A CROWDSOURCED DYNAMIC REPOSITIONING STRATEGY
FOR PUBLIC BIKE SHARING SYSTEMS

I-LIN WANG* AND CHEN-TAI HOU
Department of Industrial and Information Management
National Cheng Kung University, Tainan 704, Taiwan

ABSTRACT. Public bike sharing systems have become the most popular shared economy application in transportation. The convenience of this system depends on the availability of bikes and empty racks. One of the major challenges in operating a bike sharing system is the repositioning of bikes between rental sites to maintain sufficient bike inventory in each station at all times. Most systems hire trucks to conduct dynamic repositioning of bikes among rental sites. We have analyzed a commonly used repositioning scheme and have demonstrated its ineffectiveness. To realize a higher quality of service, we proposed a crowdsourced dynamic repositioning strategy: first, we analyzed the historical rental data via the random forest algorithm and identified important factors for demand forecasting. Second, considering 30-minute periods, we calculated the optimal bike inventory via integer programming for each rental site in each time period with a sufficient crowd for repositioning bikes. Then, we proposed a minimum cost network flow model in a time-space network for calculating the optimal voluntary rider flows for each period based on the current bike inventory, which is adjusted according to the forecasted demands. The results of computational experiments on real-world data demonstrate that our crowdsourced repositioning strategy may reduce unmet rental demands by more than 30% during rush hours compared to conventional trucks.

1. Introduction. The recent boom in the sharing economy has led to new businesses and has changed how people live in several ways. For example, the public bike systems in many metropolitan areas have provided more convenient access to the public transportation system, e.g., to MRTs, trains, or even busses. Although the business model of bike sharing systems remains questionable in terms of profitability, according to [20], by December 2020, more than 2003 cities worldwide had installed bike sharing systems, and more than 9.4 million public bikes and pedal electric cycles (pedelecs) were in use. Indeed, a bike sharing system is perfectly suitable for providing the first- and last-mile connections to public transportation systems. By deploying the rental sites at a suitable density (e.g., each site is within 300 m-500 m from another site), one can easily rent a bike from a site nearby, ride and return it to another site to take MRT, exit the MRT station and ride another bike to a site that is closer to the destination.

2020 Mathematics Subject Classification. Primary: 90B06, 90B10; Secondary: 90C10.
Key words and phrases. Bike sharing; crowdsourced; repositioning; integer programming; minimum cost network flow; random forest.

The first author is supported by MOST grant 110-2221-E-006-191-MY3.
* Corresponding author: ilinwang@mail.ncku.edu.tw.
The success of a bike sharing system depends on the quality of service in the following aspects: (1) the ease of accessing a rental site and (2) the ease of taking or returning a bike. The first aspect corresponds to a long-term strategic-level network design decision: the locations and density of rental sites should be carefully selected to provide easy accessibility to the users and efficient connections to the public transportation systems. The second aspect corresponds to a series of short-term tactical- or operational-level decisions, where bike fleets of suitable sizes should be deployed at various times. To this end, the system managers must redistribute bikes among sites to satisfy the expected bike rental or return demands. This requires the completion of several challenging tasks: (1) accurate prediction of these demands, namely, the timing and the number of bikes to be taken or returned by customers at each rental site; (2) determination of the optimal bike inventory in each site at any time; and (3) effective bike repositioning to satisfy the optimal inventory levels in (2). Failure to reposition these bikes would result in shortages in bikes or empty racks and the demands to take or return bikes whenever necessary would not be satisfied, which would damage the service quality and discourage the use of such systems.

To the best of our knowledge, most current bike sharing systems hire trucks to reposition bikes. These trucks are typically of small or medium size for easy parking and movement in metropolitan areas and can carry up to approximately 20 bikes. To load a bike onto the truck, a staff member checks out a bike from a rack and moves it to the truck. According to our survey, loading or unloading a bike takes at least 30 secs. A truck takes approximately 15 min to finish at most 30 loading/unloading tasks at each rental site in each stay. Driving to another site takes approximately 15 min on average. As a result, a truck may conduct $30 \times 2 = 60$ loading/unloading operations per hour. Assuming there are $N$ trucks and each truck works for 18 hr (e.g., 06:00 to 24:00) without rest, we estimate the upper bound on the number of loading/unloading operations within one day to be $60 \times N \times 18 = 1080 \times N$.

Considering the rental data of YouBike in 2014, for example, there are approximately 200 rental sites, $N = 10$ trucks, and approximately 5000 bikes in total. Each truck can visit $2 \times 18 = 36$ sites, and each site is visited by at most $360/200 = 1.8$ trucks on average per day, namely, at any moment, only $10/200 = 5\%$ sites can be served by trucks.

In addition, since each bike repositioning involves one loading process and one unloading process, the $1080 \times 10$ times (upper bound) of loading and unloading processes lead to roughly at most $1080 \times 10/2 = 5400$ bikes to be repositioned in one day. In 2014, YouBike had approximately 40000 daily rentals on average; hence, the effects of the 5400 repositioned bikes only contribute about $5400/40000 = 13.5\%$ (approximately 1/8) of the daily OD rental demands. Namely, if there are additional 13.5\% OD rental demands taking place, the 10 trucks can only serve those additional demands in the worst case, and leaves the original OD demands unserved.

