Hybrid Bicycle Allocation for Usage Load Balancing and Lifetime Optimization in Bike-Sharing Systems

Xiawen Yao, Xingfa Shen*, Landi Wang
School of Computer Science & Technology
Hangzhou Dianzi University, China
735649296@qq.com,shenxf@hdu.edu.cn,741510967@qq.com

Tian He
Dept. of Computer Science & Engineering
University of Minnesota, Twin Cities
tianhe@cs.umn.edu

***Abstract*—Nowadays, public bike-sharing systems are broadly adopted and deployed in many major cities, however, as public facilities, bicycles will be prone to damage and need to be replaced frequently, which results in high system maintenance costs. One of the root causes of bicycle damages is the serious load-unbalance of bicycle usage. In this paper, we propose a hybrid bicycle allocation strategy for bicycle lifetime optimization, which can effectively reduce the degree of imbalance of system load. First, we analyze and verify the load-unbalance status of bicycle usage in the existing bike-sharing system. Then, a hybrid bicycle allocation strategy is proposed, which is evaluated on Washington D.C. bike-sharing system. Furthermore, according to a bicycle lifetime model based on Weibull distribution, the proposed bicycle allocation strategy could significantly cut down the percentage of the bicycles need to be replaced in a certain period of time.**

***Keywords*-Bike Sharing System, Bike Allocation, Load Balancing, Lifetime Optimization**

I. Introduction

Nowadays, as a simple but flexible transportation mode, bike-sharing system is broadly adopted and deployed in many major cities. It is a typical application scenario on urban computing [1]. In spite of its great progress, the bike-sharing system still faces many challenges, one of them is that bicycles will be prone to damage and need to be replaced frequently, resulting in high system maintenance costs. One of the root causes of bicycle damages is the serious load-unbalance of bicycle usage. The analysis conducted based on three-month bicycle usage records of Washington D.C. bike-sharing system indicates there exists serious usage load-unbalance in the existing bike-sharing system.

To address this issue, a hybrid bicycle allocation strategy is proposed. This hybrid bicycle allocation strategy consists of two allocation models, Least Frequently Used (LFU) and Least Time Used (LTU) model. We also take other two simple contrast models, FIFO and random model, as the baselines for comparison.

In order to simulate the behavior of allocating bicycles to users utilizing the existing dataset of bicycle usage records, we propose a novel data-swapping technique by swapping key field of a trip record, *e.g.*, bicycle number, for an allocated one according to the specific allocation strategy, while keeping other part of that trip record unchanged.

We evaluate the proposed bicycle allocation strategy on Washington D.C. bike-sharing system. The results indicate that the proposed bicycle allocation strategy is effective in terms of four metrics, maximum and evenness metrics of both total usage amount and accumulative duration of use. Compared with the ground truth, two models in the hybrid bicycle allocation strategy, LFU and LTU model, can both alleviate the load unbalance of bicycle use. LFU model decreases the maximum total usage amount by 36%, and increases the evenness of total usage amount by 0.16. LTU model decreases the maximum accumulative duration of use by 36%, and increases the evenness of accumulative duration of use by 0.14.

Due to the difference in the intensity of use, types of bicycle and other factors, lifetime of public bicycle is usually shorter than ordinary bicycle, resulting in that public bicycles often need to be replaced frequently. To this end, a Weibull-distribution-based lifetime model is adopted to evaluate the impact of proposed bicycle allocation strategy on the bicycle lifetime. Results show the proposed bicycle allocation strategy could significantly cut down the percentage of the bicycles need to be replaced in a certain period of time, hence lowering the system maintenance costs.

Specially, the key contributions of this paper are as follows:

- We put forward the issue of the serious load-unbalance of bicycle use in bike-sharing system for the first time based on the analysis of the existing datasets of bicycle usage records, which is one of root causes of bicycle damage, resulting in high system maintenance costs.
- We propose a hybrid bicycle allocation strategy that consists of two models, LFU and LTU model, which can effectively optimize the system lifetime by usage load-balancing.
- In order to simulate the behavior of allocating bicycles to users utilizing the existing dataset of bicycle

* Corresponding author, E-mail: shenxf@hdu.edu.cn

2375-0324/17 $31.00 © 2017 IEEE
DOI 10.1109/MDM.2017.24

usage records, we propose a novel data-swapping technique by swapping key field of a trip record.

The rest of the paper is organized as follows. Section II introduces the related work. Section III describes the motivation of this work. Section IV proposes the hybrid bicycle allocation strategy, followed by its evaluation Section V. Section VI presents conclusive remarks.

