Utilization-Aware Trip Advisor in Bike-sharing Systems Based on User Behavior Analysis

Peng Cheng, Ji Hu, Zidong Yang, Yuanchao Shu, and Jiming Chen

Abstract—The rapid development of bike-sharing systems has brought people enormous convenience during the past decade. On the other hand, high transport flexibility gives rise to problems for both users and operators. For users, dynamic distribution of shared bikes caused by uneven user demand often leads to the check in or check out service unavailable at some stations. For operators, unbalanced bike usage comes with more bike broken and growing maintenance cost. In this paper, we consider to enhance user experiences and rebalance bicycle utilization by directing users to different stations with a higher success rate of rental and return. For the first time, we devise a trip advisor that recommends bike check-in and check-out stations with joint consideration of service quality and bicycle utilization. To ensure service quality, we firstly predict the user demand of each station to obtain the success rate of rental and return in the future. Experiments indicate that the precision of our method is as much as 0.826, which has raised by 25.9% as compared with that of the historical average method. To rebalance bike usage, from historical data, we identify that biased bike usage is rooted from circumscribed bicycle circulation among few active stations. Therefore, with defined station activeness, we optimize the bike circulation by leading users to shift bikes between highly active stations and inactive ones. We extensively evaluate the performance of our design through real-world datasets. Evaluation results show that the percentage of frequently used bikes decreases by 33.6% on usage number and 28.6% on usage time.

Index Terms—Bike-sharing, trip advisor, rebalance bicycle utilization.

1 Introduction

With the development of the economy, pollution and destruction caused by human activities to the natural environment were becoming more and more severe in recent years, and sustainable development has therefore become a consensus of the international community [1], [2]. In this circumstance, bike-sharing systems (BSS) are developed as a replacement for short vehicle journeys due to its low pollution, low energy consumption and high flexibility. In addition to the reduction of need for personal vehicle trips, public bike-sharing systems can not only extend the reach of transit and walking trips, providing people with a healthy transportation option, but also trigger greater interest in cycling, and increase cycling ridership. By the end of 2016, over 1,100 cities actively operate automated bike-sharing systems deploying an estimate of 2,000,000 public bicycles worldwide [3].

With bike-sharing systems, a user can easily rent a bike with a smart card at a nearby station and return it at another station. However, the advantages can not cover up the increasingly prominent issues. For stations, the user demand is ever-changing and unbalanced, which often leads to the check in or check out service unavailable at some stations and has a negative impact on user experience. For bikes, the usage frequency of each bike is unevenly distributed, posing a problem for both riders and system operators.

On the one hand, due to the high flexibility of bike-sharing system, the system typically ends up with an uneven distribution of bikes across the different stations (due to the uncontrolled, uneven demand), often rendering the check in or check out service unavailable at some stations where bicycle docks are either fully occupied or empty. During peak periods, user demand characteristics differ among stations in certain areas. For example, rental demand usually gets larger in workday morning near residential areas, whereas return demand gets larger near commercial districts. At present, operators perform bike redistribution based on monitor video and user complaints. However, this method has exposed the serious lag. It is usually when service unavailable events occur that operators start to give some scheduling instructions. When the vehicle arrives, service unavailable events may have passed for some time, which makes it difficult to meet the needs of users at rush hour.

To increase service availability and enhance user experience, studies have been conducted to improve these bike redistribution strategies based on bicycle mobility models and predictions. The majority of previous work focuses on bike usage patterns and rental volume forecasts for each station without considering online information (e.g., [4], [5], [6], [7]). Less attention has been devoted to demand prediction of each cluster from the view of bike flow mobility patterns

- *P. Cheng is with the School of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 China (e-mail: pcheng@zju.edu.cn).*
- *J. Hu is with the School of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 China (e-mail: hujichn@gmail.com).*
- *Z. Yang is with the School of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 China (e-mail: zdyang@zju.edu.cn).*
- *Y. Shu is with Microsoft Research, Redmond 98052, WA, United States (e-mail: yuanchao.shu@microsoft.com).*
- *J. Chen is with the School of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 China (e-mail: jmchen@ieee.org).*

Part of this paper has been presented in IEEE ICDM 2017.

which may not fit for recommending stations for users [8]. In conclusion, developing a fine-grained prediction model involving multiple factors has proven to be elusive, and has remained a largely unstudied problem. The main technical challenge is that bike traffic is not only highly dynamic and intercorrelated in both the temporal and spatial domains, but also further influenced by complex issues such as timing and meteorology. To alleviate the unbalanced demand problem, we establish a fine-grained demand forecasting model and predict check in and check out demand on a per-station basis with sub-hour granularity by using random forest algorithm. In our model, offline features such as time and weather are selected to capture the periodic patterns of user demand. Online feature is to reflect the real-time availability of the station which is helpful for abnormal traffic.

On the other hand, a small part of bikes is used much more frequently than others. Bikes that are used too much are vulnerable and hence increase repair bills and lead to potential denied service. The very first bicycle from Hangzhou BSS is reported to be rented for over 6,000 times and ridden for more than 20,000 kilometers in 3 years. Similarly, the most tireless bicycle from 2016 has been rented for 5,616 times, over 15 times on average each day. According to Hangzhou public bike-sharing company, the average life of their bicycles is less than 4 years due to longtime high load operation and lack of timely renewal and maintenance. On the contrary, the average life of private bicycles is 10 years and above. Meanwhile, the cost of repair and labor accounts for a large proportion in the overall operating expenses. In 2012, the repair cost of Hangzhou bike-sharing system was near 6 million yuan [9]. In Washington, D.C., the annual maintenance cost was $200 to $300 per bike in the year of 2012 [10].

Intuitively, operators can balance bike usage by leading users to use those unpopular bikes based on usage counts of each bike. However, directing users to rent a specific bike is not practical. Based on our analysis on a real bike-sharing dataset from Hangzhou, we observe that bikes located in some stations are much more likely to be used and moved to another active station. Hence, by introducing the station property of activeness, we transform the original problem of picking bikes to recommending check-in and check-out stations. By using the proposed trip advisor, we aim to guide users to ride bicycles between stations with different levels of activeness, therefore avoiding circumscribed circulation among active stations. For users, an advisor can not only help them choose stations with adequate bicycles, but also ensure a higher success rate when returning bikes. Also, different incentive mechanisms can be leveraged to better prompt the balancing process.

In this paper, we propose a trip advisor that recommends the optimal pair of stations to rent and return bikes. Through guiding the actions of users, it can help balance bike usage, reduce operation cost and enhance user experience. Firstly, to make sure users can find bikes and available lockers, success rates of rental and return should be predicted for each station. Different from traditional demand prediction methods, we present probabilistic forecast methods on a minute timescale instead of predicting the exact stock number on sub-hour granularity. Secondly, in order to balance bike usage through station recommendation, a station property must be associated with bike usage frequency. We define activeness for each station by exploiting the idea of PageRank. These two parts constitute the core content of the trip advisor framework.

TABLE 1
Primary fields in the bike-sharing dataset.

user_id	rent_netid	tran_date	tran_time
8601940	9926	20150601	070641
return_netid	**return_date**	**return_time**	**bike_id**
9205	20150601	071635	1708133

In summary, in this paper we propose a novel utilization-aware trip advisor to lead users to help balancing bike usage without compromising the quality of service. We highlight our key contributions as follows:

- We explore the overall characteristics of bike-sharing systems, analyze the spatial temporal patterns of user behavior and study the bike usage frequency, thus laying the foundation for trip advisor design.
- We propose a probabilistic forecast method which adopts Monte Carlo simulation and random forest model to improve prediction accuracy.
- We introduce the concept of activeness to link bike usage frequency to station property which utilizes the topological characteristics of bike sharing network and the relative check out amount of each station. Meanwhile, we dynamically update the activeness to take the effect of the advisor on the system into account.
- We present a novel framework to balance bike usage with the help of users and validate our proposed method with real-world human mobility datasets.

2 Data Preprocessing

2.1 Dataset Description

The Chinese city of Hangzhou has the world's largest public BSS with more than 3300 stations and over 84,000 shared bicycles [11]. Since deployed in May 2008, thousands of bicycles have been rented for more than 700 million times. The concept of public bicycles has since spread to 30 other provinces in China and around 175 cities nationwide.

The system is classified as a third-generation bike-sharing program due to its IT-based system, automated check-in and check-out, and distinguishable bicycles and docking stations [12]. The dataset used in this paper was collected in June 2015 from our partner who is running Hangzhou BSS. It contains 58,647 bikes and 3,329 stations. Each bike-sharing trip contains an origin and a destination with information of locations and timestamps. The primary fields of the dataset are shown in Table 1.

The meteorology dataset contains weather conditions of Hangzhou with totally $48 \times 365 = 17,520$ records. Meteorological observations were updated every half hour and the data format of each record is shown in Table 2.

2.2 Data preprocessing

Data cleaning:

TABLE 2
Fields in the meteorology dataset.

