HETEROGENEOUS MOBILE DATA ANALYTICS FOR SMART LIVING

by

YANCHI LIU

A dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

written under the direction of

Dr. Hui Xiong

and approved by

_____

_____

_____

_____

Newark, New Jersey

May 2019

ABSTRACT OF THE DISSERTATION

Heterogeneous Mobile Data Analytics for Smart Living

By YANCHI LIU

Dissertation Director: Dr. Hui Xiong

With the development of mobile, sensing, and positioning technologies, large-scale urban geographic data and human mobility data have been accumulated recently. The availability of heterogeneous mobile data and the emergence of big data technology provide unparalleled opportunities on understanding user behaviors and enabling smart living, e.g., developing livable and vibrant communities, improving energy efficiency in transportation, and enhancing urban planning. To this end, the objective of this dissertation is to exploit heterogeneous mobile data for developing data-driven solutions to enable smart living.

Along this line, we first provide a data driven solution to recommend Points-of-Interest (POIs) for the purpose of improving people's experiences for urban living. Existing approaches for POI recommendation have been mainly focused on exploiting the information about user preferences, social influence, and geographical influence. However, these approaches cannot handle the scenario where users are expecting to have POI recommendation for a specific time period. To this end, we propose a unified recommender system to integrate the user interests and their evolving sequential preferences with temporal interval assessment. As a result, the proposed system can make recommendations dynamically for a specific time period and the traditional POI recommender system can be treated as the special case of the proposed system

by setting this time period long enough.

In addition, we study the Point-of-Interest (POI) demand modeling issue in urban regions for urban planning. While some efforts have been made for the demand analysis of some specific POI categories, such as restaurants, it lacks systematic means to support POI demand modeling. To this end, we develop a systematic POI demand modeling framework, named Region POI Demand Identification (RPDI), to model POI demands by exploiting the daily needs of people identified from their large-scale mobility data.

Finally, we investigate intelligent bus routing to facilitate urban traveling. Optimal planning for public transportation is one of the keys helping to bring a sustainable development and a better quality of life in urban areas. Compared to private transportation, public transportation uses road space more efficiently and produces fewer accidents and emissions. However, in many cities people prefer to take private transportation other than public transportation due to the inconvenience of public transportation services. We focus on the identification and optimization of flawed region pairs with problematic bus routing to improve utilization efficiency of public transportation services, according to people's real demand for public transportation.

ACKNOWLEDGEMENTS

During my graduate study I have received support and encouragement from a great number of people, without whom this work would not have been possible.

I would like first to express my deepest appreciation to my advisor, Professor Hui Xiong, for his guidance, insights, patience, and immense knowledge, which are necessary to survive and thrive the graduate school and the beyond. I thank him for generously giving me motivation, support, time, understandings, assistance, opportunities, and friendship; for teaching me how to practice the art and science in data mining; and for guidance on academic research and career development.

I also sincerely thank my other dissertation committee members: Prof. Guiling Wang, Prof. Xiaodong Lin, and Prof. Thomas Lidbetter, for their great support and invaluable advice. All of them not only provide constructive suggestions and comments on my research work and this dissertation, but also offer numerous support and help in my career development.

I also want to thank my colleagues and collaborators for their continued support. I would like to thank group members at the Center of Data Mining and Business Analytics: Wenjun Zhou, Yong Ge, Zhongmou Li, Keli Xiao, Chuanren Liu, Bin Liu, Yanjie Fu, Zijun Yao, Meng Qu, Constantine Alexander Vitt, Jingyuan Yang, Hao Zhong, Farid Razzak, Qingxin Meng, Junming Liu, Mingfei Teng, Can Chen,

Pengpeng Zhao, Fuzhen Zhuang, Jingjing Gu, Xiaoyi Deng, Haochao Ying, Renjun Hu, Qi Liu, Hengshu Zhu, Chunyu Luo, Xinjiang Lu, Yanhong Guo, Yayao Zuo, Jie Liu, Huang Xu, Leilei Sun, Bowen Du, Hongting Niu, Ling Yin, Xiaolin Li, Bo Jin, Aimin Feng, Chang Tan, Xue Bai, Liyang Tang, Yuanchun Zhou, Tong Xu, Guannan Liu, for their help and valuable suggestions. Also I would like to thank Dr. Haifeng Chen, Dr. Tan Yan, Dr. Jin Cao, Dr. Xiaojing Wang, Dr. Xing Xie, and Dr. Yu Zheng, with whom I had great time during my internships.

I would like to acknowledge the Department of Management Science and Information Systems (MSIS) and Center of Data Mining and Business Analytics for supplying me with the best imaginable equipment and facilities that helped me to accomplish much of this work.

Last but not least, I would like to thank my family and friends for their love, support and understanding. This dissertation would not have been possible without their encouragement and help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

With the development of mobile, sensing, and positioning technologies, we have witnessed the huge and rapid explosion in mobile devices and their penetration into every component of everyday life. As a result, large-scale urban geographic data, human mobility data, and human behavior data have been accumulated recently. The availability of heterogeneous mobile data and the emergence of big data technology provide unparalleled opportunities on understanding user behaviors and enabling smart living, e.g., developing livable and vibrant communities, improving energy efficiency in transportation, and enhancing urban planning. Following this line, this dissertation research has been motivated by the following scenarios in real life: (i) recommending the right Points-of-Interest (POIs) to the right users for exploring new places with higher satisfactory level; (ii) modeling POI demand of each region in urban areas for urban planning and site selection; (iii) optimal planning for public transportation to bring a sustainable development and a better quality of life in urban areas. To this end, the objective of this dissertation is to exploit heterogeneous mobile data for developing data-driven solutions to enable smart living.

1.1 Research Contributions

In this dissertation, we aim to address the challenges of mobile data modeling and mining for smart living, from both theoretical and practical perspectives. More specifically, the focus of this research is utilizing big and heterogeneous data generated by diverse sources (e.g., sensors, devices, vehicles, stores, human) to address various challenges in business and city development (e.g., site selection and customer targeting), to discover and understand the relationship among residents, businesses, and urban environment, and to ultimately assist intelligent decision-making. In this dissertation, we identify several unique challenges for smart living in mobile environments, and then introduce how we use advanced data mining techniques to address them.

- **Point-of-Interest recommendation with temporal interval assessment.** We first provide a data driven solution to recommend Points-of-Interest (POIs) for the purpose of improving people's experiences for urban living. Existing approaches for POI recommendation have been mainly focused on exploiting the information about user preferences, social influence, and geographical influence. However, these approaches cannot handle the scenario where users are expecting to have POI recommendation for a specific time period. To this end, we propose a unified recommender system to integrate the user interests and their evolving sequential preferences with temporal interval assessment. As a result, the proposed system can make recommendations dynamically for a specific time period and the traditional POI recommender system can be treated as the special case of the proposed system by setting this time period long

enough. Specifically, to quantify users' sequential preferences, we consider the distributions of the temporal intervals between dependent POIs in the historical check-in sequences. Then, to estimate the distributions with only sparse observations, we develop the low-rank graph construction model, which identifies a set of bi-weighted graph bases so as to learn the static user preferences and the dynamic sequential preferences in a coherent way. Finally, we evaluate the proposed approach using real-world data sets from several location-based social networks (LBSNs). The experimental results show that our method outperforms the state-of-the-art approaches for POI recommendation in terms of various metrics, such as F-measure and NDCG, with a significant margin.

- **Point-of-Interest demand modeling with human mobility patterns.** In addition, we study the Point-of-Interest (POI) demand modeling issue in urban regions for urban planning. While some efforts have been made for the demand analysis of some specific POI categories, such as restaurants, it lacks systematic means to support POI demand modeling. To this end, we develop a systematic POI demand modeling framework, named Region POI Demand Identification (RPDI), to model POI demands by exploiting the daily needs of people identified from their large-scale mobility data. Specifically, we first partition the urban space into spatially differentiated neighborhood regions formed by many small local communities. Then, the daily activity patterns of people traveling in the city will be extracted from human mobility data. Since the trip activities, even aggregated, are sparse and insufficient to directly identify the

POI demands, especially for underdeveloped regions, we develop a latent factor model that integrates human mobility data, POI profiles, and demographic data to robustly model the POI demand of urban regions in a holistic way. In this model, POI preferences and supplies are used together with demographic features to estimate the POI demands simultaneously for all the urban regions interconnected in the city. Moreover, we also design efficient algorithms to optimize the latent model for large-scale data. Finally, experimental results on real-world data in New York City (NYC) show that our method is effective for identifying POI demands for different regions.

- **Exploiting heterogeneous human mobility patterns for intelligent bus routing.** Finally, we investigate intelligent bus routing to facilitate urban traveling. Optimal planning for public transportation is one of the keys helping to bring a sustainable development and a better quality of life in urban areas. Compared to private transportation, public transportation uses road space more efficiently and produces fewer accidents and emissions. However, in many cities people prefer to take private transportation other than public transportation due to the inconvenience of public transportation services. We focus on the identification and optimization of flawed region pairs with problematic bus routing to improve utilization efficiency of public transportation services, according to people's real demand for public transportation. To this end, we first provide an integrated mobility pattern analysis between the location traces of taxicabs and the mobility records in bus transactions. Based on the mobility

patterns, we propose a localized transportation mode choice model, with which we can dynamically predict the bus travel demand for different bus routing by taking into account both bus and taxi travel demands. This model is then used for bus routing optimization which aims to convert as many people from private transportation to public transportation as possible given budget constraints on the bus route modification. We also leverage the model to identify region pairs with flawed bus routes, which are effectively optimized using our approach. To validate the effectiveness of the proposed methods, extensive studies are performed on real world data collected in Beijing which contains 19 million taxi trips and 10 million bus trips.

CHAPTER 2

POI RECOMMENDATION WITH TEMPORAL INTERVAL ASSESSMENT

2.1 Introduction

The successful development of location-aware services, such as location-based social networks (LBSNs), has changed people's lives. For example, Foursquare[1] reported 8 billion check-ins at 65 million point-of-interests (POIs) by over 55 million users as of December, 2015. Even in one local area, there are often multiple competing POIs with similar utilities, and individual users are not capable to make fully informed choices. Based on collective intelligence, recommending the right POIs to the right users thus becomes beneficial for users exploring new places with higher satisfactory level. The POI recommendations are also essential for service providers to improve service quality and attract new customers.

First of all, some POIs need more time to plan than others due to the capacity and budget issues. Let us consider to recommend a visit to a museum, such type of POI recommendation is preferred to be provided several days in advance so as to allow the user to be more prepared (e.g., learning the background of the museum). More importantly, the users' needs and preferences often vary from time to time, and we need to capture the users' dynamic needs and evolving preferences to deliver the

[1]https://foursquare.com/

right recommendations at the right time. For instance, if one visits a museum in one day, he/she might be more interested in going for shopping instead of visiting another museum in the following day. The *right POI (e.g., a shopping center) for the coming trip* or the *right time for the next visit to a museum* could be inferred from the historical check-in records. In summary, it is critical to investigate how to recommend the right POIs for a specific time period by learning the users' evolving sequential preferences from their historical check-in records.

In the literature, various methods have been proposed for POI recommendation in recent years (Cheng, Yang, King, & Lyu, 2012; Sang, Mei, Sun, Xu, & Li, 2012; B. Liu & Xiong, 2013; B. Liu, Xiong, Papadimitriou, Fu, & Yao, 2015). Also, there are studies taking temporal factors into consideration for the purpose of improving the algorithm efficiency and/or effectiveness. For example, Yuan et al. (Q. Yuan, Cong, Ma, Sun, & Thalmann, 2013) and Gao et al. (Gao, Tang, Hu, & Liu, 2013) proposed to separate each day into different time slots and learn the users' preferences for each slot for time-aware POI recommendations. However, these methods do not consider the temporal relationship between the related check-ins by considering all check-ins as 'a bag of words'. Thus, they may recommend *museums in the morning* and *bars at night*, but the same recommendations will be provided for the same time period on different days. Indeed, these methods cannot capture the evolving changes of user preferences. The assumption of unchanging user preferences across several days may not hold due to the temporal relationship between dependent POI check-ins. More recently, the importance of the sequential relationships/patterns hidden in the historical check-in sequences has been realized for 'next' POI recommendations

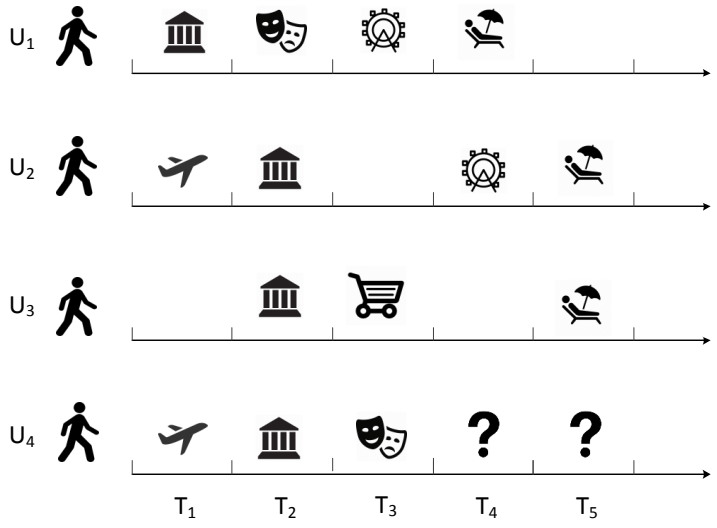


Figure 2.1. Examples of users check-in sequences

(Z. Chen, Shen, & Zhou, 2011; Cheng, Yang, Lyu, & King, 2013; Feng et al., 2015; Zhang & Chow, 2015). In particular, given the current check-in POI, the 'next' POI recommendation predicts the next interested POI which is most likely to be visited. However, these studies are unable to recommend POI for a specific time period due to the lack of modeling temporal interval information in their methods.

