

Mobility Modeling and Data-Driven Closed-Loop Prediction in Bike-Sharing Systems

Zidong Yang, Jiming Chen¹, Ji Hu, Yuanchao Shu, and Peng Cheng²

Abstract—As an innovative mobility strategy, public bike-sharing has grown dramatically worldwide. Though it provides convenient, low-cost, and environmental-friendly transportation, the unique features of bike-sharing systems give rise to problems for both users and operators. The primary issue is the uneven distribution of bikes caused by ever-changing usage and (available) supply. This imbalance necessitates efficient bike rebalancing strategies, which depends highly on bike mobility modeling and prediction. In this paper, a trace-driven simulation-based prediction approach is proposed by simultaneously taking user mobility demand and real-time status of stations into consideration. We extensively evaluate the performance of our design with the dataset from one of the world’s largest public bike-sharing systems located in Hangzhou, China, which owns more than 2800 stations. The evaluation results show an 85 percentile relative error of 0.6 for checkout and 0.4 for checkin prediction. The preliminary results on how the predictions can be used for bike rebalancing are also provided. We believe that this new mobility modeling and prediction approach can improve the bike-sharing system operation algorithm design and pave the way for rapid deployment and adoption of bike-sharing systems across the globe.

Index Terms—Bike-sharing, mobility modeling, flow prediction, Monte Carlo simulation.

I. INTRODUCTION

BIKE-SHARING System (BSS), as part of the shared transportation and shared economy, has grown tremendously in recent years. It provides a convenient, low-cost and environmental-friendly means for last mile transportation in the urban cities and attracts attention from both citizens and government. It is reported that more than 500 bike-sharing programs currently running in at least 49 countries with one million shared bikes in 2015 [1], [2].

In addition to its advantages of reducing traffic congestion and mitigating pollution, BSS features unique characteristics compared with other forms of shared-use mobility. First, bike-sharing differs from classic ride-sharing (e.g., carpooling) and ride-sourcing (e.g., Uber and Lyft) in that bikes are typically *unattended*. During vacant hours, bikes are *concentrated* at

a group of stations where operations of checkin/checkout are facilitated through a backbone network, i.e., an IT infrastructure that enables system management and monitoring. Second, unlike conventional public transit (e.g., subways and buses) which follows a regular schedule and pre-determined routes, bike-sharing provides transportation on an *on-demand* basis with a *decentralized* structure. These two distinct features, however, pose characteristic challenges in BSS management and optimization. One common problem, for example, is that the system typically ends up with an uneven distribution of bikes across the different stations (due to the uncontrolled, uneven demand), often rendering the checkin or checkout service unavailable at some stations where bicycle docks are either fully occupied or empty.

This bike unbalance problem makes it necessary for bike-sharing cities to employ costly *bike redistribution*, which is typically performed by trucks or trailers driving around the city, moving bikes among stations. To increase service availability and minimize redistribution cost, studies have been conducted to improve these bike redistribution strategies based on bike mobility models and predictions. Despite the researches on bike usage patterns and global rental volume forecasts (e.g., [3]–[7]) developing a fine-grained prediction model for the optimal number of bikes that should be redistributed has proven to be elusive, and has remained an open problem. The primary technical challenge is that bike traffic is not only highly dynamic and inter-correlated in both the temporal and spatio domains, but is also further influenced by complex issues such as timing and meteorology. In addition, users’ behavior is also coupled with the number of bikes/docks in the stations. Existing works either do not consider the temporal-spatio relationship among stations or ignore the impact of the station stock on user behavior.

In this paper, we first establish a probabilistic temporal-spatio mobility model to characterizes movements (or shifts) between different pairs of stations and present a novel fine-grained prediction framework for both checkin and checkout traffic. Our work differs fundamentally from previous approaches in that 1) we model BSS as a dynamic network and predict the traffic by jointly considering the spatio-temporal correlations among stations and additional time factors and meteorology. 2) we view each station as an “agent” and develop a simulation-based approach to capture the interaction among station status and users. Based on historical data, we first combine random forest model with inter-station transfer network to estimation the number of potential user shifts in the near future. Then, the user shifts are instantiated

Manuscript received January 31, 2018; revised August 29, 2018 and October 26, 2018; accepted November 14, 2018. This work was supported by 973 under Grant 2015CB352503 and Grant NSFC 61790571. The Associate Editor for this paper was R. Malekian. (Corresponding author: Jiming Chen.)

Z. Yang, J. Chen, J. Hu, and P. Cheng are with the School of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: zdyang@zju.edu.cn; jmchen@ieee.org; hujichn@gmail.com; pcheng@ipc.zju.edu.cn).

Y. Shu is with Microsoft Research Asia, Beijing 100000, China (e-mail: yuanchao.shu@microsoft.com).

Digital Object Identifier 10.1109/TITS.2018.2886456

as concrete trace records by sampling from the probabilistic model. Finally, a dedicated trace-based simulator is designed to predict stations' checkin/checkout number by performing discrete event simulation with the generated user shift records. By using the simulator, the mutual inference between checkout and checkin is characterized by changing the number of available bikes/docks in each station, which forms a *closed loop* between checkout and checkin events.

This paper is an extended version of the authors' conference paper [8]. We highlight the following three main contributions:

- We identify the mobility modeling problem and establish a spatio-temporal dynamic network model for BSS by taking into account the interactions among all stations;
- We proposed a data-driven closed-loop simulation-based prediction framework for bike-sharing systems, which systematically considers 1) the interaction between users and stations; 2) mutual inference between checkin and checkout;
- We evaluate the performance of mobility modeling and prediction with the world's largest public BSS with more than 2800 stations and over 103 million records [9], [10]. Compared with benchmark methods, the proposed approach provides the best performance with an 85 percentile relative error of 0.6 for checkout and 0.4 for checkin prediction.

The remainder of this paper is organized as follows. We first introduce the related works in Section II, followed by an overview of our design in Section III. We then present the proposed prediction approach along with the user mobility model in Section IV. Section V presents an in-depth evaluation of mobility modeling and prediction. In Section VI, we present some preliminary results on applying the predictions in bike rebalancing. Several insights and the future work are discussed in Section VII. We conclude the paper in Section VIII.

II. RELATED WORK

Extensive research has been done to describe the nature of bike-sharing systems, business models, how they have spread in time and space and why they have been adopted [11], [12]. For example, Shaheen *et al.* [11] reviews the history, advantages and inadequacies of bike-sharing systems across the globe. Martin and Shaheen [12] evaluates transit modal shift dynamics with the emergence of public bike-sharing. Comprehensive analysis and survey of city-scale bike-sharing systems in Paris [13], New York [14] have also been conducted.

Adoption of bike-sharing systems has motivated studies on system design optimization. The first line of work focuses on the sensing of dynamics from bike-sharing system data [5], [6], [15], which broadly consider two topics, namely clustering and prediction. Most clustering approaches identify mobility patterns in bike usage and partition the stations into clusters based on their usage profiles [7], [16]. For instance, in [3], two clustering techniques using activity statistics derived either from the evolution of station occupancy or the number of available bicycles along the day. Borgnat *et al.* [17] use graphs to describe the similarity of usage profiles

between pairs of stations for weekdays and weekends, which is then analyzed using a community detection algorithm for clustering. In contrast to clustering, the aim of prediction is to forecast the occupancy of the stations or the network state over time using time series analysis [6], Bayesian networks [3] and supervised regression model [18]. For instance, Borgnat *et al.* [6] forecasts the global rental volume, whereas Li *et al.* [7] infers the bike rental/return demand of a cluster of stations based on historical checkin and checkout data. Liu *et al.* [19] used a similar idea for prediction but they use the KNN model. Chen *et al.* [20] proposed an RNN based method for prediction. However, their method only considers the self-correlation for each station. Liu *et al.* [21] use point-of-interest (POI) data for improving the prediction accuracy. The primary difference between our work and existing demand prediction works is that in previous work, they usually ignored station stock status by implicitly assuming that all predicted users could successfully checkin/checkout. However, this assumption could not be met when the bikes/docks were not sufficient. For instance when a station is empty, no predicted user can checkout a bike from it. To take the stock information and the mutual interaction between checkin and checkout into consideration, we proposed an event-driven simulation-based approach.