Time (CST)	Temp (°F)	Dew Point (°F)
12:30 PM	100.4	69.8
Pressure (in)	Humidity (%)	Visibility (mi)
29.65	37	6.2
Wind Dir	Wind Speed (mph)	Conditions
WSW	8.9	Partly Cloudy

The data in the real world are generally incomplete and inconsistent dirty data, so data analysis cannot be directly conducted. Before analyzing the data, it is necessary to perform appropriate data cleaning to obtain high quality data and necessary information.

Actual user demand calculation:

In BSS, it often happens that a user returns the bike immediately after borrowing it at the same station, after which the user often borrows another bike. This phenomenon may be due to the user's dissatisfaction with the chair height or the current status of the bike. Therefore, if we directly count the number of records, the calculated user demand will be greater than the actual user demand.

As shown in Figure 1, the PDF curve of trip duration which begins and ends at the same station can be divided into a distinct spike and a long tail: for those real users, users at different stations could have different travel purpose, so the trip duration must be different. Due to the superimposed effects of records from all the stations, the travel time will be evenly distributed. Accordingly, the curve has a longer tail; and for the users who return the bike immediately, the trip duration is almost the same in each station, which leads to that very high spike. In the figure, the horizontal axis represents the riding time in seconds. The peak caused by the superimposed effects disappears at 120s. Therefore, records with trip duration less than 120s are treated as false records, and thus can be deleted from the original data. Finally, the actual demand can be calculated by simply accumulating the data in a half-hour unit.

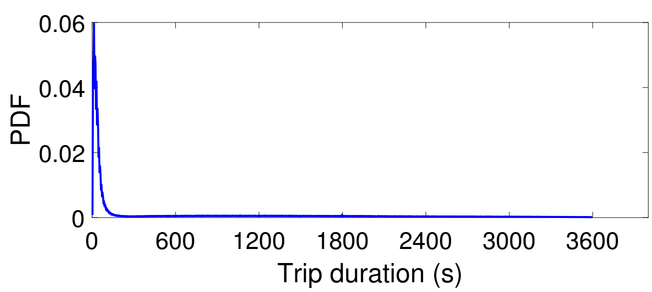


Fig. 1. PDF of trip duration which begins and ends at the same station.

Station stock calculation:

When the station is full or empty, the calculated user demand will be less than the actual demand. Therefore, we need to calculate the station stock to identify the full or empty situations.

As shown in the Figure 2, we can see that as the time goes by, the curve gradually deviates from 0, and the lowest value is even lower than -150. In real life, however, the station stock is impossible to be negative. The data excursion is a cumulative error caused by the fault transaction data. Therefore, we need to correct the excursion by considering other ancillary data.

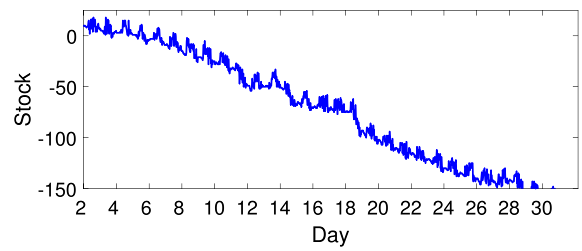


Fig. 2. Stock curve of station 4051 without correction in January 2013.

The ancillary data used here is empty and full alarm data. The steps are as follows: Firstly, get the initial state, redistribution data, empty and full alarm data of all the stations. Then, select the corresponding data as well as the transaction data for each station to be calculated. Based on the initial status data, a list of bikes in the station is generated. Combine the empty and full alarm data, redistribution data and transaction data into a single operation data table and sort by its operation time. Finally, process each row in the data table in sequence, and the list of bikes in the station could be continuously updated.

As Figure 3 shows, the stock curve of station 4051 after correction changes periodically, which means the data excursion error is eliminated.

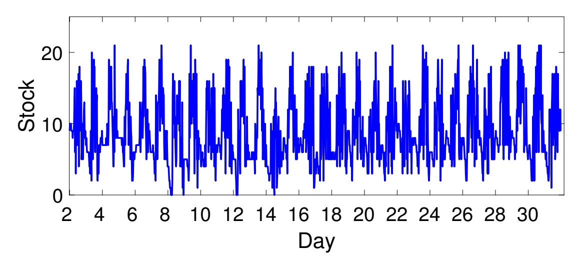


Fig. 3. Stock curve of station 4051 after correction in January 2013.

2.3 Statistics in BSS

In order to have a more intuitive view of the entire system, some statistics on public bicycle systems are given in this section, mainly about site distribution, site capacity, and so on. Based on these statistics, we can obtain the specific characteristics of Hangzhou public bike-sharing system.

Station Distribution:

Bike stations in Hangzhou are located within the urban area spanning over 600 square kilometers; the average distance to the closest neighboring station being 300 meters [12]. Figure 4 shows the probability distribution function (PDF) of the number of stations within a certain range of one station. From this figure, we notice that half the stations have more than 3 neighbors within the range of 300 meters, and typically a station may have 8 neighbors within the range of 500 meters.

Station capacity:

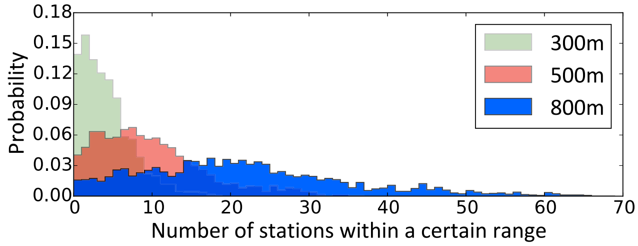


Fig. 4. Station distribution.

Station capacity is measured as the number of stocks. As shown in Figure 5, there are two types of stations in Hangzhou: normal stations with 21 docks and large station with around 33 docks. The station capacity is designed by the urban planning department whereas the actual number of docks depends on the actual situation.

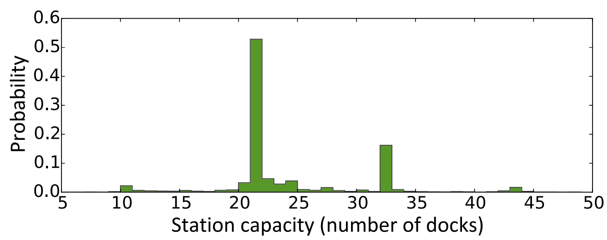


Fig. 5. Distribution of station capacity.

This provides a reference to the range settings when designing the trip advisor. If we only consider stations within a very small range, there will be few stations to be selected. Otherwise, the number of candidate stations will increase significantly but users will suffer from extra walking distance. Here, we set the range threshold to 500 meters which provides 8 stations in expectation.

3 User Behavior Analysis

In this section, we first present some statistics and spatial temporal user behavior analysis derived from the bike-sharing dataset from Hangzhou City in China. Inspired by insights obtained from the study we propose our utilization-aware trip advisor.

3.1 Overall characteristics

Figure 6 presents the distribution of monthly usage amount (i.e., check in numbers) across all 2806 stations. We find that there are more than 100 busy stations with extremely high usage amount up to 30000 (check in/month). However, the median value of the usage amount is around 2000. The skewed distribution in Figure 6 indicates a high diversity of usage amount across different stations. We observe a similar pattern for check out numbers.

Figure 7 presents CDF of empty and full hours in a month. Axis x represents full or empty hours. The calculation steps are as follows. Every half an hour, the inventory in each station is sampled once. There are 1440 samples in a month. When the inventory is less than 10% in these samples, we consider it the empty time. Similarly, if the inventory is more than 90%, we consider it the full time. This picture can be used to measure the service level of the current bike-sharing system. It can be seen that about 19% of stations have an empty status for more than 200 hours in a month, 27% of the stations have been full for more than 200 hours in a month. The empty condition appears relatively less and the full condition appears relatively more.

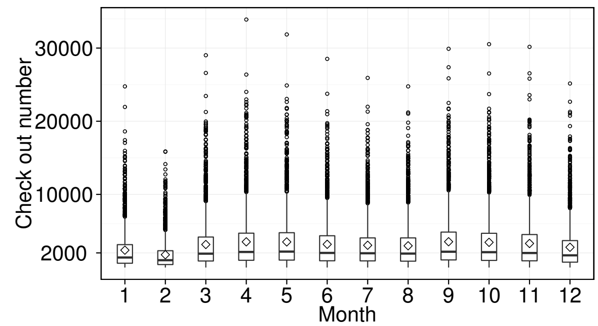


Fig. 6. Distribution of monthly usage amount.