II. Related Work

The deployment of bike-sharing system earliest occurred in Europe [2], and then spread all over the world soon [3]. Studies on bike-sharing systems can be roughly classified to three aspects as follows.

System analysis: In order to serve users better, it needs to understand the demands for bicycles of users, how the system works, and which design is reasonable. Lin *et al.* [4] investigated the issue of effective planning of bike-sharing system. Cyrille *et al.* [5] proposed multiple methodologies allowing the estimation of bike-sharing trips based on the station level data. Froehlich *et al.* [6] provided a spatio-temporal analysis of bicycle station usage.

System prediction: In order to enable users to rent or return the bicycle at any time, many researchers study on the prediction of the number of bicycles at each station to relocate bicycles between stations. Kaltenbrunner *et al.* [7] proposed a statistical model to predict the number of available bicycles at any station. Rudloff *et al.* [8] presented a model of underlying demand based on the system redistribution. Yoon *et al.* [9] built a spatio-temporal prediction system based on a modified ARIMA model. Li *et al.* [10] proposed a hierarchical prediction model to predict the number of bikes.

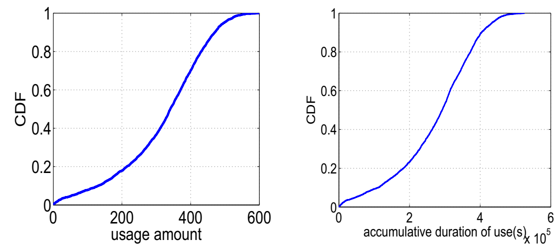
System rebalancing: To tackle the issue of unbalanced bicycle usage, system operators usually reallocate the bicycles by vehicles. Chemla *et al.* [11] considered the static rebalancing problem as a graph theory problem. Schuijbroek *et al.* [12] not only considered service requirement at each station, but also designed the optimal vehicle route to rebalance the inventory.

The above three aspects of studies on bike-sharing system are to serve users better, and the problem about load-unbalance of bicycle usage we study in this paper is to maintenance system better.

III. Motivation

In this section, we introduce the motivation of the work in this paper.

As a common transportation mode, bike-sharing system brings great convenience and flexibility to people' life. However, users often arbitrarily choose their favorite bicycles, which causes the serious load-unbalance of bicycle use. As a result, public bicycles will be easily damaged



(a) CDF of total amount of bicycle use (b) CDF of accumulative duration of bicycle use

Figure 1. Imbalance of Bicycle Use

and need to be replaced frequently, resulting in high system maintenance costs.

We analyze three-month (July-September, 2014) bicycle usage records based on Washington D.C. bike-sharing system, using two metrics, total usage amount and accumulative duration of use. Firstly, as shown in Fig.1(a), the cumulative distribution function (CDF) of total bicycle usage amount is investigated. We can obtain that there is a large span of usage amount with a gap of 600 times between the maximum usage amount and the least one.

Secondly, another metric of bicycle use, the accumulative duration of use, is investigated in a similar way and illustrated in Fig.1(b). We can obtain similar point of view using the CDF of the accumulative duration of use that the gap between the maximum and the least increases to 3285 times.

These observations described above lead us to conclude that there are serious load-unbalance situations in the existing bike-sharing system, which motivated the investigation about how to alleviate this situation. To this end, we propose a hybrid bicycle allocation strategy for usage load balancing and system life optimization.

IV. Bicycle Allocation Strategy

In this section, we first introduce the notations used in this paper and then present the hybrid bicycle allocation strategy.

A. Notations

This subsection defines the notations (Table I).

A bicycle usage record $trip = (S_o, t_o, S_d, t_d, b, D_t)$ consists of six fields, the origin station S_o, the time of departure t_o, the destination station S_d, the time of arrival t_d, the bicycle b, and the duration of use D_t that is the time difference between the time of departure t_o and the time of arrival t_d.

B. Framework

Fig.2 shows the framework of our bicycle allocation strategy with two contrast models in this paper.