Based on insights into usage patterns and bike trip demand analysis, research has also been conducted to optimize the placement of stations in bike-sharing systems [18], design strategies for bicycle rebalancing [22], [23] and provide effective bike routing [24]–[26]. For example, Chen *et al.* [18] and García-Palomares *et al.* [27] solve the station placement problem by estimating the potential trip demand using a semi-supervised learning algorithm and a GIS-based method, respectively. Raviv *et al.* [22] find truck routes by minimizing an objective function tied to both the operating cost of the vehicles as well as penalty functions relating to station imbalance. Li *et al.* [28] propose a reinforcement learning method for dynamic bike reposition. Bao *et al.* [29] use bike traces for bike lane planning. The mobility model and prediction mechanism derived in our work can be easily applied to other bike-sharing systems and lay a solid foundation for the upper layer design and optimization.

People also propose some trip-based approach for bike-sharing simulation [30]–[35]. Jian *et al.* [30] use a Poisson process to simulation the bike flow. Chemla *et al.* [31] consider users' maximal waiting time and utility function. The primary difference between our work and exiting works is that they did not consider prediction in their design. However, utilizing simulation system for prediction is not a trivial problem because it depends on accurate estimation of the users' trip intension.

In addition to bike-sharing data, traffic data also attract research from various areas [36]. Researchers analyzed human mobility based on other empirical data from taxicabs [37], buses [38] and cellular data [39], [40]. Due to the unique intrinsic properties such as the decentralized structure, on-demand usage, and unattended vehicles in BSS, our work provides a fundamentally different model from these designs.

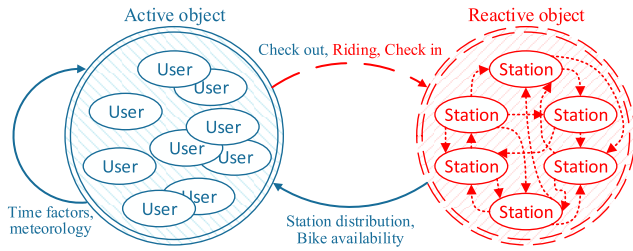


Fig. 1. Components of a bike-sharing system.

III. DESIGN OVERVIEW

This section provides a design overview, including problem formulation in Section III-A and design methodology in Section III-B.

A. Problem Formulation

Two types of entities constitute a BSS (see Figure 1): *Active objects* (users) and *Reactive objects* (bikes). Users shift bikes from checkout to checkin operations, changing the status (i.e., the number of docked bikes) of stations located at different places. Conversely, the spatio diversity of stations and bike availabilities also influence user behaviors. We call a sequence of operations - bike checkout, movement and checkin - a *shift instance* (SI). As can be seen from Figure 1, *Active objects* and *Reactive objects* are coupled in both the temporal and spatio domain. However, it is worth noting that user activities in *Active objects* are *mutually independent*, though subject to the change of time factors and meteorology.

Our objective is two-fold. First, we aim to model the mobility patterns of bikes in BSS. The mobility model characterizes the spatio-temporal transition of bikes among stations. Second, based on the mobility model, we aim to predict the number of checkin/checkout users at each station in the future. Due to the correlation among stations and the mutual influence between *Active objects* and *Reactive objects*, our design consider both parts simultaneously and propose a unified simulation-based prediction framework.

B. Design Methodology

We observe from real-world BSS that two kinds of dynamics are involved in the systems. For one thing, there exists the interaction between users' checkin/checkout behavior and stations' status (e.g., a user will fail to checkout/checkin a bike when the target station is empty/full). For another, there also exists complicated mutual influence between checkin and checkout. For example, the number of checkout bikes will certainly change the number of checkin bikes in the destination stations while the checkin number can also influence checkout by changing the number of available bikes.

As depicted in Figure 1, for one hand, despite the random bicycle checkout time and location in each SI, bikes are bound to checkin at certain stations (upper red arrow) and for another, the stock status of stations will affect users' checkout behavior conversely (lower blue arrow) and the real-time stock is mainly determined by the number of checkout/checkin users and the station capacity.

This property motivates us to model mutual influence between checkin and checkout as well as between users and

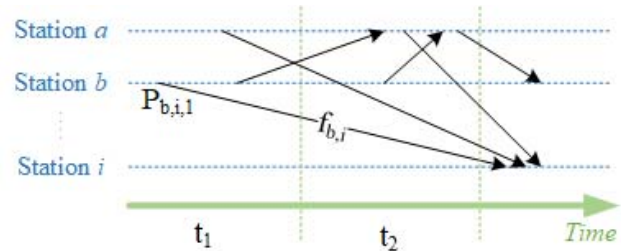


Fig. 2. Basic framework of mobility model.

stations by keeping track of every individual event. For this end, a data-driven trace-based simulation approach is proposed to perform the prediction. Firstly, we develop a probabilistic temporal-spatio model to describe the transfer relationship among stations. Secondly, we combine the transfer model (for the destination station and duration prediction) with random forest model (for the source station and departure time prediction) for synthetic traces generation. Thirdly, all generated records are fed into a dedicated simulator, which keeps track of all transition records and station stock changes. The number of checkin/checkout users are predicted by counting the number of successful checkin/checkout in the corresponding time slots. Since the synthetic records are generated randomly, we adopt the Monte-Carlo approach for final prediction. In other words, the final result is the average of multiple runs of simulation.

IV. MOBILITY MODEL AND PREDICTION

A. Bicycle Mobility Modeling

In this section, we develop a mobility model to capture the spatio-temporal transition of bikes. The model is established based on the features of bike flows between different pairs of stations in different time. For example, bikes flow into stations in working areas in the morning on weekdays and flow out in the afternoon. Therefore, we propose a statistical model, which is based on historical checkin/checkout data. The model uses a probabilistic framework to describe the spatio-temporal shifts of bikes between various pairs of stations and estimates bike checkin based on the online checkout records.

We first develop a probabilistic mobility model to characterize the transfer probability from one station to another. In other words, this model can predict a user's **destination** and **trip duration** given the station in which he/she checkout one bike. Thus, two sets of parameters are estimated in this part: 1) the transfer probability; 2) the riding time, as illustrated in Figure 2. $P_{b,i,1}$ is the probability for a bike, which checkout at station b in time slot t_1 and checkin at station i finally. $f_{b,i}$ denotes the distribution function of riding time from station b to station i.

1) *Station-Station Probability*: According to our observation of the transition data, the station to station probability may be time-variant. For example, the probability from the residential area to the business area may become significantly larger in morning rush hour on weekdays than other time periods. Due to this property, we use $P_{i,j,t}$ to denote the probability that one bike checkout from station i at time slot t will checkin at station j .

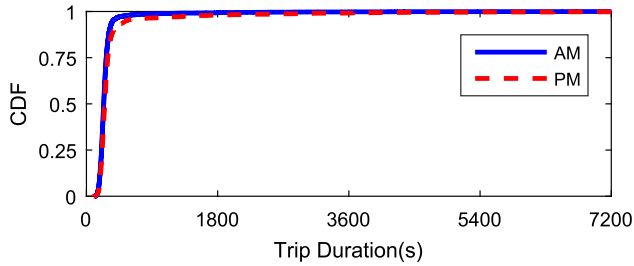


Fig. 3. CDF of trip duration between two stations.

It is also worth investigating how to efficiently update the parameter $P_{i,j,t}$. For one thing, if the time slot t is too large, the model will be less accurate since $P_{i,j,t}$ may be highly volatile within a day. For another, if t is too small, the data will be too sparse for probability estimation and also undermines the performance of prediction. In the proposed mobility model, we set the length of each time slot to one hour to get a proper tradeoff.