TABLE 3
Statistics on trip durations

$<$ 15 min	15-30 min	30-45 min
53%	27%	11%
40-60 min	$>$ 60 min	mean
5%	4%	23.31 min

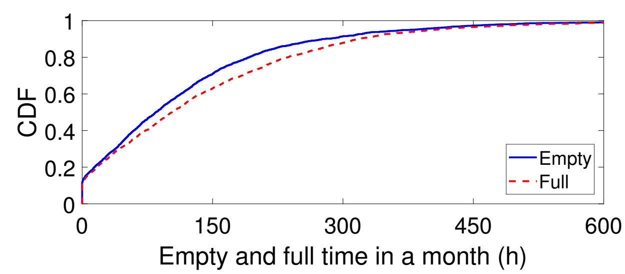


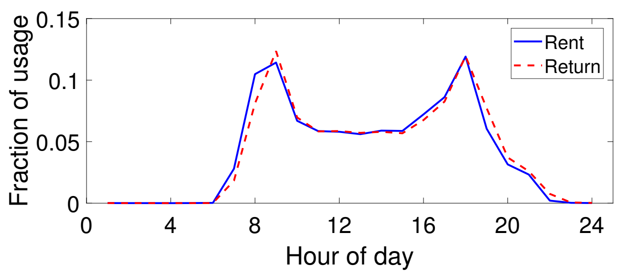
Fig. 7. CDF of empty and full hours

3.2 Temporal patterns

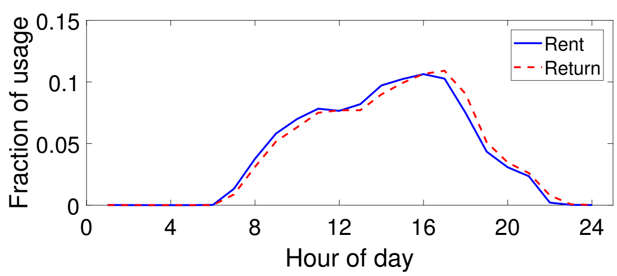
The trip duration of bike-sharing is important for analyzing the user's travel behavior. The cycling duration reflects the intention of traveling to a certain extent. Based on statistics on the data of the cycling trip length of the bicycle, the average single cycle length of the public bicycle is about 23.31 min. Table 3 shows the statistics on trip durations in 2013. In this table we found that trip durations are typically less than one hour. That's probably because of the first-hour-free policy of Hangzhou BSS.

The time distribution of the activity of rent and return is also an important part of data analysis. The time distribution of the activity of rent and return can be roughly used to understand the peak periods of borrowing public bicycles, facilitating the coordination and scheduling of bike-sharing. We select two typical days to analyze the activity of rent and return in different days respectively, which is January 1, 2013 (typical holiday) and January 8, 2013 (typical working day). The results are shown in Figure 8, plotted by time on the horizontal axis and the activity of rent and return on the vertical. The activity is calculated as the hourly usage for all sites divided by the total usage for the day.

We can find that there is an obvious characteristic of the time distribution of rent and return activity on working days, specifically the two peak periods in the morning and the evening, showing an M-shaped distribution. The peak hours of orders are concentrated around 8:00 am to 9:00 am and from 17:00 pm to 18:00 pm, and the usage during the morning and evening peak accounts for about 41% of the total daily usage. The two major peaks are mainly caused by the sudden surge of traffic in and out in rush hours. The activeness of rent and return during work hours is significantly reduced. Comparing morning rush hour with evening rush hour, rent time at evening rush hour is more dispersed, starting to rise significantly at 16:00, while the rent time at morning rush hour is more concentrated, which may be due to the user's similar attendance time and different duty-off time.



(a) Activeness distribution of rent and return in workday.



(b) Activeness distribution of rent and return in holiday.

Fig. 8. Activeness distribution of rent and return.

3.3 Spacial patterns

Figure 9 shows the spatial distribution of average usage amount in Hangzhou public bike-sharing system. Each dot represents a bicycle site. We rank the stations according to the average number of rent and return bikes at each station. The radius of the dots from large to small represent the rent and return amount from high to low. The distribution of the stations is plotted on the map. It can be observed that stations with relatively large usage amount are usually located in Xihu District, Gongshu District, Shangcheng District and Xiacheng District, while the bike usage in Xiasha District, Binjiang District and Yuhang District are much less active. The overall trend of decreasing from the center to the periphery shows that there is an obvious imbalance in the spatial distribution of the user demand. This phenomenon is consistent with the station density in various regions and the level of regional economic development.

TABLE 4
Top 10 rent and return stations

Top 10 rent stations	2015	1053	4051	7339	2040
	2105	1256	1077	2042	7358
Top 10 return stations	1053	4051	2015	2040	2105
	7339	1256	1077	2042	4049

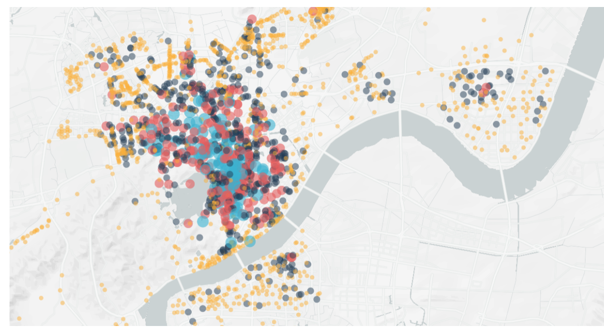


Fig. 9. Spatial distribution of bike usage amount.

In order to analyze the transaction relationship between the stations, we extract the out-degrees and in-degrees between the stations and draw the pixel maps of the station transaction behaviors on the x and y-axes. The result is shown in the figure. In Figure 10, each pixel in the pixmap represents a set of station pairs, the color of the dots depends on the number of transactions. The darker the color, the greater the average monthly trading volume.

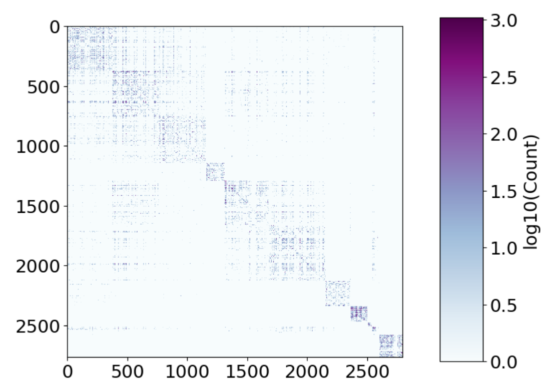


Fig. 10. Frequent pairs

We have the following observations:

Firstly, most of the regions in the figure are very light in color, indicating that the adjacency matrix formed by the transaction behavior is a sparse matrix. Most of the sites have a borrowing correlation with only a few sites in the entire network. Most nodes in this directed graph don't have many neighbors.

Secondly, the color is the deepest near the diagonal. Since the stations are sorted based on net id, it means that bikes usually tend to be rented and returned among stations with similar net id. Meanwhile, the net ids are closely related to the geographic location of the stations, so the stations that are located in a certain district are likely to have similar net ids. Therefore, we can conclude that the spatial movement of bikes is often not too far away, and that a large number of cycling records occur in the vicinity of the starting site, which is also consistent with the conclusion on riding time mentioned above.

Finally, the figure is basically diagonal symmetrical. Since the transactions we calculated are directional, that is, the records from Site A to Site B and the records from Site B to Site A are counted separately. There is no obvious trade bias between the two sites. It is not common for the system to have a large number of bikes from one site to another site while the other site does not have the returned traffic, which is beneficial for system scheduling.

After we are sure that there are enough stations to be selected near the origin and destination, we need to find out whether the stock levels of those stations are quite different from their neighbors. If the stock levels are almost the same, there is no need to predict the stock level of each station. The success rate of rental and return would be exactly the same for all the candidate stations.

Figure 11 shows the cumulative distribution function (CDF) of the number of unbalanced stations around each station in June 2015. For each station, if the difference in stock level between it and a station located within 500 meters exceeds 50%, it is considered as an unbalanced event. If the accumulated time of unbalanced events is longer than h hours in a month, the unbalanced station number increases by 1. Here, let h be 120, 180 and 240. From Figure 11, we notice that when h is set to 180, there are more than 61% of stations have at least 1 station nearby that is distinct from them in stock level. When h gets smaller, the percentage of stations that have at least 1 unbalanced station nearby is obviously increased. When h equals to 240, the corresponding percentage is 42%. According to the above analysis results, the stock level of stations within a small range could be quite different from each other, which means that it's necessary to predict the stock level and ensure that users can rent or return bikes successfully.

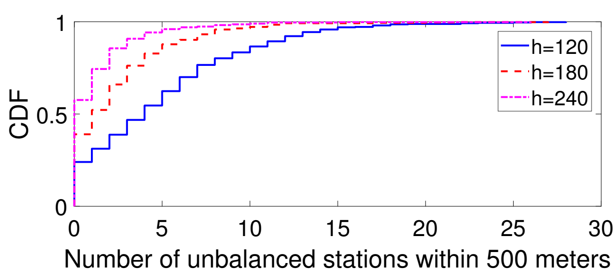
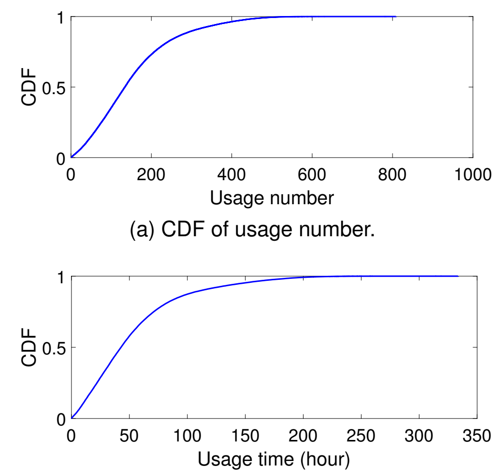


Fig. 11. Station unbalance.

