# Locally Balanced Inductive Matrix Completion for Demand-Supply Inference in Stationless Bike-Sharing Systems

Senzhang Wang, Hao Chen, Jiannong Cao, *Fellow, IEEE*, Jiawei Zhang, Philip S. Yu, *Fellow, IEEE*

**Abstract**—Stationless bike-sharing systems such as Mobike are currently becoming extremely popular in China as well as some other big cities in the world. Compared to traditional bicycle-sharing systems, stationless bike-sharing systems do not need bike stations. Users can rent and return bikes at arbitrary locations through an App installed on their smart phones. Such a convenient and flexible bike-sharing mode greatly solves *the last mile issue* of the commuters, and better meets their real bike usage demand. However, it also poses new challenges for operators to manage the system. The first primary challenge is how to accurately estimate the real bike usage demand in different areas of a city and in different time intervals, which is crucial for the system planning and operation. This paper for the first time proposes a data driven approach for bike usage demand inference in stationless bike-sharing systems. The idea is that we first estimate the demands in some regions and time intervals from a small number of observed bike check-out/in data directly, and then use them as seeds to infer the region-level bike usage demands of an entire city. Specifically, we formulate this problem as a matrix completion task by modeling the bike usage demand as a matrix whose two dimensions are time intervals of a day and regions of a city respectively. With the observation that POI distribution of a region is an important indicator to bike demand, we propose to utilize inductive matrix factorization by considering POIs as side information. As the bike usage data are highly correlated in both spatial and temporal dimensions, we also incorporate the spatial-temporal correlations as well as the balanced bike usage constraint into a joint optimization framework. We evaluate the proposed model on a large Mobike trip dataset collected from Beijing, and the experimental results show its superior performance by comparison with various baseline methods.

**Index Terms**—Bike-Sharing System, Demand Inference, Matrix Completion, Optimization

## 1 Introduction

Stationless bike-sharing systems such as Mobike are becoming extremely popular in China as well as some other big cities in the world such as Manchester and Washington DC. Unlike traditional bike-sharing systems that need to build a large number of bike stations and users have to rent/return a bike from/to one of them, there are no stations in stationless bike-sharing systems. Users can check-out a bike nearby and return it at an arbitrary location [1]. Fig. 1 shows an example of a Mobike bike, its GPS positioning module and the corresponding App installed in a smart phone. A 3G communication component and a GPS module are embedded in the lock system of each Mobike bike as shown in Fig. 1(a). Users can scan the QR-code with the pre-installed App in their smart phones to unlock a Mobike bike for a trip, and lock it manually after the trip. Fig. 1(b) shows the App installed in the smart phones. It shows all the Mobike bikes nearby to help the users find the nearest one for use. During the trip, the App will record the check-in/out locations and track the travel route of the user, and the system will charge the fee based on the trip time length and trip distance after the trip.


(a) A Mobike bike and its GPS module (b) App of Mobike

Fig. 1. An example of a Mobike bike and the App.

Stationless bike-sharing systems are designed to better solve *the last mile issue* and connect commuters to public transit networks. Compared to traditional bike-sharing systems, the advantages of stationless bike-sharing systems are: 1) **more flexible to use**, as no bike stations are needed and users can conveniently pick up and drop off their bikes at arbitrary locations with their smart phones; 2) **better meet users' need**, as the trajectories of bike users can reflect their real travel demands, and thus *the*


- *S.Z. Wang was with the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China; and the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: szwang@nuaa.edu.cn.*
- *H. Chen was with the School of Computer Science and Engineering, Beihang University, Beijing, China.*
- *J.N. Cao was with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China.*
- *J.W. Zhang was with the Department of Computer Science, Florida State University, Tallahassee, Florida, USA.*
- *P.S. Yu was with the Department of Computer Science, University of Illinois at Chicago, Chicago, USA, and Institute for Data Science, Tsinghua University, Beijing, China.*

*last mile issue* can be better solved; 3) **better relieve the traffic congestion issue**, as more commuters will choose the traveling mode of *riding bikes & taking public transit* due to the above mentioned two advantages rather than driving their private cars. It is reported by Gaode Map [23] that the traffic jams nearby over half of the subway stations in Beijing have been relieved after the stationless sharing bikes starts to be massively deployed in 2017.

Such a brand new bike-sharing mode makes people's short trips in big cities much more convenient, but on the other side of the coin it also poses new challenges for the operators to manage the system. The first primary challenge is that the real demand of the bikes in a city is hard to estimate. Accurately estimating the bike usage demand is essential for both decision making on deploying bikes in a new city and the following operation management [16]. First, when a stationless bike-sharing company decides to deploy bikes in a new city, the first problem they face is to estimate the number of needed bikes. More specifically, how many bikes are required in each region of the city and in each hour of a day. Second, accurately estimating the fine-grained bike usage demand can greatly facilitate system management and reduce the labor cost of redistributing the bikes. Compared with traditional bike sharing systems, the bikes in a stationless bike sharing system move much more randomly as users can check-out/in the bikes at arbitrary locations. Thus the distribution of the bikes can become extremely skewed from place to place and from hour to hour. It is common that the bikes in some places are usually over supplied with a large number of unoccupied bikes, while they are over demanded in some other places where users cannot find a bike for use. Thus an accurate estimation on the bike usage demand is crucial for helping the operators work out reasonable strategies on redistributing the bikes, reducing the manual-rebalance workload and improving the overall bike utilization.

Although it is practically important, currently there still lacks of a systematic study of bike usage demand inference in stationless bike-sharing systems. Traditional bike-sharing systems have been extensively studied from different aspects, including the bike rebalancing optimization [12], [29], [30], bike demand prediction [4], [5], [24], [25], [26], and bike usage patterns mining [20], [21], [22]. Chen *et al.* [24] proposed a station-level cluster based demand prediction model in bike-sharing systems, Li *et al.* [27] proposed a hybrid and hierarchical prediction model to predict the demand of bikes, and Liu *et al.* [5] proposed a functional zone based hierarchical prediction model to predict the bike usage demand for a new bike station. However, these works focus on studying traditional bike-sharing systems which have bike stations built at fixed locations with limitated number of docks. Thus, it is difficult to directly apply them to our study.

In this paper, we propose a data-driven approach to estimate the fine-grained bike usage demand in stationless bike-sharing systems. Our method enables an accurate city-wide inference with a sparse bike usage data collected from a small number of pre-deployed bikes. To perform a fine-grained inference and also facilitate effective system management in practice, we first divide a city into equally sized cell regions inspired by [38], [41], [43] and model the bike usage data in all the regions as a matrix. Then we formulate the problem as a matrix completion task by considering the regions and time intervals as the two dimensions of the bike usage demand matrix. In this matrix only a very small number of entry values are known, based on which we need to infer the remaining entry values. To solve this problem, we need not only address the data sparsity challenge, but also need to encode the spatial-temporal correlation of the bike usage, which is also challenging. To address these challenges, we first incorporate the POIs of a city based on our previous study [7] and the data analysis that the POI distribution of a region can largely reflect its bike usage pattern. An inductive matrix completion method is proposed by considering the POIs as the side information of a region. We next propose to add the spatial and temporal correlations into our model. The two correlations make the inferred bike usage demands of two geographically close regions and the same region in two successive time intervals yield to be similar. In addition, note that the entire bike usage should be balanced, which means the total number of the check-out bikes should be equal to the number of the check-in bikes. To take this constraint into consideration, we propose to cluster the regions based on the bike traffic flows among them, and incorporate an intra-cluster balanced bike usage constraint into the model. Finally, a **U**sage **B**alanced **I**nductive **M**atrix **C**ompletion model UBIMC is proposed to effectively integrate above mentioned components and more accurately infer the region-level bike usage demand in different time intervals. The data and code of this work are publicly available at *https://github.com/szwangsummer/UBIMC*.

The major contributions of this paper are as follows.

- To the best of our knowledge, we are the first to study the usage demand inference problem in stationless bike-sharing systems.
- Our data analysis shows that the bike usage is unbalanced for most regions in most time intervals of a day, and the bike usage presents strong correlations in both spatial and temporal dimensions.
- A inductive matrix factorization based framework is proposed. The proposed model can effectively incorporate POIs, the spatial-temporal correlations, as well as locally balanced bike check-in/out usage constraint into a joint optimization framework.
- We evaluate the model over a large Mobike dataset with more than 5 million bike trips in Beijng. The experimental results verify the superior performance of the proposed model by comparison with various baseline models.

The remainder of this paper is organized as follows. In Section 2, we give a formal definition of the studied problem and show the framework of our solution. Section 3 introduces the dataset we use and gives data analysis. The detail of the model is presented in Section 4. Evaluations are given in Section 5 followed by related work in Section 6. Finally, we conclude this work in Section 7.

## 2 Problem Definition and Framework

### 2.1 Problem Definition

We first give the following terminology definitions used in this paper, and then give a formal problem definition.

***Definition 1.*** **Bike Trip**. A bike trip is represented as $Tr = \{l_{ori}, l_{des}, t_{ori}, t_{des}\}$, where $l_{ori}$ denotes the starting location, consisting of latitude $l_{ori}.lat$ and longitude $l_{ori}.lon$; $l_{des}$ denotes the destination location, consisting of latitude $l_{des}.lat$ and longitude $l_{des}.lon$; $t_{ori}$ and $t_{des}$ are the starting time and the end time of the trip respectively.

To effectively manage the system, the service providers usually divide a city into small regions and assign several workers to each region who are responsible to maintain and relocate the bikes in the region. Following the real management strategy of the system, we also divide a city into regions defined as follows.

***Definition 2.*** **Region of a City**. A city can be divided into a group of equal-sized grid regions. A region $r_i$ of a city is a square area. It can be represented as $r_i = \{(lon_1, lat_1), (lon_2, lat_2)\}$, where $(lon_1, lat_1)$ is the start location and $(lon_2, lat_2)$ is the end location of $r_i$, respectively.

Although there are many griddig methods such as Kriging method, Nearest Neighbor method, and Polynomial Regression method etc., in this paper we choose to use the equal-sized gridding method due to the following reasons. First, such a gridding method can reflect the spatial correlations among the regions, and has been widely used in many spatial-temporal data mining tasks [38], [41]. Second, compared to other gridding methods, the equal-sized gridding method is more flexible because larger regions can be easily obtained by grouping a set of grid regions. Third, dividing a city into equal-sized cell regions can simplify the studied problem and is convenient to model the bike trips in all the cell regions as a matrix.

***Definition 3.*** **Bike Check-out/in Matrices.** The bike check-out/in matrices are denoted as $\mathbf{H}^{ou}/\mathbf{H}^{in} \in \mathcal{N}^{R\times T}$, where $R$ is the number of regions and $T$ is the number of time slots. Each element $h_{it}^{ou}/h_{it}^{in}$ denotes the number of bike check-out/in in region $r_i$ and in time slot $t$.

***Definition 4.*** **Bike Over-Demand in a Region.** We define a region $r_i$ as in a bike over-demand state in time slot $t$ if the following two conditions are satisfied: 1) the number of check-in bikes is equal to or slightly larger than the number of check-out bikes; 2) the average time lag between the check-in time and check-out time of a bike is less than a predefined threshold $\tau$.

***Definition 5.*** **Bike Over-Supply in a Region.** We define a region $r_i$ as in a bike over-supply state in time slot $t$ if the check-in bike number is significantly larger than that of the check-out bikes. That is $\frac{f_{it}^{in}}{f_{it}^{ou}} > \eta$, where $\eta$ is a predefined threshold.

***Definition 6.*** **Bike Real Demand/Supply Matrices.** The bike real demand/supply matrices are denoted as $\mathbf{F}^{ou}/\mathbf{F}^{in} \in \mathcal{N}^{R\times T}$. Each element $f_{it}^{ou}/f_{it}^{in}$ denotes the real bike demand/supply in region $r_i$ and in time slot $t$.