Table I
NOTATIONS

S_i	The i^{th} station
H	Number of trip records
b_k	The ID of k^{th} bicycle
m	Number of bicycles
b_k^*	The bicycle allocated to user
U_{b_k}	Usage amount of b_k
T_{b_k}	Accumulative duration of use b_k
$V_{b_k,t}$	The station the bicycle b_k is in at moment t
$C_{S_i,t}$	Bicycle set that station S_i owns at moment t
$TripSet_{b_k,t}$	Set of recent trip records for bicycle b_k at moment t

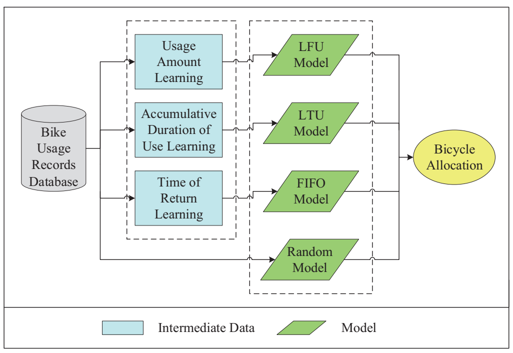


Figure 2. Framework of bicycle allocation strategy with contrast models

C. Models

In this subsection, the models used in the hybrid bicycle allocation strategy are described in details with the contrast models.

In the computer storage system, common replacement algorithms are as follows: LFU (Least Frequently Used), LRU (Least Recently Used), FIFO (First-In First-Out) and random algorithm. Inspired by the cache replacement algorithms mentioned above, two bicycle allocation models, LFU and LTU model, are proposed as core components of the hybrid bicycle allocation strategy in this paper. For comparison, we take the FIFO and random model as the contrast models.

1) LFU Model: In LFU model, the bicycle with the least usage amount in the most recent period T_{win} is allocate to the user.

Specifically, LFU model works according to the following steps:

1) Given the time of departure t_o and the origin station S_o, we count all bicycles at the station S_o in the moment t_o.
2) For all bicycles in the set C_{S_o,t_o}, we count the usage amount U_{b_k} for per bicycle in the most recent period T_{win}, and then find out the bicycle b_k^* of the least usage amount, and allocate it to the user.

Mathematical model:

$$b_k^* = \arg\min_{b_k} \quad U_{b_k}$$

$$s.t. \begin{cases} U_{b_k} = |TripSet_{b_k,t_o}| \\ TripSet_{b_k,t_o} = \{trip_i | trip_i.b = b_k, \\ \qquad t_o - T_{win} \leq trip_i.t_o \leq t_o, i = 1, 2 \cdots H\} \\ b_k \in C_{S_o,t_o} \\ C_{S_o,t_o} = \left\{b_j \left| V_{b_j,t_o} = S_o \right., j = 1, 2 \cdots m\right\} \end{cases}$$

2) LTU Model: In LTU (Least Time Used) model, we take the accumulative duration of bicycle use into account instead of the simple bicycle usage amount in LFU model. The bicycle with the least accumulative duration of bicycle use in the most recent period T_{win} is allocate to the user.

Specifically, LTU model works according to the following steps:

1) This step is same as the step 1 of LFU model described above.
2) For all bicycles in the set C_{S_o,t_o}, we compute the accumulative duration of use T_{b_k} for per bicycle in the most recent period T_{win}, and then find out the bicycle b_k^* of the least accumulative duration of use, and allocate it to the user.

Mathematical model:

$$b_k^* = \arg\min_{b_k} \quad T_{b_k}$$

$$s.t. \begin{cases} T_{b_k} = \sum\limits_{trip_i \in TripSet_{b_k,t_o}} trip_i.D_t \\ TripSet_{b_k,t_o} = \{trip_i | trip_i.b = b_k, \\ \qquad t_o - T_{win} \leq trip_i.t_o \leq t_o, i = 1, 2 \cdots H\} \\ b_k \in C_{S_o,t_o} \\ C_{S_o,t_o} = \left\{b_j \left| V_{b_j,t_o} = S_o \right., j = 1, 2 \cdots m\right\} \end{cases}$$

3) FIFO Model and Random Model: FIFO and random model are taken as contrast models to verify the effectiveness of the proposed bicycle allocation strategy.

For FIFO model, we allocate the bicycle with the earliest time of its latest arrival to the station S_o in the most recent period T_{win} to the user. For random model, a bicycle is allocated at random to the user.

V. EXPERIMENTS

A. Settings

1) Datasets: We conduct experiments on datasets of Capital Bike-share system, which is located in Washington D.C., from 1^{st} Jan. to 30^{th} Sep. in 2014. There are 2293157 bicycle usage records, and we use the most recent records to infer the station where each bicycle is in at the current moment. Owing to some bicycles may not be used for a long time, we need to deal with records of those bicycles a long time ago (specific processing refers to Section V-B).