3.4 Bike Usage

After the analysis of spatial and temporal patterns, the most essential issue is bike usage unbalance. Because historical records contain the ID of bikes, we can extract the usage characteristics by summing up the number of occurrences and trip durations of each bike. The preliminary results are depicted in Figure 12. As shown in Figure 12a, 57% of bikes are used for less than 150 times in a month, less than 5 times per day on average. However, about 10% of bikes are used more than 310 times in a month, which is twice as frequent as less used bikes. From Figure 12b, we can see that the usage time of 64.5% bikes is less than 57 hours in a month while that of 10% bikes is over 115 hours. These statistics clearly indicate that the usage of bikes is unbalanced, and a small part of bikes have much higher usage frequency and longer usage time than others, which is the leading cause of bike damage [13].



(a) CDF of usage number.

(b) CDF of usage time.

Fig. 12. Usage unbalance.

Further, we describe the usage characteristic by using the idea of the Lorenz curve. The Lorenz curve plots the percentage of total income earned by various portions of the population when the population is ordered by the size of their incomes [14]. In Figure 13, the vertical axis represents the cumulative percentage of bikes (in ascending order of usage number/time), while the horizontal axis shows cumulative percentage of bike usage number/time. We find that 60% less used bikes only contribute about 30% usage time and 33% usage number. Thus, it can be concluded that bike usage unbalanced problem does exist, and we need to design a trip advisor to guide users to help balancing bike usage.

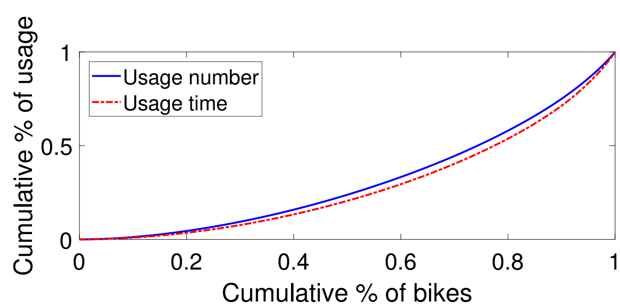


Fig. 13. Cumulative contribution rate of usage.

3.5 Insight

In this part, we offer some insights into explaining the observed bike usage unbalance problem. A direct and ef-

TABLE 5
Most frequently used bikes.

bike_id	687500	683676	687119
usage num	809	783	780
bike_id	1502964	688515	1500966
usage num	630	616	608
bike_id	687500	687119	683676
usage time (h)	333.67	319.16	314.56
bike_id	1501877	1502628	1502407
usage time (h)	259.63	258.10	257.88

fective way to explore the reasons is to identify those most frequently used bikes and observe their mobility patterns. From the historical check in and check out records, we have calculated the usage number and usage time of bikes and the results are shown in Table 5. We found that the top 3 bikes on usage number are consistent with those on usage time. The most frequently used bike with bike id 687,500 has been rented 809 times in a month with a total time of 333.67 hours.

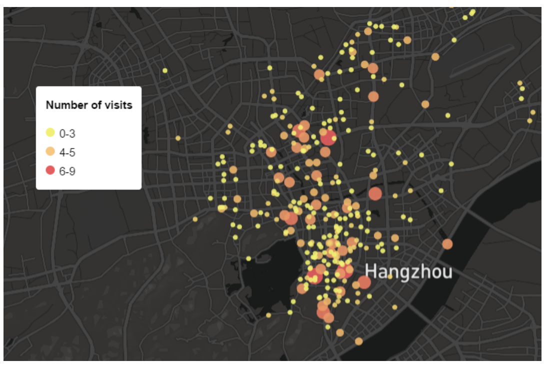


Fig. 14. Geographical distribution of stations that that top 3 frequently used bikes have been visited.

It's possible that the usage frequency of each bike has a close relationship with the stations it has been visited. Thus, the stations where top 3 frequently used bikes have been checked out are found and the amounts of visits are counted. The geographical distribution of those stations is depicted in Figure 14. From this figure, we notice that the number of visits in main urban area is much higher. The purpose of rental in the main urban areas could be going to work or school or even buying breakfast. The significant features of this kind of rental are short trip, high efficiency and quick turnover. In this case, bikes are usually rented from one station and then quickly returned to another station. After being returned, bikes are likely to be checked out again and flow to the next station quickly. Such preliminary results demonstrate that the main reason for unbalanced bike usage is the continuous circulation of bikes among active stations. On the other side, bike utilization can be balanced by introducing flows between active stations and inactive stations. How to define the activeness of stations will be elaborated in the section below.

4 Methodology Overview

In this section, we first formulate the problem of station recommendation, and then show the details of the proposed trip advisor framework.

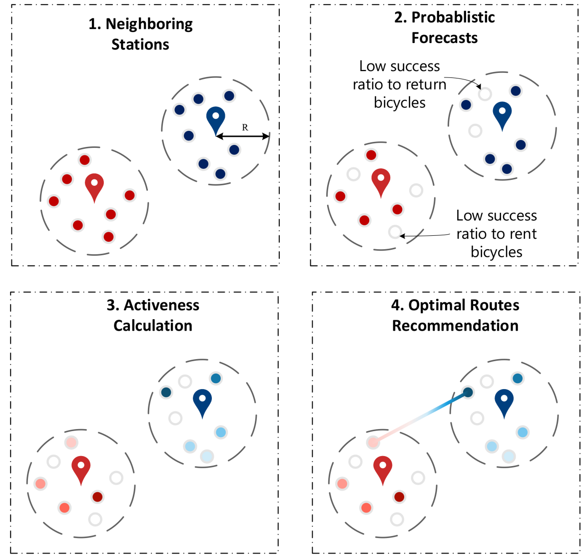


Fig. 15. Framework of the trip advisor.

4.1 Problem Definition

Considering a bike-sharing system consisting of stations, bikes and users, the inputs of trip advisor are user requests including origin location l_o, destination location l_d and leaving time t_l. The user requests are stochastic and can occur at every station at any time. Let $S_o = s_{o1}, s_{o2}, ..., s_{on}$ be a set of stations in R meters zone around the origin and $S_d = s_{d1}, s_{d2}, ..., s_{dn}$ be a set of stations near the destination. Each station has its location (e.g., latitude and longitude) and stock level r_i with sub-hour granularity, where $i \in S_o, S_d$. Based on user inputs and current status of the system, the output of trip advisor is a pair of optimal stations (s_i^*, s_j^*) for users to rent and then return a bike, where $s_i^* \in S_o$ and $s_j^* \in S_d$. The problem is dynamic because decisions can be adapted over the planning horizon. In decision making process, the first step is to filter the stations in S_o and S_d based on the success rate of rental and return. Hence, we will obtain a middle variable $S_o^{'}$ and $S_d^{'}$ representing candidate stations after probabilistic forecasts. The important notations used in this paper are listed in Table 6.

TABLE 6
Symbols and definitions.

l_o, l_d	location of origin/destination
t_l	leaving time
S_o, S_d	stations near the origin/destination
R	range
r_i	stock level of station $i \in S_o, S_d$
$S_o^{'}, S_d^{'}$	candidate stations after probabilistic forecasts

4.2 General Framework

Before leaving, users can send a query including their origin, destination and leaving time to the trip advisor and then

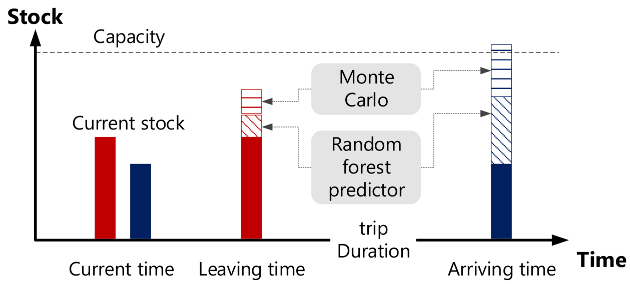


Fig. 16. The idea of probabilistic forecasts.

get the recommended stations for rental and return. The key problem is how to guide the users to balance bike usage through station recommendation while not affecting the user experience. In this section, we will introduce the framework of our method, as shown in Figure 15. The framework is comprised of two major components: probabilistic forecasts and activeness calculation.

4.2.1 Probabilistic Forecasts

In order to encourage users to use the advisor and continue to help balancing bike usage, we first need to make sure that users can rent or return bikes successfully. Therefore, the first component, probabilistic forecasts, is designed to solve the no-service problem and guarantee the higher success rate for rental and return when users arrive at the stations. No-service means the situations in which a user can't find available bikes to rent, and those in which he/she finds there's no parking spot to return. This problem is mainly caused by the asymmetric and fluctuating user demand among the stations. For users, they may know where the nearest station is, but what they really want to know is the probability of successfully renting or returning bikes when he/she arrives there. To obtain the success rate at a precise moment, simply predicting the forthcoming user demand on half-hour granularity is not enough to meet the above requirement. The component of probabilistic forecasts is needed to predict the stock level on a minute timescale and further derive success rate through the Monte Carlo method.