According to the performance analysis for the repositioning trucks that is presented above, repositioning bikes using trucks is highly ineffective. The only way of improving the effectiveness is to increase the number of trucks; however, this would cause increased air pollution and traffic jams and would contradict the underlying philosophy of bike sharing systems to reduce the use of fueled vehicles and to reduce carbon emissions. In fact, even if we equip each station with one or more trucks, the imbalance between the rental demands and bike inventory may never be canceled out for most cases. It is because any bike travel, either by the customers or by the
A CROWDSOURCED DYNAMIC REPOSITIONING STRATEGY

hired crowds, requires transportation time. As a result, any current imbalanced rental demands cannot be remedied right away if all of its adjacent stations have no bikes to be repositioned. Here we use a simple example of two stations, eight bikes, and eight customers at time 0 and 1, to demonstrate this property. Suppose initially (at time 0), stations A and B have two and six bikes, respectively. Let the travel time between A and B be 1 unit. Consider two customers to travel from A to B and six customers to travel from B to A at both times 0 and 1. Then, all the customers at A and B at time 0 can successfully rent bikes. This leads to the bike inventory at time 1 to be 6 and 2 at A and B, respectively. Then, at time 1, the two customers at A can rent bikes, which leads to 6 − 2 = 4 idle bikes there. On the other hand, at time 1, only two of the six customers at B can rent bikes, which leads to 6 − 2 = 4 unsatisfied rental demands there. In this example, even if we have sufficiently many trucks (i.e., two) standing by at A and B, the 4 idle bikes at A can not be repositioned to B in time to remedy the 4 unsatisfied customers at B. This example demonstrates the limit of service quality. That is, even with an unlimited number of trucks to conduct repositioning tasks and with perfect demand forecastings, some demands may never be satisfied. This observation intrigues us to investigate the ”minimum unsatisfied rental demands” for estimating the ”ideal service quality” for a bike sharing system. Based on this concept, in section 3, we define the ”ideal bike inventory” for each station at the beginning of each period as a target bike inventory to achieve the best possible service quality that minimizes the total unsatisfied rental demands.

To improve the repositioning performance without increasing the number of repositioning trucks, we propose a novel crowdsourced repositioning scheme for repositioning bikes with less effort, less air pollution, and lower costs. In addition, our proposed scheme may strengthen the loyalty of users to the system.

We analyze the historical rental data and use design visualization tools to help system managers understand rental trends. Based on these data, we solve an integer program (ideal inventory model, IIM) that optimizes the bike inventory level for each station in each time period (e.g., every 30 min) under the assumption that we can always reposition bikes if necessary. To predict inventory changes in the near future, we use the random forest algorithm to identify important factors for the rental trends and use these factors to more accurately predict rentals within the next time period. With more accurate predictions of the inventory trends and of the target inventory levels for each site in each time period, we formulate a linear program, namely, the voluntary rider flow model (VRFM), which is a minimum-cost flow problem for calculating the optimal bike flows for recruiting voluntary riders to satisfy the target bike inventory levels for the upcoming time period.

To recruit voluntary riders, we suggest enhancing the membership database by adding records of voluntary rides into the ordinary historical riding records for each member. To the best of our knowledge, most bike sharing systems do not even store the historical riding data for each member. The bike sharing systems should also provide easy access to the riding records for each member via websites or applications (APPs) in smart phones. Via our proposed scheme, the system can design bonus points that encourage members to ride more frequently and can offer additional bonus points when inviting voluntary riders to conduct repositioning missions. For example, the system can announce or spread information about the voluntary missions on the website or APPs and bonus points can be earned by completing each mission, which can be cashed out for additional free riding time or
gifts. By encouraging volunteer riders to conduct missions in exchange for a cashing out bonus, shared vehicles can be simultaneously repositioned at many rental sites. This crowdsourced repositioning strategy has at least 3 advantages: (1) prompt responses to the repositioning of bikes at more sites simultaneously, compared with conventional trucks; (2) cost savings in hiring trucks and staff, which, in turn, reduces the use of fueled vehicles and traffic jams; and (3) increasing the loyalty of members and improving the relationships with other companies that provide services for members in cashing out their bonus points. To the best of our knowledge, similar bonus schemes have been utilized on very few occasions (e.g., encouraging uphill riders in hilly areas); however, they have not been applied for general-purpose use. This renders our contribution more significant since we may be the first to determine voluntary repositioning OD pairs via theoretical mathematical models and algorithms, rather than via intuitive marketing techniques.

The remainder of this paper is organized as follows: Section 1 introduces background information, the drawbacks of the current repositioning scheme, and the advantages of our crowdsourced repositioning scheme. Section 2 reviews the related literature. Section 3 explains the mechanism of IIM for calculating the ideal optimal bike inventory and the details and effects of our crowdsourced repositioning scheme, namely, VRFM. Section 4 presents the data analysis and the computational tests that were conducted on our proposed model. Section 5 presents the conclusions of the paper and suggests topics for future research.

2. Literature Review. There are two major research fields in the literature that are related to bicycle sharing systems: the location problems for installing rental stations [6, 33] and the data analysis and repositioning strategies for balancing the bicycle inventories among all rental sites. Here, we focus on repositioning-related literature.

[28] proposed a stochastic network flow model with proportionality constraints on a time-expanded graph for estimating bicycle flows. Using the transit data of Singapore to simulate potential OD demands for bike sharing systems, they evaluated the performance of dynamic repositioning in improving the utilization of shared bikes.

[13] investigated how weather conditions affect the bike usage trend for the bike sharing system in Washington DC, USA. They found that riders in the rain are typically registered users or those with private bikes, but not nonregistered users. [2] presented an interactive visualization system for displaying the rental data of Boston, Washington DC, and Chicago at various times and locations. Their system can also present the busiest sites at any time. [22] list and compare the locations of rental sites, rental data, and weather conditions for 30 bike sharing systems at various times. [26] implemented a similar system but specified the usage as a percentage and grouped the rental sites according to the usage.