To better motivate this work, let us consider the example in Figure 2.1. For illustration purpose, we assume the time unit is day, i.e. T_1 is the first day and T_5 is the fifth day in the records of each user. Given users' check-ins to airport, museum, theater, shopping mall, amusement park and beach, what POIs should we recommend to user U_4 at time T_4 and T_5? Traditional POI recommender systems may recommend beach for both time periods since beach appears most frequently with museum and theater, and the existing 'next' POI recommender systems will recommend amusement park for both since amusement park appears mostly after

theater. However, intuitively, we can see amusement park should be more likely to be recommended for time T_4, and beach for T_5 because users went to amusement park mostly two days after going to museum according to the history records, and followed by going to beach. The possible reason could be that, after spending one day in museum, one would prefer to do something else to refresh the mind, e.g., by watching a show. Then, he/she would do something fun, e.g., going to amusement park, followed by relaxing on beach.

Indeed, in this chapter, we investigate how to do POI recommendations for a specific time period in LBSNs by capturing users' evolving sequential preferences from their historical check-in records. This task is much harder than the traditional POI recommendation due to the following challenges. First, it is necessary to model the sequential check-in patterns with temporal intervals between dependent POIs. Second, the sequential check-in data is very sparse. Only limited observations are available for estimating the distributions of the temporal intervals between dependent POIs in the historical check-in sequences. Finally, the third-party data, such as taxi GPS traces, are usually helpful for understanding the human mobility behaviors as well as their evolving preferences. The challenge is how to integrate these heterogeneous data sources to improve the performances of POI recommendation.

To address these challenges, we propose a unified recommender system, named 'Where and When to gO' (WWO), to integrate the static user interests and evolving sequential preferences with temporal interval assessment (Y. Liu, Liu, Liu, Qu, & Xiong, 2016). Specifically, given POI sequences consisting of check-in POIs ordered by check-in time, we first assess the temporal intervals between POIs from check-in

sequences of each user as a POI-POI transition matrix, where each item is a set of observed intervals for a POI-POI pair. Then, we develop a bi-weighted low-rank graph construction model to learn individual user's behavioral preferences by identifying a set of common graph bases. The graph is bi-weighted so that the static user interests used by the traditional recommender systems are simultaneously learned with their evolving sequential preferences.[2] Finally, the main contributions of this chapter can be summarized as follows:

- A new recommender system, named WWO, is developed for providing POI recommendations for a specific time period. WWO exploits user check-in sequential patterns with temporal interval assessment based on all historical user check-ins. Moreover, WWO is able to capture both static user interests and their evolving sequential preferences for POI recommendations.
- A bi-weighted low-rank graph construction model is proposed to estimate the distributions with only sparse observations. The model helps to identify a set of bi-weighted graph bases, which in turn can be leveraged for learning the user interests and their sequential preferences in a coherent way.
- The proposed method is flexible to utilize additional third-party information. Specifically, we propose mobile regularization methods to integrate heterogeneous human mobility data, such as taxi GPS traces, into our recommendation model to enhance the performance. The proposed regularizations are based on graph-based similarities.

[2]In this way, the traditional recommendation methods can be treated as the special cases of our approach.

- The WWO recommender system has been evaluated on large-scale real-world data for POI recommendations. The experimental results show that our method outperforms state-of-the-art methods in terms of multiple metrics such as F-measure and Normalized Discounted Cumulative Gain (NDCG).

2.2 The WWO Recommender System

In this section, we introduce our WWO (Where and When to gO) recommender system. Assume that we have M POIs denoted by the set $\mathcal{P}$ and N users. For simplicity, we let $\mathcal{P} = \{1, 2, \cdots, M\}$, i.e., we use integers to represent the POIs. For the n-th user, $n = 1, 2, \cdots, N$, we have his/her check-in records represented as a sequence of check-in events $s^n = (s_1^n, s_2^n, \cdots, s_{L_n}^n)$ with the length L_n. Each check-in event s_l^n, $l = 1, 2, \cdots, L_n$, is a tuple $s_l^n = (p_l^n, t_l^n)$ where $p_l^n \in \mathcal{P}$ is the l-th POI in the check-in records ordered by the corresponding event time t_l^n. Therefore, we have $t_{l'}^n \geq t_l^n$ when $l' > l$. Table 2.1 lists some notations used in this chapter.

Given the historical check-in records of all the users $\mathbf{S} = \{s^n | n = 1, 2, \cdots, N\}$ and a future time t, we predict the n-th user's possible check-in, e.g. $p \in \mathcal{P}$, based on not only the user interest on p but also the temporal dependency between the prediction (p, t) and the historical records $s_l^n = (p_l^n, t_l^n)$.

Formally, our idea of WWO recommendations is to maximize the likelihood of check-in $p \in \mathcal{P}$ at time t, $\Pr(p, t|s^n)$, which is computed as:

$$\Pr(p, t|s^n) \propto f^n(p) \cdot g^n(t|p). \tag{2.1}$$

Table 2.1. Descriptions of mathematical notations.

Notation	**Description**
M, N	The number of POIs and users, respectively.
$\mathcal{P} = \{1, 2, \cdots, M\}$	The set of POIs.
$s^n = (s_1^n, s_2^n, \cdots, s_{L_n}^n)$	The check-in sequence of n-th user.
$s_l^n = (p_l^n, t_l^n)$	The l-th check-in event with POI $p_l^n \in \mathcal{P}$ and time t_l^n.
d_{ij}^n	The observed temporal intervals from POI i to POI j in the sequence s^n.
$\mathcal{N}(\mu_{ij}^n, \sigma_\mu^2)$	The distribution of d_{ij}^n.
c_p^n	The observed visiting frequency to POI p in the sequence s^n.
Poisson(ν_p^n)	The distribution of c_p^n.
$\alpha \in \mathbb{R}^{N \times K}$	The approximating coefficients in low-rank graph construction.
$\beta^k \in \mathbb{R}^{M \times M}$	The edge weight matrix of the k-th graph basis, for $k = 1, 2, \cdots, K$.
$\gamma \in \mathbb{R}^{K \times M}$	The node weights of all graph bases.
$\mathcal{N}(0, \sigma_\alpha^2), \mathcal{N}(0, \sigma_\beta^2), \Gamma(\eta, \theta)$	The priors of α, β, and γ, respectively.

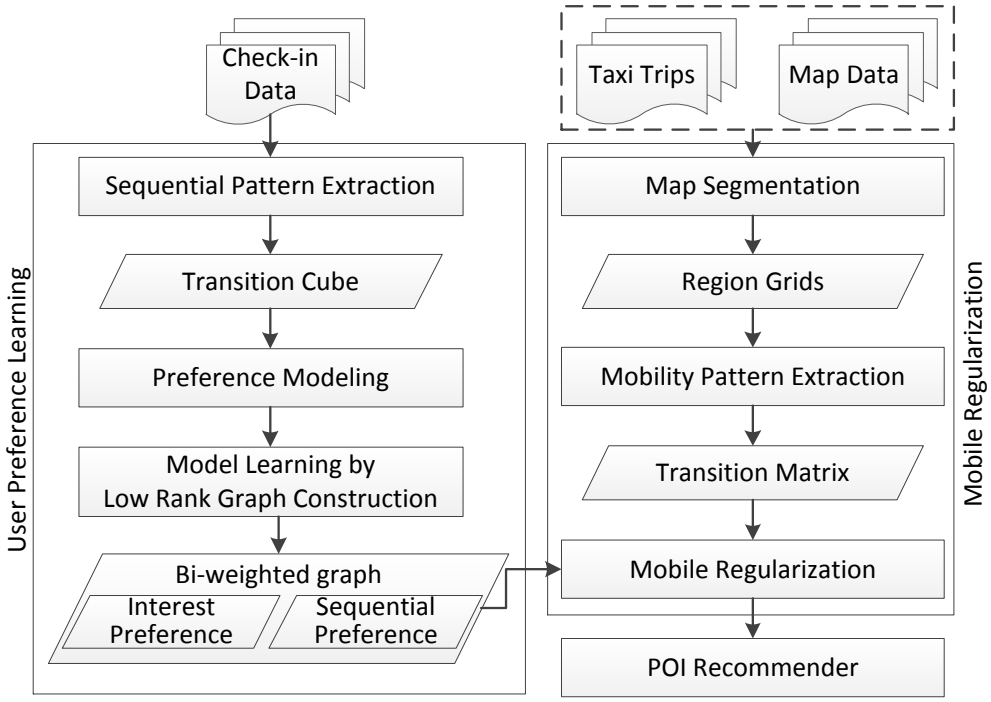


Figure 2.2. The WWO recommendation framework.

The term $f^n(p)$ computes the user's *interest* on p, which is the focus of traditional POI recommender systems. The term $g^n(t|p)$ computes the user's *evolving sequential preference*, which is the likelihood of the predicted check-in event (p, t) at time t. In other words, we aim at recommending the right POI at the right time to the right user. As such, we distribute the ongoing recommendations over time based on the historical temporal interval patterns so as to make the recommendation less disturbing and more favorable. In the following, we provide the details about computing these two terms with probability density functions.

Figure 2.2 shows the framework of our WWO recommender system. On one hand, we construct a POI transition cube by extracting sequential patterns from user check-in sequences, then the user interests and their evolving sequential preferences are

modeled and learned simultaneously with a low-rank graph construction model. On the other hand, we map taxi trips and POIs to region grids of a city, then the region transition and further POI transition matrix are constructed. Finally, by regularizing this taxi trip based POI transition matrix to the previous bi-weighted POI transition graphs, we have our WWO POI recommender system.

2.2.1 Sequential Pattern Extraction

To extract the sequential patterns from historical check-in records, we first define:

Definition 1 (Observed Temporal Intervals) *The observed temporal intervals from i to j in s^n are represented by a set:*

$$d_{ij}^n = \{ \min_{\substack{l'>l \\ s_{l'}^n = j}} t_{l'}^n - t_l^n \mid l = 1, 2, \cdots, L_n, s_l^n = i \}. \tag{2.2}$$

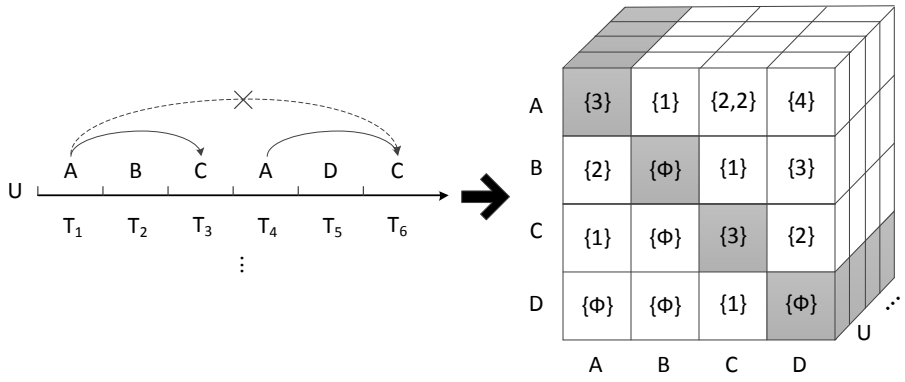


Figure 2.3. Example of observed temporal intervals. For the POI pair A to C, the temporal intervals with solid lines are counted, and we get $\{2, 2\}$ as d_{AC}^U.

Note that we use the min operator with several constraints in Equation 2.2 because both i and j can be visited by the same user multiple times. As a result, we consider only the soonest visit of j after each recent visit of i. With the temporal intervals

between POI pairs extracted for user n, a POI-POI transition interval matrix is constructed, with each item d_{ij}^n as a set of intervals from POI i to POI j of user n. By combining the transition matrices for all the users, a transition interval cube can be obtained, which unifies both user interests and their sequential preferences. To provide an intuitive understanding, an example is shown in Figure 2.3 about how to generate the transition matrix for one user from a sequence of user check-ins.

