The $B^+ - \sum_{h_i,t} p_i^+(t)$ is the remaining budget for the source incentive, and $B^- - \sum_{h_i,t} p_i^-(t)$ is the remaining budget for the destination incentive. The MDP ends either when t reaches the time-slot upper bound or the remaining budgets for both source and destination incentives are empty.

An action vector a_t for time-slot t contains the source incentive price $(p_i^+(t), i = 1, \ldots, n)$ and the destination incentive price $(p_i^-(t), i = 1, \ldots, n)$. The state transmission can be simulated via our environment model. The reward r of the incentive is constructed by rewards from source incentive $r^+(s_t, p^+)$ and destination incentive $r^-(s_t, p^-)$.

Because of the modification to the MDP, we extend the actor-critic framework in [3]. The size of the actor network is enlarged as shown in Fig. 6. The actor network 1 is used to learn the source incentive prices $p^+(t)$, and the actor network 2 is used to learn the destination incentive prices $p^-(t)$. As for the critic network, the sub-Q-value of each region h_i at step t is evaluated based on $(p_i^+(t), p_i^-(t))$ instead of just considering $p_i^+(t)$, and the estimation of Q-value changes correspondingly.

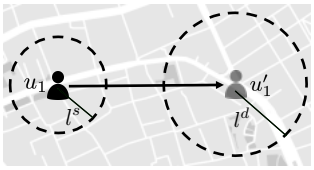


Fig. 5. Adaptively adjusting the maximum source and destination detour distance.

C. Adaptively Adjusting Source and Destination Incentives

Besides adjusting the learning framework, we also propose two ways to adjust the ratio of source and destination incentive price. One way is the budget division whose definition is shown as follows.

***Definition** 1 (Budget division):* Assume the total budget available is B, and the budget division ratio is ρ. Then the budget appointed to source incentive is ρB and the remaining $(1-\rho)B$ is used for the destination incentive.

Under this scheme, the remaining budget of the source and destination incentive at the initial state s_0 becomes:

$$B^+ = \rho B,\ B^- = (1-\rho)B$$

The overall reward during a day under policy π_θ becomes:

$$J_{\pi_\theta} = E[\sum_{k=0}^{\infty} \gamma^k (r^+(a_k, s_k) + r^-(a_k, s_k)) | \pi_\theta, s_0]$$

The other way is to adjust the ratio between detour distances of the source and destination incentives. It is achieved by adding the maximum source and destination detour constraints to users in the environment model. Let l denote the maximum detour distance that a user can accept, including source and destination detours. The value of l can be extracted from a user survey when applying the scheme in the real world. l can be split into two parts: l^s and l^d which represent the maximum source and destination detour correspondingly. Let α denote the adjust parameter between l^s and l^d.

***Definition** 2 (Detour distance division):* Given the maximum detour distance l of each user and parameter α, maximum detour under source incentive is $l^s = \alpha l$ and the detour under destination incentive is $l^d = (1-\alpha)l$.

We assume the user rejects to detouring either when his/her detour distance exceeds the limitation or he/she cannot gain profit from the detour. By setting source and destination detour distance limitation, we try to limit the source and destination incentive price.

Formally, based on α, we attempt to limit the source and destination incentives as:

$$p_i^+(t) < C + \eta\alpha l^2 \quad \text{and} \quad p_i^-(t) < C + \eta((1-\alpha)l)^2 \quad \forall t \in T$$

The budget division strictly imposes restrictions on budgets of source and destination incentives, while the detour distance division restricts the source and destination incentive price on estimation. Either kind of incentive is adaptive among

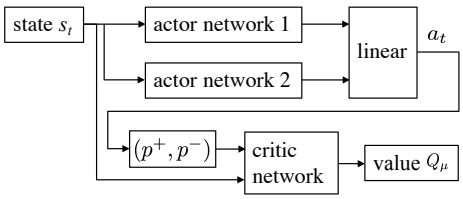


Fig. 6. The learning framework for the hybrid incentive scheme.

regions, and the sum of incentive prices cannot exceed the corresponding budget. The budget division is applied to the initial state of the MDP. The detour distance division is applied to the environment, the incentive greater than the limitation cannot bring benefits to the scheme.