[29, 30] calculated the locations for installing bikes and racks by mining the rental data. [21] proposed an algorithm for identifying the trend of inventory changes based on the rental data. [11] investigated the demand changes in weekdays versus weekends, the relation between the rental frequencies and locations, and important factors that affect the rentals. They also evaluated four methods for predicting the future demands of the next time period based on the current inventory with errors of up to 15%. [16] proposed a prediction model that is based on the autoregressive moving average (ARMA). [35] developed an ARMA-based model that further
considered the seasonal and spatial factors and claimed that this model outperforms previous ARMA and Bayesian models for the bike sharing system in Dublin. [27] determined the service level requirements (the bike inventory range) at each rental site by treating the inventory at each rental site as a finite-buffer single-server nonstationary queuing system and using the Kolmogorov forward equations to calculate the service level requirements. Then, they designed optimal truck routings for repositioning bikes via mixed-integer programming that are based on a clustering problem that decomposes the multivehicle repositioning problem into single-vehicle repositioning problems.

[25] analyzed 3 bike sharing systems in USA and identified factors that are used in the prediction model via multivariate regression analysis. [5] claimed that Bayes classifiers can outperform regression-type algorithms in terms of prediction accuracy for Citi Bike in New York and proposed the station occupancy predictor for predicting short-term future bike inventories. Recently, [34] analyzed the rental data of the bike sharing system in Hangzhou, China, and proposed a prediction model for bike rentals via the random forest algorithm (RF), which they claimed realizes higher prediction performance than several other algorithms. They use the bike rental model to estimate the bike returns, similar to the simulation models by [32]. RF has some advantages [19]: (1) applicable for data of few assumptions, (2) very efficient to quickly deal with a large amount of data, and (3) more robust against overfitting. However, it requires the observation to be independent, which fits our assumption of independent bike rentals. [1] have used RF and Least-Squares Boosting (LSBoost) as univariate regression algorithms to model the bike availability at San Francisco and reported RF has consistently better prediction performance. In this paper, we will also use RF to predict incoming demands.

The static repositioning problem [3, 5, 15, 24] investigates how to move bikes at night when there are very few or no demands for satisfying the target initial bike inventory for each site. In contrast, the dynamic repositioning problem [7, 14, 10, 31] calculates the routes of the repositioning trucks and the number of bikes to be loaded or unloaded at each site. The integer programming models for solving these repositioning problems typically cannot handle cases of more than 60 rental sites due to complexity issues. In addition, several heuristics [14, 10, 31] have been proposed; however, they perform poorly. [17] proposed a reservation scheme that enables users to reserve a bike/car and an empty rack/spot for public bike and car sharing systems. Via simulation, they concluded that introducing the reservation scheme could reduce the waiting time for renting/returning a shared vehicle, as expected. Recently, [12] proposed a moment-based model and a new hybrid approach that combines a fuzzy C-means (FCM)-based genetic algorithm (GA) with a backpropagation network for effectively predicting the rental demand. They concluded that rental demand is strongly related to the weather conditions and that the night-time demand can be predicted more accurately than the day-time demand.

Most related studies in dynamic repositioning use historical rental data as fixed OD demands to plan the inventory routing without explicitly calculating the optimal bike inventory. In practice, bike sharing systems tend to set a target bike inventory in advance for each rental site for any time period; hence, the staff who are in charge of repositioning bikes have an easier objective. [23] proposed an inventory model for a single rental site that considers user behavior. In addition, they proposed an efficient method for estimating a user dissatisfaction measure based
on a Markov chain. However, their methods assume that the bike replenishment cycle is known in advance and are not highly suitable for dynamic repositioning in which the bike inventory is also strongly related to the dynamic rental demands and the replenishment decisions that are made on other rental sites. Our proposed mathematical model, which we present in the next section, can use these factors to calculate an ideal bike inventory that is not based on Markov chain or queueing theories, but by assuming that voluntary crowds are always available to help reposition bikes.

Crowdsourced repositioning has been practiced more frequently in free floating systems such as Mobike in China. Starting in 2017, they labeled bonus bikes with a red envelope and any rider who helps reposition such a bike receives the corresponding bonus. The approach for determining whether to label a bonus bike is not clear; however, such a repositioning task only focuses on the removal of labeled bikes from their current locations, and no instructions on where to put these bonus bikes are provided. A similar scheme, namely, Bike Angels, has been evaluated experimentally on Citi bikes in New York. [9] investigated the impacts of this incentive crowdsourced repositioning program and developed a performance metric for both online and offline policies for specifying incentives.

To the best of our knowledge, [18] proposed the first crowdsourced repositioning model for dynamic repositioning. Assuming the OD demands for each site and each time period (e.g., 30 min) have been estimated, [18] incorporated possible voluntary riding OD arcs for each site in each period and solved a mixed-integer program to identify optimal voluntary riding assignments for each site in each period. However, Liaos model could not deal with real-time voluntary repositioning since it only uses historical average demands as inputs. In this paper, we have resolved this difficulty.

Based on the work of [18], we propose an ideal inventory model (IIM) that calculates the ideal inventory for each site at the beginning of each time period and a real-time voluntary rider flow model (VRFM) that optimizes the number of voluntary riders in each time period.

3. Proposed mathematical models and repositioning strategy.

3.1. Ideal Inventory Model (IIM). In this section, we try to estimate the limit of the service quality by calculating the maximum possible satisfied rental demands assuming an unlimited number of trucks to conduct repositioning tasks with perfect demand forecastings. This service quality limit can become a benchmark to evaluate the performance of any bike repositioning strategy. Based on this concept, we define and calculate the "ideal bike inventory" for each station at the beginning of each period. This ideal bike inventory can be used as a target bike inventory to achieve the best possible service quality that minimizes the total unsatisfied rental demands.