2.2.2 Preference Modeling

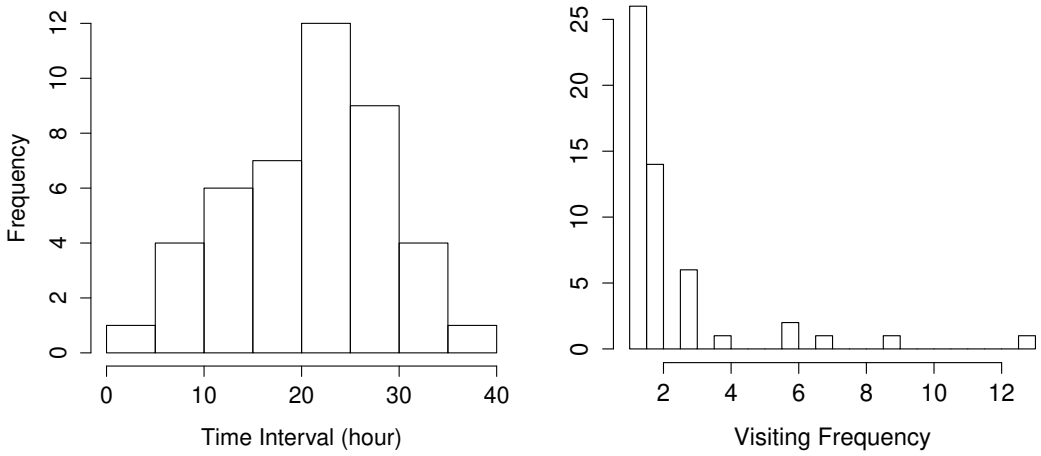
Sequential Preference: $g^n(t|p)$. We compute the sequential preference term as: $g^n(t|p) = \max_{1 \le l \le L_n} g_l^n(t|p)$, where $g_l^n(p,t)$ is the probability density of temporal interval $(t - t_l^n)$ from the check-in of p_l^n at time t_l^n to the check-in of p at time t. We use the max operator to identify the historical check-in event s_l^n having the strongest temporal dependency with the recommendation (p,t).

Generally, by letting $s_l^n = i$ and $p = j$, we want to estimate a distribution of the temporal intervals when the user would visit j after visiting i. In this chapter, we approximate this distribution using $\mathcal{N}(\mu_{ij}^n, \sigma_\mu^2)$. It follows that:

$$g_l^n(t|p) = \mathcal{N}(\delta|\mu_{ij}^n, \sigma_\mu^2) = \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp(-\frac{(\delta - \mu_{ij}^n)^2}{2\sigma_\mu^2}),$$

where $s_l^n = i, p = j$ and $t - t_l^n = \delta$. We have a histogram of one random POI pair in Figure 2.4 (a) which shapes like a Gaussian distribution and indicates our model specifications are appropriate.

Interest Preference: $f^n(p)$. To compute the interest preference term, we also



(a) Histogram of the temporal interval of one POI pair

(b) Histogram of the visiting frequency of one user

Figure 2.4. The distributions of sequential interval and visiting frequency.

fit the observed visiting frequency with a distribution, e.g., Gaussian. However, the visiting frequencies in our data is polarized as shown in Figure 2.4 (b), which does not shape like a Gaussian. As noticed by (Canny, 2004; B. Liu et al., 2015), the nature of Poisson distribution is more suitable and effective for modeling the skewed interest preference in terms of visiting counts, which provide implicit feedback for better POI recommendations. Therefore, we adopt the Poisson distribution $\text{Poisson}(\nu_p^n)$ to estimate $f^n(p)$:

$$f^n(p) = \text{Poisson}(c_p^n|\nu_p^n) = \frac{(\nu_p^n)^{c_p^n}}{c_p^n!}\exp(-\nu_p^n),$$

where c_p^n is the observed visiting frequency:

Definition 2 (Observed Visiting Frequency) *The observed visiting frequency to $p \in \mathcal{P}$ of the user in s^n is the cardinality of visiting records:*

$$c_p^n = |\{\ l \mid l = 1, 2, \cdots, L_n, s_l^n = p\ \}|. \tag{2.3}$$

2.2.3 Model Learning

Both the observed temporal intervals and visiting frequency can be very sparse. Therefore, 1) we have empty $d_{ij}^n = \emptyset$ or $|d_{ij}^n|$ is too small, for many pairs of POI (i, j); 2) we have $c_p^n = 0$ for many unobserved check-ins $p \in \mathcal{P}$. As a result, using these observations directly does not suffice to robustly estimate the temporal interval distributions and the user preferences. To address these issues, we develop the low-rank graph construction model in Section 2.3.

2.2.4 Recommendation

In practice, we aim to recommend the historically unvisited POIs within a time duration T instead of a single timestamp t. This can be achieved by ranking the options p according to the expected visiting frequency within the time duration $T = [t, t + \Delta]$:

$$\begin{aligned} &\mathbb{E}[c_p^n|\nu_p^n] \cdot \max_{1\le l\le L_n} \int_t^{t+\Delta} g_l^n(t|p)\mathrm{d}t \\ =&\nu_p^n \cdot \max_{1\le l\le L_n} \left(\Phi(\frac{t+\Delta-t_l^n-\mu_{s_l^n p}^n}{\sigma_\mu}) - \Phi(\frac{t-t_l^n-\mu_{s_l^n p}^n}{\sigma_\mu}) \right) \end{aligned} \tag{2.4}$$

where $c_p^n \sim \text{Poisson}(\nu_p^n)$ and the integral term can be easily computed with the cumulative distribution function (CDF) $\Phi(\cdot)$ of standard $\mathcal{N}(0, 1)$. The top ranked options can then be recommended to the user. It's worthy to note that, when $t \to t_{L_n}^n$ and $\Delta \to \infty$, the computation of Equation 2.4 will be dominated by the interest preference ν_p^n. In other words, our unified recommender system is a proper generalization of the existing approaches.

2.3 Low-Rank Graph Construction

Since the observations for specific users are sparse, we estimate the distribution parameters μ^n and ν^n for all users $n = 1, 2, \cdots, N$ collaboratively. To this end, we define the bi-weighted graph $G^n = \langle \mathcal{P}, E, \mu^n, \nu^n \rangle$ with $\mathcal{P}$ as graph nodes, where each graph node $p \in \mathcal{P}$ is weighted by ν_p^n and each graph edge $(i, j) \in E$ is weighted by μ_{ij}^n. In other words, the bi-weighted graph G^n is a unified representation of two aspects in the POI check-in behaviors of the n-th user: 1) The node weight vector $\nu^n \in \mathbb{R}^M$ characterizes the distribution of user interests. 2) The edge weight matrix $\mu^n \in \mathbb{R}^{M \times M}$ characterizes the distribution of the temporal intervals between the check-in events.

With this definition, we estimate the distribution parameters by collaboratively constructing the bi-weighted graphs for all the users. Specifically, we assume there are K graph bases B^k for $k = 1, 2, \cdots, K$ and each user-specific graph G^n can be approximated by:

$$G^n \leftarrow \sum_{k=1}^{K} \alpha_{nk} B^k.$$

Equivalently, by letting the edge weight matrix and the node weight vector of the graph basis B^k be $\beta^k \in \mathbb{R}^{M \times M}$ and $\gamma^k \in \mathbb{R}^M$, respectively, it follows that:

$$\mu^n = \sum_{k=1}^{K} \alpha_{nk} \beta^k, \tag{2.5}$$

$$\nu^n = \sum_{k=1}^{K} \alpha_{nk} \gamma^k. \tag{2.6}$$

Here, α_{nk} are the approximating coefficients. Note that, the number of graph bases,

K, can be deemed the rank of the graph construction, which is set to be much smaller than the number of user-specific graphs, i.e., $K \ll N$.

Now we can compute the probability density of the observations in d_{ij}^n and c_p^n:

$$\begin{aligned}
&\Pr(d, c|\alpha, \beta, \gamma, \sigma_\mu) \\
=&\prod_{n=1}^{N}\prod_{i=1}^{M}\prod_{j=1}^{M}\prod_{\delta\in d_{ij}^n}\frac{1}{\sqrt{2\pi}\sigma_\mu}\exp(-\frac{(\delta-\mu_{ij}^n)^2}{2\sigma_\mu^2}) \\
&\prod_{n=1}^{N}\prod_{p=1}^{M}\frac{(\nu_p^n)^{c_p^n}}{c_p^n!}\exp(-\nu_p^n)
\end{aligned}$$

Therefore the log-likelihood is:

$$\begin{aligned}
&\ln\Pr(d, c|\alpha, \beta, \gamma, \sigma_\mu) \\
=&-\frac{1}{2\sigma_\mu^2}\sum_{n=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\delta\in d_{ij}^n}(\delta-\mu_{ij}^n)^2 \\
&-\frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{t\in d_{ij}^n}\ln(2\pi\sigma_\mu^2) \\
&+\sum_{n=1}^{N}\sum_{p=1}^{M}(c_p^n\ln(\nu_p^n)-\nu_p^n)-\sum_{n=1}^{N}\sum_{p=1}^{M}\ln(c_p^n!)
\end{aligned}$$

Maximizing this log-likelihood is equivalent to minimize the following objective function:

$$\mathcal{L}(\alpha,\beta,\gamma)=\frac{1}{2\sigma_\mu^2}\sum_{n=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\delta\in d_{ij}^n}(\delta-\mu_{ij}^n)^2-\sum_{n=1}^{N}\sum_{p=1}^{M}(c_p^n\ln(\nu_p^n)-\nu_p^n),$$

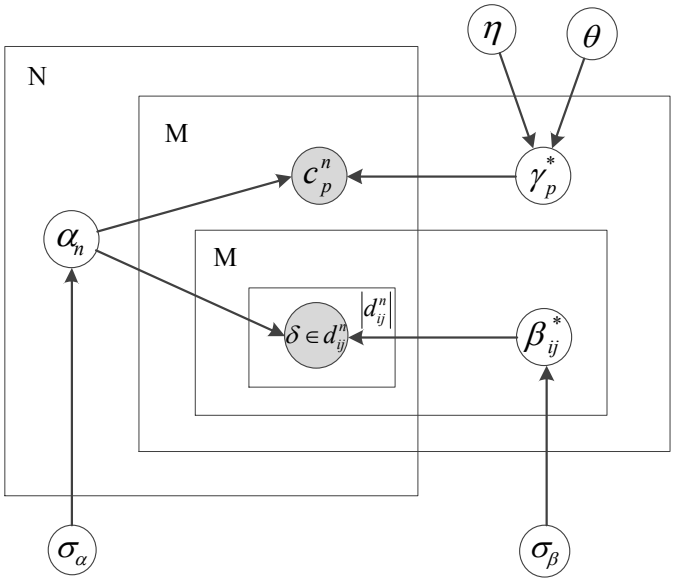


Figure 2.5. The probabilistic graphical model.

where the hyperparameter σ_μ is fixed. With the low-rank graph construction Equation 2.6, it follows that

$$\begin{aligned}\mathcal{L}(\alpha,\beta,\gamma) =& \frac{1}{2\sigma_\mu^2}\sum_{n=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\delta\in d_{ij}^n}(\delta-\sum_{k=1}^{K}\alpha_{nk}\beta_{ij}^k)^2 \\ & -\sum_{n=1}^{N}\sum_{p=1}^{M}(c_p^n\ln(\sum_{k=1}^{K}\alpha_{nk}\gamma_p^k)-\sum_{k=1}^{K}\alpha_{nk}\gamma_p^k).\end{aligned}$$

Indeed, the above factorization can be demonstrated as a probabilistic graphical model shown in Figure 2.5. Moreover, in order to reduce the generalization error (accuracy on unseen data), priors of the latent variables (α,β,γ) can be used as regularizations:

$$\alpha\sim\mathcal{N}(0,\sigma_\alpha^2),\ \ \beta\sim\mathcal{N}(0,\sigma_\beta^2),\ \ \gamma\sim\Gamma(\eta,\theta). \tag{2.7}$$

In particular, we use the Gamma distribution $\gamma\sim\Gamma(\eta,\theta)$ since it is the conjugate

one with Poisson. Then the joint log-likelihood to be minimized is as follows:

$$\begin{aligned}\mathcal{L}(\alpha,\beta,\gamma) =& \frac{1}{2\sigma_\mu^2}\sum_{n=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\delta\in d_{ij}^n}(\delta-\sum_{k=1}^{K}\alpha_{nk}\beta_{ij}^k)^2 \\ &- \sum_{n=1}^{N}\sum_{p=1}^{M}(c_p^n\ln(\sum_{k=1}^{K}\alpha_{nk}\gamma_p^k)-\sum_{k=1}^{K}\alpha_{nk}\gamma_p^k) \\ &+ \frac{1}{2\sigma_\alpha^2}\|\alpha\|_F^2+\frac{1}{2\sigma_\beta^2}\sum_{k=1}^{k}\|\beta^k\|_F^2 \\ &- \sum_{k=1}^{K}\sum_{p=1}^{M}((\eta-1)\ln\gamma_p^k-\theta\gamma_p^k), \end{aligned} \tag{2.8}$$

where $\alpha \in \mathbb{R}^{N\times K}$ is the matrix of approximating coefficients α_{nk}, $\beta^k \in \mathbb{R}^{M\times M}$ is the edge weight matrix of graph basis B^k, and $\gamma \in \mathbb{R}^{K\times M}$ is the matrix of node weights γ_p^k in all graph bases.