The process is illustrated in Figure 16. At the beginning, the stock levels of candidate stations near the origin/destination are known. The forecasts consist of two parts. The first part is coarse-grained prediction using random forest model; the second part is fine-grained prediction based on Monte Carlo method.

Here, we take predicting return success rate at arriving time as an example to elaborate on the details. Let $[t]$ represent the rounded time of t to the nearest 30 minutes before. At the rounded current time $[t_{now}]$, we already know the stock status r_i of station i within R meters of the destination. Firstly, we predict the base check in and check out demand at each station with sub-hour granularity by using random forest model. Random forests are an ensemble learning method for regression which operate by constructing a multitude of decision trees with different samples and different initial variables. The final output is the mean prediction of the individual trees. We apply the random forest theory to model and predict the users behaviors with a joint consideration of time factors, meteorology and real-time bike availability [15]. Let $CI_i(t)$ and $CO_i(t)$ be the predicted check in and check out number of station i within a temporal window $(t, t+T)$, where $i \in S_d$ and $T = 30min$. The coarse-grained prediction of stock level at the rounded arriving time $[t_a]$ is as follows:

$$Stock_i([t_a]) = r_i + \sum_{t=[t_{now}]}^{[t_a]-T} (CI_i(t) - CO_i(t)) \quad (1)$$

Then, to get a more accurate stock number, we adopt the Monte Carlo method to simulate the bike rental and return process at the temporal window $([t_a], t_a)$. The general method of Monte Carlo is to obtain numerical results through repeated random sampling. We assume that the number of bikes rented or returned in the predicted time window follows a Poisson distribution. Given the station i with the predicted bike check in and check out number $CI_i([t_a])$ and $CO_i([t_a])$ in the time window $([t_a], [t_a] + T)$, we divide time delta into T small consecutive time intervals $\delta t = 1min$. The number of bikes returned to this station in each δt, noted as x, follows a Poisson distribution with mean parameter $\lambda = CI_i([t_a])/T$:

$$P(x = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, ... \quad (2)$$

For each simulation, we generate a stochastic sequence Q_{+i} from the return distribution to simulate the bike return events of each station. Similarly, we generate a stochastic sequence Q_{-i} for the bike rental events. Afterward, we randomly arrange the return and rental events based on the two sequences and update the stock number over time. If the stock number exceeds the capacity of the station, we mark it as an over-demand station and stop the process.

We repeat the simulation for M times to count the over-demand occurrences U. In the end, we estimate the probability of successfully returning bikes at arriving time as the rate:

$$p = 1 - \frac{U}{M}. \quad (3)$$

The success rate for bike rental at leaving time can be calculated in a similar manner.

In summary, the main idea of probabilistic forecasts is to simulate the probabilistic process of check in and check out and derive the probability of success-of-service across a sufficiently large number of simulations. We choose the stations as candidate stations $S_o^{'}$, $S_d^{'}$ on the basis of whether its success rate is larger than a threshold P, which is set as 0.8 in our work.

4.2.2 Activeness Calculation

For the candidate stations $S_o^{'}$, $S_d^{'}$, we need to further decide which is the best pair of stations to recommend. Our ultimate goal is to balance bike usage and extend their lifespan, but we can only lead users to a station instead of recommending a specific bike. Therefore, we have to concern about how to link up the bike usage characteristic with a certain property of the station, such as activeness.

According to the previous analysis, active stations are characterized by the following properties: (1) Bikes returned to this station are easily checked out and flow to many other stations; (2) The stations that those bikes flowed to are also very active. These properties remind us of the way to measure a web page's importance. PageRank is an algorithm used by Google Search to rank websites in their search engine results [16]. According to Google: PageRank works by evaluating the quality and quantity of links to a web page to determine a relative score of that page's importance. The idea that PageRank brought up is that more important websites are likely to receive more links from other websites.

In bike-sharing systems, activeness can be defined to measure the active level of bike usage for each station based on the idea of PageRank. We begin by picturing the station network as a directed graph, with nodes represented by stations and edges represented by the bike flow (rent to return) between them. The underlying assumption is that more active stations in the network are likely to send more links to other stations. This makes sense because according to the analysis in Section 3, bikes do tend to be checked out extensively to many other stations at active stations and the bike usage in stations with more links out are usually more frequent. But this is only a start: the bikes must continue to flow to active stations so they can enter a high-speed circulation and be repeatedly used. This leads to the next assumption that stations which are themselves active weigh more heavily and help to make the stations that link to them active. If bikes rent from one station to stations with lower activeness, the bikes are likely to stay there and it will take a long time for them to be checked out again. Therefore, this station may have low activeness as well. Finally, the activeness of station i is given as

$$A(i) = \frac{1-\alpha}{N} + \alpha \sum_{j \in out(i)} \frac{n(i,j)A(j)}{n_{in}(j)} \tag{4}$$

where $A(i)$ is the activeness of station i, α is a damping factor which can be set between 0 and 1, N is the number of stations, $n(i,j)$ is the number of bikes rent from i and return to j, $n_{in}(j)$ is total number of bikes return to j and $out(i)$ is the set of stations that have bikes rent from i.

We can see that the activeness of station i is recursively defined by the activeness of those stations which are linked to by station i. If station i links to a lot of stations, the common belief is that station i is active. The activeness of station j which station i links to does not influence the activeness of station i uniformly. Within this algorithm, the activeness of a station j is always weighted by $n(i,j)/n_{in}(j)$. This means that the more return bikes station j has, the less will station i benefit from the link to station j. In addition, if a node has no ingoing edges, it cannot transfer its activeness to any other stations. Therefore, a damping factor is added for giving each node a probability that a bike can be returned to this station from any other station, each station has $1/N$ probability of being the source.

In the above formula, flow patterns in the station network is the main consideration, but the rental scale of each station has to be concerned as well. Stations with a large amount of rentals will certainly affect the mobility of more bikes. Bikes in those stations are usually easier to spread to more stations which is an expression of high activeness. So we adopt the normalized relative check out number to indicate the rental scale and suppose that stations with large rental scale are more active. Therefore, we rewrite the activeness of station i as following:

$$\begin{aligned} A(i) &= (1-\alpha)r_i + \alpha \sum_{j \in out(i)} \frac{n_{in}(i,j)A(j)}{n(j)} \\ r_i &= \frac{n_{out}(i)/c(i)}{\sum_{j=1}^{N} n_{out}(j)/c(j)} \end{aligned} \tag{5}$$

where $c(i)$ is the capacity of station i and $n_{out}(i)$ is the absolute check out number of station i. In this way, bikes are more likely to come from stations with higher relative check out number. By introducing this prior distribution, this method provides a more comprehensive measure of the activeness of stations.

Finally, to obtain the optimal pair of stations (s_i^*, s_j^*), we select stations according to the following equation:

$$(s_i^*, s_j^*) = \arg\max |A(s_i) - A(s_j)| \tag{6}$$

where $s_i \in S_o^{'}, s_j \in S_d^{'}$. If users strictly follow the advisor, the activeness of stations could have a distinct change due to the altered user behaviors. Taking into account this counteraction of the advisor to the network, we update the activeness each hour using the check in and check out records within the last hour.

5 Evaluation

In this section, we empirically evaluate the performance of our proposed method. We conduct experiments on the dataset of Hangzhou bike-sharing system in June 2015. There are 10,190,841 records, which contains 58,647 bikes and 3,329 stations. The data format is presented in Table 1. As mentioned in Section 2, the records that check out and check in at the same station with a trip duration less than 2 minutes are considered as noise data and removed from the original records.

5.1 Probabilistic Forecasts

In our experiments, we use the results of probabilistic forecasts as a condition for filtering stations, so we evaluate the probabilistic forecasts step as a classification problem and the metrics are as follows:

Precision and Recall: Given the results of whether stations will be over-demand, precision and recall are defined as:

$$Precision = \frac{|N_{pre-od}| \cap |N_{real-od}|}{|N_{pre-od}|} \tag{7}$$

$$Recall = \frac{|N_{pre-od}| \cap |N_{real-od}|}{|N_{real-od}|} \tag{8}$$

where N_{pre-od} represents the number of events that are predicted to be over-demand, and N_{pre-od} represents the number of events that are really over-demand.

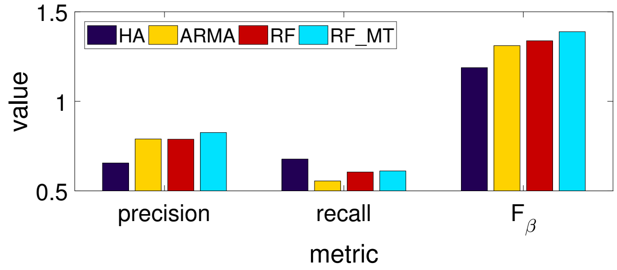


Fig. 17. Precision, recall and F-measure for probabilistic forecasts.