Note that the real demand/supply of bikes in a region and in a time slot is usually larger than the corresponding observed check-out/in bike number. The bike usage in different regions and time intervals is rather unbalanced, and in most cases a region is either in an over-demand or in over-supply state.

***Definition 7.*** **Region-POI Matrix.** The Region-POI matrix is denoted as $\mathbf{X} \in \mathcal{N}^{R\times M}$, where $M$ is the size of Point of Interest (POI) categories. Each element $x_{im}$ denotes the number of $m$-th type of POIs located in region $r_i$.

We briefly introduce the problem setting as follows. Before the stationless bike operators deploy new bikes in a city, they first need to estimate the demand/supply of the bikes in different regions of the city and different time intervals of a day. To do that, they will first estimate the real-demand/supply of bikes in some regions and time slots based on the current available bike trip data, based on which the real-demand/supply of the other regions and time slots are inferred. Note that in the real-world scenario, the bike deployment to a new city is not one shot but incremental. The bike usage patterns may change when more bikes are deployed. Thus the algorithm should be updated when more data are available. Based on the problem setting and the terminology definitions, we formally define the studied problem as follows.

***Problem Definition: Bike Usage Demand-Supply Inference:*** Given the incomplete demand and supply matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ constructed from a set of bike trips $Tr$, the regions of the city and the POI matrix $\mathbf{X}$, our goal is to infer the hourly real bike demand and supply for all the regions in the city. That is, we aim to complete the matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ simultaneously.

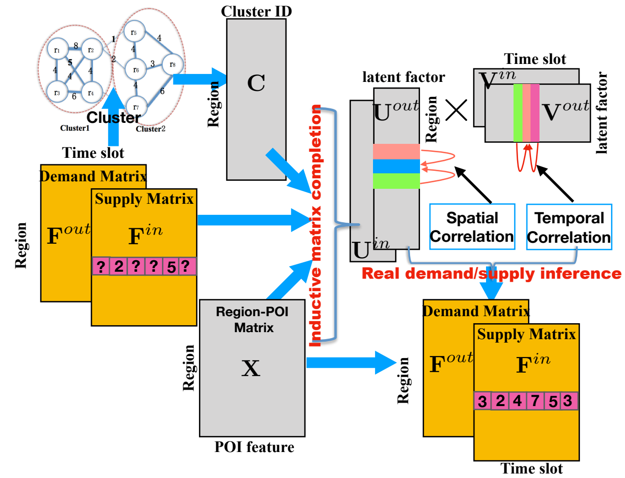
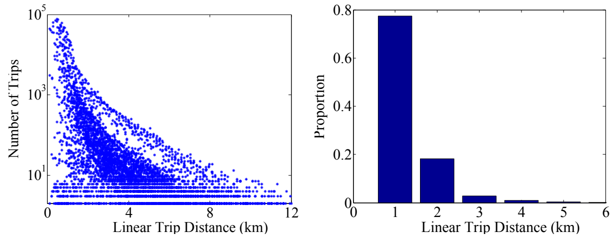



Fig. 2. Framework of the proposed model

## 2.2 Framework

The model framework is shown in Fig. 2. The model inputs are the bike usage demand matrix $\mathbf{F}^{ou}$, the supply matrix $\mathbf{F}^{in}$, and the POI feature matrix $\mathbf{X}$. Note that $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ are both incomplete due to the insufficient bikes deployed in the early stage. Our goal is to complete the two matrices through matrix factorization. Solely factorizing the two matrices may not achieve promising performance due to the very limited information. Thus we also incorporate the POI feature matrix and each region is associated with a set of POI features. To take the POI features into consideration, we propose to utilize an inductive matrix factorization model which can effectively encode the POI features into the learned latent factor matrices. As the bike usages in geographically close regions and in close time intervals are highly correlated, we also add the spatial and temporal correlations into our model. In addition, the bike usage should be balanced, namely the total demand should be equal to the total supply. However, directly incorporating this global constraint is difficult to solve. To simplify this problem, we propose to first cluster the regions based on the bike traffic flows, and then relax the global constraint to some local constraints that the bike usage in each cluster should be balanced. The local balanced bike usage constraints are also considered for a joint matrix factorization. Finally, the demand matrix $\mathbf{F}^{ou}$ is factorized into two latent factor matrices $\mathbf{U}^{ou}$ and $\mathbf{V}^{ou}$, and the supply matrix $\mathbf{F}^{in}$ is factorized into two latent factor matrices $\mathbf{U}^{in}$ and $\mathbf{V}^{in}$. With the learned low-dimensional latent factor matrices, the demand matrix $\mathbf{F}^{ou}$ and the supply matrix $\mathbf{F}^{in}$ can be easily inferred by multiplying the corresponding two latent factor matrices. More details of the model will be elaborated in the following sections step by step.




(a) # of trips *vs* trip distance (b) Trip proportation *vs* trip distance

Fig. 3. Linear distance distribution of the bike trips.

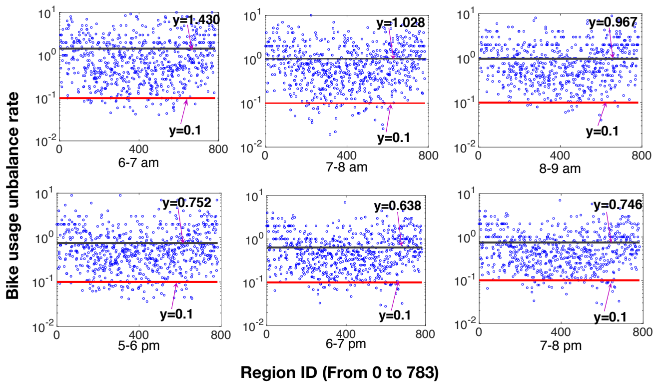



Fig. 4. Bike usage unbalance rate in different hours of a day.

## 3 Dataset and Data Analysis

In this paper, we use the public Mobike trip dataset of Beijing for data analysis. This dataset is released by Mobike in June 2017 for the Mobike bike travel destination prediction challenge [1]. In total there are 5.3 million Mobike bike trip data collected in 18 days from May 14 to May 31 2017, and about 3.2 million trips are training data while the remaining 2.1 million trips are testing data. In total we have 5.3 million bike check-out data and 3.2 million check-in data. Each trip contains the following information: *orderID*, *userID*, *bikeID*, *bikeType*, *startTime*, *startLocation*, and *endLocation*. Each testing trip contains all the above information except for the end location of the trip.

Fig. 3 shows the distance distribution of the bike trips. As the dataset only has the bike check-out and check-in locations, we cannot calculate a very accurate trip distance. Instead, we calculate the linear distance which is smaller than the real trip distance but can roughly reflect the real trip distance. Fig. 3(a) shows the distance *vs* trip number. Each blue point in Fig. 3(a) shows the number of trips with the equal trip distance. One can see that with the increase of trip distance, the corresponding trip number drops quickly, which means that only a small number of trips have long distance and most trips are short. Fig. 3(b) more clearly shows the distribution of the trip distances. One can see that the linear distance of nearly 80% trips is less than 1 kilometer, and the linear distance of more than 95% trips is less than 2 kilometers.

To further study whether the usage of bikes in different regions of Beijing is balanced, namely whether the check-in bike number is roughly equal to the check-out number, we show the bike usage unbalance rate of the regions in rush hours of a day in Fig. 4. The unbalance rate of region $r_i$ in $t$-th time slot is defined as $ur_{it} = \frac{|h_{it}^{in} - h_{it}^{ou}|}{max(h_{it}^{in}, h_{it}^{ou})}$. A small $ur_{ij}$ means the check-in bike number is close to the check-out bike number and thus the usage is balanced, while a large $ur_{ij}$ means that the bike usage is unbalanced. The red line $y = 0.1$ in each figure shows the unbalanced rate 0.1, which means that the check-in number is very close to the check-out number. One can see that only a very small number of points are below the line $y = 0.1$. The black line shows the average unbalance rate for all the regions. In 6:00-7:00 am and 7:00-8:00 am, the average unbalance rates are both larger than 1, which means that the average check-in bike number is two times of the check-out number, or vice versa. This figure shows that the bike usage in rush hours are highly unbalanced for most regions.

1. https://www.biendata.com/competition/mobike/

## 4 Usage Balanced Inductive Matrix Completion

In this section we will introduce the proposed **U**sage **B**alanced **I**nductive **M**atrix **C**ompletion (UBIMC) model. In this paper, we use bold uppercase characters for matrices, bold lowercase characters with a subscript for row vectors in matrices, and lowercase characters with a subscript for the scalar elements in matrices. For example, $\mathbf{F}^{ou}$ represents the demand matrix, $\mathbf{f}_i^{ou}$ is the $i$-th row of $\mathbf{F}^{ou}$ which is a vector, and $f_{ij}^{ou}$ is the $i$-th row and $j$-th column entry of $\mathbf{F}^{ou}$ which is a scalar.

Before elaborating the model, we first introduce how to initialize the real demand and supply matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$. If the bike usage is over-supplied in region $r_i$ in hour $t$, namely $h_{it}^{in}$ is significantly larger than $h_{it}^{ou}$, we can use $h_{it}^{ou}$ as the real demand $f_{it}^{ou}$ directly. We consider $h_{it}^{in}$ as the real bike supply for region $r_i$ in hour $t$ if the real demands of all $r_i$'s neighbor regions are known. This is because the supply of region $r_i$ is mostly composed of the check-out bikes of all $r_i$'s neighbor regions due to the fact that most trips are within one kilometers for linear distance.

### 4.1 The Basic Model

Low rank matrix completion (MC) is widely explored in various data mining tasks including recommendation [35], clustering [36] and feature learning [37]. The goal is to recover the incomplete matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ by factorizing them into low rank small matrices, which can be typically formulated as follows:

$$min\mathcal{L}_{\{\mathbf{U}^{in},\mathbf{V}^{in}\}} = \ell(\mathbf{F}^{in}, \mathbf{U}^{in}(\mathbf{V}^{in})^T) + \frac{\lambda_1}{2}(||\mathbf{U}^{in}||_F^2 + ||\mathbf{V}^{in}||_F^2) \tag{1}$$

$$min\mathcal{L}_{\{\mathbf{U}^{in},\mathbf{V}^{in}\}} = \ell(\mathbf{F}^{ou}, \mathbf{U}^{ou}(\mathbf{V}^{ou})^T) + \frac{\lambda_1}{2}(||\mathbf{U}^{ou}||_F^2 + ||\mathbf{V}^{ou}||_F^2) \tag{2}$$

where the first term is the loss function and the second term is the regularization term. $\mathbf{U}^{in} \in \mathcal{R}^{R\times L^{in}}$ and $\mathbf{V}^{in} \in \mathcal{R}^{T\times L^{in}}$ are the latent representations for the bike supply number in the region and time dimensions respectively, and $\mathbf{U}^{ou} \in \mathcal{R}^{R\times L^{ou}}$ and $\mathbf{V}^{ou} \in \mathcal{R}^{T\times L^{ou}}$ are the latent representations for the bike demand number in the two dimensions respectively.