The optimal bike inventory for each rental site may vary with time and depends on the dynamic rental demands. To simplify the problem, we consider 30-min time periods (e.g., 06:00 - 06:30 - 01:00 - 01:30 -- 23:30 - 24:00), and we assume that the optimal bike inventory remains the same over each time period. Assume there are $N$ stations, $B$ bikes, $T$ time periods, and $U_i$ empty bikes for site $i$. Let $A$ denote the set of possible OD pairs. For each site $i$ in period $t$, let $b^t_i$ and $r^t_i$ represent numbers of bikes to be taken and returned, respectively, within period $t$. Assume there are at most $R^t$ voluntary riders available at the beginning of period $t$ and that each bike reposition by a voluntary rider must be completed within that period. The
assumption of same-period bike rental and repositioning accords with most real-world practices since most such systems encourage short-term rental (within 30 min) and voluntary repositioning.

We would like to determine the following variables: in each time period $t$, $x_{ij}^t$ represents the optimal voluntary rider flow for OD pair $(i, j) \in A$; for each site $i$, $I_i^t$ denotes the optimal bike inventory level at the end of period $t$, $\Delta U_i^t$ is the optimal number of bikes that exceed the capacity (the total number of racks), and $\Delta L_i^t$ is the optimal bike shortage. Let $\varepsilon$ represent a very small number. The ideal bike inventory model (IIM) can be formulated as follows:

$$\min \sum_{t=1}^{T} \sum_{i=1}^{N} (\Delta U_i^t + \Delta L_i^t) + \varepsilon \sum_{t=1}^{T} \sum_{(i,j) \in A} x_{ij}^t \quad (\text{IIM})$$

subject to

$$I_i^t = I_i^{t-1} - b_i^t + r_i^t - \sum_{(i,j) \in A} x_{ij}^t + \sum_{(j,i) \in A} x_{ji}^t - \Delta U_i^t + \Delta L_i^t \quad \forall t = 1, \ldots, T; \ i = 1, \ldots, N \quad (1)$$

$$\sum_{i=1}^{N} I_i^0 = B \quad (2)$$

$$\sum_{(i,j) \in A} x_{ij}^t \leq R^t \quad \forall t = 1, \ldots, T \quad (3)$$

$$0 \leq I_i^t \leq U_i, \ \Delta U_i^t \geq 0, \ \Delta L_i^t \geq 0 \quad \forall t = 1, \ldots, T; \ i = 1, \ldots, N \quad (4)$$

$$x_{ij}^t = 0 \quad \forall (i, j) \in A, \ t \leq \Delta ij \quad (5)$$

$$x_{ij}^t \geq 0 \quad \forall t = 1, \ldots, T; \ (i, j) \in A \quad (6)$$

The objective function aims at minimizing the mismatched demands (the surplus $\Delta U_i^t$ or the shortage $\Delta L_i^t$ of bikes). The second term, namely $\varepsilon \sum_{t=1}^{T} \sum_{(i,j) \in A} x_{ij}^t$, in the objective function, is used to reduce unnecessary voluntary rider flows. Constraints (1) define the flow balance relation for the bike inventory at the end of each period and at each site. Constraints (2) conserve the total number of bikes, whereas constraints (3) limit the total number of voluntary riders in each period. Constraints (4), (5), and (6) define the ranges of variables.

If unlimited voluntary riders are available anytime and anywhere, then we should always meet the optimal target inventory ($I_i^t$) anytime, anywhere. We call this target inventory the ideal inventory since such an inventory would satisfy the most rental demands, regardless of the repositioning costs. Thus, we will use this ideal inventory as a target inventory value for each site and each period in the real-time crowdsourced repositioning model, namely, VRFM.

3.2. Voluntary Rider Flow Model (VRFM). Since IIM assumes the rental data are all known deterministic parameters, we propose the voluntary rider flow model (VRFM), which can be regarded as a rolling-horizon version of partial IIM that is decomposed by periods, for dealing with the dynamic repositioning based on the real-time rental demand.

At the beginning of period $t$, for each site $i$, let $\bar{b}_i^t$ and $\bar{r}_i^t$ represent the predicted numbers of bikes to be taken and returned, respectively, which are based on historical rental data $b_i^t$ and $r_i^t$, and are adjustable via machine learning techniques such as
RF; let $\tilde{I}^t_i$ denote the real-time current bike inventory and $I^t_i$ the target optimal bike inventory that is calculated from IIM; and let $\tilde{x}_{ji}$ be the number of voluntary riders who are currently on the way from site $j$ and are expected to arrive at site $i$ during this period. Other parameters are the same as those of IIM.