IV. Experiment

A. Experiment Setup

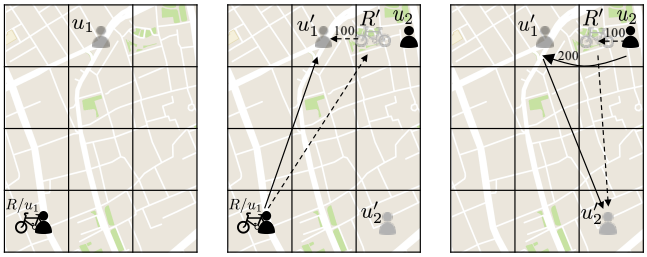
We use the data published by *Mobike* to construct our real-world dataset. We use a set of one-month history trip data from 8/1/2016 to 9/1/2016. Our Mobike dataset contains more than 100k trip records of Shanghai. The record of each trip includes trip duration (in seconds), trip start (end) time and date, start (end) latitude and longitude, etc.

In our experiment, the environment model is built on *OpenAI Gym*, a toolkit for comparing reinforcement learning algorithms. Specifically, a day is temporally divided into 24 time-slots and the Shanghai city is spatially divided into 20×40 regions. The effective area of the city is bounded by [30.841°N, 31.477°N] and [120.486°E 121.971°E]. Users' request time, locations and destinations are extracted from the Mobike trace data. Through the statistic of unique bike ID, totally there are 79,063 bikes used in the dataset. Considering the retirement of broken bikes, the actual number of bikes may be less than that amount.

When training the hybrid incentive scheme, the Adam algorithm [9] is used to optimize both actor and critic networks. The learning rates for training both parts are set as 10^{-4}. In each step, to explore the more action space, Gaussian noise is added to each action generated from the actor network. Although [8] proposed to add Uhlenbeck-Ornstein noise to actions, the Gaussian noise is used for simplicity. The discount factor γ in the MDP is chosen as 0.99.

In the first set of experiments, we compare the performance of our algorithm with others under different budgets. In the experiment, the budget is varied from 1,000 to 2,000 and the performance is quantified by the Decreased Unserviced Ratio (DUR) defined in [3]. The number of unserviced user increases by one if a user cannot find a bike and he/she is not satisfied with any source incentives offered by the system. Let N_1 denote the number of unsatisfied users without incentive, and N_2 denote the number of unsatisfied users with incentive. Then, the corresponding DUR is defined as $(N_1 - N_2)/N_1$.

The second set of experiments focus on the number of satisfied users under different budgets. We assume the BSS operator gets a reward of 1 for each user who rents the bike. The cost of users is bounded by 5 by setting C and η in the cost model. The profit of the operator can be calculated by



(a) t=0 (b) t=1 (c) t=2

Fig. 7. Illustration of combining the source and destination incentives.

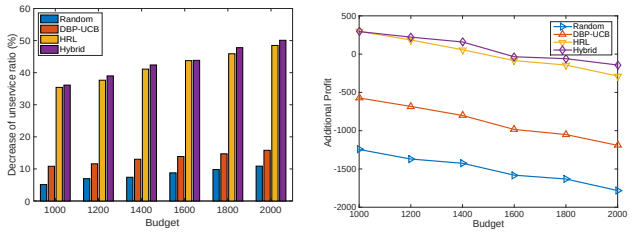
subtracting the budget spent for the incentive from the overall income of a day. We also conducted a set of experiment to test the influence of initial bike amount. If the initial bikes are sufficient enough, then each user can find a bike without a detour and the maximum service level is achieved. However, the number of bikes are limited in each region. Therefore, wisely spending the budget to achieve a better service level is important. The last set of experiment is focus on the rebalance performance across multiple days.

We first compare our approach with the source incentive scheme proposed in [3], and their approach is denoted as HRL. The second comparison algorithm is the DBP-UCB proposed in [4], which is one of the state-of-the-art bike rebalancing approaches based on user incentive. A randomized inventive scheme is used as a baseline.

B. Results

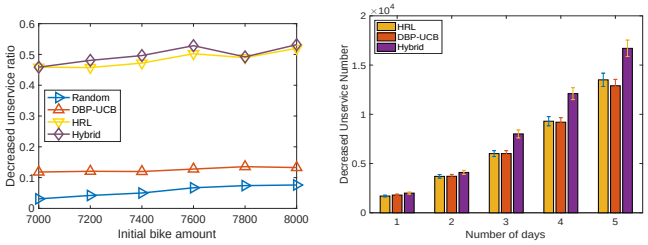
The decreased unserviced ratio under different budgets is shown in Fig. 8(a). From the figure, we can conclude that the performance of our hybrid approach achieves better performance than others. Comparing with the HRL that just considers the source incentive, we can conclude that adaptively allocating incentive on source detour and destination detour can bring additional benefits to the service level. It is reasonable since the source and destination incentives are included in the action spaces of the hybrid incentive scheme. By comparing HRL and DBP-UCB we can conclude that using the reinforcement learning can greatly improve the service level since it considers further reward when choosing the action for each state. The performance trend of all approaches shows that more user requests can be satisfied with a higher budget, even for the randomized policy.