We add the two parts together and form a unified objective function as follows:

$$\begin{aligned} &min\mathcal{L}_{\{\mathbf{U}^{ou},\mathbf{V}^{ou},\mathbf{U}^{in},\mathbf{V}^{in}\}} = \\ &\ell(\mathbf{F}^{ou}, \mathbf{U}^{ou}(\mathbf{V}^{ou})^T) + \ell(\mathbf{F}^{in}, \mathbf{U}^{in}(\mathbf{V}^{in})^T) + \\ &\frac{\lambda_1}{2}(||\mathbf{U}^{ou}||_F^2 + ||\mathbf{U}^{in}||_F^2 + ||\mathbf{V}^{ou}||_F^2 + ||\mathbf{V}^{in}||_F^2) \end{aligned} \tag{3}$$


Bike check-in/out | Residential area | Office building | Malls

Fig. 5. The heat maps of bike check-in/out data *vs* various POIs (Residential area, office buildings, and malls) in Beijing

### 4.2 Inductive Matrix Completion By Incorporating POIs

The usage demand of bikes is highly correlated with people's activities, which can be largely reflected by the distribution of POIs. Two regions with similar POI distributions may present similar bike usage patterns. To show the correlations between bike usage pattern and the distribution of POIs in a region, we give the heat maps of bike trip data and three types of POIs including *residential areas*, *office buildings* and *malls* in Fig. 5. One can see that Area1 is a large area where the usage demand of bikes is high. Area1 is located in *Chaoyang District* which is close to the center of Beijing. *Chaoyang District* has the largest number of office buildings, malls and residential areas among all the districts of Beijing. Area2 is also a large area presenting high bike usage demand. Area2 is located in *Haidian District* which has a large number of office buildings of IT corporations. Area3 is a large residential area called *Tian Tong Yuan*, which is the biggest residential area in the suburbs of Beijing. Area4 and Area5 are far from the central area of Beijing and outside the *Fifth Ring Road*. However, the subway Line1 crosses the two area. People tend to ride bikes to Line1 from their homes or return homes after work.

In this paper we consider POIs as the features of the regions. Each region $r_i$ is associated with a POI feature vector $\mathbf{x}_i$ with each entry $x_{ij}$ denoting the number of the $j$-th category of POI. To incorporate the POI features, we apply the recently proposed inductive matrix factorization model [2], [3]. Inductive matrix completion is initially proposed in recommender systems to integrate the features of users and items, and generally it can be also applied in other matrix completion tasks with side information. In a recommender system, each user can be associated with a set of features $\mathbf{x}_i^T$ such as gender, age, occupation, *et al.*, and each item is also associated with a set of features $\mathbf{y}_j$ such as price, category, size, *et al.* To incorporate the features of users and items for recommendation, usually one can model the rating matrix $\mathbf{A}_{ij}$ as $\mathbf{A}_{ij} = \mathbf{x}_i^T\mathbf{W}\mathbf{y}_j$, where $\mathbf{W}$ can be further factorized as two low rank matrices $\mathbf{W} = \mathbf{U}\mathbf{V}^T$. Thus $\mathbf{x}_i^T\mathbf{U}$ is the low rank latent representation of a user, and $\mathbf{V}\mathbf{y}_j$ is the low rank latent representation of an item. Motivated by this idea, we consider the POIs in each region as its features and utilize inductive matrix factorization model to incorporate such features as follows.

$$\begin{aligned} &min\mathcal{L}_{\{\mathbf{U}^{ou},\mathbf{V}^{ou},\mathbf{U}^{in},\mathbf{V}^{in}\}} = \\ &\ell(\mathbf{F}^{ou},\mathbf{X}\mathbf{U}^{ou}(\mathbf{V}^{ou})^T)+\ell(\mathbf{F}^{in},\mathbf{X}\mathbf{U}^{in}(\mathbf{V}^{in})^T)+ \\ &\frac{\lambda_1}{2}(||\mathbf{U}^{ou}||_F^2+||\mathbf{U}^{in}||_F^2+||\mathbf{V}^{ou}||_F^2+||\mathbf{V}^{in}||_F^2) \end{aligned} \tag{4}$$

where $\mathbf{X}\mathbf{U}^{ou}$ is the low rank latent representation of the regions for the bike check-out data, and $\mathbf{V}^{ou}$ is the low rank latent representation of the time slots. Note that here we have $\mathbf{U}^{ou} \in \mathcal{R}^{M\times L^{ou}}$, $\mathbf{V}^{ou} \in \mathcal{R}^{T\times L^{ou}}$, $\mathbf{U}^{in} \in \mathcal{R}^{M\times L^{in}}$, and $\mathbf{V}^{in} \in \mathcal{R}^{T\times L^{in}}$.

A significant advantage of inductive matrix completion is that the POI features are embedded into the factorized low rank matrices $\mathbf{U}^{in}$ and $\mathbf{U}^{ou}$. Given a new region $r_m$ with no bike check-in and check-out data at all, its real demand can be directly estimated by $\mathbf{x}_m\mathbf{U}^{in}(\mathbf{V}^{in})^T$ or $\mathbf{x}_m\mathbf{U}^{ou}(\mathbf{V}^{ou})^T$.

For simplicity, we let $\mathbf{Q}^{in} = \mathbf{X}\mathbf{U}^{in} \in \mathcal{R}^{R\times L^{in}}$ and $\mathbf{Q}^{ou} = \mathbf{X}\mathbf{U}^{ou} \in \mathcal{R}^{R\times L^{ou}}$. Then the inductive matrix factorization model can be rewritten as

$$\begin{aligned} &min\mathcal{L}_{\{\mathbf{U}^{ou},\mathbf{V}^{ou},\mathbf{U}^{in},\mathbf{V}^{in}\}} = \\ &\ell(\mathbf{F}^{ou},\mathbf{Q}^{ou}(\mathbf{V}^{ou})^T)+\ell(\mathbf{F}^{in},\mathbf{Q}^{in}(\mathbf{V}^{in})^T)+ \\ &\frac{\lambda_1}{2}(||\mathbf{Q}^{ou}||_F^2+||\mathbf{Q}^{in}||_F^2+||\mathbf{V}^{ou}||_F^2+||\mathbf{V}^{in}||_F^2) \end{aligned} \tag{5}$$

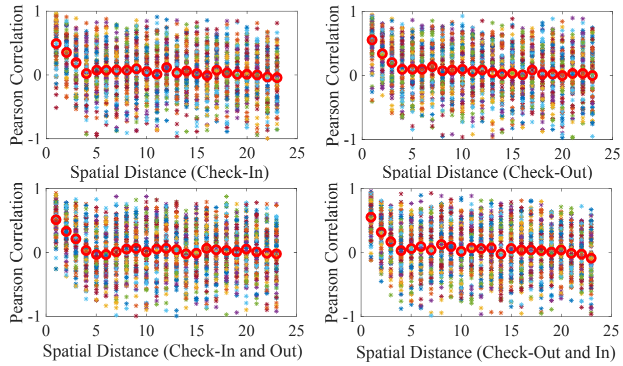
### 4.3 Matrix Completion by Incorporating Spatial-Temporal Correlations

**Incorporating Spatial Correlations.** Two regions close to each other may present similar bike usage patterns due to their spatial correlations. To study the spatial correlations of the bike usage, we compute the Pearson Correlations of the bike check-out and check-in numbers between each pair of regions in the hours from 6:00 am to 22:00 pm. For each region $r_i$, we first average the numbers of check-out and check-in bikes in each hour of a day and form two time serious data $Out_i = \{n_i^1, n_i^2, ...\}$, $In_i = \{m_i^1, m_i^2, ...\}$. Then compute the spatial distance between each pair of regions $r_i$, $r_j$ and their Pearson Correlations $p_{r_i,r_j} = \frac{\sum_l(n_i^l-\bar{n}_i)(n_j^l-\bar{n}_j)}{\sqrt{\sum_l(n_i^l-\bar{n}_i)}\sqrt{\sum_l(n_j^l-\bar{n}_j)}}$ of the bike usage time serious in a day. The spatial distance is measured by the number of regions across which region $r_i$ can reach $r_j$.

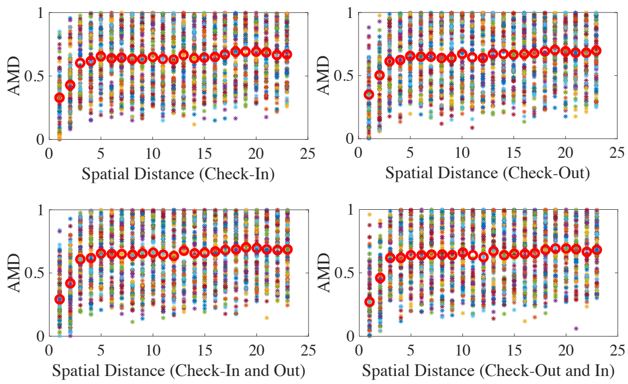
Pearson Correlation can show the linear relationship between two sets of data, but it cannot reflect the quantitive similarity between the two sets of data. For example, the Pearson Correlation between $(1, 2, 4)$ and $(2, 4, 8)$ is 1 since they present the same increase trend. However, the quantity of the two data sets are not quite similar. To address this issue, we define Absolute Mean Difference (AMD) as follows to measure the quantitive difference of two time serious data $Out_i$ and $Out_j$.

$$AMD(Out_i, Out_j) = \frac{1}{m}\sum_{k=1}^{m}\frac{|n_i^k - n_j^k|}{max(n_i^k, n_j^k)}$$

The value of AMD is in $[0, 1]$. A larger AMD means a larger difference in the bike check-out or check-in numbers between two regions. Compared to Pearson Correlation that can measure whether two time serious data present similar trend, AMD can show the average quantity difference of the two time serious data.




(a) The Pearson Correlations of bike usage data for each pair of regions with different spatial distances




(b) The Absolute Mean Difference (AMD) of bike usage data for each pair of regions with different spatial distances

Fig. 6. The spatial correlation of Mobike bike usage data

Fig. 6 shows the results. The x-axis is the spatial distance between two regions, and the y-axis is the Pearson Correlation (Fig. 6(a)) and the AMD (Fig. 6(b)), respectively. Each point in the figures represents a Pearson Correlation or AMD between two regions and the red circle represents the average value of all the region pairs with the same spatial distance. The left upper figure in Fig. 6(a) shows the check-in bike number correlation between each pairs of regions, the right upper figure shows the check-out bike number correlations, and the lower two figures shows the correlation between the check-in bike number of a region with the check-out bike number of another region. From Fig. 6(a) one can see that two neighbor regions with distance 1 have the largest Pearson Correlations in all the four cases. The whole trends are that the correlations first drop with the increase of the spatial distance and then become stable. When the spatial distance is too large, say 4, the correlation is so small that we can consider there is no apparent correlations between them. Fig. 6(b) shows the very similar trend as Fig. 6(a). Fig. 6(b) shows that in all the four cases, the AMD value of two regions with the spatial distance less than 2 is much smaller than that of two regions far away from each other. It also verifies the high spatial correlation of the bike check-out and check-in quantities in different regions. From the two groups of figures one can conclude that two regions that are spatially close to each other are much more likely to present both similar bike usage trend and bike usage quantity. This motivates us that incorporating such spatial correlations could potentially help us better perform the bike usage demand inference problem.

To take the spatial correlation into consideration, we add the following penalty term.

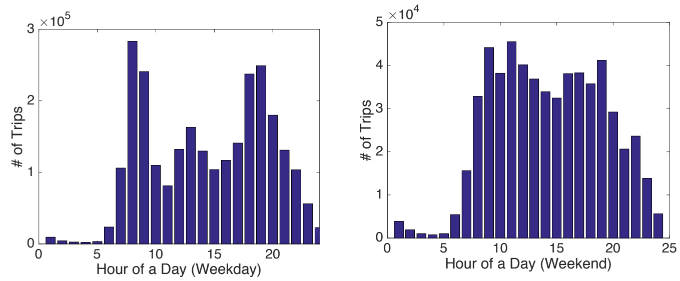



Fig. 7. Mobike bike trip distribution in different hours of weekdays and weekends.

$$\sum_{i,j=1}^{n} f(d(r_i, r_j))(||\mathbf{q}_i^{in} - \mathbf{q}_j^{in}||_F^2 + ||\mathbf{q}_i^{ou} - \mathbf{q}_j^{ou}||_F^2) \quad (6)$$

where $\mathbf{q}_i^{ou}$ is the $i$-th row of the matrix $\mathbf{Q}^{ou}$, and $\mathbf{q}_i^{in}$ is the $i$-th row of the matrix $\mathbf{Q}^{in}$. This penalty term aims to control the similarity of the learned latent factors in the region dimension based on the spatial distance of the two regions $r_i$ and $r_j$. $f(d(r_i, r_j))$ is an exponential decaying function controlled by the distance between $r_i$ and $r_j$, and a larger distance $d(r_i, r_j)$ will lead to a smaller $f(d(r_i, r_j))$. The idea is that a small spatial distance between $r_i$ and $r_j$ leads to a large value of $f(d(r_i, r_j))$, and thus a large penalty is assigned to it for making them closer to each other in the latent space. A large spatial distance between $r_i$ and $r_j$, on the contrary, means that they might have low bike usage pattern correlations and thus small penalty should be assigned.