In summary, we first group all trip records according to their source station, destination station and time slot. Let $N_{i,j,t}$ denote the number of shifts from station i to station j at time slot t in the historical data. Then $P_{i,j,t}$ is calculated by

$$P_{i,j,t} = \frac{N_{i,j,t}}{\sum_j N_{i,j,t}} \quad (1)$$

Since $P_{i,j,t}$ may be different under different conditions (e.g., weather and holiday). We compute $P_{i,j,t}$ for sunny workdays, sunny holidays, rainy workdays and rainy holidays receptively.

2) *Trip Duration*: Although the trip duration may be pretty different from one user to another, it is observed from the data that the distribution of trip duration is quite stable regardless of the time slot t . As an example, the empirical CDF of trip time from one station (9702) to another (9706) in the morning and afternoon are shown in Figure 3 respectively. As we can see, two curves are quite close.

Thus, from trip duration, we first group the SIs according to their source and destination. Then, we estimate the distribution function f_{ij} for each group.

3) *Model Visualization*: In this part, we present a visualization result based on our probabilistic transfer model to justify its effectiveness. Here We present the spatio characters of bikes flows according to our model. In Figure 4, we draw the primary flow directions at 7:00 am according to the probability and station locations. As can be seen from this figure, most bike flows go to the downtown area of the city.

B. Potential Trip Generation

In addition to the transfer model described in Section IV-A, in this section, we propose a mechanism to generate user potential trip demand in the near future, which lays fundamentals for our simulation-based Monte Carlo prediction.

The purpose of this section is to predict the entire records in the near future. Each transition record consists of checkout/checkin station id and the timestamp. In principle, the checkout demand is determined by the users traffic desire, but in contrast, the checkin demand can be viewed as the

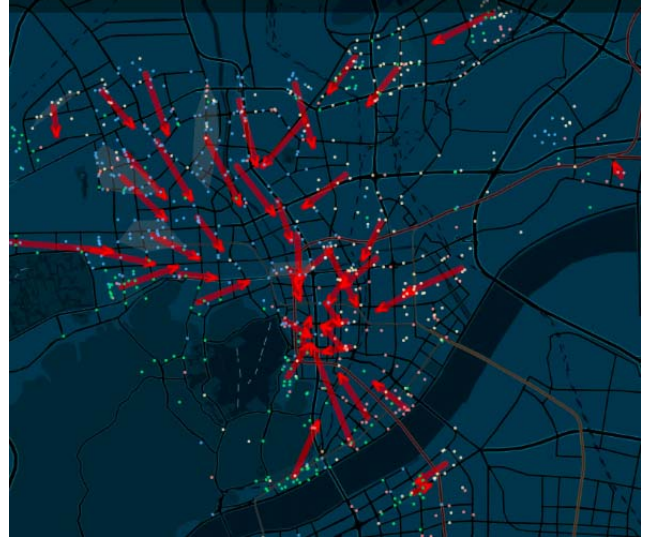


Fig. 4. Flow direction at 7:00 A.M.

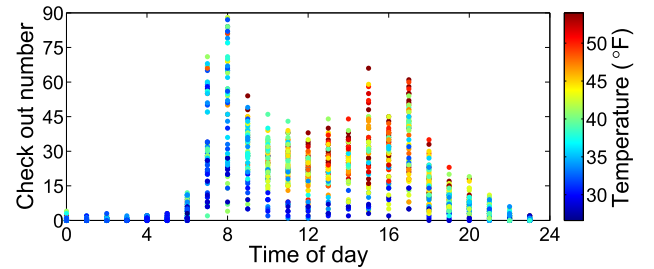


Fig. 5. Checkout number by time and temperature.

consequence of users' checkout behavior. According to this intrinsic property of BSS (each checkout is paired with a checkin), we first predict users' potential checkout demand at each station using the Random Forest model along with external features. Then, the destination and trip duration for each checkout are randomly sampled from the mobility model.

1) *Checkout Prediction*: We apply *random forest theory* [41] to model and forecast the checkout behaviors based on historical SIs along with important features. We utilize both online and offline feature to gain a more accurate prediction result. The difference between offline and online features is that most offline features can be gained in advance (e.g., time and temperature) while online features can only be accessed in real time (e.g., checkout/checkin number in last time slot). The features we adopt are as follows:

a) *Time factors*: Although characteristics of checkout actions differ among stations, they are all closely related to time factors and show unique temporal patterns. We select the four most significant time factors: day of week, time of day, weekday and holiday.

b) *Meteorology*: Meteorology condition has a huge influence on user behaviors in BSS [7], [18], [42]. As is shown in Figure 5, checkout numbers grow when people feel more comfortable at higher temperatures. Similar patterns exist for other meteorological conditions such as humidity, visibility and wind speed. Nevertheless, these patterns vary with time and across stations. For example, users' checkout behaviors are less influenced by weather conditions during peak hours.

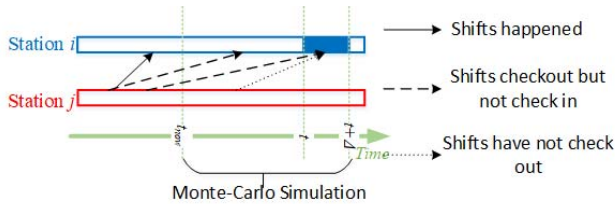


Fig. 6. Trace based prediction.

c) *Online checkout number*: Despite the tight correlations between users' checkout behaviors and time factors as well as the meteorological data, there exist checkout anomalies that differ significantly from the results observed at the same hours on other days. We notice that these anomalies are caused by sudden changes of bike availabilities at stations. For instance, bikes in surrounding areas run out very quickly with large audiences coming from a stadium after a soccer game. Also, the bike-sharing service becomes unavailable when dock (or even station) failures happen. Since anomalies usually last for a short period, we adopt an online feature of checkout number from the previous time window. By incorporating this feature, the model is capable of adapting to unexpected events that dramatically change the bike availability at stations.

We combine all offline and online features mentioned above to generate a feature vector f_t for each time window t . Denote the ground truth of checkout number for each time window as r_t , we combine feature f_t and r_t into a big vector $x_t = (f_t, r_t)$ to train the model. The predicted number is rounded to the nearest integer for the following record generation.

2) *Entire Trip Generation*: After generating the checkout number for a specific station in a time window (e.g., 5 checkouts from 8:00 a.m. to 8:30 a.m.), the exact leaving time of each user is assumed to be independently and follows uniform distribution (e.g., uniformly sample from 8:00 a.m. to 8:30 a.m.).

Similarly, the destination and duration for one trip are generated by sampled from the mobility model. For example, if one user checking out from station i in hour t , his/her destination is first sampled from $P_{i,j,t}$. Once determining the target station j , the trip duration is sampled from f_{ij} . The expected arriving time is calculated by adding up the leaving time and the trip duration.

C. Trace Based Prediction

Finally, the synthetic records generated above is fed into a trace-based simulator to perform predictions.

We use an example to demonstrate how the simulator works. To predicted the future checkin/checkout demand in time window $[t, t + \Delta]$, the simulator is initialized with stations' stock level at time t_{now} .

Without loss of generality, we consider shifts from station j to station i as shown in Figure 6. All events can be classified into three classes according to their time stamps. Here, we should consider two kinds of events: the real events that checkout before t_{now} but checkin after time t and the synthetic events that checkout during $[t, t + \Delta]$.

The destination and trip time of both events are estimated from the mobility model but for the former events,

the checkout time and station are actual while for the latter events, this information is predicted by random forest model.

The simulator then processes the events in time order. However, not all user demand can be satisfied (e.g., the station can be empty for checkout or full for checkin). We use an underlying user behavior model to cope with unexpected events.

- If a user fails to checkout a bike, he/she will wait for at most p seconds. Then the user will leave the BSS.
- If a user fails to checkin a bike, he/she will keep waiting until there is a vacant dock.

The simulator follows the settings described above and can keep track of all successful checkin/checkout events and rejected users for each station.