The additional profit brought by the incentive is illustrated in Fig. 8(b). As stated in [3], the HRL can bring additional benefits to the BSS operator when the budget is not too large. The hybrid incentive scheme also can gain profits from the incentive which is arguably one of the most important features to BSS operators. However, when budgets increase, the profit decreases. It illustrates that the number of satisfied users increases more slowly with the increasing budget. That is to say, it is not necessary for BSS operators to fully rebalance the system. The totally rebalanced system means that all user requests can be satisfied. The DBP-UCB and randomized scheme can bring additional profits to the system with a budget less than 1,000 in our test.



(a) Comparison on DUR. (b) Comparison on additional profit.

Fig. 8. Comparison on DUR and additional profit with of varying budget.



(a) Comparison on bike amount. (b) Long-term comparison.

Fig. 9. Cumulative density function and long-term performance comparison.

Fig. 9(a) shows the influence of the initial bike amount. With more bikes placed in the city, incentive schemes are more likely to achieve better performance. The increase ratio in the figure is not significant. The reason could be that the additional bikes are uniformly distributed among the city. Adding bikes to regions with nearly no user request may cause the waste of bikes. If the distribution of the initial bike could fit user request distribution, increasing initial bike amounts may greatly decrease the unservice ratio.

Fig. 9(b) shows the rebalance performance over multiple days. We count the number of reduced unservice events. The difference between the HRL and the hybrid incentive scheme increases when the number of days increases. It shows that the hybrid incentive scheme can keep a better bike distribution than the HRL. As we show in previous sections, the destination incentive scheme is more likely to place bikes on regions with more requests. These bikes are more likely to be used when the number of time-slots increases. It may explain that the advantage of the hybrid scheme is more obvious when the number of time-slots is larger.

Although our algorithm has better performance on the selected Mobike dataset, its benefits in the real world dock-less BSS is still untested. [3] shows the possibility of applying the HRL algorithm to deal with the high dynamics of the system and makes it feasible to learn the incentive price on dockless BSSs. We further enhance their framework by considering both source and destination incentives and provide more flexibility to the system. We show the benefit brought by the destination incentive as well as its combination with source incentive. Although we cannot be certain that the performance of combination always performs better than source incentive only or destination incentive only, it gives the BSS operator a chance to adjust the incentive policy in a different area according to the specific user dynamic in the area.

V. Related Work

With the booming of the bike sharing, more and more researchers have devoted their effort to related issues including user demand prediction [10–12], bike rebalance strategy [4, 13, 14], station location optimization [15, 16], bike lane planning [17], suggestion of user's journeys [18, 19]. We focus on the studies that have been conducted on rebalance strategy, which are closely related to our work.

Rebalancing strategies designed for docked BSSs are have been widely studied. Typically, there are two major approaches which are the truck-based and the user-based approach. The truck-based approach such as [20, 21] means the BSS operator hires a fleet of trucks to transport bikes from overflow stations to underflow stations. Liu et al. [14] proposed a method that first clusters bike stations according to geographic information and station status, and then assigns a truck to each cluster. The routing for each truck used in rebalancing is modeled as an integer programming problem.

As for user-based approach like [4, 22], the BSS operator gives incentive to users and encourages them to rent or return bikes at certain stations. User-based approaches expect that the BSS can achieve self-balance. They improve the overall service level by controlling user's dynamics through incentive. Designing the pricing mechanism is the key problem in these approaches. However, these approaches are focused on rebalancing docked BSSs and cannot be directly used in our dock-less BSSs rebalancing problem.

As for dock-less BSS, besides the source incentive scheme based on reinforcement learning proposed by Pan et. al [3]. Caggiani et. al [23] proposed a dynamic bike rebalance method including a prediction scheme of the number and position of bikes and a relocation decision system. Our hybrid scheme is a end-to-end system and the incentive price can be given without demand prediction.