**Temporal Correlation.** In addition to the spatial correlation, the bike usage also presents high temporal correlation. We plot the distributions of the bike trip number in different hours of weekdays and weekends in Fig. 7, respectively. One can see that there are two significant peaks from 7:00 am to 9:00 am and from 17:00 pm to 19:00 pm on weekdays. The peaks on weekend is not that significant as people do not go to work, but there is still a significant increase from 6:00 am to 7:00 am and a significant decrease from 20:00 pm to 21:00 pm. Overall, in the day time the curve of weekend is smoother than that of weekdays. Except for the peaks in the morning and night, the bike usage in neighbor hours changes smoothly, which presents remarkable temporal correlation. Thus different from previous work [38] that uses the temporal correlation directly, we need to consider both the two peaks and the smoothly changing trends of the bike usage time serious data of a day. In this paper, we make the assumption that the latent factors of a region in two successive slots should be similar except that one time slot is in the rush hour and the other is not. To achieve this end, we add the following constraint to capture the temporal correlation of the bike usage.

$$\sum_{i=1}^{T-1} I(i, i+1)(||\mathbf{v}_i^{in} - \mathbf{v}_{i+1}^{in}||_F^2 + ||\mathbf{v}_i^{ou} - \mathbf{v}_{i+1}^{ou}||_F^2) \quad (7)$$

where $\mathbf{v}_i^{ou}$ is the $i$-th row of the matrix $\mathbf{V}^{ou}$, and $\mathbf{v}_i^{in}$ is the $i$-th row of the matrix $\mathbf{V}^{in}$. $I(i, j)$ is such an indicator function,

$$I(i, i+1) = \begin{cases} 0 & i \in \{6:00, 8:00, 16:00, 18:00\} \\ 1 & otherwise \end{cases}$$

This indicator function aims to penalize the latent factor difference of a region in two successive time slots $t_i$ and $t_j$ if they are both in rush hours or non-rush hours. If one of the two time slots is in rush hours and the other is not or vice versa, the value of the indicator function is zero and no penalty is assigned.

### 4.4 Matrix Completion with Locally Balanced Bike Usage Constraint

The bike usage should be balanced, namely the total number of check-out bikes should be equal to the total number of check-in bikes. However, it is intractable to find the optimal solution that satisfies the globally balanced bike usage constraint. Instead we relax the global constraint to several local constraints. The idea is that we first cluster the regions based on the bike traffic flows among them. Inter-cluster bike flows will be of a relative small proportion and can be neglected compared with the large-volume of intra-cluster bike flows. Then we assume the bike usage in each region cluster is balanced. The motivation of doing this is that, 1) the bike usage is not evenly distributed in a city and present several hot areas where the bike demands are much larger than in the other areas; 2) a locally balanced constraint is much easier to solve; and 3) most bike trips are short in distance.

**Clustering regions based on bike traffic flow.** We apply a hierarchical clustering method to cluster the regions based on the available bike traffic flows among the regions. Specifically, we first construct a bike traffic flow graph $G$. Each node of $G$ denotes a region and each edge $e_{ij}$ between two regions $r_i$ and $r_j$ denotes the corresponding bike traffic flow. Note that graph $G$ is a directional graph. $e_{ij}$ denotes how many bike trips start from region $r_i$ and end at region $r_j$, while $e_{ji}$ denotes how many bike trips start from region $r_j$ and end at region $r_i$. Based on the bike traffic flow graph $G$, we utilize an agglomerative clustering method to cluster the regions. Here we define the similarity between two clusters as follows by considering both the bike check-in and check-out flows of the two region clusters.

$$Sim(\mathbf{c}_i, \mathbf{c}_j) = \frac{F_{\mathbf{c}_i->\mathbf{c}_j}}{F_{\mathbf{c}_i->\bullet}} + \frac{F_{\mathbf{c}_i->\mathbf{c}_j}}{F_{\bullet->\mathbf{c}_j}} + \frac{F_{\mathbf{c}_j->\mathbf{c}_i}}{F_{\mathbf{c}_j->\bullet}} + \frac{F_{\mathbf{c}_j->\mathbf{c}_i}}{F_{\bullet->\mathbf{c}_i}} \quad (8)$$

In formula (8), $F_{\mathbf{c}_i->\mathbf{c}_j}$ denotes the number of bike trips starting from region cluster $\mathbf{c}_i$ and ending at region cluster $\mathbf{c}_j$, $F_{\mathbf{c}_i->\bullet}$ denotes the number of all the bike trips starting from region cluster $c_i$, and $F_{\bullet->\mathbf{c}_j}$ denotes the number of all the bike trips ending at the region cluster $\mathbf{c}_j$. This similarity measurement means that region clusters $\mathbf{c}_i$ and $\mathbf{c}_j$ are similar if the bike flow between them is large while the flow between the two clusters to the other region clusters is small. The pseudocode of the clustering algorithm is given Algorithm 1.

**Algorithm 1** Region Clustering based on Bike Traffic Flows

**Input:** The bike flow graph $G$ and the cluster number $K$
**Output:** The region cluster matrix $\mathbf{C}$

```
Initialize each region as a cluster;
Construct the cluster similarity matrix S based on formula (8);
while (t < K) do
    Pick a pair of clusters c_i and c_j with the largest similarity using S;
    Merge cluster c_i and c_j;
    Delete rows/columns i and j from S and create a new row and column for the new merged cluster;
    Update the similarity matrix S;
end while
return C
```

The difference between the check-in and check-out bike numbers in the $i$-th region cluster can be represented as

$$\mathbf{c}_k(\mathbf{Q}^{in}\mathbf{V}^{in} - \mathbf{Q}^{ou}\mathbf{V}^{ou}) \quad (9)$$

where $\mathbf{c}_k$ is the $i$-th row of the region cluster matrix $\mathbf{C}$. The check-in and check-out difference in all the region clusters can be calculated by

$$\sum_{k=1}^{K} \mathbf{c}_k(\mathbf{Q}^{in}\mathbf{V}^{in} - \mathbf{Q}^{ou}\mathbf{V}^{ou}) \quad (10)$$

As the bike traffic flows within the same cluster are large and among different clusters are small, the number of check-out bikes should be close to the check-in number in the same cluster. Thus we need to minimize the difference between supply and demand in each cluster as given in formula (10),

By taking both the spatial-temporal correlation and the locally balanced bike usage constraint into consideration, we can write the final objective function as follows.

$$\begin{aligned}
&min\mathcal{L}_{\{\mathbf{Q}^{in},\mathbf{Q}^{ou},\mathbf{V}^{in},\mathbf{V}^{ou}\}} = \underbrace{\ell(\mathbf{F}^{in}, \mathbf{Q}^{in}\mathbf{V}^{in}) + \ell(\mathbf{F}^{ou}, \mathbf{Q}^{ou}\mathbf{V}^{ou})}_{inductive\ matrix\ completion} \\
&+ \underbrace{\frac{\lambda_1}{2}\sum_{i,j=1}^{R} f(d(r_i, r_j))(||\mathbf{q}_i^{in} - \mathbf{q}_j^{in}||_F^2 + ||\mathbf{q}_i^{ou} - \mathbf{q}_j^{ou}||_F^2)}_{spatial\ correlation} \\
&+ \underbrace{\frac{\lambda_2}{2}\sum_{i=1}^{T-1} I(i, i+1)(||\mathbf{v}_i^{in} - \mathbf{v}_{i+1}^{in}||_F^2 + ||\mathbf{v}_i^{ou} - \mathbf{v}_{i+1}^{ou}||_F^2)}_{temporal\ correlation} \\
&+ \underbrace{\frac{\lambda_3}{2}\sum_{i=1}^{K} ||\mathbf{c}_k(\mathbf{Q}^{in}\mathbf{V}^{in} - \mathbf{Q}^{ou}\mathbf{V}^{ou})||_F^2}_{balanced\ bike\ usage\ constraint} \\
&+ \underbrace{\frac{\lambda_4}{2}(||\mathbf{Q}^{in}||_F^2 + ||\mathbf{Q}^{ou}||_F^2 + ||\mathbf{V}^{ou}||_F^2 + ||\mathbf{V}^{in}||_F^2)}_{regularization\ term} \\
&\text{s.t. } \mathbf{Q}^{in}, \mathbf{Q}^{ou}, \mathbf{V}^{in}, \mathbf{V}^{ou} \geq 0
\end{aligned} \quad (11)$$

The first term in the objective function is the inductive matrix completion which utilizes the POI features of the regions to help complete the sparse matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$, the second term is the spatial correlation term, the third term is the temporal correlation term, the fourth term is the balanced bike usage constraint term to connect the bike demand data and bike supply data, and the final term is the regularization term to avoid overfitting.

One can see that our model forces two regions with similar POI features (similar $\mathbf{x}$) and spatially close to each other (similar $\mathbf{q}$) to have similar bike demand in the same time intervals (similar $\mathbf{v}$), which is consistent with our data observation. The bike demand and supply data in the first three terms of the model can be optimized separately. However, the demand and supply of bikes should not be estimated independently as the usage of bike should be balanced. Thus we have the fourth term to add the constraint that the total demand should be roughly equal to the total supply in a region cluster. Based on the factorized low ranked matrices $\mathbf{Q}^{ou}$ and $\mathbf{V}^{ou}$, the real bike demand $f_{ij}^{ou}$ in region $r_i$ and time slot $t_j$ can be inferred by $f_{ij}^{ou} = \mathbf{q}_i^{ou}\mathbf{v}_j^{ou}$, where $\mathbf{q}_i^{ou} = \mathbf{x}^i\mathbf{U}^{ou}$. A significant advantage of our method is that it can efficiently predict the bike usage demand for a new region $r_{new}$ with the POI features $\mathbf{x}^{new}$ with the above steps.