2.4 Mobile Regularization

Another research question that we aim to answer is whether the mobile transaction data can be used to improve the POI recommendations. To answer this question, we consider the taxi transactions in the city scale. Each transaction is a tuple $tr = (tr_o, tr_o^t, tr_d, tr_d^t)$ corresponding to a taxi trip with origin $tr_o \in \mathbb{R}$ and destination $tr_d \in \mathbb{R}$, leaving at time tr_o^t and arriving at time tr_d^t. Here $\mathbb{R}$ is the set of geographical grid regions (e.g., $100m \times 100m$) discretizing the whole geospatial space of the city.

With all the mobile transactions in the data base, we extend our learning model with mobile regularization. Specifically, we derive several regularizers to constrain the POI recommendations by exploiting the mobile connectivities between the POIs. As noted, however, the transitions are recorded with origin/destination regions in $\mathbb{R}$

instead of explicitly POIs in $\mathcal{P}$, and there are typically multiple POIs located in each region $r \in \mathbb{R}$. Therefore, we first need to estimate the mobile connectivities between POIs in $\mathcal{P}$ based on the mobile transactions between regions in $\mathbb{R}$.

2.4.1 Mobile Connectivity

The mobile connectivities between regions in $\mathbb{R}$ can be estimated as

$$W^{\mathbb{R}}_{r_1 r_2} = \Pr(r_2|r_1) = \frac{|\{tr|tr_o = r_1, tr_d = tr_2\}|}{|\{tr|tr_o = r_1\}|},$$

which is the transition probability from region r_1 to region r_2.

To estimate the mobile connectivities between POIs in $\mathcal{P}$, we compute:

$$\begin{aligned} W_{p_1 p_2} &= \Pr(p_2|p_1) \propto \Pr(r_1|p_1)\Pr(r_2|r_1)\Pr(p_2|r_2) \\ &= \Pr(r_1|p_1) W^{\mathbb{R}}_{r_1 r_2} \Pr(p_2|r_2) \end{aligned} \tag{2.9}$$

where the regions r_1 and r_2 contains p_1 and p_2, respectively. In this equation, the first and last two terms translating the mobile connectivity of regions to that of POIs can be estimated using the check-in records. First, for each region r and each POI $p \in r$, it is easy to estimate $Pr(p|r)$ by counting the check-in frequency of POI p within its region r. Second, with the Bayes' rule,

$$Pr(r|p) = \frac{\Pr(p|r)\Pr(r)}{\Pr(p)},$$

where $\Pr(r)$ and $\Pr(p)$ can be estimated with check-in records.

2.4.2 Mobile Resemblance

With the mobile connectivities between POIs in $\mathcal{P}$ as edge weight, we can construct the mobile connectivity graph $G = \langle \mathcal{P}, E, W \rangle$. However, this graph is not directly feasible to regularize the POI recommendations. First of all, the mobile connectivity is different with the temporal dependency estimated with temporal intervals. It is possible that two temporally dependent POIs are weakly connected in G. Moreover, the observed mobile connectivity may not reveal the real life as a whole, since data collected is sparse and biased. We need to derive robust measures to quantify the relationship between POIs with the observed graph.

To address these issues, our idea is to quantify the resemblance between the graph nodes (POIs) based on the graph topology. Then, for two POI pairs (i_1, j_1) and (i_2, j_2), if both (i_1, i_2) and (j_1, j_2) are positioned similarly in the graph G, we assume the optimal $\beta^k_{i_1 j_1}$ and $\beta^k_{i_2 j_2}$ in the graph basis B^k are also similar. Specifically, suppose the the mobile resemblance between the nodes of G is ρ, i.e., the ρ_{ij} quantify the similarity between POI i and j. Based on our assumption, we regularize the objective function of low-rank graph construction as follows:

$$\mathcal{J}(\alpha, \beta, \gamma) = \mathcal{L}(\alpha, \beta, \gamma) + \lambda \cdot \Omega(\beta),$$

where

$$\Omega(\beta) = \frac{1}{2} \sum_k \sum_{i_1 i_2} \sum_{j_1 j_2} (\beta^k_{i_1 j_1} - \beta^k_{i_2 j_2})^2 \rho_{i_1 i_2} \rho_{j_1 j_2}.$$

Note that, the degree of regularization λ can be tuned in a data-driven manner, as

we will show in Section 2.6.5.

In the following, we show several methods computing the mobile resemblance ρ based on the mobile connectivity W.

Cosine Similarity

The cosine similarity has been used to measure the closeness between two objects in high-dimensional space. With the mobile connectivity graph W, the cosine similarity between two POIs (i, j) in $\mathcal{P}$ is defined as

$$\rho_{ij} = \frac{1}{2}\left(\frac{\langle W_{i*}, W_{j*}\rangle}{|W_{i*}||W_{j*}|} + \frac{\langle W_{*i}, W_{*j}\rangle}{|W_{*i}||W_{*j}|}\right) \in [0, 1],$$

since W is asymmetric.

Pearson Correlation

The cosine similarity may not work well when comparing POIs with different popularities. To compute the unbiased similarities, we can use the Pearson correlation coefficient, defined as

$$\rho_{ij} = \frac{1}{4}\left(\frac{\langle \widehat{W}_{i*}, \widehat{W}_{j*}\rangle}{|\widehat{W}_{i*}||\widehat{W}_{j*}|} + \frac{\langle \widehat{W}_{*i}, \widehat{W}_{*j}\rangle}{|\widehat{W}_{*i}||\widehat{W}_{*j}|}\right) + \frac{1}{2} \in [0, 1]$$

where $\widehat{\cdot}$ is the centering operator subtracting the empirical mean.

SimRank

SimRank is an intuitive graph-theoretic model computing the similarities between the nodes of a graph (Jeh & Widom, 2002). With the asymmetric mobile connectivity

graph, we use the extension proposed in (Blondel, Gajardo, Heymans, Senellart, & Van Dooren, 2004), which iteratively computes:

$$\rho \leftarrow \frac{W\rho W' + W'\rho W}{\|W\rho W' + W'\rho W\|}.$$

The initial of the iterative process is set to be $\rho_{ij} = 1$, for all $i, j \in \mathcal{P}$.

The refined POI recommendation model with the above mobile regularization methods are WWO_Cos, WWO_Cor, WWO_Sim for Cosine Similarity, Pearson Correlation, and SimRank, respectively. While WWO represents the POI recommendation model with only check-in data.

2.5 Learning Algorithm

We use an alternative algorithm to solve α, β, and γ iteratively. Here, we let

$$D_{ij}^n = \sum_{\delta \in d_{ij}^n} \delta, C_{ij}^n = |d_{ij}^n|.$$

Also, at the beginning of each iteration, we compute μ and ν with Equation 2.6 with current solution (α, β, γ). Then we have

$$\begin{aligned}
\frac{\partial \mathcal{J}}{\partial \alpha_{nk}} &= \frac{\sum_{ij}(D_{ij}^n - C_{ij}^n \mu_{ij}^n)\beta_{ij}^k}{-\sigma_\mu^2} - \sum_p (\frac{c_p^n}{\nu_p^n} - 1)\gamma_p^k + \frac{\alpha_{nk}}{\sigma_\alpha^2} \\
\frac{\partial \mathcal{J}}{\partial \beta_{ij}^k} &= \frac{\sum_n (D_{ij}^n - C_{ij}^n \mu_{ij}^n)\alpha_{nk}}{-\sigma_\mu^2} + \frac{\beta_{ij}^k}{\sigma_\beta^2} + \lambda \cdot \frac{\partial \Omega}{\partial \beta_{ij}^k} \\
\frac{\partial \mathcal{J}}{\partial \gamma_p^k} &= -\sum_n (\frac{c_p^n}{\nu_p^n} - 1)\alpha_{nk} - (\frac{\eta - 1}{\gamma_p^k} - \theta)
\end{aligned}$$

Although it is straightforward to update one entry at a time with arbitray initlization for all entries in α, β, and γ, we improve the robustness of the alternative algorithm using: 1) the pre-clustering initialization based on spherical KMeans; and 2) the block-coordinate updating (Richtárik & Takáč, 2014).

2.5.1 Pre-clustering Initialization

Specifically, we first propose a novel clustering method to initialize the graph bases and the approximating coefficients. Note that, since the graph bases are used to approximate the distribution parameters μ and ν, the bases can be initialized with the representative observed distribution samples. Therefore, we aggregate and cluster the observed temporal intervals and visiting frequencies. To aggregate the observations, we construct the feature matrix $X \in \mathbb{R}^{N\times(M^2+M)}$, where each row X_n corresponds to one sequence s^n. The first M^2 entries of X_n encodes the average temporal interval:

$$X_{n,(i-1)M+j} = D^n_{ij}/C^n_{ij}.$$

The last M entries of X_n encodes the observed visiting frequency:

$$X_{n,M^2+p} = c^n_p.$$

Then, we apply the spherical KMeans algorithm to group the rows of X into K clusters, where K is the number of graph bases for the low-rank graph construction. Here, we use spherical KMeans with cosine similarity instead of Euclidean distance since the dimension of X is high.

Although the results of this pre-clustering may not be the true graph bases, we believe the clustering results can provide positive guidance for the graph basis learning algorithm. Thus, given the centroid $\widehat{X}_k$ of the k-th cluster in the pre-clustering solution, we initialize the graph basis B^k by reformatting the first M^2 entries of $\widehat{X}_k$ as the matrix β^k and the last M entries as the vector γ^k. Moreover, we initialize the approximating coefficients α with the cosine similarity between X_n and $\widehat{X}_k$ used by the spherical clustering:

$$\alpha_{nk} = \frac{\cos(X_n, \widehat{X}_k)}{\sum_{k'} \cos(X_n, \widehat{X}_{k'})}.$$

2.5.2 Block-coordinate Updating

With the above pre-clustering initialization, we then iteratively update rows in α where each row corresponds to one user. Then we update each graph basis B^k with the edge weight matrix β^k and the node weight vector γ^k. For each updating, the learning rate ϵ is determined to make sure the objective function $\mathcal{J}(\alpha, \beta, \gamma)$ is decreasing. The overall algorithm repeats until the value of $\mathcal{J}(\alpha, \beta, \gamma)$ keeps stable or the limitation on the number of iterations is reached. More details are shown in Algorithm 1.

Algorithm 1 Block-coordinate Optimization

Initialize α, β, and γ with pre-clustering.
repeat
  for $n = 1, \cdots, N$ **do**
    Update the row $\alpha_n \leftarrow \alpha_n - \epsilon \frac{\partial \mathcal{J}}{\partial \alpha_n}$.
  end for
  for $k = 1, \cdots, K$ **do**
    Update the base $\beta^k \leftarrow \beta^k - \epsilon \frac{\partial \mathcal{J}}{\partial \beta^k}$.
  end for
  for $k = 1, \cdots, K$ **do**
    Update the ror $\gamma_p^k \leftarrow \gamma_p^k - \epsilon \frac{\partial \mathcal{J}}{\partial \gamma_p^k}$.
  end for
until Convergence

2.6 Experimental Results

In this section, we evaluate the performances of the proposed WWO recommender system.

2.6.1 The Experimental Data

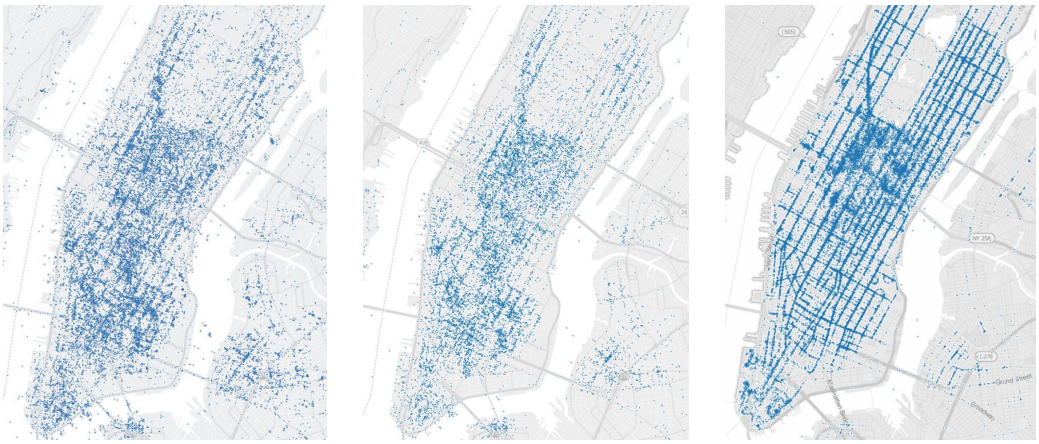
All the experiments were performed on real-world datasets including two LBSN datasets collected from Foursquare and Gowalla[3] as well as one taxi trip dataset from the New York City (NYC)[4]. Note that all the datasets shown in Figure 2.6 include the records from the same period (Year 2010).