Since the trip destination and duration sampled from the probabilistic mobility model involve randomness, we utilize the Monte-Carlo approach to gain a stable and reliable prediction. Specifically, we run the simulation for multiple times (usually 5) and calculate the averaged number of successful checkin/checkout as our prediction result.

V. EVALUATION

Extensive data-driven experiments were conducted in this section to evaluate the performance of the proposed model. In the following, we first describe the dataset, baselines and evaluation metrics. Then, the performance of both checkin and checkout prediction are shown with comprehensive analysis.

A. Dataset Description

We conduct the evaluation on Hangzhou Bike-Sharing System dataset from June 2015 to August 2015. Hangzhou Bike-Sharing System is one of the largest bike-sharing systems worldwide, and as for our dataset, it consists of 25,336,206 shift records, 63,968 bikes, and 3,390 stations. Each record contains the user's id, checkin/checkout timestamps, checkin/checkout station ids. Other auxiliary datasets include the static information of stations (such as location, capacity), weather data of the same period in hourly granularity. In this evaluation, the records of the first 60 days are used as training data while the remaining 32 days are used as testing data. For predictions, each algorithm is given the necessary information before specific time t (e.g., 8 a.m.) and asked to predict the checkin/checkout amount in the following time window $t + \Delta$ (e.g., from 8:00 a.m. to 8:30 a.m.). The time window is set to 30 minutes unless explicitly stated.

B. Baseline Approaches

We first introduce prediction techniques that comprise the baselines of our model.

- **Historical Mean (HM)** uses the average of historical observations of the same time and location to forecast the future data [43]. Specifically, when conducting prediction for a period, we first find out historical periods with the same time of day and day of week, and then average the checkin/checkout results from these periods.
- **Auto-Regressive and Moving Average (ARMA)** is widely used for time series prediction and was adopted for BSS checkin estimation [5]. It leverages

checkin/checkout information of the most recent p time windows for future prediction. Parameters are determined using historical data with the least squares method. Both HA and ARMA can be viewed as approaches without considering environmental/external factors.

- **Random Forest Model (RF)** is a typical machine learning approach for regression tasks and shown to be effective in the demand prediction of BSS [41]. In this evaluation, temperature, weather, hour of day, day of weekday, holiday information, the checkin/checkout demand of the previous time interval are chosen as prediction features.
- **Probabilistic Flow Model (PFM)** is the most recent work of traffic prediction in bike-sharing system [8] where the checkin demand is predicted by considering the transfer probability among different pairs of stations and the number of real-time checkout users and calculating the expected number. The difference between PFM and our approach is that PFM does not consider the available bikes/docks in each station.
- **Potential Demand (PD)** is a simplified approach which predicts the future demand by aggregating the generated synthetic records in the prediction period.

The simulation-based prediction approach proposed is abbreviated as SP.

For performance metrics, we adopt CDFs of both absolute and relative prediction error. The absolute error is the difference between ground truth and prediction while the relative error is absolute error divided by ground truth. Also, we use Root Mean Squared Error (RMSE) for evaluation. RMSE is computed as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2} \quad (2)$$

where \hat{y} and y are the prediction and the ground truth respectively while n is the number of predictions.

C. Checkout Prediction

Figure 7 presents the overall prediction performance across all stations. We divide each day into 48 30-minute intervals. All algorithms are required to predict the traffic flow for each 30-minute interval. Figure 7(a) demonstrates the CDF of absolute error while Figure 7(b) shows the CDF of relative error.

It is worth noting that in Figure 7(a), the CDF of SP and PD appear as the step function. This is because both SP and PD only predict integer number, so the absolute errors are also integers. Thus, for SP and PD, we should only pay attention to the CDF values when $error = 0, 1, \dots$. We can see that 90% of the absolute error is less than 2 in our approach, outperforming three baselines. Table I also gives evidence of the advantages of our approach while the RMSE of SP is as low as 1.91. Similar patterns exist for relative error. We note that in relative error metric, the performance of SP and RF are similar. This is because in relative error we only consider time intervals with more than five checkouts in a 30 minutes prediction period. Since the checkout demand is generated

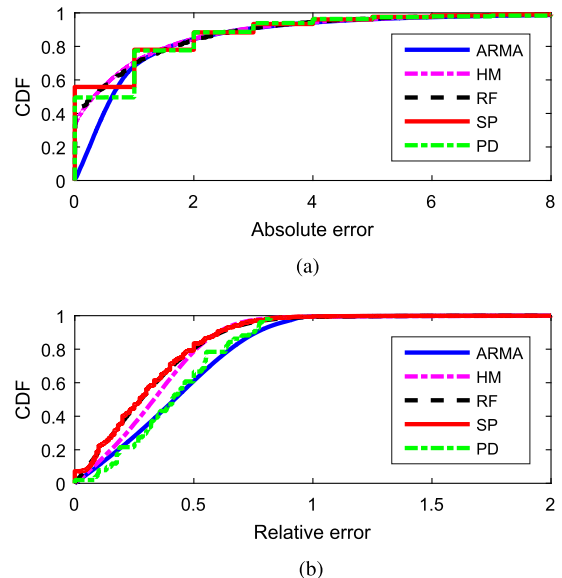


Fig. 7. Overall performance of checkout prediction. (a) CDF of absolute error. (b) CDF of relative error.

TABLE I
RMSE OF CHECKOUT PREDICTION

ARMA	HM	RF	PD	SP
2.18	2.07	1.96	2.53	1.91

based on RF prediction and five checkouts in prediction interval imply the high activeness level of the station, most predicted users will be able to checkout one bike, resulting in the similar performance of SP and RF.

D. Checkin Estimation

In this section, we evaluate the effectiveness of the proposed approach by demonstrating the results of checkin estimations.

We also evaluate the overall performance of the checkin estimation of all approaches. The CDFs of absolute error and relative error are depicted in Figure 8, while RMSE results are presented in Table II.

As one can see from Figure 8(a), for absolute error, SP outperforms competitors, demonstrating the advantage of considering station status. For example, more than 92% of errors are less than 2 in SP while it is 86% and 84% in PD and HM. Similar performance improvement can be observed from Figure 8(b). 92% of the relative error is less than 0.5 in SP. However, different from Figure 7(a), we notice that the performance of SP and RF are no longer closed. This is because the checkin prediction mechanisms of SP and RF are completely different and the result shows the effectiveness of taking track of available docks in stations (e.g., SP prevents users from checking in bikes to stations that are already full). Table II also convinces that SP has the smallest prediction error and it has a 25% performance gain over RF.

E. Impact of Settings

To better understand the performance of the proposed simulation-based approach in different settings, we further

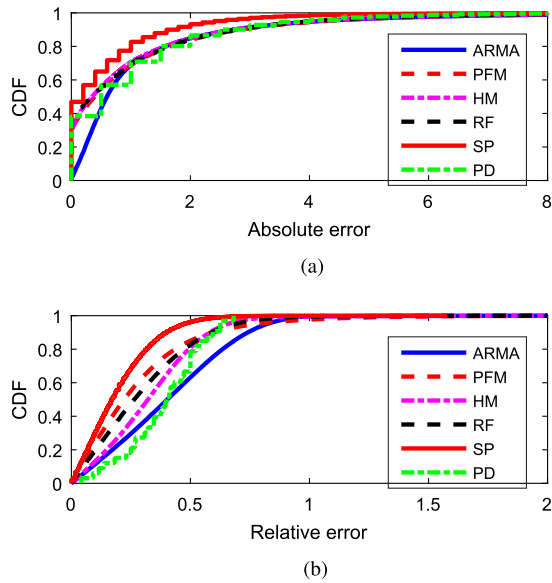


Fig. 8. Overall performance of checkin estimation. (a) CDF of absolute error. (b) CDF of relative error.