### 4.5 Optimization Algorithm

To make the objective function (11) easy to optimize, we first rewrite the second term and the third term as the matrix form. For the second term, we first define the distance closeness matrix as $\mathbf{D}$ with each entry $d_{ij} = f(d(r_i, r_j))$. Then the second term can be rewritten as

$$\begin{aligned}
&\frac{\lambda_1}{2}\sum_{i,j=1}^{R} f(d(r_i,r_j))(||\mathbf{q}_i^{in}-\mathbf{q}_j^{in}||_F^2+||\mathbf{q}_i^{ou}-\mathbf{q}_j^{ou}||_F^2)\\
=&\frac{\lambda_1}{2}\sum_{i,j=1}^{R} d_{ij}(||\mathbf{q}_i^{in}-\mathbf{q}_j^{in}||_F^2+||\mathbf{q}_i^{ou}-\mathbf{q}_j^{ou}||_F^2)\\
=&\frac{\lambda_1}{2}\sum_{i,j=1}^{R} d_{ij}[(\mathbf{q}_i^{in}-\mathbf{q}_j^{in})((\mathbf{q}_i^{in})^T-(\mathbf{q}_j^{in})^T)+\\
&(\mathbf{q}_i^{ou}-\mathbf{q}_j^{ou})((\mathbf{q}_i^{ou})^T-(\mathbf{q}_j^{ou})^T)]\\
=&\lambda_1[tr((\mathbf{Q}^{in})^T(\mathbf{Z}-\mathbf{D})\mathbf{Q}^{in})+tr((\mathbf{Q}^{ou})^T(\mathbf{Z}-\mathbf{D})\mathbf{Q}^{ou})]\\
=&\lambda_1(tr((\mathbf{Q}^{in})^T\mathbf{L}_{spatial}\mathbf{Q}^{in})+tr((\mathbf{Q}^{ou})^T\mathbf{L}_{spatial}\mathbf{Q}^{ou}))
\end{aligned}$$

where $\mathbf{Z} \in \mathcal{R}^{R\times R}$ and $z_{ii} = \sum_{j=1}^{F} d_{ij}$. $\mathbf{L}_{spatial} = \mathbf{Z} - \mathbf{D}$ is the Laplacian matrix. Similarly, the third term can be rewritten as

$$\lambda_2(tr(\mathbf{V}^{in}\mathbf{L}_{temporal}(\mathbf{V}^{in})^T)+tr(\mathbf{V}^{ou}\mathbf{L}_{temporal}(\mathbf{V}^{ou})^T))$$

With the above deductions, the matrix form of the objective function (11) is as follows.

$$\begin{aligned}
&min\mathcal{L}_{\{\mathbf{Q}^{in},\mathbf{Q}^{ou},\mathbf{V}^{in},\mathbf{V}^{ou}\}} = \underbrace{\ell(\mathbf{F}^{in},\mathbf{Q}^{in}\mathbf{V}^{in})+\ell(\mathbf{F}^{ou},\mathbf{Q}^{ou}\mathbf{V}^{ou})}_{inductive\ matrix\ completion}\\
&+\underbrace{\lambda_1(tr((\mathbf{Q}^{in})^T\mathbf{L}_{spatial}\mathbf{Q}^{in})+tr((\mathbf{Q}^{ou})^T\mathbf{L}_{spatial}\mathbf{Q}^{ou}))}_{spatial\ correlation}\\
&+\underbrace{\lambda_2(tr(\mathbf{V}^{in}\mathbf{L}_{temporal}(\mathbf{V}^{in})^T)+tr(\mathbf{V}^{ou}\mathbf{L}_{temporal}(\mathbf{V}^{ou})^T))}_{temporal\ correlation}\\
&+\underbrace{\frac{\lambda_3}{2}||\mathbf{C}(\mathbf{Q}^{in}\mathbf{V}^{in}-\mathbf{Q}^{ou}\mathbf{V}^{ou})||_F^2}_{balanced\ bike\ usage\ constraint}\\
&+\underbrace{\frac{\lambda_4}{2}(||\mathbf{Q}^{in}||_F^2+||\mathbf{Q}^{ou}||_F^2+||\mathbf{V}^{ou}||_F^2+||\mathbf{V}^{in}||_F^2)}_{regularization\ term}\\
&\text{s.t. } \mathbf{Q}^{in},\mathbf{Q}^{ou},\mathbf{V}^{in},\mathbf{V}^{ou} \geq 0
\end{aligned} \tag{12}$$

The pseudocode of the algorithm is given in Algorithm 2.

**Algorithm 2** Usage Balanced Inductive Matrix Completion

**Input:** Incomplete bike demand/supply matrix $\mathbf{F}^{ou}$/$\mathbf{F}^{in}$, the POI features of the regions matrix $\mathbf{X}$, and region-cluster matrix $\mathbf{C}$.
**Output:** Low rank matrices $\mathbf{U}^{ou}$, $\mathbf{U}^{in}$, $\mathbf{V}^{ou}$, $\mathbf{V}^{in}$.

```
Initialize the matrices U^ou, U^in, V^ou, V^in with small random values, and let Q^in = XU^in, Q^ou = XU^ou
Set γ as the learning rate
while (t < IterMax and L^t − L^{t+1} > ε) do
    Calculate the gradients ∇_{Q^in}L, ∇_{Q^ou}L, ∇_{V^in}L, ∇_{V^ou}L.
    Update Q^in = Q^in − γ∇_{Q^in}L
    Update Q^ou = Q^ou − γ∇_{Q^ou}L
    Update V^in = V^in − γ∇_{V^in}L
    Update V^ou = V^ou − γ∇_{V^ou}L
    Update t and L
end while
Get U^ou, U^in based on the learned Q^ou, Q^in
return U^ou, U^in, V^ou, V^in
```

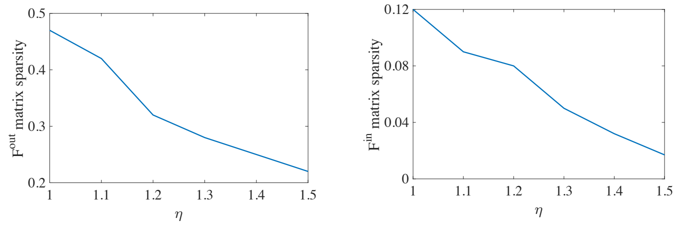



Fig. 8. Impact of parameter $\eta$ on the sparsity of the two matrices.

## 5 EVALUATION

We evaluate the effectiveness of the proposed UBIMC through answering the following two questions. (1) Whether UBIMC is more effective than previous bike demand prediction models in the stationless bike-sharing systems? (2) Whether incorporating the spatial-temporal correlations and the balanced bike usage constrain can improve the performance of the model? If it does, to what extent it improves?

### 5.1 Dataset and Experiment Setup

**Dataset.** We use the publicly available Mobike trip dataset for evaluation. The details of the dataset are described in Section 3. We also use more than 0.18 million POIs of Beijing. For each POI, we extract the location and the POI type such as *mall*, *restaurant* and *parks*. We categorize all the POIs into 17 classes. Some similar POI categories are grouped for simplicity. The statistics of the POI data is shown in Table 1. One can see that *Office*, *residential*, *Mall & Shopping*, and *restaurant* are four major types of POIs, and they together account for more than 60% of the total number of POIs.

We partition the main urban area (within the *Fifth Ring Road*) of Beijing into 28×28 regions, and the area of each region is about 1 square kilometer. Then we map the check-out and check-in locations of each trip data to the corresponding regions. We further extract the check-in/out hours of each trip, and group the trips of a region that are in the same hour of a day. Then we average the hourly check-in/out numbers in each region for all the days in our dataset, and construct the check-out matrix $\mathbf{H}^{ou}$ and the check-in matrix $\mathbf{H}^{in}$, based on which we construct the real demand and supply matrices $\mathbf{F}^{ou}$, $\mathbf{F}^{in}$. Given the bike check-out number $h_{ij}^{ou}$ of region $r_i$ in hour $t_j$, we consider $h_{ij}^{ou}$ is the real demand $f_{ij}^{ou}$ if region $r_i$ in $t_j$ is in over-supply state based on $Definition$ 3. Given a region $r_j$ in hour $t_j$, if all its neighbor regions are in over-supply state, we consider the check-in bike number $h_{ij}^{in}$ is the real supply $f_{ij}^{in}$. In this way, we obtain the real demand and supply matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ for evaluation.

Note that the parameter $\eta$ in *Definition* 3 can affect the sparsity of the two matrices. We show the sparsity change curves of the two matrices under different settings of $\eta$ in Fig. 8. One can see that a larger $\eta$ leads to sparser $F^{ou}$ and $F^{in}$. In this paper we set $\eta = 1.2$, which means a region is over-supplied if the check-in bike number is 1.2 times larger than the check-out bike number. Although a larger $\eta$ will let us have a denser matrix $F^{ou}$ and more ground truth data, it will lead to misclassify more over-supplied regions as not over-supplied. Under such a setting, around 32% entries of the demand matrix $F^{ou}$ and 8% entries of the supply matrices $F^{in}$ have ground truth values. For the entries without ground truth values due to data sparsity or unavailability, out model can still give estimations. But we do not evaluate the estimations for these entries due to the lack of ground truth data.

TABLE 1
Statistics of the POI in Beijing

| Category of POI | Number | Percentage |
|---|---|---|
| Government agency | 6,997 | 3.78% |
| Life services | 13,282 | 7.19% |
| Education & Training | 5,876 | 3.18% |
| Parks | 955 | 0.52% |
| Transportation facilities | 6,351 | 3.43% |
| Automobile service | 11,187 | 6.06% |
| Scenic spots | 1,572 | 0.85% |
| Cultural media | 624 | 0.34% |
| Entertainment | 4,576 | 2.48% |
| Restaurant | 20,457 | 11.08% |
| Residential | 35,135 | 19.02% |
| Hotel | 1,479 | 0.80% |
| Mall & Shopping | 26,283 | 14.23% |
| Exercise & Fitness | 1,845 | 1% |
| Financial | 5,196 | 2.81% |
| Medical & Hospitals | 4,289 | 2.32% |
| Office | 38,577 | 20.89% |
| Total | 184,681 | 100% |

**Experiment Setting.** We evaluate the performance of UBIMC on the following two settings. For the first setting, we randomly partition the regions into two groups. One group of regions are used for training, and the other group of regions are used for testing. We conduct experiments under such a setting to examine whether UBIMC can make accurate predictions for the new regions. For the second setting, we randomly sample some entries from the demand matrix $\mathbf{F}^{ou}$ and supply matrix $\mathbf{F}^{in}$, and assume that the other entries are unknown and need to be inferred. We choose to use the data in the time period from 6 am to 22 pm for evaluation. In the following experiments, we perform two groups of experiments with different data sparsity ratios based on the above two experiment settings. In the first group of experiments, we randomly select 50% data for training and the remaining 50% data for testing. In this case, 84% entries of the demand matrix and 96% entries of the supply matrix are missing, respectively. In the second group of experiments, 30% data are selected for training and the remaining 70% are for testing. In this case, the demand and supply matrices are even sparser with 91% and 97.6% missing entry values, respectively. Following previous matrix factorization works [2], [32], we run our algorithm 5 rounds with different initializations of the low dimensional matrices, and average the results. The results show that the deviation of the results in different rounds is not significant and small than 0.01 for ER. Thus we present the average of 5-round results as the final results in the following experiments. As the bike usage patterns on weekdays and weekends are different, we evaluate our model on weekdays and weekends separately. Specifically, we train two models for weekdays and weekends by using the corresponding data respectively, and then evaluate the results.

### 5.2 Baselines and Evaluation Metrics

**Baselines.** We compare UBIMC with both state-of-the-art matrix completion and bike demand prediction models that are proposed for traditional bike-sharing systems, and variations of UBIMC. Specifically, we first compare UBIMC with the existing methods as follows to answer the first question.

- **Context-Aware Matrix Factorization (CAMF) [31].** Context-aware matrix factorization technique is widely explored recently to address the data sparsity issue in matrix completion [31], [32]. As solely factorizing the demand matrix $\mathbf{F}^{ou}$ and supply matrix $\mathbf{F}^{in}$ cannot achieve promising performance due to data sparsity, we can factorize them together with the POI feature matrix by assuming they share a latent feature represent matrix.
- **Boosted Inductive Matrix Completion (BIMC) [2].** BIMC is proposed to address the blog recommendation task in the microblogging site of Tumblr. The main difference between BIMC and our proposal is that BIMC combines both standard matrix completion and inductive matrix completion to reduce the noise in the input data as well as to incorporate side information. Although BIMC was proposed to solve a totally different task, we also compare with it since the ideas of the two models are similar. For the check-out and check-in matrices, we use two BIMC models to complete them separately.
- **LinUOTD [39].** LinUOTD is a linear regression model proposed recently to predict the Unit Original Taxi Demand (UOTD) [39]. A spatial-temporal regularization is added into the model to make it fit the spatial-temporal data prediction task. In this paper, we construct the following features for LinUOTD and use it for our inference task: number of different types of POIs in the region, time slot, average bike usage of neighbor regions, region cluster, number of subway and bus stations, etc.
- **Single Region Level Estimation (SRLE).** The single region level estimation performs the bike usage demand estimation for each single region [27] rather than estimating the demands of all the regions like a whole as the proposed matrix completion method. Such model generally first extracts a set of global features and then feeds them into a regression model like Random Forest (RF) or K-Nearest Neighbor Regressor (KNN).
- **Functional Zone based Random Forest Regressor (FZ+RF)[5].** FZ+RF is a recent work that predict the station-level bike demand for bike system expansion. It proposes a bi-clustering model on POIs and stations by utilizing POI characteristics and geographical distances as features [5]. This method is designed for traditional bike-sharing systems and cannot be applied to our case directly. To make it comparable, we modify this method and consider each region as a Voronoi Region as proposed in [5], and then use the POI distribution in each region and the distance among regions to cluster the regions into functional regions. Then we predict the Function-to-Function and Region-to-Region bike transitions based on their method. Finally, we predict the bike demand in each region by using their proposed Random Forest Regressor.