The Foursquare dataset includes 2,932 users for 4,194 POIs with 121,678 check-in observations. Each user checked into 41.5 POIs on average. Each check-in contains the user ID, check-in time, venue ID and the venue's geo-coordinates.

The Gowalla dataset includes 1,313 users for 2,196 POIs with 45,410 check-in observations. Each user checked into 34.58 POIs on average. Note that, for Foursquare and Gowalla data, we remove those POIs with less than 10 users and remove users with less than 10 check-ins.

The taxi trip dataset includes 10 million transactions from about 50,000 taxis in NYC in 2010. Each transaction is associated with a pick-up point, a drop-off point, timestamp, fare, and number of passengers. Here we only focus on the trip origins and destinations.

[3]https://snap.stanford.edu/data/loc-gowalla.html
[4]http://www.nyc.gov/html/tlc/html/home/home.shtml



(a) Foursquare (b) Gowalla (c) Taxi trips

Figure 2.6. Data visualization

2.6.2 Evaluation Metrics

In our experiments, the first 80% check-ins are used as the training data, and the other 20% are used as the testing data. We learn the models and obtain the recommendation POI list for a future time period $T_n = [t_n + \delta, t_n + \Delta]$ for each user n, where t_n is the latest check-in time in the training data. Note that if there is no check-in existing in the T_n period, this user will not be included for evaluation. Also, the visited POIs in the training data have been removed in the testing data since recommending new POIs is our target. Finally, the evaluation metrics include F-measure and the NDCG of the newly visited POIs.

F-measure. F-measure combines precision and recall together with a harmonic mean. Here we use the $F_{0.5}$ measure which puts more emphasis on precision than recall (B. Liu et al., 2015),

$$F_{0.5}@P = (1 + 0.5^2)\frac{Precision@P \times Recall@P}{0.5^2 \cdot Precision@P + Recall@P}. \tag{2.10}$$

Given a top-P recommendation list S_{rec} sorted in a descending order based on the prediction values, precision and recall can be obtained as follows:

$$Precision@P = \frac{S_{rec} \bigcap S_{new}}{P}, Recall@P = \frac{S_{rec} \bigcap S_{new}}{S_{new}},$$

where S_{new} are the POIs a user newly visited in the test data. The precision and recall for the entire recommender system are computed by averaging all the precision and recall values of all the users.

Normalized Discounted Cumulative Gain (NDCG). Given a top-P recommendation list sorted in a descending order of the prediction values, NDCG (Järvelin & Kekäläinen, 2002) is defined as

$$NDCG@P = \frac{1}{IDCG} \times \sum_{i=1}^{N} \frac{2^{rel_i} - 1}{log(i+1)}, \tag{2.11}$$

where IDCG is the maximum possible DCG for a given set of recommendations, and rel_i is 1 if the recommended POI at position i is visited by the user and 0 otherwise. NDCG measures the ranking quality of the recommender system based on a graded relevance scale of recommendations. The NDCG for the entire recommender system are computed by averaging all the NDCG values of all the users.

2.6.3 Baselines

Probabilistic Matrix Factorization (PMF) (Mnih & Salakhutdinov, 2007) is a generalized matrix factorization model, which has been widely used for recommendation tasks.

Factorized Personalized Markov Chains (FPMC) (Rendle, Freudenthaler, & Schmidt-Thieme, 2010) embeds users' preferences and their personalized Markov chain to provide next basket item recommendation. An extended factorized personalized Markov chain with localized region constraint (FPMC-LR) (Cheng et al., 2013) was proposed for POI recommendation while considering only the neighborhood locations. Since we only focus on POI recommendation and try to explore the time interval influence, so we do not consider the location constraint here.

Purchase Interval based Matrix Factorization (PIMF) (Zhao, Lee, Hsu, & Chen, 2012) shows that the intervals between user purchases have an influence on a user's purchase decision, and thus PIMF incorporates this purchase interval factor into matrix factorization for recommendation.

2.6.4 POI Recommendation Performances

In this subsection, we present the performance comparison on the recommendation accuracy between WWO and baselines. Also, the refined models with mobile regularization are compared with WWO. The results are based on Foursquare and Gowalla data by setting the latent dimensions to $K = 10$.

The Performance of WWO

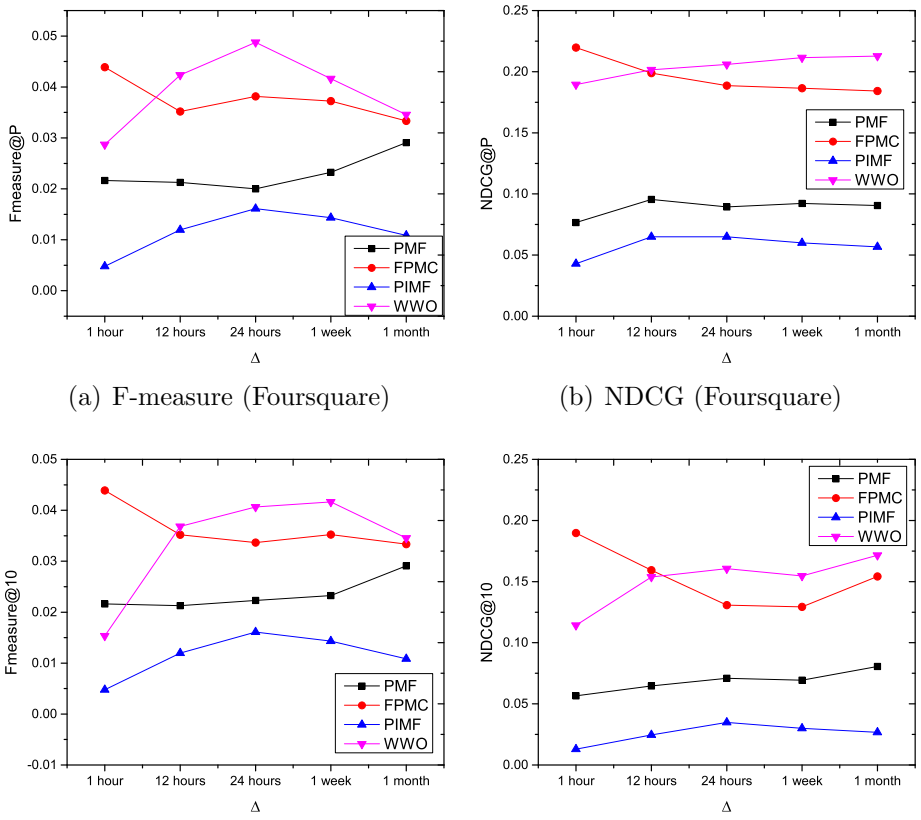
We first investigate the performance of our method without considering the human mobility information encoded in the taxi trip data. When $\delta = 0$ (recommending for the next Δ time period), the performances in terms of F-measure and NDCG with respect to Δ (from 1 hour to 1 month) on Foursquare and Gowalla data are shown in Figure 2.7. In the figures, we can see that FPMC and WWO obtain better

overall performances than PMF and PIMF. For a short next time period, such as 1 hour, FPMC outperforms others significantly. When Δ is increased to 24 hours and 1 week, WWO achieves the best performance. This implies that the temporal interval information between check-ins can be helpful for identifying the temporal relationship between POIs, especially better captured for a granularity of several days. This result is consistent with the observation for the time interval histogram shown in Figure 2.8, where we can see that the intervals between POIs are mostly distributed in the first 7 days. Moreover, as the increase of Δ, the performance of PMF is getting closer to FPMC and WWO. This is because PMF does not consider sequential preferences. POI recommendation for a long term is mainly determined by user preferences and the temporal factor is no long significant. Finally, PIMF does not work well in POI recommendation, mainly because check-in data include higher level of noise and are very sparse compared to purchase data. In fact, PIMF cannot capture the intervals between POIs without considering the interval distribution.

The recommendation results for one specified day (i.e., $\Delta - \delta = 24$ hours) with respect to various δ are shown in Figure 2.9. We illustrate the results of δ from 0 to 6 because these are the most effective time periods to recommend. In the figure, we can see that WWO obtains better performances than others. The reason is that all the other methods cannot make recommendations for a specific time period.

Mobile Regularization

Here, we show the impact of integrating heterogeneous human mobility data into our recommendation model. Figure 2.10 shows the results of three different integration



(a) F-measure (Foursquare)

(b) NDCG (Foursquare)

(c) F-measure (Gowalla)

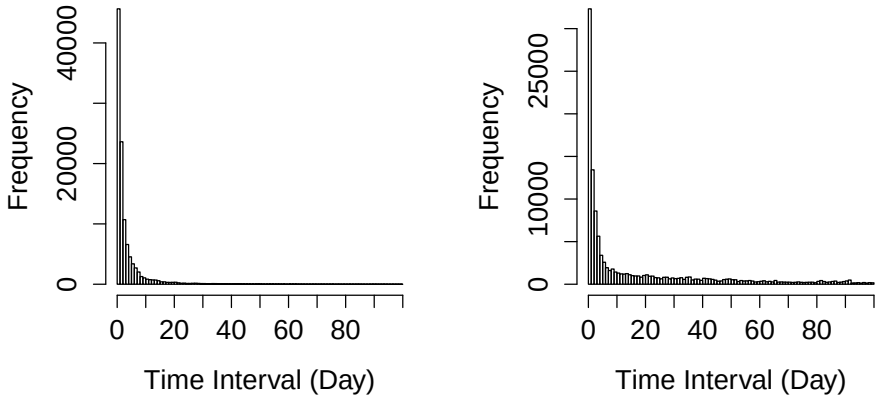
(d) NDCG (Gowalla)

Figure 2.7. The performances for different Δ, with $\delta = 0$.

methods. All these methods can enhance the recommendation performances, however *WWO_Sim* achieves the most performance improvement. Surprisingly, the improvement for a next short time period is higher than a longer time period, especially from 1 hour to 24 hours. That is because taxi trips mostly happen within the same day, and the short-term mobility patterns are more accurately captured.

2.6.5 Parameter Selection

For simplicity, we only present the parameter selection results on Foursquare data. First, we use the empirical standard deviation in $\cup_{n,i,j} d_{ij}^n$ as σ_μ in $\mathcal{N}(\mu_{ij}^n, \sigma_\mu^2)$. Then, we determine the number of graph bases K based on the recommendation performances.

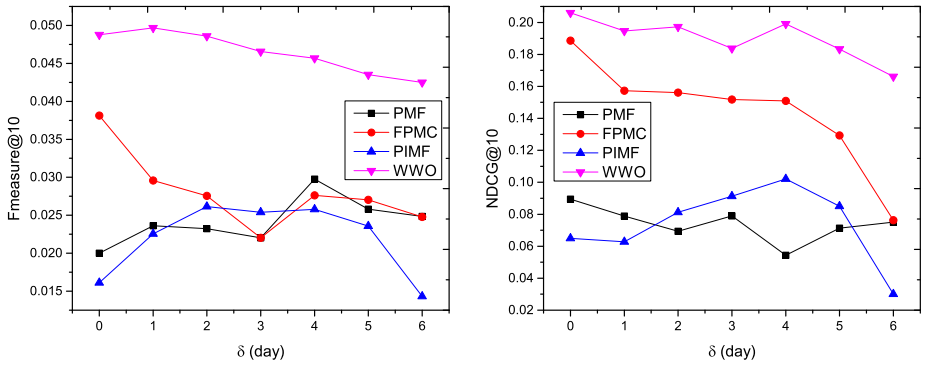


(a) Foursquare (b) Gowalla

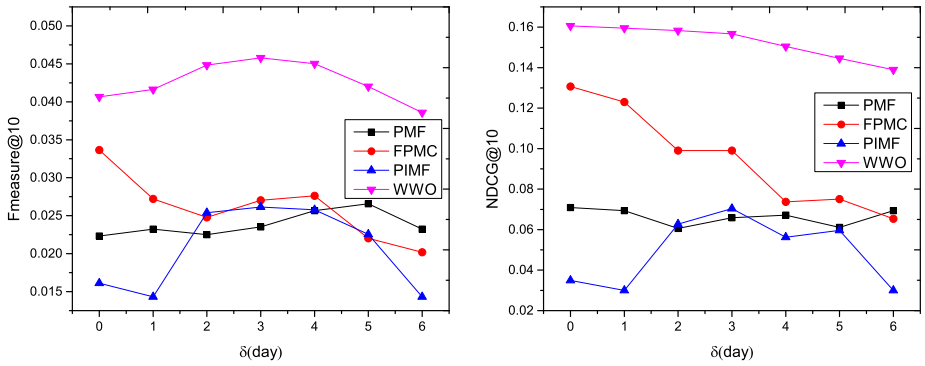
Figure 2.8. The histogram of all user check-in intervals

As shown in Figure 2.11 (a), we plot the results with increasing number of graph bases. It is worthwhile to note that the performance might not increase with more bases. The reason is that more bases imply higher modeling complexity and may lead to overfitting in the training data and decreasing generality of the identified graph bases. From Figure 2.11 (a), we see that $K = 10$ is a feasible trade off between the modeling complexity and the empirical accuracy. Thus, we choose $K = 10$.