F-measure: F-measure is a weighted average of the precision and recall. We use F_β which weighs precision higher than recall by setting $\beta = 0.5$:

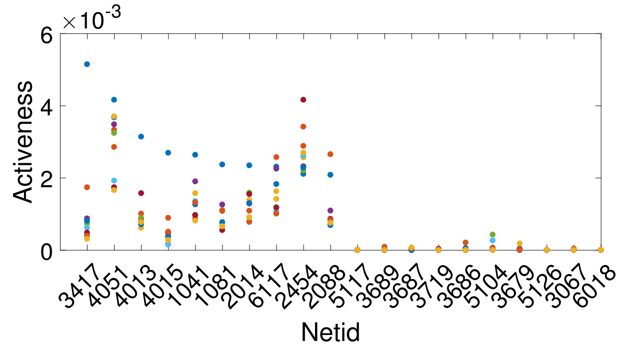
$$F_\beta = (1+\beta)^2 \frac{Precision \cdot Recall}{\beta^2 Precision + Recall} \tag{9}$$

We compare our proposed probabilistic forecasts method with the following three algorithms:

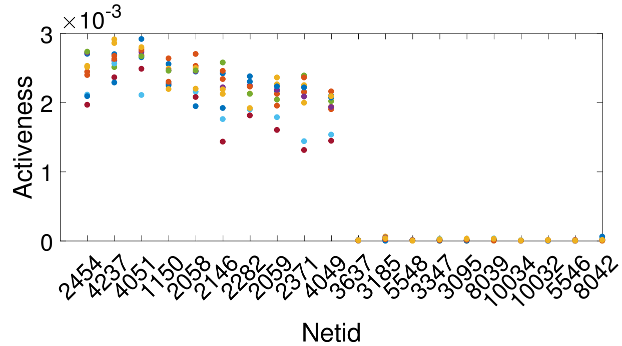
- **Historical average (HA)** predicts the usage demand by averaging the historical values for the same day and time [17]. For instance, the check-out number of Monday 08:00 a.m. equals to the average of check out numbers of Monday 08:00 a.m. in the history and check out number of 08:00 a.m. last day.
- **Auto-Regressive and Moving Average (ARMA)** belongs to time series analysis methods and has been applied in demand prediction in [6]. It captures the temporal patterns of rental and return by leveraging check in/out information of the most recent p time windows.
- **Random forest (RF)** is the basic model where fine-grained prediction is not considered. Therefore, this method directly gives the prediction of stock number instead of probabilistic results for each station.
- **Random forest and Monte Carlo method (RF_MT)** is the proposed probabilistic forecasts method in this paper.

For the experiment setup, we divide the historical records into two parts: the first 20 days for training and last 10 days for testing. We extract over-demand events by comparing the predicted stock with the threshold η multiplying the capacity. η equals to 0.2 for check out prediction and 0.8 for check in prediction.

The results are shown in Figure 17. As one can see from Figure 17, the precision of RF_MT method is as much as 0.826, which is 25.9% more than the HA method. ARMA and RF methods have relative higher precision but the recall of ARMA is only 0.55, which is the lowest among the three methods. On the other hand, we observe that the recall of HA is significantly larger than other methods. This is because HA method tends to predict more over-demand events, which makes most of the real over-demand events can be predicted successfully. Due to this characteristic, HA method is low in precision. Among all the approaches, RF_MT method demonstrates the best performance both in terms of precision and F-score.



(a) Activeness changes within 10 hours.



(b) Activeness changes within 10 days.

Fig. 18. Activeness changes with the time.

5.2 Activeness Changes

In the simulation, we notice that the activeness of stations has different characteristics under different time granularities. The results are shown in Figure 18. Figure 18a reflects the activeness changes of Top 10 active and inactive stations within 10 hours. Different colors represent different hours/days. Since check out number in one hour is uncertain and random, the activeness of active stations fluctuates wildly. Meanwhile, the difference between active and inactive station looks rather small due to the short time interval. Figure 18b reflects the activeness changes of Top 10 active and inactive stations within 10 days. It shows relatively smooth changes of activeness for active stations and there are deep gaps between active and inactive ones. In the simulation, we update the activeness of stations for each hour because the activeness changes can be more obvious among hours especially when only small part of users follow the advisor.

5.3 Bike Usage Distribution

In this section, we evaluate the effectiveness of the proposed trip advisor by demonstrating the results of bike usage distribution.

5.3.1 *Case Studies*

We first present a case study of bike 1400,865, which is the most tireless bicycle on June 22. This bike has been rented for 38 times in the original records, while the number of records in the simulation is 13 times. As shown in Figure 19, the blue circle represents bike usage in the simulation, the orange rhombus represents that in the original records.

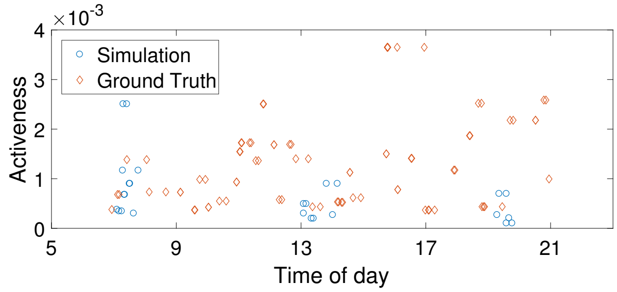


Fig. 19. Activeness of stations that bike 1400845 passed.

From the figure, we can see that bike usage frequency has been significantly reduced, especially in the later stage. After arriving at some relatively inactive stations, the bike will stay there for a while instead of quickly flowing to the next active station. On average, the visited stations are less active than those in the original records.

5.3.2 Overall Performance

To study the overall performance on bike usage distribution, we adopt PDFs of both usage number and usage time of bikes as performance metrics. In addition, we also use average (AVG) and standard deviation (STD) of usage number and usage time for evaluation. As shown in Figure 20 and Table 7, we compare situations when different proportions of the users, with 100, 75, 50, 25 and 0 percent, respectively, follow the advisor. We have two observations. Firstly, we can see from Figure 20a that compared with 0%, the percentage of less used bikes whose usage number belongs to [0,5] increases by 14.8% and the percentage of frequently used bikes whose usage number belongs to [15,40] decreases by 33.6% when the user proportion is 100%. We find out that the average usage number per day for each bike decreases from 7.656 to 6.901 when 50% of the users listen to the advisor. When the percentage rises to 100%, the average usage number is 6.625 which is down by 13.5%. The reason is that the advisor tends to use bikes that are rarely or never used more frequently. Since the total user demand stays the same with the original records, the more bikes are used, the smaller the average usage number will be. Secondly, the average usage time per day becomes more balanced as shown in Figure 20b, especially for the bikes with usage time larger than 6 hours per day. The percentage of frequently used bikes whose usage number belongs to [6,15] decreases by 28.6% when the user proportion is 100%. These results prove that the proposed method can help to balance both bike usage number and usage time. In addition, with the proportion of users grows, the effect of usage balancing gets better.

5.4 Impact of Range Settings

Experimental results for the advisor derived in this paper show high performance, demonstrating the potential of the approach. To better understand the performance of the proposed method, we further conduct an evaluation by varying the range parameter in the model. The range R is the distance allowed between stations and the origin/destination, which is set from 500m to 1000m and 200m.

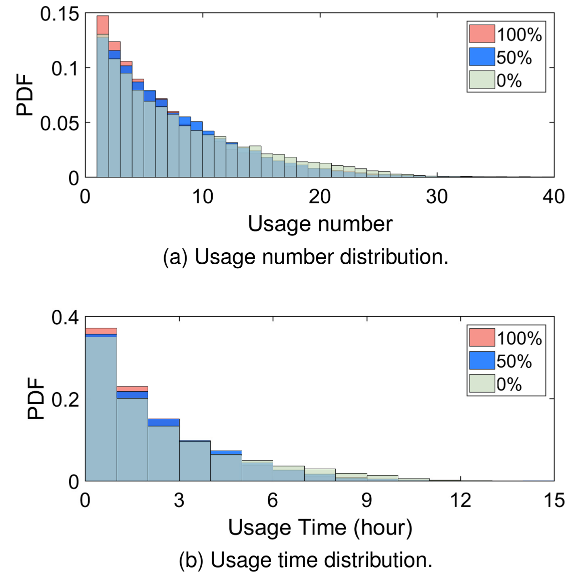


Fig. 20. Usage distribution under different proportions of the users.

TABLE 7
AVG and STD usage under different proportions of the users.