2) Metrics: The metrics we adopt to evaluate the results are as follows:

Maximum usage amount: $\max_U = \max\{U_{b_k}\}$

Maximum accumulative duration of use: $\max_T = \max\{T_{b_k}\}$

Besides, we propose two evenness metrics based on the two basic metrics mentioned above.

Evenness of usage amount: $e_U = \frac{\text{average usage amount}}{\text{maximum usage amount}}$

Evenness of accumulative duration of use: $e_T = \frac{\text{average accumulative duration of use}}{\text{maximum accumulative duration of use}}$

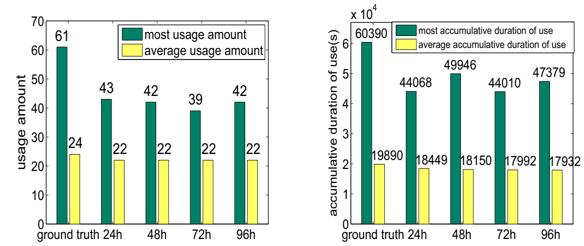
B. Data Processing

1) Data Swapping Method: This section introduces the data-swapping method used in the proposed bicycle allocation strategy in this paper. There are huge historical datasets of bicycle use records in the large-scale bike-sharing system, which are the ground truth we have. Ideally, in order to get the actual bicycle usage record after allocating a bicycle to a user, we should apply our bicycle allocation strategy to the real-life bike-sharing system. However, it needs to stop the running of the bike-sharing system and get our bicycle allocation strategy programmed into both system servers and the users' smart phones. Without such conditions, we evaluate our bicycle allocation strategy using an off-line way on the huge historical bicycle usage record datasets.

Our purpose is to get the bicycle use records after allocating a bicycle to a user, and we can achieve this goal by swapping key field of a trip record, for an allocated one according to the specific allocation strategy. Specifically, for the bicycle use record $trip = (S_o, t_o, S_d, t_d, b_k, D_t)$, to allocate a bicycle to a user, we just need to swap the bicycle number b_k of the trip, without need to change the user's trip origin, destination, time and others.

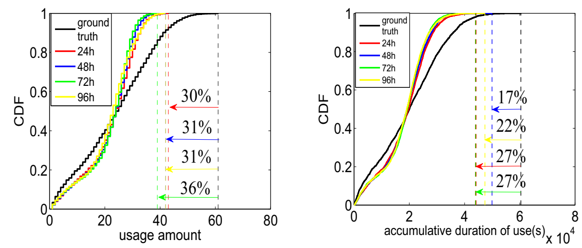
2) Data Preprocessing: Many large-scale bicycle sharing systems can predict the traffic volume of the whole system in real-time and thus conduct effective bicycle scheduling in order to meet the needs of the users. We use the bicycle usage records based on Bike-sharing System in Washington, D.C., which has bicycle scheduling. However, we do not know exactly how the system forecasts the traffic volume and conducts the specific bicycle scheduling. The situation that there is no bicycle when a user arrives at a station can not produce a bicycle usage record, while there are available bicycles at this station and the bicycle usage record will be produced in real-life system due to its bicycle scheduling. In order to generate the bicycle usage records of the same size as those of ground truth, we delete the bicycle usage records that could not be regenerated in the simulating process from the datasets of ground truth.

In order to allocate the bicycle in the specific station after knowing the time of departure t_o and the origin



(a) Usage amount (b) Accumulative duration of use

Figure 3. Impact of training time



(a) Usage amount (b) Accumulative duration of use

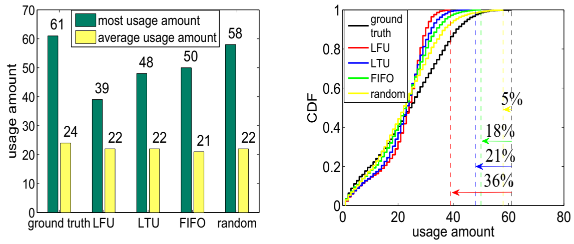
Figure 4. Impact of training time in CDF

station S_o, we need to obtain the station V_{b_k,t_o} where each bicycle b_k is in at the moment t_o. To obtain that for each bicycle b_k, we traverse all bicycle usage records in a reverse-time order to find the last docked station of bicycle b_k before t_o, which is the station where the bicycle is in at the current moment t_o. We choose the time period from January 1^{st}, 2014 to moment t_o as the search region, whose range is quite wide because that it needs a mass of bicycle usage records, due to some bicycles may not be used for a long time.