With K chosen, we further determine the appropriate degree of mobile regularization λ, again, according to the recommendation performance. Figure 2.11 (b) shows the performance with increasing λ, where $\lambda = 0.01$ gives the optimal results. Note that, we obtain a positive optimal $\lambda > 0$, which signifies the benefits of the mobile regularization based on the human mobility data (more details in Section 2.6.4). Due to space limitations, we omit the similar procedures for tuning the prior parameters such as σ_α, σ_β, η, and θ.



(a) F-measure (Foursquare) (b) NDCG (Foursquare)



(c) F-measure (Gowalla) (d) NDCG (Gowalla)

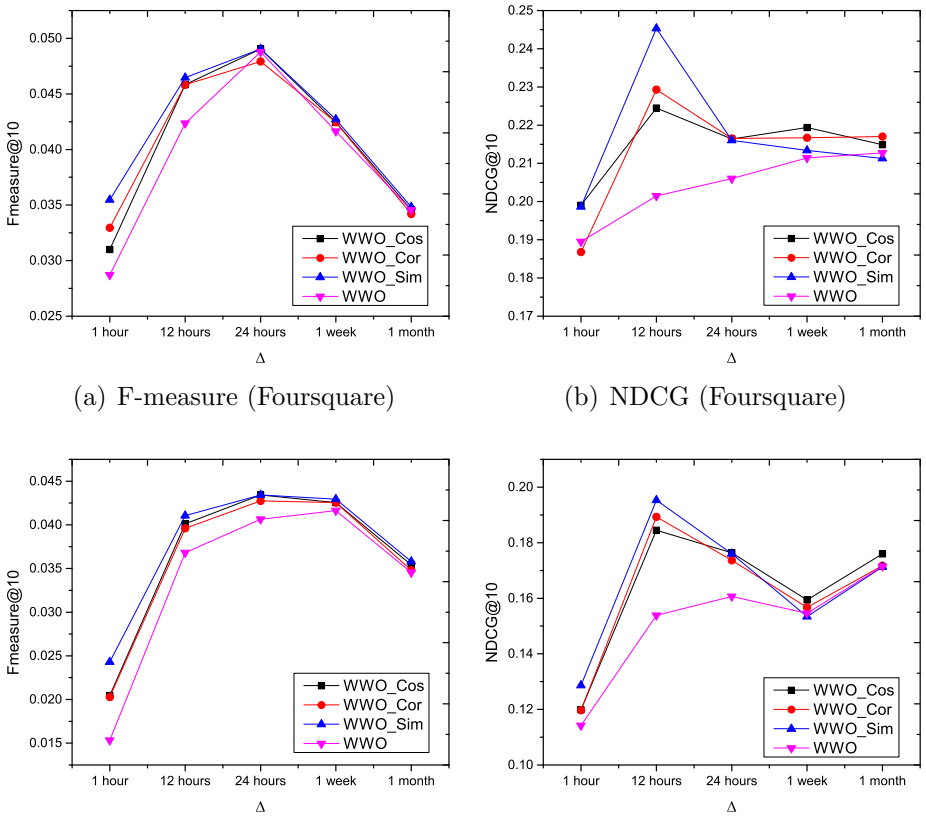
Figure 2.9. The performances for different δ, with $\Delta - \delta = 24hours$.

Table 2.2. The running time (seconds).

Data	**PMF**	**FPMC**	**PIMF**	**WWO**
Foursquare	79.34	1431.67	27416.5	1378.03
Gowalla	17.96	273.97	4301.62	287.56

2.6.6 Time Complexity

The overall asymptotic computational complexity of Algorithm 1 is $O(NM^2K^2T)$ where T is the number of iterations. Although the complexity is higher than the conventional static recommender systems, it is comparable with other state-of-the-art approaches utilizing the temporal patterns. Moreover, due to the pre-clustering



(a) F-measure (Foursquare)

(b) NDCG (Foursquare)

(c) F-measure (Gowalla)

(d) NDCG (Gowalla)

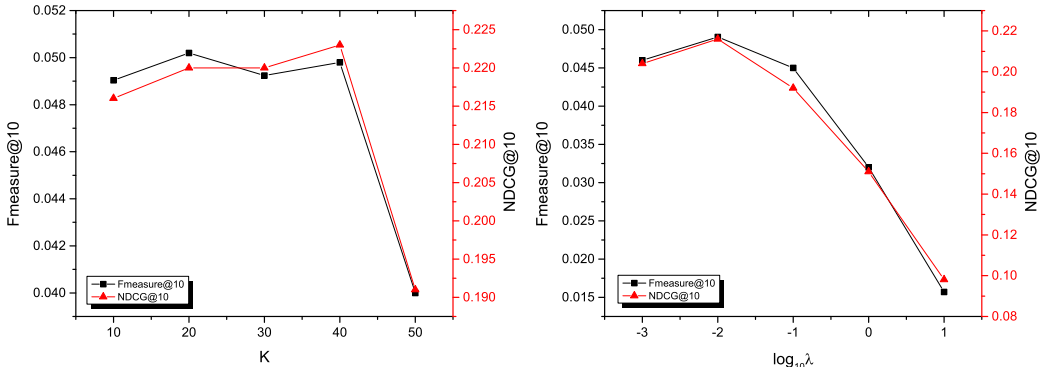
Figure 2.10. The performances for different Δ with $\delta = 0$ using mobile regularization initialization proposed in Section 2.5.1, our method converges quickly. Table 2.2 reports the running time of all methods.

2.7 Related Work

In this section, we will first introduce relevant studies on POI recommendation, followed by the general recommendation tasks with sequential information.

2.7.1 POI Recommendation

POI recommendation, targeting at recommending the right POIs to the target users, has been an important task in recent years (V. W. Zheng, Cao, Zheng, Xie, & Yang, 2010; Lian et al., 2014). Unlike other recommender systems based on explicit user



(a) Number of Graph bases K (b) Regularization factor λ

Figure 2.11. The performances at different parameters when $\delta = 0, \Delta = 24hours$

feedback, such as user ratings, POI recommendation is developed based on implicit user feedback, such as the check-in frequency. Recently, other implicit information, such as check-in locations, check-in time, and transition between POIs, have been exploited for POI recommendations.

Previous studies often used collaborative filtering (CF) algorithm to fuse the check-in information, e.g., user interest preferences, social influence, temporal influence and geographical influence (Q. Yuan et al., 2013; Gao et al., 2013). In (Ye, Yin, & Lee, 2010; Ye, Yin, Lee, & Lee, 2011), Ye et al. considered the social influence under the framework of a user-based CF model, and modeled the geographical influence by a model-based method (a Bayesian CF algorithm). To exploit the social influence, the authors made use of the users' friends for recommendation rather than all the users. On the other hand, to explore the geographical influence, they assumed that the probability that a user visited two POIs was determined by their distance, the larger the distance the smaller the probability. Moreover, Yuan et al. (Q. Yuan et al., 2013) and Gao et al. (Gao et al., 2013) introduced temporal preference to enhance the

algorithm efficiency and effectiveness. The authors separated a day into different time slots and user preferences were learned for each slot, thus POIs can be recommended according to different times of a day. Cheng et al. (Cheng et al., 2012) considered more comprehensive information, such as the multi-center of user check-in patterns, and the skewed user check-in frequency. However, this work lacked an integrated consideration of factors that can influence POI recommendation. To improve the ad hoc integration between them, Liu et al. (B. Liu, Fu, Yao, & Xiong, 2013) proposed a geographical probabilistic factor analysis framework to analyze the joint effects of multiple factors by considering user preference for locations as a multiplication of interest in the locations, location popularity and distance between user and POIs.

The above works focus on evaluating the relationships between POIs and check-in features, such as location or time, while the relationship between POIs has rarely been considered. Recent works have shown the fact that human movement exhibits sequential patterns (Gonzalez, Hidalgo, & Barabasi, 2008a; Song, Qu, Blumm, & Barabási, 2010; Cho, Myers, & Leskovec, 2011), suggesting that users usually follow some sequential behaviors when visiting POIs. In light of this, Cheng et al. (Cheng et al., 2013) considered the task of next POI recommendation, in an attempt to recommend POIs to users for their next visits. This work took user check-in sequential information into account, assuming the next check-in is dependent on last check-in. Zhang et al. (Zhang, Chow, & Li, 2014) further extended the next POI recommendation with sequential information by considering not only the latest visited POI but also the earlier visited POIs with a n-order Markov chain, and integrating geographical and social influence into the proposed method. Feng et al. (Feng et al., 2015) also

proposed a metric embedding model to learn the personalized sequential information for next POI recommendation. However, these methods did not consider the temporal interval information between dependent POIs, thus cannot capture the sequential interval patterns to make recommendations for a specified future time period.

Different from the above works, we consider the sequential preferences with temporal interval assessment between check-ins. And, the temporal interval distribution for different users and POI pairs are estimated with a novel factorization method, which enables our method to make recommendation for a specific time period.

2.7.2 Recommendation Tasks with Sequential Information

Most recommender systems rely on statistical models that use the event history of users on items to produce recommendations. Also, as an important information source for understanding user preferences, the information about user sequential behaviors have often been utilized in recommender systems. Indeed, sequential information is critical for those time-sensitive recommendation scenarios, such as travel package recommendations (Q. Liu, Ge, Li, Chen, & Xiong, 2011; Q. Liu et al., 2014; Ge, Liu, Xiong, Tuzhilin, & Chen, 2011), since users' preferences change over time in those application scenarios.

Recent works have shown that sequential patterns can be utilized to improve personalized recommendations at the right time. For example, Rendle et al. (Rendle et al., 2010) proposed a factorized personalized Markov chain (FPMC) model to recommend the products which user will probably buy in the next visit. Specifically, a transition matrix from product to product was constructed for each user, thus a

transition cube is formed for all the users. Then, the transition tensor was estimated by a factorization model to propagate information among similar users, similar items and similar transitions. Additionally, Wang et al. (Wang & Zhang, 2013) proposed an opportunity model to estimate the follow-up purchase probability of a user at a specific time. To recommend the best next-items to each target user, Yap et al. (Yap, Li, & Philip, 2012) learned user-specific sequential knowledge through personalized sequential pattern mining. Moreover, the time interval information between purchase transactions was used to improve the performance of next-product recommendations (Zhao et al., 2012).

Similarly, POI recommendation is also a time-sensitive task, since people's interests for POIs are drifting over time. Therefore, the POI recommendations will be more effective if the recommender system can capture the users' evolving sequential preferences. In order to make recommendations at the right time, the interval information between check-ins should be considered. Along this line, this chapter focuses on POI recommendations for a specific time period by considering users' evolving sequential preferences with temporal interval assessment.

2.8 Summary

In this chapter, we investigated how to exploit both user interests and their evolving sequential preferences for recommending POIs for a specific time period. Along this line, we first proposed a unified framework to integrate user interests and their sequential preferences. Specifically, the distributions of the temporal intervals between dependent POIs were studied for measuring users evolving sequential preferences

based on their historical check-in sequences. Here, to address the challenge of estimating the distributions with only sparse observations, we developed a bi-weighted low-rank graph construction model to identify a set of bi-weighted graph bases, which in turn can be leveraged for learning user interests and their sequential preferences in a coherent way. Furthermore, we provided a mobile regularization method to effectively incorporating human mobility patterns captured by user mobility data to improve the performances of POI recommendations. As shown in the experimental results on real-world data, unlike existing POI recommender systems, the proposed WWO recommender system can provide effective POI recommendations for a future specified time period by capturing user evolving sequential preferences.

CHAPTER 3

POI DEMAND MODELING WITH HUMAN MOBILITY PATTERNS

3.1 Introduction

Identification of POI demands in spatially differentiated regions is fundamental for governments, also it is critical for the survival of local businesses. In some cases, failing to estimate demands properly is enough to force a company to go out of business. The demand analysis helps the entrepreneur and the authority in making decisions for the efficient allocation of limited resources, such as business site selection, real estate investment and land use planning. Take site selection for example, a business will have high chance to success if it is placed in a highly-demanded region, otherwise it may result in serious business risk and even failure. In this chapter, we aim to provide a systematic POI demand analysis for urban regions with the help of large-scale human mobility data.