TABLE II
RMSE OF CHECKIN PREDICTION

ARMA	HM	RF	PFM	PD	SP
2.12	2.03	1.89	2.08	1.75	1.41

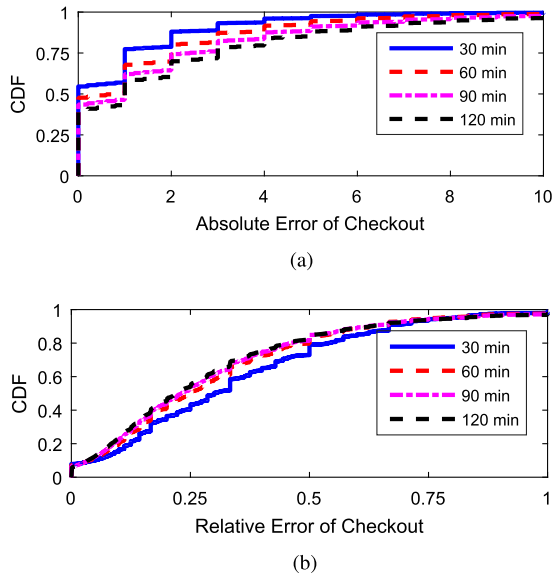


Fig. 9. Impact of prediction horizon for checkout. (a) CDF of absolute error. (b) CDF of relative error.

conduct two sets of evaluation by varying several critical parameters in the evaluation.

1) *Prediction Horizon*: In the above evaluation, we only test the performance for the prediction horizon of half an hour. However, our model can predict flow traffic in arbitrary length of prediction horizon. In this part, we change the prediction horizon from 30 minutes to 120 minutes with a step of 30 minutes. The absolute and relative error of checkout are shown in Figure 9. As we can learn from this figure, the 30-minute prediction has the best performance for absolute error while it comes to relative error 30-minute prediction is

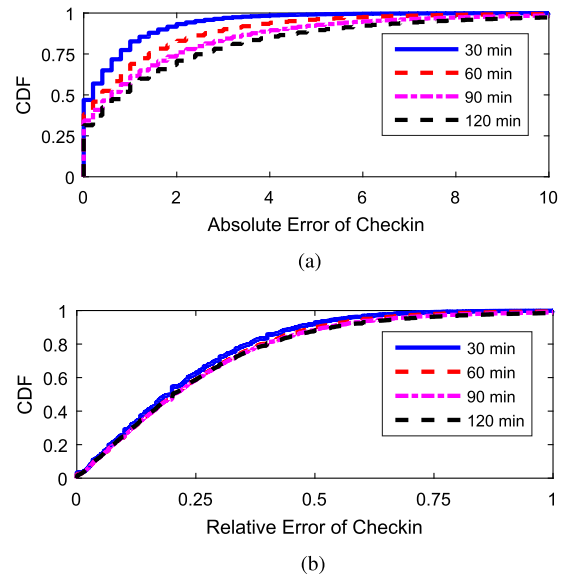


Fig. 10. Impact of prediction horizon for checkin. (a) CDF of absolute error. (b) CDF of relative error.

slightly worse. The absolute and relative error of checkin are shown in Figure 10. This time, the 30-minute prediction is the best choice for both absolute error and relative error. From both figures, we can see the performance for predicting relative error is very similar. Since we remove interval with small checkin/checkout number while calculating the relative error, we infer that a shorter prediction horizon has better prediction performance for small values.

This is because the prediction can benefit a lot from accurate online checkout information in the previous time slot. However, since the duration of most trips are short (99% of the riding time of the trip is shorter than 60 minutes.), when the prediction horizon is longer, the algorithm gets less benefit from online information and relies much more on the synthetic records.

However, we also notice an interesting result that in Figure 9(b), the 30-minute prediction has the worse prediction accuracy. The reason is that for relative error, we only consider time slots with more than five checkout records. Since 30 minutes are a shorter period, the time slots satisfying above condition usually concentrate in rush hours, which have larger volatility.

2) *Number of Runs of Simulation*: Our prediction approach is based on Monte Carlo simulation, which means result is the averaged value of multiple runs of simulation. In this part, we evaluate the impact of different number of simulation runs. We test the performance for the number from 1 to 5.

The result for checkout is shown in Figure 11. The performance for different numbers of runs are very close to each other regarding both absolute error and relative error. This is because the checkout demand is largely determined by the Random Forest model and it is identical for all runs of simulation. For checkin prediction, the performance of five runs is better than that of one run as shown in Figure 12. The reason is that the events in the synthetic dataset are generated at random. Thus, there can exist some rare events when running just one or two simulations. Instead, by running multiple

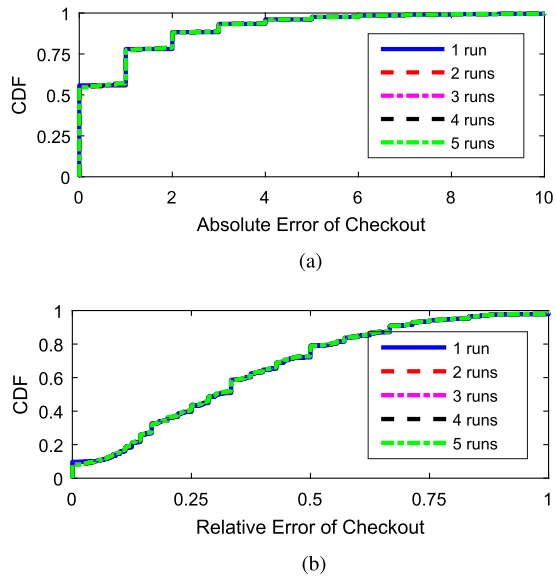


Fig. 11. Impact of different runs of simulation on checkout. (a) CDF of absolute error. (b) CDF of relative error.

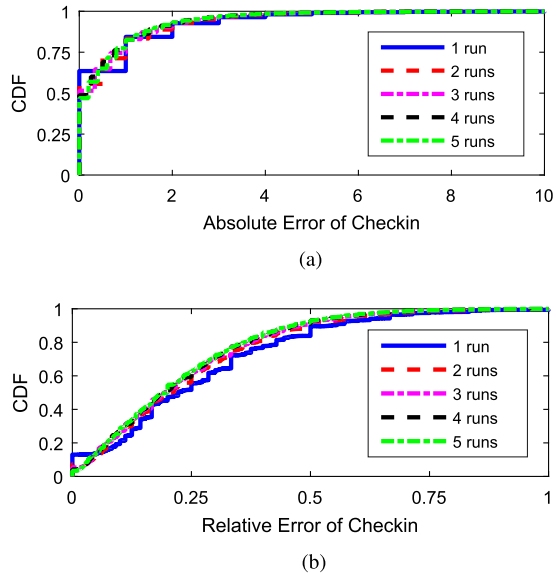


Fig. 12. Impact of different runs of simulation on checkin. (a) CDF of absolute error. (b) CDF of relative error.

simulations (with different generated data) and averaging the result, we may avoid some rare events and get a better result. So we believe more runs of simulation usually grants more stability. One may also notice that the results of one run of simulation appear in a step function. This is because when we only use one dataset, the prediction value of a station is an integer. Since the ground truth value is also an integer, the difference is an integer as well.

VI. APPLICATION: BIKE REBALANCE

As mentioned in the introduction, the primary issue in BSS is the unbalanced usage among stations. For example, stations near the business area are likely to run out of bikes during evening rush hours, which stops future users to checkout bikes. To migrate this issue, the BSS operators usually use trucks to reposition bikes among stations (e.g., bring bikes from

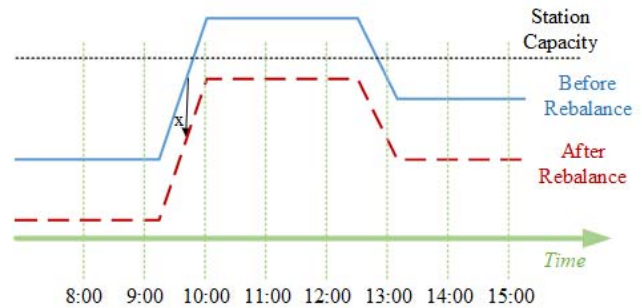


Fig. 13. Optimal rebalance number.

residential area to the business area in evening rush hours). Currently, this kind of rebalancing is usually inefficiently for the fact that the operation (e.g., how many bikes to pick up or drop off in a specific station) is determined according to the operators' prior knowledge and experience without any real-time prediction result. Thus, we believe the prediction approach presented in this paper is promising for improving the rebalancing efficiency. In the section, we conduct a preliminary study on how to rebalance bikes according to simulation-based prediction results and show the impact of different prediction horizons.