We aim at determining $\tilde{x}_{ij}$, which is the optimal voluntary rider flow for each OD pair $(i, j) \in A$, and $\tilde{I}^t_i$, which is the planned ending inventory in the current period $t$, such that $\tilde{I}^t_i$ is as close to $I^t_i$ as possible with minimal repositioning efforts. At the beginning of period $t$, we can define the following linear program, which we denote as VRFM$^i$:

$$\min \sum_{i=0}^{N} |\tilde{I}^t_i - I^t_i| + \varepsilon \sum_{(i, j) \in A} \tilde{x}_{ij}$$

(VRFM$^i$)

subject to

$$\tilde{I}^t_i = \tilde{I}^{t-1}_i - \tilde{b}_i^t + \tilde{r}_i^t - \sum_{(i, j) \in A} \tilde{x}_{ij}^t + \sum_{(j, i) \in A} \tilde{x}_{ji}^t \quad \forall i = 1, ..., N$$

(7)

$$\sum_{(i, j) \in A} \tilde{x}_{ij}^t \leq R^t \quad \forall t = 1, ..., T$$

(8)

$$0 \leq \tilde{I}^t_i \leq U_i \quad \forall i = 1, ..., N$$

(9)

$$\tilde{x}_{ij}^t \geq 0 \quad \forall (i, j) \in A$$

(10)

The linear program VRFM$^i$ corresponds to the following minimum cost flow problem: Draw $N$ nodes from the left, with indices $i_L^t$, $i = 1, ..., N$, for representing each site at the current time $t$; draw another $N$ nodes from the right, with indices $i_R^t$, $i = 1, ..., N$, for representing each site in the next period. We associate each left node $i_L^t$, $i = 1, ..., N$ with a value $\tilde{I}^{t-1}_i - \tilde{b}_i^t + \tilde{r}_i^t$, which is its expected inventory prior to repositioning. We associate each right node $i_R^t$, $i = 1, ..., N$, with a value $I^t_i$, which is its target optimal inventory after repositioning. Then, for each possible OD pair $(i, j) \in A$, we construct a repositioning arc $(i_L^t, j_R^t)$. We also construct $N$ inventory arcs $(i_L^t, j_R^t)$, $i = 1, ..., N$.

Figure 1 illustrates a small VRFM example in which there are 4 rental sites with expected bike inventory vector $(3, 10, 1, 0)$ and target optimal inventory vector $(5, 6, 2, 4)$. There will be a shortage of 3 bikes since $(3+10+1+0) - (5+6+2+4) = -3$. Suppose we can only find voluntary riders to move from site 1 to site 2, 1 to 3, 2 to 4, 3 to 1, and 4 to 1. Figure 1(a) illustrates the original graph, which can be converted to the graph in Figure 1(b) via the following steps:

1. Add a new dummy node $S$ as a source (with supply $\sum_{i=1}^{N} (\tilde{I}^t_i - I^t_i)$) or a sink (with demand $\sum_{i=1}^{N} (I^t_i - \tilde{I}^t_i)$).
2. Add $2N$ dummy arcs $(j_R^t, S)$ and $(S, j_R^t)$, for $j = 1, ..., N$.
3. Associate each left node $i_L^t$ with a supply $\tilde{I}^t_i$ and each right node $i_R^t$ with a supply $I^t_i$.
4. Associate each dummy arc with cost 1, each inventory arc with cost 0, and each repositioning arc with cost $\varepsilon$.

The transformed graph represents a minimum cost flow problem, where flows are sent from supply nodes (the left nodes or $S$) to demand nodes (the right nodes or $S$) via uncapacitated arcs with minimum total costs. The optimal solution to this minimum cost flow problem corresponds to an optimal repositioning strategy.

The above network transformation ignores constraints (8), namely, the availability of voluntary riders. If the optimal solution to the minimum cost flow problem
requires more than $R^t$ voluntary riders to be hired, then we can randomly remove
the extra riders from the solution without affecting the optimality. Consider Figure 1 as an example: An optimal solution may reposition 4 bikes from sites 2 to
4, thereby leaving shortages of 2 bikes and 1 bike at sites 1 and 3, respectively.
However, if we can hire at most 3 voluntary riders, then we may simply lay off 1
rider from the previous arrangement.

3.3. Proposed crowdsourced dynamic repositioning strategy based on I-
IM and VRFM. Here, we summarize our online crowdsourced dynamic reposition-
ing mechanism. Based on historical bike rental data $b^t_i$ and $r^t_i$ for each rental
site $i$ and each period $t$, we first calculate the ideal inventory $I^t_i$ using our IIM mod-
el. Then, for each rental site $i$ at the beginning of period $t$, which has actual bike
inventory $\bar{I}^t_i$, predicted bikes to be rented $\bar{b}^t_i$, and predicted bikes to be returned $\bar{r}^t_i$, we can solve our VRFM$^t$ model to calculate the optimal crowd assignment $\tilde{x}^t_{ij}$ such that the resultant bike inventory $\tilde{I}^t_i$ at the end of this period is as close to
the ideal bike inventory $I^t_i$ as possible. We can repeat this procedure by solving
the VRFM$^t$ model for each period $t$ in a rolling-horizon fashion and by using the
real-time data to conduct the crowdsourced dynamic repositioning.

4. Data analysis and evaluation.

4.1. Random forest (RF) and inputs for VRFM. Any real-time repositioning
decision requires an accurate prediction of the bike rentals in the near future. To
this end, we use RF, which was proposed by Breiman [4] in machine learning, to
construct multiple decision trees for the classification of important features and
regression on the rental demands. Here, RF is simply used as a tool for obtaining
more accurate forecasted rental demands. We only implemented a generic RF, yet
our evaluation results demonstrate that our RF yields more accurate predictions
than other methods that we implemented, such as linear regression and ARIMA.

Using the 10-month historical YouBike rental data in 2014 as the dataset, the
classification errors converge after more than 150 decision trees have been construct-
ed. We construct 500 trees since this yields almost the same errors as 3000 trees
but requires much less time. Based on our evaluation results, we have selected 7
important features that affect the prediction: holiday or not, weekday or not, day
(Monday, Sunday), current hour, numbers of checkouts and returns in the previous period, temperature, and rainfall. For each site, we construct a random forest, which can output predicted values for the numbers of checkouts ($\bar{b}_i$) and returns ($\bar{r}_i$) with the 7 specified parameters.