To answer the second question, we also compare UBIMC with the following variations.

- **UBIMC-Spa.** UBIMC-Spa is a variation of UBIMC which does not consider the spatial correlation term. We choose it as a baseline to test whether the spatial correlations is helpful to the studied task.
- **UBIMC-Tem.** Similar to UBIMC-Spa, UBIMC-Tem is a variation of UBIMC which does not consider the temporal correlation term. This baseline is used to test whether the temporal correlation is helpful.

- **UBIMC-Ban.** UBIMC-Ban removes the intra-cluster balanced bike usage constraint. This baseline is used to test whether incorporating the balanced bike usage constraint can improve the performance.

**Evaluation metrics.** We use two types of metrics for evaluation. The first metric is the estimation accuracy. Based on the fact that the real demand should be no less than the number of observed check-out bikes and the real supply should be no less than the number of observed check-in bikes, we use the inference accuracy defined as follows to perform a coarse grained evaluation. Given the check-out bike number $h_{ij}^{ou}$ of region $r_i$ in time $t_j$ and the estimate real demand $f_{ij}^{ou}$, if $f_{ij}^{ou} \geq h_{ij}^{ou}$ we say the estimation is accurate; otherwise we say the estimation is wrong.

$$Acc = \frac{\sum_{i,j} I(f_{ij}^{ou} \geq h_{ij}^{ou})}{NT}.$$

Note that $Acc$ can be used to evaluate all the regions in all the time slots since ground truth is not needed in this evaluation. The second type of metrics are *Error Rate* (ER) and *Root Mean Squared Logarithmic Error* (RMSLE) [27].

$$ER = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{T}\frac{|f_{ij} - \bar{f}_{ij}|}{f_{ij}}$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{T}(log(f_{ij}+1) - log(\bar{f}_{ij}+1))^2}$$

where $f_{ij}$ is the real value while $\bar{f}_{ij}$ is the estimation. As ER and RMSLE are more quantitive evaluation metrics and need the ground truth demand/supply bike number, we only evaluate the entries of $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ which have the ground truth values as described in Section 5.1.

#### 5.2.1 Parameter Study

There are four parameters in UBIMC to control the importance of different components. As different parameter settings can significantly affect the model performance, in this subsection we study the sensitivity of UBIMC to the parameters $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$. We use the following 2-round tuning method to find the best parameter setting. In the first round, we first fix the values of the three parameters to 1, and then tune the value of the fourth one to find the value that achieves the best performance. In the second round, we fix the three parameters with their values found in the first round, and then tune the fourth one to find the best value setting again. Here we use ER as the evaluation metric. The result is shown in Fig. 9. One can see that the model performance is sensitive to all the four parameters, which implies that the spatial correlation, the temporal correlation, the balanced bike usage constraint and the regularization term are all important to the studied problem and should be carefully considered. A proper parameter setting in Fig. 9 is $\lambda_1 = 10$, $\lambda_2 = 1000$, $\lambda_3 = 0.1$, and $\lambda_4 = 100$. Thus in the following experiments, we use this parameter setting for evaluation.

#### 5.2.2 Results over Inference Accuracy

We first present the inference accuracy of various models. We perform the evaluation through two tasks under two experiment settings described in Section 5.1. $Task1$ corresponds to the first experiment setting and $Task2$ corresponds to the second experiment setting. As the bike usage patterns in weekdays and weekends can be different significantly, we also conduct the evaluation on weekdays and weekends separately.

TABLE 2
Accuracy comparison of various methods on the two tasks

| Methods | $train = 30\%, Test = 70\%$ | | | |
|---|---|---|---|---|
| | $T1$ $(Wee)$ | $T1$ $(Wke)$ | $T2$ $(Wee)$ | $T2$ $(Wke)$ |
| CAMF | 0.685 | 0.654 | 0.724 | 0.722 |
| BIMC | 0.712 | 0.687 | 0.732 | 0.713 |
| LinUOTD | 0.654 | 0.647 | 0.636 | 0.656 |
| SRLE | 0.675 | 0.654 | 0.710 | 0.713 |
| FZ-RF | 0.702 | 0.686 | 0.842 | 0.804 |
| UBIMC-Spa | 0.726 | 0.702 | 0.846 | 0.823 |
| UBIMC-Tem | 0.715 | 0.705 | 0.834 | 0.815 |
| UBIMC-Ban | 0.726 | 0.712 | 0.845 | 0.822 |
| UBIMC | **0.745** | **0.724** | **0.862** | **0.826** |
| Methods | $train = 50\%, Test = 50\%$ | | | |
| | $T1$ $(Wee)$ | $T1$ $(Wke)$ | $T2$ $(Wee)$ | $T2$ $(Wke)$ |
| CAMF | 0.735 | 0.714 | 0.825 | 0.802 |
| BIMC | 0.745 | 0.723 | 0.822 | 0.824 |
| LinUOTD | 0.587 | 0.612 | 0.634 | 0.635 |
| SRLE | 0.744 | 0.697 | 0.812 | 0.810 |
| FZ-RF | - | - | - | - |
| UBIMC-Spa | 0.808 | 0.793 | 0.914 | 0.903 |
| UBIMC-Tem | 0.793 | 0.784 | 0.906 | 0.886 |
| UBIMC-Ban | 0.816 | 0.784 | 0.916 | 0.895 |
| UBIMC | **0.836** | **0.804** | **0.935** | **0.910** |

[1] $T1$: Task 1, $T2$: Task 2.
[2] $Wee$: weekday, $Wke$: weekend.

Table 2 shows the experiment results. The best result is highlighted with bold font. Note that FZ-RF is proposed to predict the bike usage demand for new regions. So we only apply it for $Task1$. Based on Table 2, we can have the following conclusions. First, it is obvious that more training data leads to better performance. One can see the performance of the methods with 50% training data consistently outperforms their corresponding performance with only 30% training data. Second, the bike flows in weekends are harder to predict than that in weekdays. On average, the inference accuracy in weekdays is 2%-5% higher that than in weekends for different methods. This is mainly because 1) the bike usage pattern in weekend is more complex and less regular compared to in weekdays, and 2) we only have 5 days data in weekend which is sparse. Third, $Task1$ is much harder than $Task2$ as the regions for prediction have no data in all the time intervals. Fourth, UBIMC outperforms all the baselines in all the cases, which verifies its effectiveness. Compared to CAFM, BIMC, LinUOTD, SRLE, and FZ-RF, the performance improvement of UBIMC is significant. Among the three baselines, FZ-RF performs best because it considers both the spatial and POI features for prediction, but it is still inferior to UBIMC. CAMF performs bad as it ignores the spatial and temporal correlations and the balanced bike usage constraint. BIMC performs better than CAMF because BIMC also uses inductive matrix factorization to integrate the POI features. The performance of LinUOTD is undesirable, and it is even inferior to CAFM. This is mainly because it is a linear regression model and cannot effectively incorporate other information and constrains. UBIMC also outperforms the three variations UBIMC-Spa, UBIMC-Tem, and UBIMC-Ban. It demonstrates that these components all contribute to generating a more accurate inference, and ignoring any one of them can lead to performance drop.

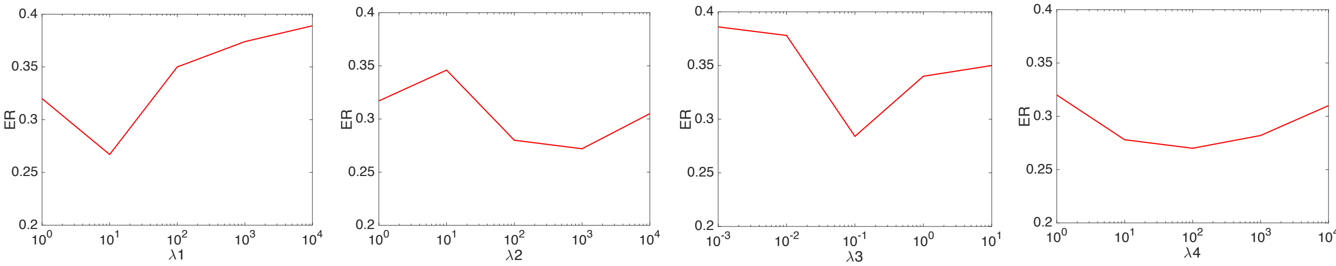



Fig. 9. Impact of the four parameters on the model performance for demand inference

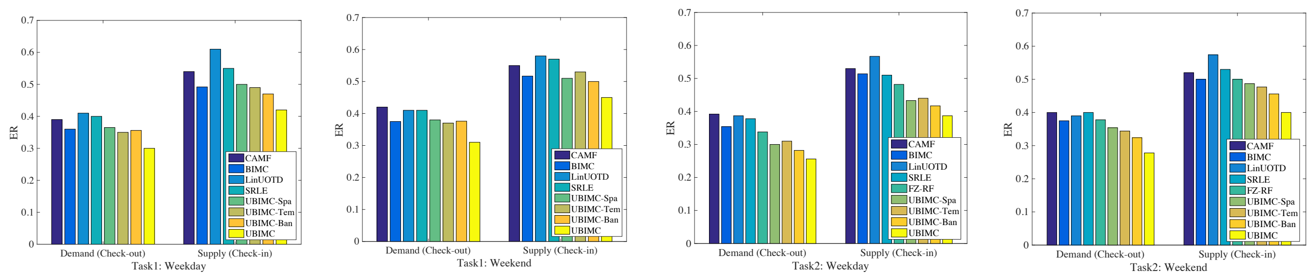



Fig. 10. ER comparison of various methods on the two tasks (50% training data)

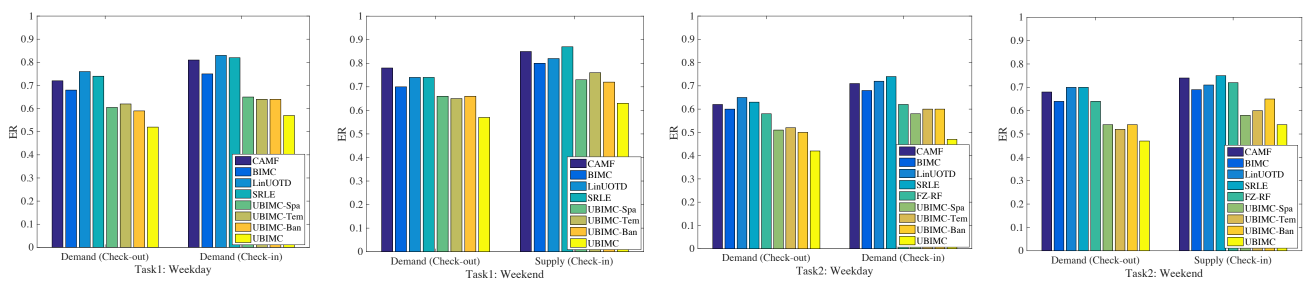



Fig. 11. ER comparison of various methods on the two tasks (30% training data)

#### 5.2.3 Results over RMSLE in Rush Hours

Table 3 shows the RMSLE comparison of various methods, and a lower RMSLE means a better performance. In this experiment, we focus on evaluating the performance of the methods on rush hours from 7 am to 9 am and from 17 pm to 19 pm in weekdays, since the bike usage demand in rush hours are much larger in other hours. We use 50% of the entire data for training and the remaining 50% for testing. We also perform the two tasks and for each task we evaluate the inference performance for bike demand and supply separately. One can see that UBIMC outperforms all the baselines in almost all the cases. For $Task1$, the average RMSLE achieved by UBIMC is 0.33, while the RMSLE of the state-of-the-art model FZ-RF is 0.405, which is a significant performance improvement reducing RMSLE by nearly 20%. It shows that UBIMC is much more effective than FZ-RF in bike usage demand inference for new regions. For $Task2$ which is a relatively easier task, UBIMC outperforms all the baselines in all the cases, and performance improvement is even more significant. On average, UBIMC reduces RMSLE by 46% compared to CAMF and 40% compared to SRLE. It demonstrates that simply factorizing the demand and supply matrices $\mathbf{F}^{ou}$ and $\mathbf{F}^{in}$ cannot achieve desirable performance. Similar to the results in Table 2, the results shown in Table 3 demonstrate BIMC outperforms CAMF, but is inferior to our proposal. LinUOTD performs worst among all the methods. Table 3 also verifies that UBIMC outperforms thee variations UBIMC-Spa, UBIMC-Tem, and UBIMC-Ban. It further demonstrates that the three components are all helpful to the task. Additionally, the three variations also outperforms other baselines, which also shows the power of the proposed model.