A. Background and Scenario

In BSS, the operator usually uses one or more trucks to redistribute bikes among different stations to prevent stations from being full or empty. The performance of rebalancing is determined by two factors: the demand predictions and the rebalancing capability. Optimal bike rebalancing with limited rebalancing capability (such as, truck capacity and rebalancing cost) is complicated problem and attracts the attention of researchers [22], [44]. Solving this problem completely is out of the scope of this paper and the primary purpose of this section is to show the potential improvement on rebalancing when taking prediction result into consideration. Thus we assume the rebalance capability to be large enough (i.e., the operator can redistribute an arbitrary number of bikes). So the limitation of the rebalancing capability is removed and the rebalancing performance is only determined by the predictions. However, as we will show later, the number of bikes redistributed in our method is close to that of current rebalancing operation.

B. Determine Optimal Rebalancing Number

By taking advantage of simulation-based prediction, we know when and where a checkin/checkout event happens, which grants us with the possibility to gain a fine-grained stock curve as shown in Figure 13. This curve is calculated in the following way: Starting with the initial stock, we add one to it when there is a checkin event and subtract one from it when there is a checkout event. It is worth noting that here we ignore the limitation of station capacity.

The truck may take x bikes from one station at the very beginning according to the prediction result. Note that x can be both positive and negative. When x is a positive/negative number, we pick up/drop off specific bikes in this station.

Since the goal to prevent the station from being empty/full, we choose a proper x such that the total time of being higher than capacity or lower than zero is minimized. For example, as shown in Figure 13, we pick up some bikes at the beginning, so that whole curve lies between zero and station capacity.

C. Result and Analysis

The experiment is conducted in the following way. Starting from 8:00 a.m., we first use our simulation-based approach to predict bike usage in the following t hours, calculate the optimal number and change the number of bikes in each station accordingly.

To evaluate the performance, we conduct simulation with a special dataset. This dataset contains two kinds of records. i) Real usage records: the record includes a users checkout station, checkout time, checkin station and checkin time; ii) Complementary records: these records are generated for the stations being empty during certain time windows to account for the users who do not show up in the historical record due to unavailability of bikes.

The performance metric includes the number of rejected users (i.e., users failed to checkin/checkout bikes) and the out-of-service time (i.e., the station is full/empty). These metrics are computed by the system simulator. Transition records (including user checkout station, checkout time, checkin station, checkin time) are input into the simulator and it is responsible for keeping track of the number of bikes in all stations. For the number of rejected users, if a user fails to checkout a bike, this user is assumed to choose an alternative transportation means and marked as a rejected user. For the out-of-service time, if a station is empty or full for time period t during the simulation, its out-of-service time is t . The sum of t over all stations is treated as the out-of-service time of the system.

To ensure a fair comparison among different prediction horizon, we compute the hourly averaged value of both metrics. The prediction horizon t is set to 3 hours, 6 hours and 9 hours. We compared our results with two baselines: 1) Experience-based rebalancing—The rebalance performed by operators; 2) Null rebalancing—Doing nothing for rebalancing. We conduct this evaluation on Hangzhou BSS from August 20 to August 31, 2015. The result is shown in Figure 14.

Rebalance based on predictions can effectively reduce the out-of-service time and the number of rejected user. Specifically, compared with experience-based rebalance, the out-of-service time is reduced by 86% when the prediction hour is 3 hours. The result is an optimistic estimation because we do not consider truck capacity, but it shows that prediction-base rebalance will be a promising approach. When comparing different prediction horizon, for out-of-service time, the performance becomes worse when the prediction is longer. This is caused by two factors: 1) When the prediction horizon becomes longer, the accumulated prediction error also becomes larger; 2) It is more difficult to choose a proper rebalancing number for a longer time. Note that the optimization object is the out-of-service time and the reduction of rejected users is a side effect. As we can learn from Figure 14(b), the averaged rejected user remains a stable

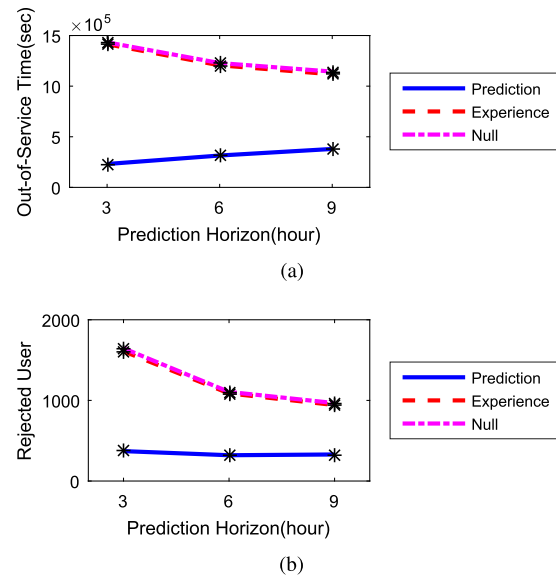


Fig. 14. Impact of prediction horizon for rebalance. (a) Out-of-service time. (b) Rejected user.

number for different prediction horizons. In addition, when the prediction horizon t is set to 9 hours, we find the number of bikes picking up/dropping off (on average 18572.7) in our proposal is close to that of the experience-based rebalancing (on average 15385.2), which means that it may be feasible to redistribute these bikes in real world.

VII. DISCUSSION AND FUTURE WORK

We provide several insights into the modeling and prediction results and provide directions for future work in this part.

A. Insights

We first give some insights from both different scenarios and modeling approaches.

1) *Variation Among Different Scenarios*: We conduct studies in different scenarios to get a better understand of human mobility in BSS. CDFs of relative error of checkin prediction are presented in Section 15. The results of checkout prediction are similar and omitted due to space limitation.

In Figure 15(a), we compare prediction results between stations in the business area and the tourist area. As we can see, stations in the business area are more predictable due to users' regular mobility patterns (e.g., from home to office). Differences between rainy days and sunny days are shown in Figure 15(b). Consistent with our intuition, fewer people use public bikes on rainy days, which increases randomness and degrades prediction performance. As can be seen from Figure 15(c), workdays own better prediction results than holidays or weekends for similar reasons. Finally, stations with high utilization exhibit high predictability over stations with low throughputs in Figure 15(d).

B. Open Research Issues

We summarize some open research issues related to the emerging bike-sharing systems.