User proportion	AVG of usage number	STD of usage number	AVG of usage time	STD of usage time
100%	6.56	5.60	2.11	1.99
75%	6.64	5.14	2.15	1.93
50%	6.83	5.45	2.22	2.04
25%	7.17	5.70	2.36	2.21
0%	7.57	6.16	2.50	2.39

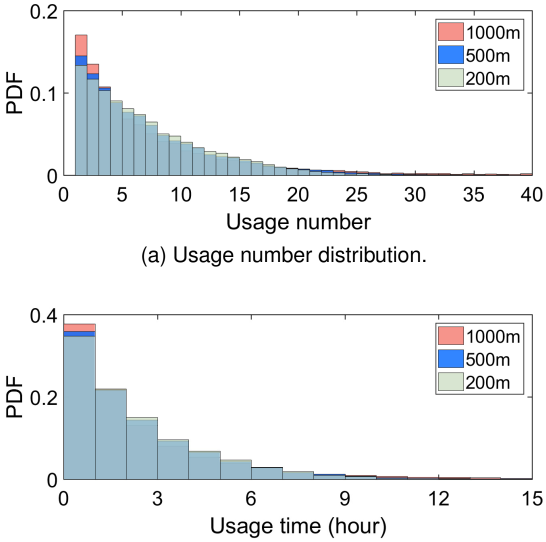
Here, we assume that all the users follow the advisor. The bike usage distribution under different range settings are shown in Figure 21. When the range is set to 200m, usage number between 5 and 15 per day take the large proportion compared with other settings which have benefit effect on usage balancing. However, there are only few stations to be chosen when $R = 200m$ and the simulator failed to offer a suggestion for more than 15,000 times per day. When the range is set to be 1000m, the experiment results have been improved, but too large range settings will cause added walking distance of users and seriously impact user experience.

6 Discussion

In this part, we provide some insights into the proposed framework, and provide directions for future work.

6.1 Reward Design

Although the advisor can improve the success rate of rental and return to a certain extent, it may also bring additional distance cost to users when realizing the goal of balancing bike usage. For the sake of keeping users' enthusiasm, recent works have investigated reward mechanisms to guide the use of shared bikes. An incentive scheme is proposed in [18] that dynamically sets incentives based on model predictive control to change the endpoint of customers' journey to



(a) Usage number distribution.

(b) Usage time distribution.

Fig. 21. Usage distribution under different range settings.

alternative nearby stations. [19] encourages users to return bikes to the less loaded station between two neighboring ones to improve the empty-station condition with price incentives. [20] designed a dynamic pricing mechanism based on an efficient and provably near-optimal on-line learning framework under given budget constraints.

As for our work, we can extend the last research to take additional cycling distance cost and individual user behavior into account, which is illustrated in Algorithm 1. In each time interval of a day, the budget B_k is calculated based on the forecasted number of trips N_k. When given a list of available prices that the mechanism offers to define the value of users' additional walking time, the incentive level p_{m^n} can be selected based on the current budget B^n and the optimistic estimate $\widetilde{F^n_{u,m}}$ on the *"current cost curve"* $F^n_{u,m}$, which denotes the cumulated user-level acceptance rate when given p_m and can be computed iteratively with the accept times $N^n_{u,m}$. The reward r^n can then be calculated with the incentive level and the additional cycling cost that users have to pay, where $s_i \in O$, $s_j \in D$. Then, the mechanism transforms the reward of the users into a discount of their public transit cards. Detailed design and evaluation of such reward mechanism are beyond the scope of the paper, and there are many references on this subject. Through this way, users are motivated to help balancing bike usage and its beneficial to build intelligent and self-sustainable transportation systems.

6.2 Other Objective Functions

In practical applications, the advisor enables system operators to design other objective functions, thus achieving flexible resource scheduling. For example, we could advise users to rent bikes from active stations and still return them to active stations. Therefore, the aging process of a small part of bikes will be accelerated, allowing the regular upgrades of bikes in the system. Otherwise, it's unacceptable to the normal operation of the systems that a large number of bikes need replacing at the same time.

Algorithm 1 Incentive Mechanism

Input: start time t_0, user u, daily budget B, number of trips in each time interval $\{N_0, ..., N_k, ..., N_h\}$, number of total trips N, available prices offered $\{p_0, ..., p_m, ..., p_q\}$;

Output: Price r^n at iteration n;

Initialization:
- **First time interval.** $n = 0$; $h^0 = h(t_0)$;
- **Budgets.** $B_k = \frac{N_k \cdot B}{N}$, $\forall k \in [0, h]$; $B = B_{k^n}$, $B^n = B$;
- **Value estimates.** $N^n_{u,m} = 0$, $F^n_{u,m} = 0$, $\forall m \in [0, q]$;

for each request at time t **do**
  if $k^n \neq k(t)$ **then**
    $k^n = k(t)$;
    $B^n = B^n + B_{k^n},\ B = B^n$;
  end if
  $\widetilde{F^n_{u,m}} = F^n_{u,m} + \sqrt{\frac{2 \cdot ln(n)}{N^n_{u,m}}}$;
  $m^n = \underset{m \in [0,q]}{arg\ max} \left\{ min \left(\widetilde{F^n_{u,m}}, \frac{B}{N \cdot p_m} \right) \right\}$ s.t. $p_m \leq B^n$;
  return $r^n = p_{m^n} \cdot [dist(l_o, s^*_i) + dist(s^*_j, l_d) - min(dist(l_o, s_i) + dist(s_j, l_d))] + f \cdot [dist(s^*_i, s^*_j) - min(dist(s_i, s_j))]$;
end for

Feedback: Observe acceptance decision y^n;

Update Variables:
- $B^{n+1} = B^n - r^n \cdot y^n$; $F^{n+1}_{u,m^n} = F^n_{u,m^n} + \frac{y^n - F^n_{u,m^n}}{N^n_{u,m^n} + 1}$;
- $N^{n+1}_{u,m^n} = N^n_{u,m^n} + 1$; $k^{n+1} = k^n$; $n = n + 1$;

6.3 Application in Car-sharing Systems

Similar to bike-sharing systems, rebalancing car utilization is a significant issue in car-sharing systems. The proposed framework can also help by directing users but there are still some differences between these two systems. On the one hand, there are only about one hundred car rental stations in Hangzhou by the year of 2018 which is far less than bike stations [21]. On the other hand, users of car-sharing systems are willing to pay more compared to users of bike-sharing systems, so it's easier to use the price mechanism to guide the use of cars. Therefore, we need to further improve the proposed framework for application in car-sharing systems in the future.

7 Related Work

Due to the increasing importance and rapid development of bike-sharing systems, a great deal of attention has been focused on a variety of problems that relate to bike-sharing. There are various interesting research questions concerning the establishment, operation and strategic problems of bike-sharing systems [12], [22], [23], [24], [25]. For example, Shaheen *et al.* [12], [22], [23] studied the history, business models and the social and environmental benefits of bike-sharing in Europe, the Americas and Asia. Parkes *et al.* [24] explored systems' location, evolution, and their adoption. In addition, a novel use case of the heterogeneous urban open data, namely bike-sharing station placement, was proposed in [25] while similar problem like electric vehicle charging station placement has been studied in [26].

Another important research direction concerns user demand prediction. Several papers firstly analyzed user behavior patterns and then proposed predictive models to forecast bike usage demand or stock level of stations in the future period [4], [5], [6], [7], [8]. The prediction methods are summarized into two categories: station-centric model and cluster-centric model. The station-centric model predicts demand for each station individually. For instance, Froehlich *et al.* [4] used four basic prediction models to predict available bikes in each station: last value, historical mean, historical trend and Bayesian network. Kaltenbrunner *et al.* [5], Borgnat *et al.* [7] and Vogel *et al.* [6] distinguished typical usage patterns and predicted the hourly user demand in the bike-sharing systems of Barcelona, Lyon and Veinna, respectively, by using time series analysis method. However, these methods show their limitation on prediction performance, especially when predicting the traffic under unusual situations. For cluster-centric model, it usually partitions the stations into clusters and predicts the total demand of each cluster [8], [27]. For example, Yexin Li *et al.* [8] proposed a hierarchical prediction model, which contains a bipartite clustering algorithm, a multi-similarity-based inference model, and a check-in inference algorithm, to predict the number of bikes that will be rent from/returned to each cluster, but the geographical granularity of this method is too sparse for trip advisor design.

Based on insights into usage demand analysis, the allocation of resources, bikes and empty places, has to be managed by the operator. To balance the stock level, methodologies in [28], [29], [30], [31] tackled the problem of finding truck routes and decided the number of bikes to move between stations that minimizes the distance traveled by trucks. Raviv *et al.* [28] presented two mixed integer linear program formulations to solve the static repositioning problem which assumes that the repositioning is during the night when the usage rate of the system is negligible. Authors in [29] introduced a dynamic public bike-sharing balancing problem when the status of the system is rapidly changing. Redistribution can also be done by users through a crowdsourcing mechanism that incentivizes the users in the bike repositioning process [19], [20]. In addition to bike-sharing data, researchers studied about balancing power demand through EV mobility in [32]. Similar method has also been applied into vehicle sharing systems in [33]. Both dynamic vehicle redistribution and online price incentives were considered in [18]. Different from the above methods, we establish a framework aiming at balancing the usage of bikes instead of the stock level of stations.

8 Conclusion

In this paper, based on the analysis of general characteristics, spatial temporal patterns and bike utilization in bike-sharing, we propose a novel architecture of a utilization-aware trip advisor which engages users to balance bike usage and prolong the maintenance intervals of bikes. Starting from ensuring users' success rate of rental and return, the advisor is designed to dynamically recommend the optimal stations based on their current activeness of bike usage. We evaluated the proposed system through extensive simulations using historical records from the world's largest bike-sharing system, confirming the effectiveness of our framework.