#### 5.2.4 Results over ER

Fig. 10 and Fig. 11 show the experimental result over the metric ER of different methods with 50% and 30% training data, respectively. The results are similar to previous results. UBIMC performs best among all the methods in all the cases. As shown in Fig. 10, for $Task1$ the average ER of UBIMC is around 0.35 on weekdays, and for $Task2$ the average ER of UBIMC is around 0.26 on weekdays, which verifies again that $Task1$ is harder to infer than $Task2$. As shown in Fig. 11, the average ER of UBIMC is around 0.57 for $Task1$ and 0.42 for $Task2$ on weekdays. Similar to the RMSLE comparison experiment, the inference performance for the supply (check-in) is inferior to the demand (check-out) due to the sparser data of the real supply data on the metric ER.

From the above experiments, one can conclude that 1) UBIMC is much more effective than previous bike demand prediction models in stationless bike-sharing systems, and 2) incorporating the spatial-temporal correlations and the balanced bike usage constraint can significantly improve the performance of the model.

TABLE 3
RMSLE comparison in rush hours

| $Task1$: $train = 50\%$, $Test = 50\%$ | | | | | |
|---|---|---|---|---|---|
| Methods | demand (check-out) | | supply (check-in) | | Average |
| | 7-9 am | 17-19 pm | 7-9 am | 17-19 pm | |
| CAMF | 0.615 | 0.562 | 0.574 | 0.547 | 0.575 |
| BIMC | 0.587 | 0.546 | 0.562 | 0.554 | 0.562 |
| LinUOTD | 0.617 | 0.608 | 0.614 | 0.587 | 0.606 |
| SRLE | 0.480 | 0.478 | 0.464 | 0.446 | 0.467 |
| FZ-RF | 0.420 | 0.348 | 0.434 | 0.416 | 0.405 |
| UBIMC-Spa | 0.367 | 0.356 | 0.335 | 0.346 | 0.351 |
| UBIMC-Tem | 0.412 | 0.387 | 0.367 | 0.356 | 0.381 |
| UBIMC-Ban | **0.320** | 0.338 | 0.344 | 0.366 | 0.342 |
| UBIMC | 0.327 | **0.322** | **0.325** | **0.345** | **0.330** |
| $Task2$: $train = 50\%$, $Test = 50\%$ | | | | | |
| Methods | demand (check-out) | | supply (check-in) | | Average |
| | 7-9 am | 17-19 pm | 7-9 am | 17-19 pm | |
| CAMF | 0.512 | 0.468 | 0.520 | 0.515 | 0.534 |
| BIMC | 0.518 | 0.502 | 0.492 | 0.474 | 0.496 |
| LinUOTD | 0.524 | 0.546 | 0.522 | 0.498 | 0.522 |
| SRLE | 0.446 | 0.454 | 0.513 | 0.476 | 0.47 |
| FZ-RF | - | - | - | - | - |
| UBIMC-Spa | 0.325 | 0.344 | 0.323 | 0.315 | 0.327 |
| UBIMC-Tem | 0.332 | 0.320 | 0.314 | 0.325 | 0.323 |
| UBIMC-Ban | 0.320 | 0.298 | 0.324 | 0.326 | 0.317 |
| UBIMC | **0.262** | **0.275** | **0.284** | **0.314** | **0.284** |

## 6 Related Work

As a convenient and green transportation mode, bike-sharing system has attracted increasing research interests since the first system was deployed in Europe in 1965 [19]. Current researches can be roughly summarized into the following three categories: *system planning* [14], [15], [16], [18], [25], [28], *system prediction* [5], [13], [20], [24], [27], [34], and *system operation* [8], [9], [10], [11], [12], [17], [30]. Recently, with the great success of deep learning techniques, deep learning models have been widely used to various spatial-temporal data mining tasks including crowd flow prediction [41], [44], [45], [49] and traffic flow prediction [46], [48]. Next we will review the related works.

Before setting up a bike-sharing system, the first task is to determine the number, capacity and locations of the bike stations. Luigi *et al.* [18] proposed a methodology for calculating the potential demand for bike use and the willingness to pay of future users. A location model for fixing the bike pick-up and drop-off stations is also proposed. Lin and Yang [16] addressed the strategic planning of public bike sharing system by considering the interests of both users and investors. Their model can determine the number and locations of bike stations, the network structure of bike paths connected between the stations, and the travel paths for users between each pair of origins and destinations. Chen *et al.* [15] proposed to leverage heterogeneous urban open data to address the bike station place problem. Carlos et al [28] proposed a GIS-based method to calculate the spatial distribution of the potential demand for bike trips, locate the bike stations using location-allocation models and determine the capacities of the bike stations.

Froehlich *et al.* [13] adopted a Bayesian network to predict station status based on the current time and current available dock number. Andreas *et al.* [20] made a short term prediction of the number of available bikes in stations via the analysis of cyclic mobility patterns. Li *et al.* [27] proposed a hybrid and hierarchical prediction model to predict the number of bikes that will be rent from/returned to each station cluster in the early future . Chen *et al.* [24] proposed a station cluster-level over demand prediction model in bike-sharing systems by considering the correlation among the stations. Liu *et al.* [5] tried to predict the station-level bike demand for bike system expansion. They proposed a bi-clustering model on POIs and stations by utilizing POI characteristics and geographical distances as features. Then they predict the Function-to-Function and Region-to-Region bike transitions by a Random Forest Regressor. Yoon et al [34] proposed a personal journey advisor application for helping people to navigate the city using the available bike-sharing system.

A primary task in system operation is to re-balance or re-allocate the bikes from time to time for the unbalanced bike usage. Chemla *et al.* [12] formulated this task as an optimization problem and proved it is NP-hard. They proposed a branch-and-cut algorithm for solving a relaxation of the problem. Forma *et al.* [11] further proposed a 3-step mathematical programming based heuristic for this problem. Liu *et al.* [30] provided an integer nonlinear programming formulation of multiple capacitated bike routing problem with the objective of minimizing the total travel distance. Some other works studied how to rebalance the bike demand with minimum operation cost through making corresponding reservation policies [10] and pricing mechanisms [8], [9]. Bao *et al.* [6] made the first attempt to plan bike lanes based on the massive Mobike trajectory data. Aeschbach et al [17] studied various methods to balance bike-sharing systems by actively engaging customers in the balancing process. They discovered that by appropriately sending “control signals” to customers requesting them to slightly change their intended journeys, bike-sharing systems can be balanced without using staffed trucks.

Recently, deep leaning models have enjoyed considerable success in various spatial-temporal data mining tasks due to their powerful hierarchical feature learning ability. A line of studies applied CNN to capture the spatial correlation by treating the entire city's traffic as images. Ma *et al.* [42] utilized CNN on images of traffic speed for the speed prediction problem. Zhang *et al.* [44], [45] proposed to use residual CNN on the images of traffic flow. These methods simply use CNN on the whole city and use all the regions for prediction. The major limitation of these method is that although they used historical traffic images in previous time slots for prediction, they did not explicitly model the temporal sequential dependency. Another line of research is combining CNN model and RNN model to capture both spatial and temporal correlations. Yao *et al.* [46] proposed a Spatial-Temporal Dynamic Network (STDN) model for road network based traffic prediction. Cheng *et al.* [47] proposed the DeepTransport model which combined CNN and RNN to capture the spatial-temporal traffic data within a transport network. All these models are basically designed for one-step prediction, which means they focus on predicting the traffic data in the next time slot like 20 minutes or half an hour. [48] was the first recent work that studied multi-step taxi passenger demand prediction. [48] proposed to use the attention-based neural network which combined encoder-decoder framework and ConvLSTM to predict the passenger pickup/dropoff demands for the mobility-on-demand services. All these works focused on predicting the future urban traffic or crowd flows based on a large number of observed historical data, which is quite different from the research purpose of this work.

## 7 Conclusion and Future Work

This paper studied the novel problem of inferring the bike usage demand in stationless bike-sharing systems and proposed a data-driven approach to effectively address it. We first utilized a relatively small number of the bike check-out and check-in data to identify the real bike demand and supply in some regions and time intervals for constructing the sparse demand and supply matrices. Then we tried to complete the two matrices by matrix factorization. To more accurately perform matrix completion, we also incorporated POIs, bike usage correlations among regions and intra-cluster balanced bike usage constraint into the model. Finally a usage balanced inductive matrix factorization model UBIMC was proposed. Experiment results on a large Mobike trip dataset in Beijing verified the effectiveness of UBIMC.

For the future work it would be interesting to further study the impact of some other external factors on the usage of bikes such as weather [33], population distribution in a city and the future urban planning [40]. For example, the bike usage demand might be significantly affected if a new subway station will be built in a region in the future. Taking such information into consideration may further improve the model performance. We are also interested in bike usage prediction based on the bike trip data [?]. Real time bike usage prediction is also essential for efficient system management especially for rebalancing the system. We will study whether the spatial-temporal correlation and the locally balanced bike usage constraint studied in this work are also important for bike traffic prediction.

## Acknowledgment

This work was supported by the NSFC (Grant No. 61602237), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20171420), and Hong Kong Scholar Program.

## References

[1] *https://en.wikipedia.org/wiki/Mobike*, July 2017.

[2] Donghyuk Shin, Suleyman Cetintas and Kuang-Chih Lee, *Tumblr Blog Recommendation with Boosted Inductive Matrix Completion*, In Proceedings of ACM Conference on Knowledge Discovery and Data Mining, 2015.

[3] Prateek Jain and Inderjit S. Dhillon, *Provable Inductive Matrix Completion*, CoRR, abs/1306.0626, 2013.

[4] Bei Chen, Fabio Pinelli, Mathieu Sinn, Adi Botea and Francesco Calabrese, *Uncertainty in urban mobility: predicting waiting times for shared bicycles and parking lots*, In Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems, 2013.

[5] Junming Liu, Leilei Sun, Qiao Li, Jingci Ming, Yanhua Li and Hui Xiong, *Functional Zone Based Hierarchical Demand Prediction For Bike System Expansion*, In Proceedings of 23rd ACM Conference on Knowledge Discovery and Data Mining, 2017.

[6] Jie Bao, Tianfu He, Sijie Ruan, Yanhua Li and Yu Zheng, *Planning Bike Lanes based on Sharing-Bikes' Trajectories*, In Proceedings of 23rd ACM Conference on Knowledge Discovery and Data Mining, 2017.

[7] Senzhang Wang, Xiaoming Zhang, Jianping Cao, Lifang He, Leon Stenneth, Philip S. Yu, Zhoujun Li and Zhiqiu Huang, *Computing Urban Traffic Congestions by Incorporating Sparse GPS Probe Data and Social Media Data*, ACM Transactions on Information Systems 35(4): 40:1-40:30 (2017).