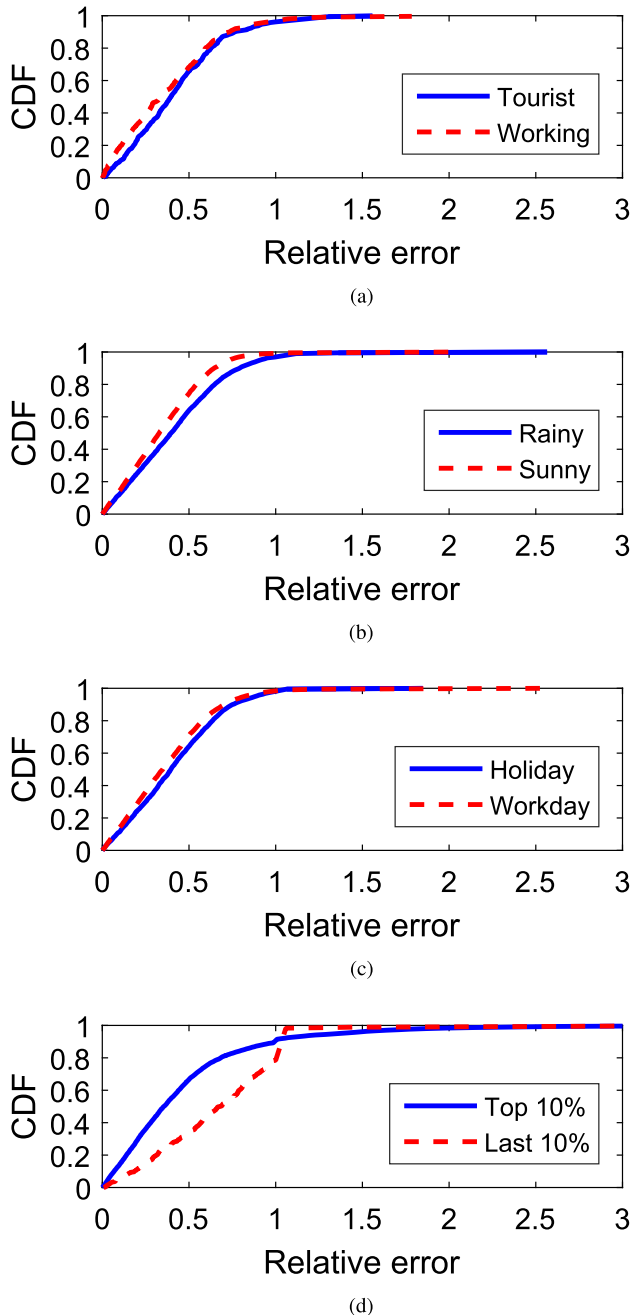


Fig. 15. Overall checkin prediction performance in various scenarios. (a) Different areas. (b) Different weathers. (c) Different days. (d) Different stations.

1) *Mobility Model Fusion With Multi-Source Data*: Study of human mobility has drawn significant attention in the mobile community. One intuitive idea is to improve the existing models by integrating bike-sharing data. Zhang *et al.* [45] have demonstrated the reduced bias of mobility modeling by exploiting the inherent diversities from multi-source data (i.e., taxi, bus, subway and smartphone CDR).

2) *Bike Rebalancing*: Though we investigate how system rebalancing can benefit from the prediction results in Section VI, designing a rebalancing system with rebalancing capability constraints remains an unsolved problem. An interesting and practical problem is how to achieve balancing between system performance and operation cost. Apart from

the operator’s “passive” rebalancing, how to devise an incentive and price mechanism enabling user-based “proactive” rebalancing is also an interesting subject to pursue

3) *Service Optimizations*: In addition to bike rebalancing, future work on service optimization includes station location optimization, service hour optimization, pricing strategy design, bicycle utilization balancing, etc. From a customer perspective, prompt bike stock information delivery and user-friendly interaction design are also of great help.

Many mobile techniques are promising to improve BSS service efficiency. RFID positioning and tracking systems, such as TrackT [46], may be able to track the bikes in the roads and provide additional online information of bike flows, which should be helpful for deploying new stations or designing dedicate bike lanes. We may also use RFID to monitor bikes in the stations rather than using docks, which may not only increase the station capacity but also reduce the infrastructure cost. In addition, mobile computing can also be used in user navigation [47] and system maintenance [48].

VIII. CONCLUSION

This paper proposes a data-driven simulation-based prediction approach for the bike-sharing systems. Firstly, based on historical bike sharing data, we first use statistical methods to model the spatio-temporal shifts of bikes between stations. Then, users’ potential demand is estimated by combining the random forest model and mobility model, which generates multiple synthetic records for simulation. Finally, by feeding the artificial records into a dedicated simulator, the model can predict users’ checkin/checkout demand in a long prediction horizon. Experiment on real-world dataset shows an 85 percentile relative error of 0.6 for checkout and 0.4 for checkin. We also ingestive to what extent prediction results can benefit system rebalancing. Evaluation results show that provided sufficient rebalancing capability, prediction based rebalancing can reduce the system out-of-service time by 86% compared with experience based rebalancing. Potential limitations of our approach is that its accuracy highly depends on the estimation of transfer probability and trip duration. But in some cases, these numbers are error prone. This may be caused by data sparsity, environmental changes or improper approximation of $\delta_{i,j}$. The model should be updated periodically to migrate this problems. Besides, from a practitioners perspective, multiple runs of simulation and online data collection usually bring computation and communication overhead.

ACKNOWLEDGMENT

The authors thank their partner BSS operator, Hangzhou Public Bicycle Transport Service Development Co., Ltd.

REFERENCES

- [1] L. MetroBike, *The Bike-Sharing Blog*. Accessed: Dec. 1, 2017. [Online]. Available: <http://bike-sharing.blogspot.com/2015/01/the-bike-sharing-world-2014-year-end.html>
- [2] *List of Bicycle-sharing Systems*. Accessed: Dec. 1, 2017. [Online]. Available: https://en.wikipedia.org/wiki/List_of_bicycle-sharing_systems
- [3] J. Froehlich, J. Neumann, and N. Oliver, “Sensing and predicting the pulse of the city through shared bicycling,” in *Proc. IJCAI*, Jul. 2009, pp. 1420–1426.