Therefore, with the current bike inventory $I_{t-1}^i$, the optimal target inventory $I_t^i$, the predicted numbers of checkouts $\bar{b}_i$ and returns $\bar{r}_i$, and the estimated total number of available voluntary riders $R_t^i$, we can solve the VRFM for the optimal voluntary rider flows $\tilde{x}_{ij}^t$, that we try to hire at the beginning of this period. Then, we repeat the same procedures for each period until the end of the day.

4.2. Simulations of repositioning strategies. To evaluate the performance of our crowdsourced repositioning strategy, we use two simulation settings: (1) sampled real daily data (e.g., the rental data of 2014.03.05 or 2014.03.11) and (2) a sample from a set of mixed real data (e.g., if we mix the 85 sunny weekdays in the period of 2014.01-05, then each rental record has a probability of 1/85 of being selected; we repeat this sampling 100 times for an entire day). Note that we pick up weekdays instead of weekends (and holidays) for our tests because we focus more on the rental demands of commuting purpose rather than the leisure purpose. The weekday rental demand patterns are more fluctuating with a few peaks in rush hours (e.g., 07:30-09:00, 11:30-13:30, 17:00-19:00, 20:30-21:30), whereas the weekend rental demand patterns are smoother. Since the operational challenges take place on weekdays more often than weekends, we decide to simulated weekday demands for testing.

The first simulation uses the random forest algorithm to predict the rental demands since the environmental information, such as the temperature or rainfall, is also available as input. Nevertheless, the second simulation can only use the historical average statistics (e.g., the average checkouts/ins at each site and in each period) for rental demand prediction since the random sampling would violate the data consistency that is required for using the random forest algorithm (e.g., previous rental records may correspond to different dates, which might differ substantially in terms of environmental data, and, thus, would not be applicable for the random forest algorithm). We use the optimal initial and ending bike inventories that are calculated using IIM in each period $t$ to set the initial and target optimal bike inventories in each period for simulation.

To simulate the crowdsourced repositioning strategy, we assume that we can always find voluntary riders if necessary, which yields an estimate for $R_t^i$, namely, the maximum number of available voluntary riders in each period $t$. In our evaluations, we find that by inviting $\sum_{t=1}^T R_t^i = 4458$ daily voluntary riders, the simulation would satisfy 94% 98% of the daily rental demands. Note that our mathematical model can easily deal with "limited" voluntary riders by adding a capacity constraint. Yet we conduct this experiment with unlimited crowd assumption to point out three observations: Firstly, the current practice (i.e., repositioning by trucks) is very ineffective; secondly, we show how to achieve the best possible service quality (i.e., repositioning by sufficiently many crowds); and thirdly, we explain how such an "ideal" objective could be pursued to some extent in practice.

For example, suppose each hired crowd only works in one period (i.e., 30 min) at one station and then leaves the system. Based on our experiments, the best result would suggest hiring 4458 riders in total to serve 200 stations over 36 periods in one day. It implies we may, on average, hire only $4458/(200 \times 36) \approx 0.62$ riders in
each period for each station if the OD rentals take place in a uniformly distributed manner over space and time. A more appropriate strategy would distribute those 4458 riders to the most active 20% stations (e.g., top 40 stations) at the busiest ten periods (e.g., 11:00-13:00, 17:00-19:00, and 21:00-22:00). Then we would only hire $4458/(40 \times 10) \approx 11.15$ riders on average at each rush-hour period for each busy station. If we view each engaged crowd as a virtual truck of capacity 1, we may distribute up to 12 virtual trucks to 12 different stations. Note that this is not equivalent to the strategy of hiring only 1 truck of capacity 20, because the latter strategy can only serve at most 1 station, yet the former strategy might serve 12 stations at one time. This explains the nature of the primary difficulty for bike repositioning: the key to reducing the bike imbalance relies on the OD-pair coverage rather than the node-only coverage. In other words, even if we hire one truck for each of the n stations, which is already too expensive and impractical, the $O(n)$ reposition trips in one period are likely insufficient to compensate for the imbalance caused by the $O(n^2)$ OD demands. Moreover, even if we do hire $O(n)$ trucks, the truck capacity utilization might be low. For example, if we split out the reposition of 11.15 bikes to 4 adjacent stations by 4 trucks, each truck only needs to carry $11.15/4=3$ bikes. Thus we think such repositioning tasks are more suitable by crowdsourcing.

To compare with our crowdsourced repositioning strategy, we also implement a truck repositioning strategy, where each truck works on repositioning within a service zone. Intuitively, we hope a service zone to be self-contained so that a truck inside a zone can pick up and deliver bikes among stations in the same zone. In other words, the number of clusters equal to the number of trucks (e.g., in our tests, that is 10 clusters). Theoretically, the service zone may change over periods for better service. In practice, the service zone does not change over time and are often decided by political districts. Here in our tests, we first use a K-means clustering algorithm to partition the rental sites into disjoint service zones such that each service zone has a similar amount of total net rental data (the difference between the total numbers of checkouts and returns) and the rental sites in each service zone are close to one another. The repositioning tasks within each service zone are conducted by a unique repositioning truck. We also assign repositioning missions every 30 min based on the real practices of YouBike (10-15 min to move to another site in the same service zone and 10-15 min to load/unload bikes). We assume a truck can carry at most 20 bikes and that every 30 min it will select a site that requires the most loading or unloading operations. For example, for a truck that is currently carrying 3 bikes at a site, if there are 4 other sites in the same service zone that require +5, -2, +10, and -14 bikes (+ denotes that a site requires bikes to be added, whereas - denotes that a site requires bikes to be removed), then the truck would go to the 4th site to move 14 bikes onto the truck. A similar process is conducted every 30 min.