[8] Ariel Waserhole and Vincent Jost, *Pricing in vehicle sharing systems: optimization in queuing networks with product forms*, EURO Journal on Transportation and Logistics, 5(3), pp 293–320, 2016.

[9] Adish Singla, Marco Santoni, Ga bor Barto k, Pratik Mukerji, Moritz Meenen and Andreas Krause, *Incentivizing Users for Balancing Bike Sharing Systems*, In Proceedings of The Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[10] Mor Kaspi, Tal Raviv and Michal Tzur, *Parking reservation policies in one-way vehicle sharing systems*, Transportation Research Part B, 62, pp 35–50, 2014.

[11] Iris A. Forma, Tal Raviv and Michal Tzur, *A 3-step math heuristic for the static repositioning problem in bike-sharing systems*, Transportation Research Part B, 71, pp 230–247, 2015.

[12] Daniel Chemla, Frédéric Meunier and Roberto Wolfler Calvo, *Bike sharing systems: Solving the static rebalancing problem*, Discrete Optimization, 10, pp 120–146, 2013.

[13] Jon Froehlich, Joachim Neumann and Nuria Oliver, *Sensing and Predicting the Pulse of the City through Shared Bicycling*, In Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence, 2009.

[14] Ines Frade and Anabela Ribeiro, *Bike-sharing stations: A maximal covering location approach*, Transportation Research Part A: Policy and Practice, 82, pp 216–227, 2015.

[15] Longbiao Chen, Daqing Zhang, Gang Pan, Xiaojuan Ma, Dingqi Yang, Kostadin Kushlev, Wangsheng Zhang, Shijian Li, *Bike sharing station leveraging heterogeneous urban open data*, In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2015.

[16] Jenn-Rong Lin and Ta-Hui Yang, *Strategic design of public bicycle sharing systems with service level constraints*, Transportation Research Part E: Logistics and Transportation Review, 47(2), pp 284–294, 2011.

[17] Philipp Aeschbach, Xiaojing Zhang and Angelos Georghiou, *Balancing bike sharing systems through customer cooperation - a case study on London's Barclays Cycle Hire*, In Proceedings of IEEE 54th Annual Conference on Decision and Control (CDC), 2015.

[18] Dell'Olio Luigi, Ibeas Angel and Moura Jose Luis, *Implementing bike-sharing systems*, the Institution of Civil Engineers: Municipal Engineer, pp 89–101, 2011.

[19] Runde Sache, *Readers Digest Deutschland (in German)*, pp 74–75, 2011.

[20] Kaltenbrunner Andreas, Meza Rodrigo, Grivolla Jens, Codina Joan and Banchs Rafael, *Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system*, Pervasive & Mobile Computing, 6(4), pp 455–466, 2010.

[21] Come Etienne and Latifa Oukhellou, *Model-Based Count Series Clustering for Bike Sharing System Usage Mining: A Case Study with the Velib' System of Paris*, ACM Transactions on Intelligent Systems and Technology, 5(3), pp 1-21, 2014.

[22] P. Vogel, T. Greiser and Mattfeld D.C., *Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns*, Procedia-Social and Behavioral Sciences, 20, pp 514–523, 2011.

[23] Gaode, *2017 Q2 the traffic analysis report of major cities in China*, Gaode Map and Tsinghua University & Daimler Sustainable Transportation Joint Research Centre, 2017.

[24] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Zhaohui Wu, Gang Pan, Thi-Mai-Trang Nguyen and Jérémie Jakubowicz, *Dynamic Cluster-based Over-demand Prediction in Bike Sharing Systems*, In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016.

[25] Nicolas Gast, Guillaume Massonnet and Daniel Reijsbergen, *Probabilistic Forecasts of Bike-Sharing Systems for Journey Planning*, In Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015.

[26] Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen and Thomos Moscibroda, *Mobility Modeling and Prediction in Bike-Sharing Systems*, In Proceedings of the 14th ACM International Conference on Mobile Systems, Applications, and Services, 2016.

[27] Yexin Li, Yu Zheng, Huichu Zhang and Lei Chen, *Traffic Prediction in a Bike-Sharing System*, In Proceedings of 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2015.

[28] Juan Carlos, Garca-Palomares, Javier Gutirrez and Marta Latorre, *Optimizing the location of stations in bike-sharing programs: A GIS approach*, Applied Geography (35)1-2: 235-246, Nuvember 2012.

[29] Panagiotis Angeloudis, Jun Hu and Michael G. H. Bell, *A strategic repositioning algorithm for bicycle-sharing chemes*, Transportmetrica A: Transport Science, 10, pp 759–774, 2014.

[30] Junming Liu, Leilei Sun, Weiwei Chen and Hui Xiong, *Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization*, In Proceedings of 22nd ACM Conference on Knowledge Discovery and Data Mining, 2016.

[31] Vincent W. Zheng, Yu Zheng, Xing Xie and Qiang Yang, *Collaborative location and activity recommendations with GPS history data*, In Proceedings of the 19th international conference on World wide web, 2010.

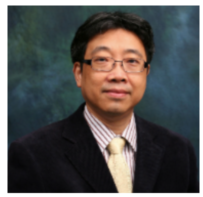
[32] Jingbo Shang, Yu Zheng, Wenzhu Tong, Eric Chang and Yong Yu, *Inferring gas consumption and pollution emission of vehicles throughout a city*, In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014.

[33] Kyle Gebhart and Robert B. Noland, *The impact of weather conditions on bikeshare trips in Washington, DC*, Transportation (2014) 41: 1205. https://doi.org/10.1007/s11116-014-9540-7.

[34] Ji Won Yoon, Fabio Pinelli and Francesco Calabrese, *Cityride: A Predictive Bike Sharing Journey Advisor*, In Proceedings of 13th International Conference on Mobile Data Management, 2012.
[35] Zhao Kang, Chong Peng and Qiang Cheng, *Top-N Recommender System via Matrix Completion*, In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[36] Ehsan Elhamifar, *High-Rank Matrix Completion and Clustering under Self-Expressive Models*, In Proceedings of 30th Conference on Advances in Neural Information Processing Systems, 2016.
[37] Cho-Jui Hsieh, Nagarajan Natarajan and Inderjit S. Dhillon, *PU Learning for Matrix Completion*, In Proceedings of the 32nd International Conference on Machine Learning, 2015.
[38] Junbo Zhang, Yu Zheng and Dekang Qi, *Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction*, In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[39] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye and Weifeng Lv, *The Simpler The Be er: A Unified Approach to Predicting Original Taxi Demands based on Large-Scale Online Platforms*, In Proceedings of of the 23rd ACM SIGKDD international conference on Knowledge discovery and data mining, 2017.
[40] Jeffrey Kenworthy, P.W.G. Newman and T.J. Lyons, *Urban Planning And Traffic Congestion*, May 2008, Urban Policy and Research 7(2):67-80, DOI10.1080/08111148908551389.
[41] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye and Zhenhui Li, *Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction*, In Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[42] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang and Yunpeng Wang, *Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction*, Sensors (2017) 17(4), 818:1-16.
[43] Senzhang Wang, Lifang He, Leon Stenneth, Philip S. Yu and Zhoujun Li, *Citywide traffic congestion estimation with social media*, In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2015.
[44] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiqing Zhang and Xiuwen Yi, *DNN-based prediction model for spatio-temporal data*, In Proceedings of The ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2016.
[45] Junbo Zhang, Yu Zheng and Dekang Qi, *Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction*, In Proceedings of AAAI Conference on Artificial Intelligence, 2017.
[46] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, Yanwei Yu and Zhenhui Li, *Modeling Spatial-Temporal Dynamics for Traffic Prediction*, arXiv preprint arXiv:1803.01254, 2018.
[47] Xingyi Cheng, Ruiqing Zhang, Jie Zhou and Wei Xu, *DeepTransport: Learning Spatial-Temporal Dependency for Traffic Condition Forecasting*, arXiv:1709.09585, 2017.
[48] Xian Zhou, Yanyan Shen, Yanmin Zhu and Linpeng Huang, *Predicting Multi-step Citywide Passenger Demands Using Attention-based Neural Networks*, In Proceedings of ACM International Conference on Web Search and Data Mining, 2018.
[49] Bowen Du, Hao Peng, Senzhang Wang, Md Zakirul Alam Bhuiyan, Lihong Wang, Qiran Gong, Lin Liu and Jing Li, *Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction*, IEEE Transactions on Intelligent Transportation Systems (2019), DOI: 10.1109/TITS.2019.2900481, pp 1-14.
[50] Hao Chen, Senzhang Wang, Zengde Deng, Xiaoming Zhang and Zhoujun Li, *FGST: Fine-Grained Spatial-Temporal Based Regression for Stationless Bike Traffic Prediction*, In Proceedings of The Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2019.

**Senzhang Wang** received the B.Sc. and Ph.D. degree in Southeast University, Nanjing, China in 2009 and Beihang University, Beijing, China in 2016 respectively. He is currently an associate processor at College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, and also a "Hong Kong Scholar" Postdoc Fellow at Department of Computing, The Hong Kong Polytechnic University. His main research focus is on data mining, social computing, and urban computing. He has published more than 60 referred conference and journal papers.

**Hao Chen** is currently a second year Ph.D. student in the school of computer science and engineering at Beihang University. He received B.Sc. degree in 2015 from Nankai University. His main research focus on sequential data mining.



**Jiannong Cao** received the B.Sc. degree in computer science from Nanjing University, China, in 1982, and the M.Sc. and Ph.D. degrees in computer science from Washington State University, USA, in 1986 and 1990 respectively. He is currently a Chair Professor of Department of Computing at The Hong Kong Polytechnic University, Hong Kong. His research interests include parallel and distributed computing, wireless networks and mobile computing, big data and cloud computing, pervasive computing, and fault tolerant computing. He has published over 500 papers in major international journals and conference proceedings. He is a fellow of IEEE, a distinguished member of ACM, a senior member of China Computer Federation (CCF).

**Jiawei Zhang** received the B.Sc. degree in computer science from Nanjing University, and the Ph.D. degree in computer science from the University of Illinois at Chicago. He is an assistant professor in the Department of Computer Science at Florida State University. His principal research interest is on data mining, deep learning and machine learning, with a focus on developing general methodologies for information fusion and mining, which will be shown to work well for a diverse set of applications. He has published more than 50 referred papers in books, journals, and major conferences.

**Philip S. Yu** received the B.S. Degree in E.E. from National Taiwan University, the M.S. and Ph.D. degrees in E.E. from Stanford University, and the M.B.A. degree from New York University. He is a Distinguished Professor in Computer Science at the University of Illinois at Chicago and also holds the Wexler Chair in Information Technology. Before joining UIC, Dr. Yu was with IBM, where he was manager of the Software Tools and Techniques department at the Watson Research Center. His research interest is on big data, including data mining, data stream, database and privacy. He has published more than 1,100 papers in refereed journals and conferences. He holds or has applied for more than 300 US patents. Dr. Yu is a Fellow of the ACM and the IEEE. Dr. Yu is the recipient of ACM SIGKDD 2016 Innovation Award for his influential research and scientific contributions on mining, fusion and anonymization of big data, the IEEE Computer Societys 2013 Technical Achievement Award for pioneering and fundamentally innovative contributions to the scalable indexing, querying, searching, mining and anonymization of big data, and the Research Contributions Award from IEEE Intl. Conference on Data Mining (ICDM) in 2003 for his pioneering contributions to the field of data mining. He also received the ICDM 2013 10-year Highest-Impact Paper Award, and the EDBT Test of Time Award (2014). He was the Editor-in-Chiefs of ACM Transactions on Knowledge Discovery from Data (2011-2017) and IEEE Transactions on Knowledge and Data Engineering (2001-2004).