- [4] 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 Mobile Comput.*, vol. 6, no. 4, pp. 455–466, 2010.
- [5] P. Vogel and D. C. Mattfeld, "Strategic and operational planning of bike-sharing systems by data mining—A case study," in *Proc. Int. Conf. Comput. Logistics*, 2011, pp. 127–141.
- [6] P. Borgnat, É. Fleury, C. Robardet, and A. Scherrer, "Spatial analysis of dynamic movements of Vélo^v, Lyon's shared bicycle program," in *Proc. ECCS*, Jul. 2009, pp. 1–6.
- [7] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2015, p. 33.
- [8] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroda, "Mobility modeling and prediction in bike-sharing systems," in *Proc. ACM Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, Jun. 2016, pp. 165–178.
- [9] S. A. Shaheen, H. Zhang, E. Martin, and S. Guzman, "China's Hangzhou public bicycle: Understanding early adoption and behavioral response to bikesharing," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2247, no. 1, pp. 33–41, Jan. 2011.
- [10] Wikipedia. *Hangzhou Public Bicycle*. [Online]. Available: https://en.wikipedia.org/wiki/Hangzhou_Public_Bicycle.
- [11] S. A. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in europe, the americas, and asia past, present, and future," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2143, pp. 159–167, Oct. 2010.
- [12] E. W. Martin and S. A. Shaheen, "Evaluating public transit modal shift dynamics in response to bikesharing: A tale of two U.S. cities," *J. Transp. Geography*, vol. 41, pp. 315–324, Jul. 2014.
- [13] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić, "Large-scale vehicle sharing systems: Analysis of Vélib," *Int. J. Sustain. Transp.*, vol. 7, no. 1, pp. 85–106, Jun. 2013.
- [14] "Bike-Share. Opportunities in New York City," Dept. City Planning, New York, NY, USA, Tech. Rep. PTCPO8D00.G05, 2009.
- [15] A. Schlote, B. Chen, and R. Shorten, "On closed-loop bicycle availability prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1449–1455, Jun. 2015.
- [16] C. Ome and O. Latifa, "Model-based count series clustering for bike sharing system usage mining: A case study with the Vélib^v system of Paris," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–21, Oct. 2014.
- [17] P. Borgnat, C. Robardet, P. Abry, P. Flandrin, J.-B. Rouquier, and N. Tremblay, "A dynamical network view of Lyon's Vélo^v shared bicycle system," *Dyn. Complex Netw.*, vol. 2, pp. 267–284, Apr. 2013.
- [18] L. Chen *et al.*, "Bike sharing station placement leveraging heterogeneous urban open data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2015, pp. 571–575.
- [19] J. Liu, L. Sun, W. Chen, and H. Xiong, "Rebalancing bike sharing systems: A multi-source data smart optimization," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1005–1014.
- [20] P.-C. Chen, H.-Y. Hsieh, X. K. Sigalingging, Y.-R. Chen, and J.-S. Leu, "Prediction of station level demand in a bike sharing system using recurrent neural networks," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.
- [21] J. Liu, L. Sun, Q. Li, J. Ming, Y. Liu, and H. Xiong, "Functional zone based hierarchical demand prediction for bike system expansion," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 957–966.
- [22] T. Raviv, M. Tzur, and I. A. Forma, "Static repositioning in a bike-sharing system: Models and solution approaches," *J. Transp. Logistics*, vol. 2, no. 3, pp. 187–229, 2013.
- [23] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari, "Dynamic vehicle redistribution and online price incentives in shared mobility systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1567–1578, Aug. 2014.
- [24] J. Hrnčák, P. Žilecký, Q. Song, and M. Jakob, "Practical multicriteria urban bicycle routing," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 493–504, Mar. 2017.
- [25] J. Hu, Z. Yang, Y. Shu, P. Cheng, and J. Chen, "Data-driven utilization-aware trip advisor for bike-sharing systems," in *Proc. IEEE ICDM*, Nov. 2017, pp. 167–176.
- [26] P. Cheng, J. Hu, Z. Yang, Y. Shu, and J. Chen, "Utilization-aware trip advisor in bike-sharing systems based on user behavior analysis," *IEEE Trans. Knowl. Data Eng.*, to be published.
- [27] J. C. García-Palomares, J. Gutiérrez, and M. Latorre, "Optimizing the location of stations in bike-sharing programs: A GIS approach," *Appl. Geogr.*, vol. 35, nos. 1–2, pp. 235–246, Nov. 2012.
- [28] Y. Li, Y. Zheng, and Q. Yang, "Dynamic bike reposition: A spatio-temporal reinforcement learning approach," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2018, pp. 1724–1733.
- [29] J. Bao, T. He, S. Ruan, Y. Li, and Y. Zheng, "Planning bike lanes based on sharing-bikes' trajectories," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1377–1386.
- [30] N. Jian, D. Freund, H. M. Wiberg, and S. G. Henderson, "Simulation optimization for a large-scale bike-sharing system," in *Proc. Winter Simul. Conf. (WSC)*. Piscataway, NJ, USA: IEEE Press, Dec. 2016, pp. 602–613.
- [31] D. Chemla, F. Meunier, T. Pradeau, R. W. Calvo, and H. Yahiaoui. *Self-Service Bike Sharing Systems: Simulation, Repositioning, Pricing*. Accessed: Dec. 1, 2017. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00824078/en>
- [32] Á. Lozano, J. F. De Paz, G. Villarrubia González, D. H. Iglesia, and J. Bajo, "Multi-agent system for demand prediction and trip visualization in bike sharing systems," *Appl. Sci.*, vol. 8, no. 1, p. 67, Jan. 2018.
- [33] V. C. Pérez. *Simulation of a Public e-Bike Sharing System*. Accessed: Dec. 1, 2017. [Online]. Available: <https://upcommons.upc.edu/handle/2117/103128?locale-attribute=en>
- [34] J. P. Romero, J. L. Moura, A. Ibeas, and B. Alonso, "A simulation tool for bicycle sharing systems in multimodal networks," *Transp. Planning Technol.*, vol. 38, no. 6, pp. 646–663, Jun. 2015.
- [35] T. Preisler, T. Dethlefs, W. Renz, W. Kersten, T. Blecker, and C. Ringle, "Data-adaptive simulation: Cooperativeness of users in bike-sharing systems," in *Proc. Int. Conf. Logistics (HICL)*, vol. 20, Aug. 2015, pp. 201–228.
- [36] X. Zheng *et al.*, "Big data for social transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 620–630, Mar. 2016.
- [37] R. Ganti, M. Srivatsa, A. Ranganathan, and J. Han, "Inferring human mobility patterns from taxicab location traces," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2013, pp. 459–468.
- [38] S. Bhattacharya, S. Phithakkitnukoon, P. Nurmi, A. Klami, M. Veloso, and C. Bento, "Gaussian process-based predictive modeling for bus ridership," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, Sep. 2013, pp. 1189–1198.
- [39] X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting mobile big data: Sources, features, and applications," *IEEE Netw.*, vol. 31, no. 1, pp. 72–79, Jan./Feb. 2017.
- [40] X. Cheng, L. Fang, L. Yang, and S. Cui, "Mobile big data: The fuel for data-driven wireless," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1489–1516, Oct. 2017.
- [41] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [42] E. O. Mahony and D. B. Shmoys, "Data analysis and optimization for (city)bike sharing," in *Proc. 29th AAAI Conf. Artif. Intell.*, Jan. 2015, pp. 687–694.
- [43] N. Gast, G. Massonnet, D. Reijsbergen, and M. Tribastone, "Probabilistic forecasts of bike-sharing systems for journey planning," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2015, pp. 703–712.
- [44] G. Erdoğan, M. Battarra, and R. W. Calvo, "An exact algorithm for the static rebalancing problem arising in bicycle sharing systems," *Eur. J. Oper. Res.*, vol. 245, no. 3, pp. 667–679, Sep. 2015.
- [45] D. Zhang, J. Huang, Y. Li, F. Zhang, C. Xu, and T. He, "Exploring human mobility with multi-source data at extremely large metropolitan scales," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 201–212.
- [46] Z. Wang, N. Ye, R. Malekian, F. Xiao, and R. Wang, "Trackt: Accurate tracking of rfid tags with mm-level accuracy using first-order Taylor series approximation," *Ad Hoc Netw.*, vol. 53, pp. 132–144, Dec. 2016.
- [47] P. Aeschbach, X. Zhang, A. Georghiou, and J. Lygeros, "Balancing bike sharing systems through customer cooperation—A case study on London's barclays cycle hire," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Dec. 2015, pp. 4722–4727.
- [48] S. Wang *et al.*, "BRAVO: Improving the rebalancing operation in bike sharing with rebalancing range prediction," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, p. 44, Mar. 2018.



Zidong Yang received the B.Sc. degree from Zhejiang University, China, where he is currently pursuing the Ph.D. degree in control science and engineering. His research interests include data mining, machine learning, intelligent traffic systems, and mobile computing.



Jiming Chen received the B.Sc. and Ph.D. degrees in control science and engineering from Zhejiang University, China. He was a Visiting Researcher at the University of Waterloo from 2008 to 2010. He is currently a Changjiang Chair Professor with the College of Control Science and Engineering and the Deputy Director of the State Key Laboratory of Industrial Control Technology, Zhejiang University. His research interests include sensor networks, IoT, networked control, and cyber security. He serves/served as an Associate Editor for several

international journals, including the IEEE TII, IEEE TPDS, and ACM TECS. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society from 2015 to 2018.



Yuanchao Shu received the Ph.D. degree from Zhejiang University in 2015, and was also a joint Ph.D. Student with the EECS Department, University of Michigan, Ann Arbor, MI, USA. He is currently a Researcher with the Mobility and Networking Research Group, Microsoft Research Asia. His research interests include 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, and USENIX Security. He received the IEEE WCNC Best Paper Award, the IEEE INFOCOM Best Demo Award, the ACM China Doctoral Dissertation Award, and the IBM Ph.D. Fellowship. He served as the Award Chair of ACM TURC 2017 and 2018, the Publication Co-Chair of SmartCom 2017, and a member of the TPC of IEEE Globecom 2017 and 2018, and ICCCN 2017.



Ji Hu is currently pursuing the master's degree with the College of Control Science and Engineering, Zhejiang University. Before coming to ZJU, she majored in automation at Southeast University. Her research focuses on data analysis and data mining.



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 for the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, *Wireless Networks*, and *International Journal of Communication Systems*. He served as the Guest Editor for the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS and 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 the ACM MobiHoc 2015, and the Publicity Co-Chair for the IEEE MASS 2013.