C. Parameter Selection

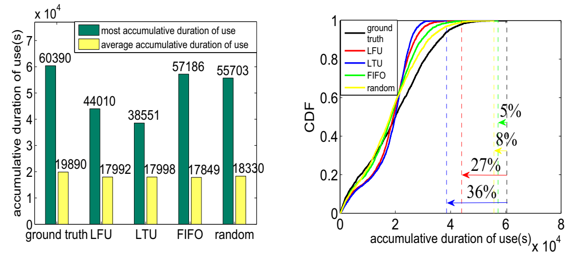
In terms of parameters of the models, there are two time parameters that need to be considered, one is duration of conducting the bicycle allocation strategies, called test time, the other is the time parameter in LFU, LTU and FIFO model, *i.e.*, the period of time (called training time) in which the usage amount, the accumulative duration of use and the time of last return will be computed when allocating a bicycle. We conduct experiments for parameter selection on LFU model.

We choose 24 hours, 48 hours, 72 hours, and 96 hours as training time to train our models respectively. In the meantime, we choose one week (September 24^{th}-30^{th}) as the test time. The results show that every choice of test time could decrease the maximum, and increase the evenness both in usage amount and accumulative duration of use (Fig.3 and Fig.4). So we can obtain that different choices of train time would not have significant impact on results, and after considering effect, we choose 72



(a) metrics (b) CDF

Figure 5. Results of total usage amount



(a) metrics (b) CDF

Figure 6. Results of accumulative duration of use

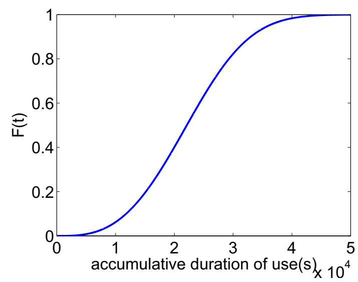


Figure 7. Lifetime distribution of bicycle

hours as training time. The selection process and results of test time parameter is similar with that of train time. We choose one week, two weeks, three weeks, and one month as test time, specifically, September 24^{th}-30^{th}, September 17^{th}-30^{th}, September 10^{th}-30^{th} and September 1^{st}-30^{th}, respectively. In the meantime, we choose 72 hours as training time. And according to results we choose one week as test time.

D. Experiment Results

Based on the analysis above, we choose one week (September 24^{th}-30^{th}) as test time, and 72 hours as training time to evaluate our bicycle allocation strategy.

1) Results of Usage Amount: As shown in Fig.5(a), compared with the ground truth, the maximum total usage amount of the contrast models decreases by small amplitude while our bicycle allocation strategy could greatly decreases the maximum total usage amount, and the average usage amount of all models decreases due to there are bicycles not be used in the real life while they are used in our experiment during test time.

Fig.5(b) shows the CDF of usage amount, and we can see that compared with the ground truth, not only the curve span of every model becomes smaller, but also the curve of every model becomes more steep, that indicates the distribution is more concentrated.

The specific difference calculated values about usage amount are showed in table II, and the evaluation results show that our bicycle allocation strategy could effectively reduce the degree of imbalance of bicycle system load in terms of usage amount. Compared with the ground truth, the maximum total usage amount decreases by 36% and the evenness of usage amount increases by 0.16 using LFU model.

2) Results of Accumulative Duration of Use: There is a similar result of accumulative duration of use to that of usage amount (Fig.6(a)), and we can obtain the same conclusion of accumulative duration of use as the usage amount from Fig.6(b).

The specific difference calculated values about accumulative duration of use are showed in table II. Compared with the ground truth, the maximum accumulative duration of use decreases by 36% and the evenness of accumulative duration of use increases by 0.14 using LTU model.

E. Lifetime Model

As bicycle lifetime and riding distance are positively correlated, we suppose bicycle lifetime and accumulative duration of use are positively correlated and the bicycle lifetime follows Weibull distribution. Because the test time of our bicycle allocation strategy is too short, and there is no bicycle whose lifetime is arrived and could not be used, we make the variable of the accumulative duration of use in the lifetime model reduced by 1000 times, which does not affect the distribution trend of the model itself.

Specifically, we use a lifetime model based on Weibull distribution that represents the lifetime distribution of bicycle as follow:

$$F(t) = 1 - e^{-\left(0.4\times10^{-7}\times10^{3}\times t\right)^{3}}$$

As shown in Fig.7, the vertical axis represents the probability of the bicycles reaching its lifetime limit, and can also be considered as the damage rate in the system since the damage rate is more close to 100% when the probability of the bicycles reaching its lifetime is more close to 1.