Note that the proposed truck repositioning strategy (calculated using an algorithm) and the crowdsourced repositioning strategy (calculated using integer programming, IP) in this section are not optimized at the same level. We meant not to route trucks by IP because most IP-based dynamic repositioning strategies can only deal with small scale problems (up to around 60 stations for 10 trucks), according to literature ([7, 14, 15, 18, 28, 27]). In practice, most bike sharing systems hire trucks to cover fewer than 10% of stations at the same time due to budget
constraints. As a result, hiring more trucks would be impractical and computationally infeasible. This explains why we route trucks by heuristics only. Moreover, we believe the key to providing better service quality relies on providing more OD-pair than node-only coverage of rental demands. Hiring $O(n)$ trucks is too costly, yet hiring $O(n^2)$ crowds might still be possible to some extent in practice, should the manager propose some innovative demand planning strategies, as suggested in the end of section 1, to encourage crowd repositioning. Here we simply demonstrate how those target OD repositioning tasks can be allocated.

Using the first simulation setting of 18 real daily rentals, according to Figure 2, the failure percentage of prediction via the random forest algorithm is approximately 0.8%. According to Figure 3, the percentage of unmet demand has improved to 1-3.5% using the crowdsourced repositioning strategy compared to repositioning by 10 to 25 trucks for the 18 actual daily rental data. These results indicate that more accurate prediction (e.g., via the random forest algorithm) improves the service quality. According to Figure 4, crowdsourcing repositioning leads to 6% unmet demands on average for each day, which is approximately 2.8% lower compared to truck repositioning (approximately 8.8%). To analyze how this 2.8% improvement is realized over a day, Figure 5 breaks down the hourly distribution of unmet demands for both repositioning strategies. According to Figure 5, during rush hour (e.g., near dinner time at 18:00), the crowdsourced repositioning strategy can help satisfy approximately 32.5% more unmet demands than the truck repositioning strategy, namely, even if the crowdsourced repositioning provides an average 2.8% daily service quality improvement, most of the performance increase (approximately 32.5% improvement) occurs during the rush hours. This analysis provides solid evidence for the satisfactory performance of crowdsourced repositioning.
5. Conclusions. To reduce the unmet rental demands in bike sharing systems, we propose a novel crowdsourced repositioning scheme and show that it is more effective than the current truck repositioning strategy via mathematical formulations and simulations. Although a similar approach has been discussed for years,
to the best of our knowledge, our work is the first to present detailed mathematical models and conduct numerical experiments on how to implement the crowdsourced repositioning strategy.

We discuss the drawbacks of truck repositioning and propose a mathematical programming model (IIM) for calculating the ideal optimal bike inventory for each site in each time period under the assumption of unlimited availability of voluntary riders. Then, calculated ideal bike inventories for each time period $t$ are used as target values in our second simplified linear programming model (VRFM$_t$), which we have shown to be a minimum cost flow problem. At the beginning of a period, given the current bike inventory, the estimated numbers of bikes to be checked out and returned, and the estimated maximum number of available voluntary riders in this period, one can solve VRFM$_t$ to calculate the optimal numbers of voluntary riders for specified OD pairs (how many voluntary riders to be assigned for repositioning bikes from the origin site to the destination site in each pair). To estimate the rental demands with higher accuracy, we have used the random forest algorithm to identify important factors and parameters. To evaluate the performance improvement in terms of service quality of crowdsourced repositioning, we have conducted two simulation experiments using the 10-month real rental data that were collected from YouBike in Taiwan. The results demonstrate that more accurate rental demand predictions could improve the service quality by up to 0.8% and that our crowdsourced repositioning strategy can improve the average daily service quality by up to 3.5% and 2.8% compared to the truck repositioning strategy in our two simulation experiments. By further analyzing the hourly service quality improvement, we observe the crowdsourced repositioning strategy can contribute up to a 32.5% reduction in unmet demands in rush hours. The substantial service quality improvement during the rush hours that is realized by the crowdsourced repositioning strategy demonstrates that our proposed strategy should be seriously considered in practice, at least for the rush hours, and our proposed methodologies are designed to realize this objective.
For future research, we suggest investigating more accurate rental demand prediction models and marketing strategies for encouraging voluntary riders.

Acknowledgments. I-Lin Wang was partially supported by Sayling Wen Cultural & Educational Foundation and Service Science Society of Taiwan, as well as the Ministry of Science and Technology of Taiwan under Grant MOST 105-2221-E-006-182-MY3, 108-2221-E-006-058, 109-2221-E-006-150, and 110-2221-E-006-191-MY3.

REFERENCES

[1] H. I. Ashqar, M. Elhenawy, M. H. Almanna, A. Ghanem, H. A. Rakha and L. House, Modeling bike availability in a bike-sharing system using machine learning, in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, IEEE, 2017.

[2] A. Bargar, A. Gupta, S. Gupta and D. Ma, Interactive visual analytics for multi-city bikeshare data analysis, in *The 3rd International Workshop on Urban Computing (UrbComp 2014)*, New York, USA, 2014.

[3] M. Benchimol, P. Benchimol, B. Chappert, A. de la Taille, F. Laroche, F. Meunier and L. Robinet, Balancing the stations of a self service “bike hire” system, *RAIRO - Operations Research*, 45 (2011), 37–61.

[4] L. Breiman, *Random forests*, *Machine Learning*, 45 (2001), 5–32.