Currently, local businesses and governments largely rely on labor-intensive surveys to inform their decision-making. For example, to understand region demands, a series of research has been done based on survey data (Pilinkienė, 2008a, 2008b). However, the information obtained through the surveys may not be sufficient and timely enough. Recently, the wide availability of Information Communications Technology (ICT) has enabled unprecedented opportunities to collect large-scale human mobility data, e.g.,

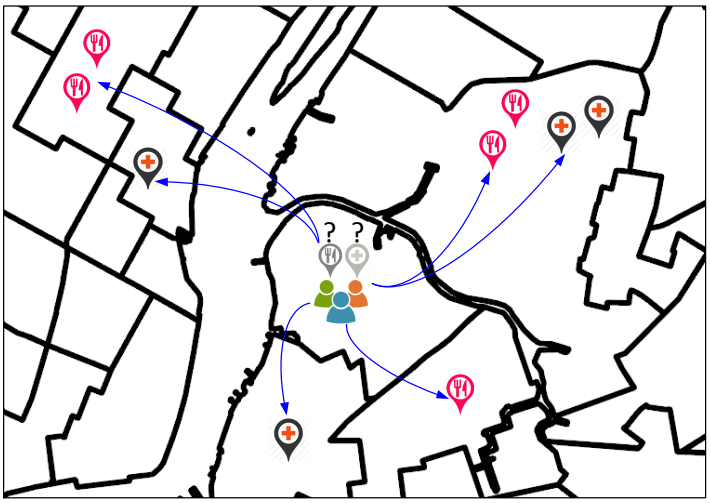


Figure 3.1. Identification of region POI demand

taxi GPS traces, which is able to cover the whole urban area with fine-grained time and location information. It reflects the underlying dynamics of residents in the city, which is much more detailed, and has larger scale than survey data. While more and more efforts have been put into analyzing the large-scale human mobility data to understand urban dynamics, few of them provide a systematic POI demand modeling but focus on specific POI categories such as restaurant and gas station (Karamshuk, Noulas, Scellato, Nicosia, & Mascolo, 2013; Niu, Liu, Fu, Liu, & Lang, 2016). Actually, *people vote with their feet*, which is an important economic logic (Tiebout, 1956), indicates that people have the ability to choose what they need by traveling. For example, as illustrated in Figure 3.1, if people from one region frequently travel to other regions for restaurants and doctors, it is much likely people need new or improved restaurants and health services in this region. Therefore, the human mobility patterns can be utilized to identify daily needs of people, and provide governments and local businesses with opportunities to better understand

POI demand and formulate future planning.

Indeed, in this chapter, we investigate how to identify POI demands for urban regions by modeling region POI preferences, region POI supplies, and region demographic features. To the best of our knowledge, this is the first attempt to identify POI demands over different categories in the city scale. The main challenges of this work include: 1) Human mobility data is highly skewed between different regions, even no human mobility data collected for underdeveloped regions which in fact need more attention. Take NYC as an example, Staten Island is one borough of NYC but separated from other boroughs by New York Bay, which makes it difficult to travel by taxi to downtown NYC but by ferry. As shown in Figure 3.3 (a), the number of taxi trips in Staten Island is much less than other boroughs. 2) How to integrate region POI profiles and demographic data together with human mobility data to better model the POI demand of urban regions in a holistic way. 3) How to learn the demands simultaneously for all the regions in a city with community opinions considered.

To this end, in this chapter, we develop a systematic POI demand modeling framework, named Region POI Demand Identification (RPDI), to model POI demands by exploiting the daily needs of people identified from their large-scale mobility data (Y. Liu, Liu, Lu, et al., 2017). Specifically, in our proposed framework, we first partition the urban space into spatially differentiated neighborhood regions formed by many small local communities. Then, a Bayesian model (Gong, Liu, Wu, & Liu, 2016) considering context POI information like distance, rating, and popularity is exploited to infer trip activities. Besides, we aggregate the trips that origin from the

same region and identify the underlying demands of all the regions simultaneously using a latent factor model, which integrates region preferences, POI supplies, and demographic features. Furthermore, we apply the identified demands for region POI demand ranking. Finally, the main contributions of this chapter can be summarized as follows:

- A systematic framework, named RPDI, is developed to identify POI demands for urban regions in a city. RPDI is able to identify POI demands over different categories in the city scale with large-scale human mobility data.
- A latent factor model is proposed to integrate region preferences, POI profiles, and demographic features for POI demand modeling. The model helps to identify a set of latent factors, which in turn can be leveraged for learning region demands in a coherent way.
- The RPDI framework has been evaluated on large-scale real-world data for region POI demand identification. The experimental results show that our method outperforms baseline methods in terms of multiple metrics such as Normalized Root-Mean-Square Error (NRMSE), F-measure and Normalized Discounted Cumulative Gain (NDCG).

3.2 The Region POI Demand Identification Framework

In this section, we first formally introduce the problem of region POI demand identification, and then provide an overview of our proposed Region POI Demand Identification framework.

Table 3.1. Mathematical notations

Notation	**Description**
N, M, C, T, D	The number of regions, POIs, POI categories, time slots, dimension of region features, respectively.
$\mathbb{R}, \mathcal{P}, \mathcal{C}, \mathcal{T}$	The set of regions, POIs, POI categories, and time slots, respectively.
$\mathbf{F}$	The matrix of region features, including POI profiles and demographic features.
$\mathbf{X}, \mathbf{Y}$	The region activity cube at POI level and category level, respectively.
K	The dimension of latent factors.
α, v	The latent demand patterns at POI category level, and POI level, respectively.
β	The region coefficients for features.
u	The latent region pattern coefficients.

3.2.1 Problem Statement

Assume that we have M POIs denoted by the set $\mathcal{P}$ and N regions in the set $\mathbb{R}$. For simplicity, we let $\mathcal{P} = \{1, 2, \cdots, M\}$, i.e., we use integers to represent the POIs. For the n-th region in time slot $t \in \mathcal{T}$, where $n = 1, 2, \cdots, N$, and $t = 1, 2, \cdots, T$, we have its activity records represented as a vector of visiting probabilities to POIs $x_t^n = (x_{1t}^n, x_{2t}^n, \cdots, x_{Mt}^n)$, which indicates needs of people. Thus all the activities for all the regions in all time slots form a region activity cube $\mathbf{X} = \{x_{pt}^n\}$. However, $\mathbf{X}$ is very sparse since not all the POIs are visited, and moreover, not all the activities are recorded for certain regions. Formally, our idea of demand identification is to recover the visiting probabilities of all the regions using community opinions with region preferences and region supplies considered. With recovered $\hat{x}_{pt}^n$ from the demand

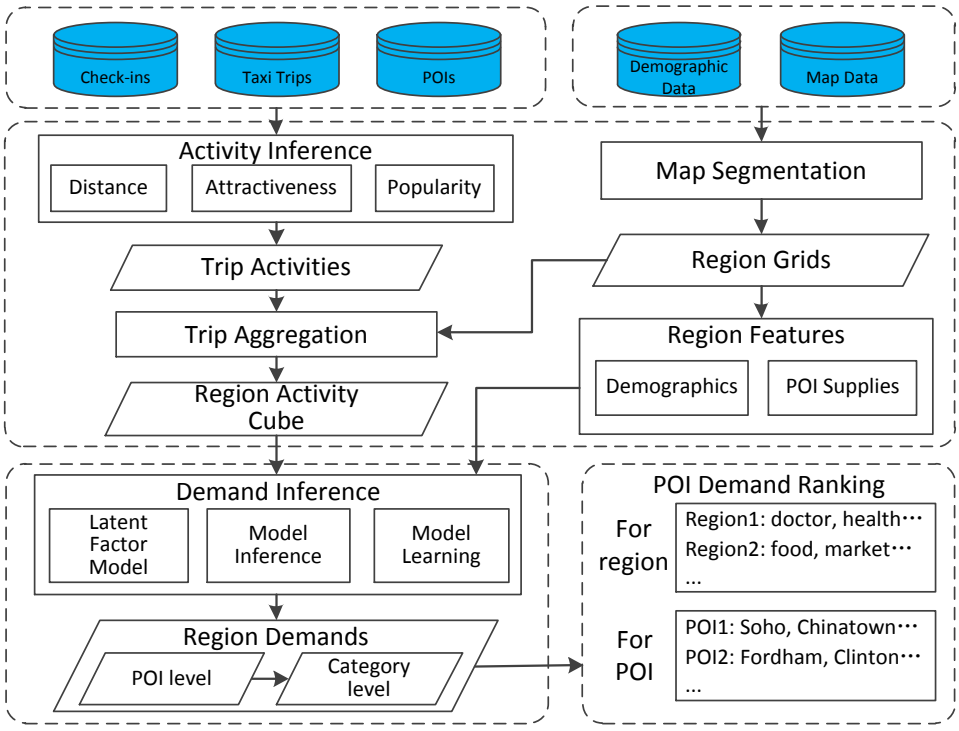


Figure 3.2. An overview of RPDI framework

inference model, we further derive the region demand d^n at POI level and category level. Then we apply the identified demand for region POI demand ranking, which is one of many applications leveraging POI demand. Table 3.1 lists some notations used in this chapter.

3.2.2 Framework Overview

Figure 3.2 shows the proposed Region POI Demand Identification framework. This framework first segment an urban area into regions which are shared by local communities. Then with the help of collected geographic and demographic data, we are able to extract POI profiles and demographic features for each region. In the mean time, we infer trip activities from taxi GPS traces with context information considered,

then propagate trips to the region level. We treat the activities performed outside the region as the underlying POI demand of this region. With the region activity cube and region features obtained, we model the region demand with a latent factor model, which integrates region preferences, region supplies and demographics. With the proposed demand model, we aim to infer region activities with partial activity information given. We also develop efficient algorithms to optimize the latent model with large-scale data. Finally, we derive the region demands from the learned region activities and further apply them for POI demand ranking, which is useful for various applications such as business site selection and real estate investment.

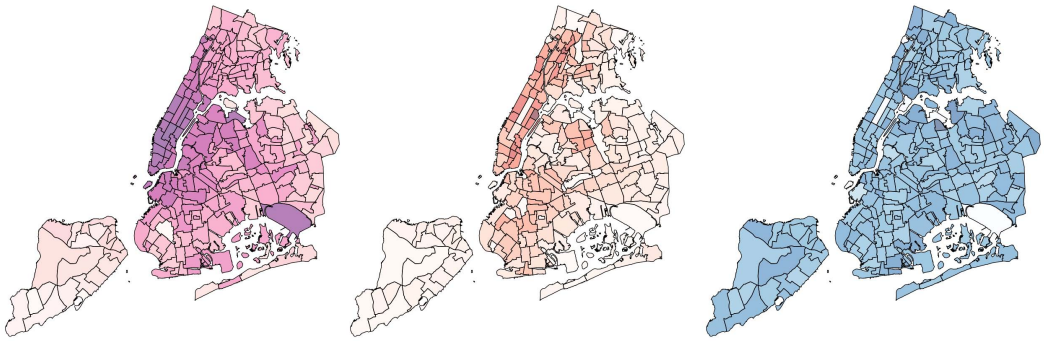
3.3 Preliminaries

In this section, we first introduce the region partition for urban area, followed by the details of demographics and POI profiles in regions. At last we introduce how to extract human mobility patterns from taxi GPS traces and POIs.

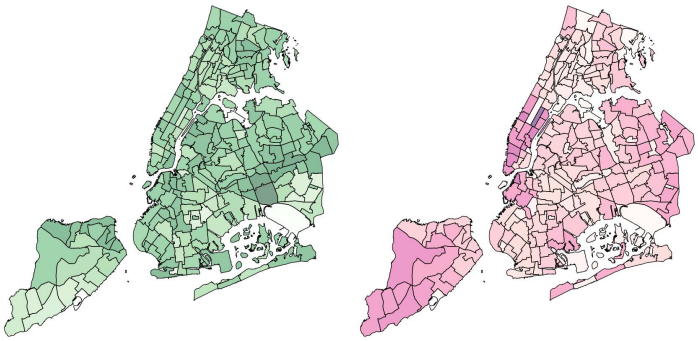
3.3.1 Region Partition

The urban area can be partitioned into regions with different methods, e.g., grid-based, road network-based (Y. Zheng et al., 2014). However, these methods do not take the socioeconomic factor into account. Instead, we use Neighborhood Tabulation Areas (NTAs)[1] provided by New York City government as our region partition method, which is shown in Figure 3.3. NTAs are created to project populations at a small area level for the long-term sustainability plan for New York City. NTAs are a valuable summary level for use with both the 2010 Census and the American

[1]https://www.nyc.gov



(a) Trip Number (b) Population Density (c) Median Age



(d) Population Composition (e) Household Income

Figure 3.3. Features of regions. Note that darker color stands for a higher value in that region.

Community Survey (ACS). These geographic areas offer a good compromise between the very detailed data for census tracts (2,168) and the broad strokes provided by community districts (59). As a result, we obtain 195 neighborhood regions which are shared by local communities.

3.3.2 Region Demographics

The POI demand indeed is the demand of people. The different distributions of people in a region will affect the demands significantly. Therefore, the demographic information could be a complement to human mobility patterns to estimate region demands. To this end, we integrate the demographic information collected from the US Census data[1] into our proposed model.