VI. Conclusion

In this paper, we show the underflows and overflows caused by imbalanced bike distribution in the dock-less BSS, which may decrease the service level of BSSs or bring congestions to the city. To avoid the negative impacts, we propose to rebalance the dock-less BSS via adaptive source and destination incentives with an objective that maximizes the service level over a day. The problem is modeled by a Markov decision process, and we adapt the deep reinforcement learning framework in [3] which only considers the source incentive. The combination of source and destination detours provide the system operator with a more flexible approach to rebalance the dock-less BSSs according to varied user dynamics across different areas. The experiments are conducted based on real-world trace data extracted from *Mobike* dataset. The experiments show that our adaptive approach can achieve a higher service level in comparison with the state-of-the-art approaches including the original one that only considers the source incentive.

Acknowledgement

This research was supported in part by NSF grants CNS 1824440, CNS 1828363, CNS 1757533, CNS 1629746, CNS 1651947, and CNS 1564128.

REFERENCES

[1] P. DeMaio, "Bike-sharing: History, impacts, models of provision, and future," *Journal of Public Transportation*, vol. 12, no. 4, p. 3, 2009.

[2] S. A. Shaheen, E. W. Martin, A. P. Cohen, N. D. Chan, and M. Pogodzinski, "Public bikesharing in north america during a period of rapid expansion: Understanding business models, industry trends & user impacts, mti report 12-29," 2014.

[3] L. Pan, Q. Cai, Z. Fang, P. Tang, and L. Huang, "Rebalancing dockless bike sharing systems," *arXiv preprint arXiv:1802.04592*, 2018.

[4] A. Singla *et al.*, "Incentivizing users for balancing bike sharing systems." in *Proc. of AAAI*, 2015, pp. 723–729.

[5] P. Dayan and G. E. Hinton, "Feudal reinforcement learning," in *NIPS*, 1993, pp. 271–278.

[6] T. Dean and S.-H. Lin, "Decomposition techniques for planning in stochastic domains," in *Proc. of IJCAI*, vol. 2, 1995, p. 3.

[7] R. Ashar, "Hierarchical learning in stochastic domains," Ph.D. dissertation, Citeseer, 1994.

[8] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[10] L. Chen *et al.*, "Dynamic cluster-based over-demand prediction in bike sharing systems," in *Proc. of ACM Ubicomp*, 2016, pp. 841–852.

[11] Y. Li *et al.*, "Traffic prediction in a bike-sharing system," in *Proc. of ACM SIGSPATIAL*, 2015, p. 33.

[12] Z. Liu, Y. Shen, and Y. Zhu, "Inferring dockless shared bike distribution in new cities," in *Proc of ACM WSDM*, 2018, pp. 378–386.

[13] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proc. ACM WWW*, 2013, pp. 1167–1178.

[14] J. Liu, L. Sun, W. Chen, and H. Xiong, "Rebalancing bike sharing systems: A multi-source data smart optimization," in *Proc. ACM SIGKDD*, 2016, pp. 1005–1014.

[15] J. Liu, Q. Li, M. Qu, W. Chen, J. Yang, H. Xiong, H. Zhong, and Y. Fu, "Station site optimization in bike sharing systems," in *Proc. of IEEE ICDM*, 2015, pp. 883–888.

[16] L. Chen, D. Zhang, G. Pan, X. Ma, D. Yang, K. Kushlev, W. Zhang, and S. Li, "Bike sharing station placement leveraging heterogeneous urban open data," in *Proc. of ACM Ubicomp*, 2015, pp. 571–575.

[17] J. Bao, T. He, S. Ruan, Y. Li, and Y. Zheng, "Planning bike lanes based on sharing-bikes' trajectories," in *Proc. of ACM SIGKDD*, 2017, pp. 1377–1386.

[18] J. W. Yoon *et al.*, "Cityride: a predictive bike sharing journey advisor," in *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*. IEEE, 2012, pp. 306–311.

[19] J. Zhang, P. Lu, Z. Li, and J. Gan, "Distributed trip selection game for public bike system with crowdsourcing," in *INFOCOM*. IEEE, 2018.

[20] D. Chemla, F. Meunier, and R. W. Calvo, "Bike sharing systems: Solving the static rebalancing problem," *Discrete Optimization*, vol. 10, no. 2, pp. 120–146, 2013.

[21] S. Ghosh, M. Trick, and P. Varakantham, "Robust repositioning to counter unpredictable demand in bike sharing systems," 2016.

[22] A. Waserhole and V. Jost, "Pricing in vehicle sharing systems: Optimization in queuing networks with product forms," *EURO Journal on Transportation and Logistics*, vol. 5, no. 3, pp. 293–320, 2016.

[23] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, "A modeling framework for the dynamic management of free-floating bike-sharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 159 – 182, 2018.