[5] L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza and X. Xiao, Predicting critical conditions in bicycle sharing systems, *Computing*, 99 (2016), 39–57.

[6] J. X. Cao, C. C. Xue, M. Y. Jian and X. R. Yao, Research on the station location problem for public bicycle systems under dynamic demand, *Computers & Industrial Engineering*, 127 (2019), 971–980.

[7] L.-C. Chang, *Design and Management of Urban Bike Sharing Systems*, Master’s thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan, 2010.

[8] D. Chemla, F. Meunier and R. W. Calvo, Bike sharing systems: Solving the static rebalancing problem, *Discrete Optimization*, 10 (2013), 120–146.

[9] H. Chung, D. Freund and D. Shmoys, Bike angels: An analysis of citi bike incentive program, in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, article 5.

[10] C. Contardo, C. Morency and L. Roux, Balancing a dynamic public bike-sharing system, *Centre Interuniversitaire de Recherche sur les Rseaux d’Entreprise, la Logistique et le Transport*, 4.

[11] J. Froehlich, J. Neumann and S. Oliver, Sensing and predicting the pulse of the city through shared bicycling, *International Joint Conference on Artificial Intelligence*, 9 (2009), 1420–1426.

[12] X. Gao and G. M. Lee, Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning, *Computers & Industrial Engineering*, 128 (2019), 60–69.

[13] K. Gebhart and R. B. Noland, The impact of weather conditions on bikeshare trips in washington, DC, *Transportation*, 41 (2014), 1205–1225.

[14] C.-P. Hung, *Optimal Station Allocation and Dynamic Bike Repositioning Strategies for Public Bike Sharing Systems*, Master’s thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan, 2011.

[15] A. A. Kadri, I. Kacem and K. Labadi, A branch-and-bound algorithm for solving the static rebalancing problem in bicycle-sharing systems, *Computers & Industrial Engineering*, 95 (2016), 41–52.

[16] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina and R. Banchs, Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system, *Pervasive and Mobile Computing*, 6 (2010), 455–466.

[17] M. Kaspi, T. Raviv, M. Tzur and H. Galili, Regulating vehicle sharing systems through parking reservation policies: Analysis and performance bounds, *European Journal of Operational Research*, 251 (2016), 969–987.
[18] M.-T. Liao, A Strategic Study on Managing Public Bike Sharing Systems by Demand Profile and Temporary Manpower Allocation, Master’s thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan, 2012.
[19] W.-Y. Loh, Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1 (2011), 14–23.
[20] R. Meddin and D. DeMaio, The meddin bike-sharing world map, Available from: http://bikesharingworldmap.com.
[21] R. Montoliu, Discovering mobility patterns on bicycle-based public transportation system by using probabilistic topic models, in Ambient Intelligence - Software and Applications, Springer Berlin Heidelberg, (2012), 145–153.
[22] O. O’Brien, J. Cheshire and M. Batty, Mining bicycle sharing data for generating insights into sustainable transport systems, Journal of Transport Geography, 34 (2014), 262–273.
[23] T. Raviv and O. Kolka, Optimal inventory management of a bike-sharing station, IIE Transactions, 45 (2013), 1077–1093.
[24] T. Raviv, M. Tzar and I. A. Forma, Static repositioning in a bike-sharing system: models and solution approaches, EURO Journal on Transportation and Logistics, 2 (2013), 187–229.
[25] R. A. Rixey, Station-level forecasting of bikesharing ridership, Transportation Research Record: Journal of the Transportation Research Board, 2387 (2013), 46–55.
[26] A. Sarkar, N. Lathia and C. Mascolo, Comparing cities’ cycling patterns using online shared bicycle maps, Transportation, 42 (2015), 541–559.
[27] J. Schuijbroek, R. Hampshire and W.-J. van Hoeve, Inventory rebalancing and vehicle routing in bike sharing systems, European Journal of Operational Research, 257 (2017), 992–1004.
[28] J. Shu, M. C. Chou, Q. Liu, C.-P. Teo and I.-L. Wang, Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems, Operations Research, 61 (2013), 1346–1359.
[29] P. Vogel, T. Greiser and D. C. Mattfeld, Understanding bike-sharing systems using data mining: Exploring activity patterns, Procedia - Social and Behavioral Sciences, 20 (2011), 514–523.
[30] P. Vogel and D. C. Mattfeld, Strategic and operational planning of bike-sharing systems by data mining – a case study, in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, 127–141.
[31] P. Vogel, B. A. N. Saavedra and D. C. Mattfeld, A hybrid metaheuristic to solve the resource allocation problem in bike sharing systems, in Hybrid Metaheuristics, Springer International Publishing, (2014), 16–29.
[32] I.-L. Wang and T.-L. Wu, A simulation study on the value and impact of exploiting rental information to the metropolitan bike sharing systems, Technical report.
[33] S. Yan, J.-R. Lin, Y.-C. Chen and F.-R. Xie, Rental bike location and allocation under stochastic demands, Computers & Industrial Engineering, 107 (2017), 1–11.
[34] Z. Yang, J. Hu, Y. Shu, F. Cheng, J. Chen and T. Moscibroda, Mobility modeling and prediction in bike-sharing systems, in Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, ACM, 2016
[35] J. W. Yoon, F. Pinelli and F. Calabrese, Cityride: A predictive bike sharing journey advisor, in 2012 IEEE 13th International Conference on Mobile Data Management, 2012.

Received July 2020; 1st revision October 2020; Final revision January 2021; Early access November 2021.

E-mail address: ilinwang@mail.ncku.edu.tw
E-mail address: vjack070707@gmail.com