Owing to areas, types of bicycle and other factors, lifetime of public bicycle is shorter than ordinary bicycle. We suppose an ordinary bicycle can not be used when its damage rate is up to 100% while a public bicycle needs to be replaced when its damage rate is not necessary up to 100%. When a public bicycle needs to be replaced, we

Table II
DIFFERENCE BETWEEN SIMULATED RESULTS AND GROUND TRUTH

bicycle allocation strategy and contrast models	LFU model	LTU model	FIFO model	random model
Difference percent of $\max_U$	-36%	-21%	-18%	-5%
Difference of e_U	+0.16	+0.06	+0.02	-0.02
Difference percent of $\max_T$	-27%	-36%	-5%	-8%
Difference of e_T	+0.08	+0.14	+0.02	+0.00

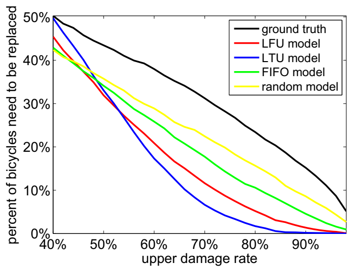


Figure 8. Percentage of the bicycles need to be replaced

call its damage rate upper damage rate. Fig.8 shows the percentage of the bicycles need to be replaced with the variation of upper damage rate.

From Fig.8, we can conclude that our bicycle allocation strategy could significantly cut down the percentage of the bicycles that need to be replaced, especially when upper damage rate is 70% or more, LTU model of our bicycle allocation strategy could decrease the percentage of the bicycles need to be replaced to 10% or less.

VI. CONCLUSION

In this paper, we put forward and investigate the issue about the serious load-unbalance of bicycle use in bike-sharing system by analyzing the historical bicycle usage records. To reduce the imbalance of bicycle use load, we propose a hybrid bicycle allocation strategy for life optimization by allocating a specific bicycle to a user, which consists of two models, LFU and LTU model. We evaluate our bicycle allocation strategy on Washington D.C. bike-sharing system, which can effectively reduce the load-unbalance of bicycle use that LFU model is the best choice in terms of usage amount and LTU model is the best choice in terms of accumulative duration of use. In the end, we use a lifetime model based on Weibull distribution to represent the lifetime distribution of bicycle. From the perspective of bicycle lifetime, our bicycle allocation strategy could significantly cut down the percentage of the bicycles need to be replaced. In the future, we would like to consider more aspects that influence load-unbalance of bicycle use, and reduce the load-unbalance of bicycle use from more perspectives.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation of China (NSFC) under Grants No.61190113, 61401135, 61473109, 61672198 and Zhejiang Provincial Natural Science Foundation under Grant No.LY14F020047.

REFERENCES

[1] Ouri Wolfson Hai Yang Yu Zheng, Licia Capra. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 2014.

[2] P. DeMaio. Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, 12(4):41–56, 2009.

[3] S. Guzman S.A. Shaheen and H. Zhang. Bikesharing in europe, the americas, and asia past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board*, (2143):159–167, 2010.

[4] J.-R. Lin and T.-H. Yang. Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review*, 47(2):284–294, 2011.

[5] Cyrille Mdard de Chardon and Geoffrey Caruso. Estimating bike-share trips using station level data. *Transportation Research Part B: Methodological*, 78(C):260–279, 2015.

[6] J. Neumann J. Froehlich and N. Oliver. Sensing and predicting the pulse of the city through shared bicycling. In *Proc. of the 21st IJCAI*, 2009.

[7] A. Kaltenbrunner. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466, 2010.

[8] C. Rudloff and B. Lackner. Modeling demand for bike-sharing systems neighboring stations as source for demand and reason for structural breaks. *Transportation Research Record*, (2430):1–11, 2014.

[9] F. Pinelli J. W. Yoon and F. Calabrese. Cityride: a predictive bike sharing journey advisor. In *Proc. of the 13th IEEE ICMDM*, 2012.

[10] H. Zhang Y. Li, Y. Zheng and L. Chen. Traffic prediction in a bike-sharing system. In *SIGSPATIAL*, 2015.

[11] F. Meunier D. Chemla and R. W. Calvo. Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, 10(2):120–146, 2013.

[12] W.-J. Van Hoeve J. Schuijbroek, R. C. Hampshire. Inventory rebalancing and vehicle routing in bike sharing systems. *Tepper School of Business*, page 1491, 2013.