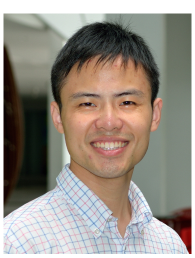
Acknowledgment

We would like to thank the anonymous editor and the reviewers. The work was partially supported by the National Basic Research Program (973 Program) under grant 2015CB352503. The corresponding author of this paper is Jiming Chen.

References

[1] P. DeMaio, "Bike-sharing: History, Impacts, Models of Provision, and Future," *Journal of Public Transportation*, vol. 12, no. DeMaio 2004, pp. 41–56, 2009. [Online]. Available: http://www.transitinformatics.org/test/nctr/wp-content/uploads/2010/03/JPT12-4DeMaio.pdf
[2] P. Midgley, "Bicycle-sharing schemes: enhancing sustainable mobility in urban areas," *United Nations, Department of Economic and Social Affairs*, pp. 1–12, 2011.
[3] L. MetroBike, "2016 Year-end wrap-up will appear at the end of January," http://bike-sharing.blogspot.com/2017/01/2016-year-end-wrap-up-will-appear-at.html.
[4] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and Predicting the Pulse of the City through Shared Bicycling," in *IJCAI*, 2009.
[5] 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*, vol. 6, no. 4, pp. 455–466, 2010.
[6] P. Vogel and D. C. Mattfeld, "Strategic and Operational Planning of Bike-Sharing Systems by Data Mining - A Case Study," in *Computational Logistics*, 2011, pp. 127–141.
[7] P. Borgnat, E. Fleury, C. Robardet, and A. Scherrer, "Spatial Analysis of Dynamic Movements of Vélo'v, Lyon's Shared Bicycle Program," in *European Conference on Complex Systems (ECCS)*, 2009.
[8] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic Prediction in a Bike Sharing System," in *ACM SIGSPATIAL*, 2015.
[9] Z. Online, "Hangzhou will add three public bicycle maintenance bases," http://zjnews.zjol.com.cn/system/2013/09/23/019608857.shtml.
[10] MSA, "BIKE SHARE PROGRAM," http://www.michigansuburbsalliance.org/wp-content/uploads/MSA-CAP-Strategy-Library_Bike-Share-Program_20211017.pdf.
[11] Wikipedia, "Hangzhou Public Bicycle," https://en.wikipedia.org/wiki/Hangzhou_Public_Bicycle.
[12] S. Shaheen, H. Zhang, E. Martin, and S. Guzman, "China's hangzhou public bicycle: Understanding early adoption and behavioral response to bikesharing," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2247, pp. 33–41, 2011.
[13] dayoo, "Difficulty of renting public bicycles due to many damaged bikes," http://hainan.ifeng.com/news/detail_2014_06/18/2451605_0.shtml.
[14] J. L. Gastwirth, "A general definition of the lorenz curve," *Econometrica: Journal of the Econometric Society*, pp. 1037–1039, 1971.
[15] Z. Yang, J. Hu, Y. Shu, P. 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*, ser. MobiSys '16. New York, NY, USA: ACM, 2016, pp. 165–178. [Online]. Available: http://doi.acm.org/10.1145/2906388.2906408
[16] S. Brin and L. Page, "Reprint of: The anatomy of a large-scale hypertextual web search engine," *Computer networks*, vol. 56, no. 18, pp. 3825–3833, 2012.
[17] N. Gast, G. Massonnet, D. Reijsbergen, and M. Tribastone, "Probabilistic forecasts of bike-sharing systems for journey planning," in *ACM CIKM*, 2015.
[18] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari, "Dynamic vehicle redistribution and online price incentives in shared mobility systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1567–1578, 2014.

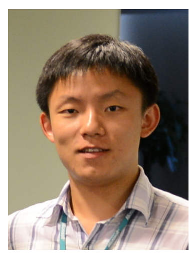
[19] C. Fricker and N. Gast, "Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity," *Euro journal on transportation and logistics*, vol. 5, no. 3, pp. 261–291, 2016.
[20] A. Singla, M. Santoni, G. Bartók, P. Mukerji, M. Meenen, and A. Krause, "Incentivizing users for balancing bike sharing systems." in *AAAI*, 2015, pp. 723–729.
[21] Panshi, "Hangzhou Micro Bus," http://www.myhz.com/?p=97.
[22] S. a. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the Americas, and Asia," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2143, pp. 159–167, 2010.
[23] S. a. Shaheen, A. P. Cohen, and E. W. Martin, "Public Bikesharing in North America: Early Operator Understanding and Emerging Trends," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2387, pp. 83–92, 2013. [Online]. Available: http://trb.metapress.com/openurl.asp?genre=article{&}id=doi:10.3141/2387-10
[24] S. D. Parkes, G. Marsden, S. A. Shaheen, and A. P. Cohen, "Understanding the Diffusion of Public Bikesharing Systems: Evidence from Europe and North America," *Journal of Transport Geography*, vol. 31, pp. 94–103, 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0966692313001130
[25] 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 *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 883–888.
[26] X. Wang, C. Yuen, N. U. Hassan, N. An, and W. Wu, "Electric vehicle charging station placement for urban public bus systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 128–139, 2017.
[27] E. O. Mahony and D. B. Shmoys, "Data Analysis and Optimization for (Citi) Bike Sharing," in *AAAI*, 2015.
[28] T. Raviv, M. Tzur, and I. Forma, "Static Repositioning in a Bike-sharing System: Models and Solution Approaches," *EURO Journal on Transportation and Logistics*, vol. 2, no. 3, pp. 187–229, 2013. [Online]. Available: http://dx.doi.org/10.1007/s13676-012-0017-6
[29] Contardo, Claudio, C. Morency, and L.-M. Rousseau, "Balancing a Dynamic Public Bike-sharing System," Tech. Rep., 2012.
[30] J. Schuijbroek, R. Hampshire, and W.-J. van Hoeve, "Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems," Tech. Rep., 2013.
[31] S. Wang, T. He, D. Zhang, Y. Shu, Y. Liu, Y. Gu, C. Liu, H. Lee, and S. H. Son, "Bravo: Improving the rebalancing operation in bike sharing with rebalancing range prediction," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 44, 2018.
[32] R. Yu, W. Zhong, S. Xie, C. Yuen, S. Gjessing, and Y. Zhang, "Balancing power demand through ev mobility in vehicle-to-grid mobile energy networks," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 79–90, 2016.
[33] 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.



Peng Cheng received the B.E. degree in automation and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. He is currently a Professor with the College of Control Science and Engineering, Zhejiang University. His research interests include networked sensing and control, cyber-physical systems, and control system security. He serves as an Associate Editor of the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, Wireless Networks, and the International Journal of Communication Systems. He has served as the Guest Editor of the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS and the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. He served as the TPC Co-Chair for the IEEE IOV 2016, the Local Arrangement Co-Chair for ACM MobiHoc 2015, and the Publicity Co-Chair for the IEEE MASS 2013.

Ji Hu is a master student in College of Control Science and Engineering at Zhejiang University. Before coming to ZJU, She majored in Automation in Southeast University. Her research focuses on data analysis and data mining.

Zidong Yang is a Ph.D. student from Control Science and Engineering, Zhejiang University, China. Before that, he received B.Sc degree from Zhejiang Univeristy. His research interest includes data mining, machine learning, intelligent traffic system and mobile computing.



Yuanchao Shu is currently a Researcher with Mobility and Networking Research Group at Microsoft Research. His research interests lie broadly in mobile and wireless systems, networked control and optimization, and mobile security and privacy. His previous research results have been published at top-tier venues including ACM MobiCom, MobiSys, Ubicomp, SenSys, IEEE INFOCOM, JSAC, TMC, TPDS, USENIX Security etc. He served as Award Chair of ACM TURC 17/18, Publication Co-Chair of SmartCom 17, and member of TPC of IEEE Globecom 18/17, ICCCN 17 etc. He won IEEE WCNC Best Paper Award, IEEE INFOCOM Best Demo Award, and was the recipient of ACM China Doctoral Dissertation Award and IBM PhD Fellowship. Dr. Shu received Ph.D. from Zhejiang University in 2015, and was also a joint Ph.D. student in the EECS Department at the University of Michigan, Ann Arbor.

Jiming Chen received B.Sc degree and Ph.D degree both in Control Science and Engineering from Zhejiang University. He was a visiting researcher at University of Waterloo from 2008 to 2010. Currently, he is a Changjiang Chair professor with College of control science and engineering, and deputy director of the State Key laboratory of Industrial Control Technology at Zhejiang University, China. He serves/served associate editors for several international Journals including IEEE TII, IEEE TPDS, ACM TECS, etc. His research interests include sensor networks, IoT, networked control, cyber security.He is a Distinguished lecturer of IEEE Vehicular Technology Society 2015-2018.