For each neighborhood, population density, sex, age, composition are used to depict population information in that region. Moreover, household income, household type, housing occupancy, and housing tenure are used to depict household information. In total we have 25 attributes to depict the demographic features $df^n = (df_1^n, df_2^n, \cdots, df_{25}^n)$ for region r_n, and some typical ones are shown in Figure 3.3. Note that we describe population composition using eight attributes with one for each race type (e.g., white, black, asian, etc.), but in Figure 3.3 (c) we only show the entropy of these eight attributes for visualization, where a higher value means a more mixed population compostion.

3.3.3 Region POI Profiles

Existing POIs in one region indicate the POI supply of this region provided, which is on the other side of region demand. As long as the supply and demand can be balanced, there is no need to add new POIs to increase the supply. From this point of view, what we try to estimate in this chapter is the region POI demand cannot be fulfilled locally.

For the n-th region r_n, the number of POIs in each POI category can be counted. The frequency density of POI category c in region r_n is calculated by:

$$pf_c^n = \frac{|\{p|p \in r_n, c\}|}{Area\ of\ r_n}$$

and the region POI feature vector of region r_n is denoted by $pf^n = (pf_1^n, pf_2^n, \cdots, pf_C^n)$, where C is the number of POI categories. The region feature vector $f_n = (df_n, pf_n) \in$

F, consists of demographic and POI features, is regarded as the metadata of each region.

3.3.4 Trip Activity Inference

The original human mobility from taxi GPS traces is described by trip origin and destination, where the activities participated for the trips are unknown. Fortunately, efforts have been made to infer the activities involved for each trip (Kumar, Mahdian, Pang, Tomkins, & Vassilvitskii, 2015; Fu et al., 2016). Here, we leverage the trip context information, i.e., POIs around destination, to describe the trip activities instead of one destination point. Specifically, we utilize the Bayesian activity inference model proposed by Gong et al. (Gong et al., 2016), which take both spatial and temporal constraints into consideration, to estimate the probabilities of possible destination POIs for each trip. After we get the probabilities of POIs for each trip, we aggregate the trips for regions. Gong et al. (Gong et al., 2016) show this model can effectively infer trip activity at the aggregation level.

Given trip $tr = (tr.ori, tr.dest, tr.time)$, the inference process first chooses a set of candidate POIs $\mathcal{PC} \subset \mathcal{P}$ within the walking distance δ of the trip destination, then assigns different visiting probabilities by considering influences of 1) distance decay, further the distance from POI to destination smaller the chance of visiting; 2) popular time, POI categories have different popularities at different time of day; and 3) attractiveness, POIs with higher ratings can attract more people to visit.

As we know, the region demands may vary over time of day, e.g., more people visiting bar in the evening than in the morning. We thus segment time into multiple

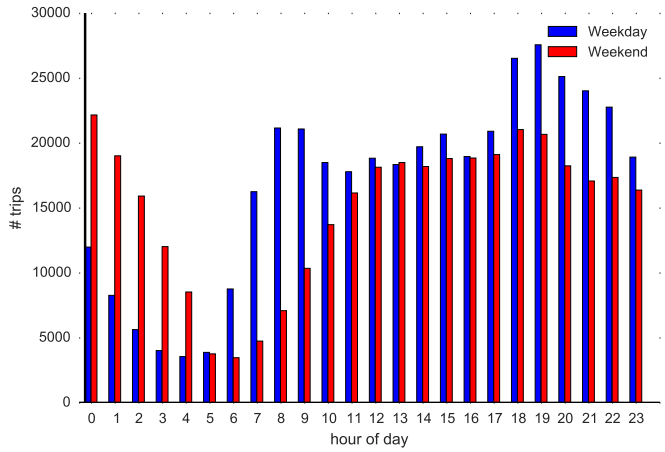


Figure 3.4. Number of trips at different hours of a day

segments in terms of T defined time slots. For example, we first segment days by weekday and weekend, then segment each day into 24 slots with each hour as a slot, and finally we get 48 time slots. The daily average number of trips for each time slot is shown in Figure 3.4.

Specifically, in this model, the probability of user choosing a POI $p \in \mathcal{PC}$ for a trip $tr, tr.t \in t$ is formulated as follows.

$$\Pr(p|tr) = \frac{A_p \cdot dist(tr.d, p)^{-\lambda} \cdot pop(p|t)}{\sum_{q \in \mathcal{PC}} A_q \cdot dist(tr.d, q)^{-\lambda} \cdot pop(q|t)} \tag{3.1}$$

where $dist(a, b)$ is the distance between two points a and b, and $dist \leq \delta$, A_p is the attractiveness of p, which is represented by the rating from Google POI data, $pop(p|t)$ is the popularity of p at time slot t which can be derived from Foursquare check-in data with

$$pop(p|t) = \frac{1}{|\{p \in c\}|} \cdot \frac{|\{checkin \in c, t\}|}{\sum_{c \in \mathcal{C}} |\{checkin \in c, t\}|}.$$

$Pr(p|tr)$ ranges from 0 to 1, and the sum of the visiting probabilities of all the

candidate POIs for one trip equals to 1. According to (Gong et al., 2016), we choose $\delta = 200m, \lambda = 1.5$ as the parameters. Finally, for each trip tr, we extract the *activity* tuple: $act(tr) = (tr.ori, tr.time, tr.pois)$.

3.4 Demand Inference

In this section, we introduce how to derive the POI demands for all the regions by integrating human mobility records, POI and region profiles. We start with the semantic aggregation of the mobile activities of each region.

3.4.1 Region Activity Aggregation

After we infer the trip activities, next we aggregate the trips from the same region to obtain region activities. Specifically, for the n-th region r_n and time slot t, we aggregate the intentions of activities originating from r_n as follows:

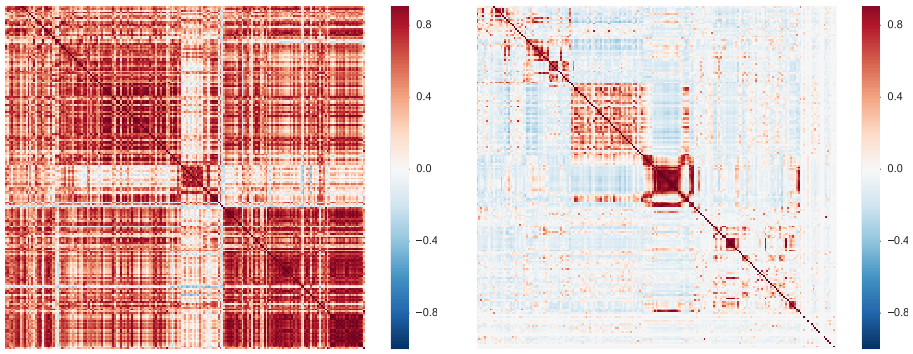
$$pr_{pt}^{n} = \sum_{tr:\substack{tr.ori=r_n \\ tr.time=t}} \Pr(p|tr),$$

where $p \in \mathcal{P}$ is the index of the POIs. Moreover, we aggregate and normalize the probability score as

$$x_{pt}^{n} = \frac{pr_{pt}^{n}}{\sum_{q\in\mathcal{P}} pr_{qt}^{n}},$$

and obtain the region activity cube $\mathbf{X} = \{x_{pt}^{n}\}$ at the POI level.

Other than that, the region activities can also be aggregated to POI category level, which is commonly used in literature (Fu et al., 2016). Specifically, with the probabilities of POIs visited by people from region r_n inferred, we can further summarize



(a) (b)

Figure 3.5. Correlation map (a) features of regions, (b) human mobility patterns of regions.

the visiting probability of those POIs per category $c \in \mathcal{C}$ and obtain the category-level aggregated visit probability as $y^n_{ct} = \sum_{p \in c} x^n_{pt}$. In this way, we construct the representation of POI visit as an aggregated visiting probability vector over different POI categories. Similarly, we obtain the region activity cube $\mathbf{Y} = \{y^n_{ct}\}$ at the POI category level.

Figure 3.5 shows the correlations of extracted features and human mobility patterns between regions, from which we can see several region clusters formed indicating that we may learn from peers. Please note our latent factor model can be applied at both the POI level and the POI category level to infer the POI demand.

3.4.2 Latent Factor Model

As aforementioned, human mobility data can tape the "foot voting", i.e., where people go is for what they need but cannot be fulfilled locally. Therefore, one straightforward way to infer the region demand is to directly aggregate the probabilities of the destination POIs. However, the distribution of the mobility data is highly skewed, thus some regions may not have enough or even no observations to recover the demand by

activity aggregation. Moreover, one trip may visit multiple POIs and the POIs can compete with each other, but the simple aggregation cannot take these factors into account. In the following, we develop a latent factor model, which considers profiles of regions and POIs, to learn the region demand with skewed mobility records. In other words, the regions with enough human mobility data can help the regions with few observations in the modeling process.

Intuitively, a person starting an activity from a region first needs to decide which demand category (e.g., shopping, eating, recreation) to be fulfilled. If the demand cannot be fulfilled locally, which cost the least amount of time and energy, then the person needs to decide which POIs in which regions to go. Along this line, we model the *demand patterns* at both POI category level (α) and POI level (v). Given a time slot t, in each column vector $\alpha_{ct} \in \mathbb{R}^K$ and $v_{pt} \in \mathbb{R}^K$ are the *pattern coefficients* for one category (c) and one POI (p), respectively. Similarly, the latent variables in the matrix u encode the *pattern coefficients* of the regional POI demands. The column vector $u_n \in \mathbb{R}^K$ is for the n-th region. In this way, the observed activity can be modeled as:

$$x_{pt}^n \sim u_n' v_{pt}.$$

The structure of the proposed region demand inference model is shown in Figure 3.6. Note that, the region demands may vary over time of the day, e.g., more people visiting bar in the evening than in the morning. We thus segment time everyday into multiple time slots indexed by $t = 1, 2, \cdots, T$ and learn the demand patterns for each time slot.

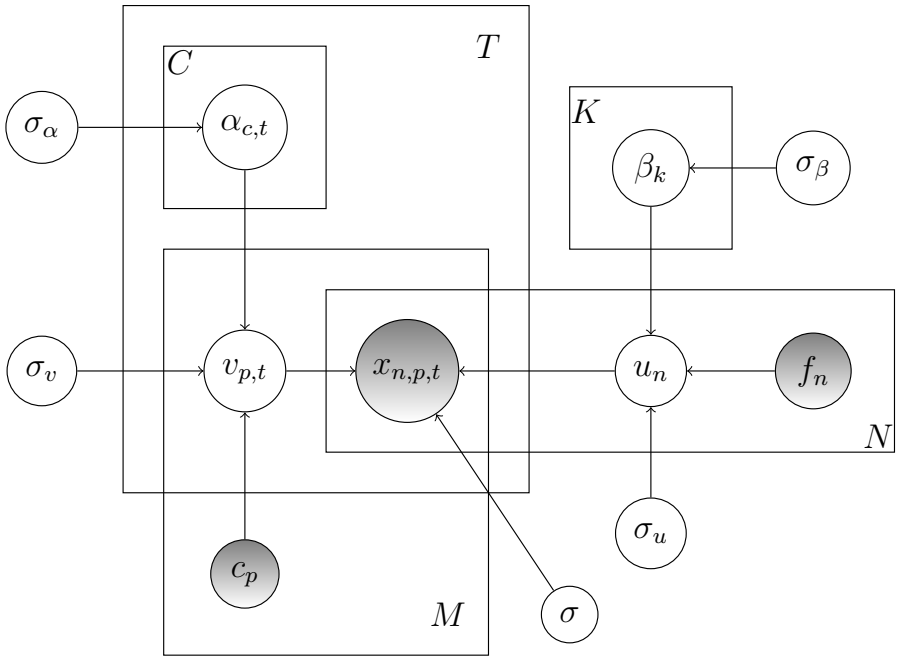


Figure 3.6. The demand inference model

As shown in the graphical model (Figure 3.6), we also use side-information (e.g., POI category, regional POI supply and demographic data) to enhance the model. Specifically, suppose we use K latent demand patterns to model the demand portfolio of the region. The pattern coefficients $u_n \in \mathbb{R}^K$ and the demographic features $f_n \in \mathbb{R}^D$ of the n-th region is modeled as:

$$u_n \sim \beta' f_n,$$

where the matrix $\beta \in \mathbb{R}^{D \times K}$ will be learned in the modeling process. Similarly, we have:

$$v_{pt} \sim \alpha_{c(p)t},$$

if the p-th POI is in the $c(p)$-th POI category.

We put the above modeling processes in an unified probabilistic framework, with the following distribution specifications:

$$x_{pt}^{n} \sim \mathcal{N}(u_n' v_{pt}, \sigma) \in \mathbb{R},$$

$$u_n \sim \mathcal{N}(\beta' f_n, \sigma_u I_K) \in \mathbb{R}^K,$$

and

$$v_{pt} \sim \mathcal{N}(\alpha_{c(p)t}, \sigma_v I_K) \in \mathbb{R}^K,$$

where σ, σ_u, and σ_v are standard deviations of the normal distributions, respectively. Now, we use the negative log-likelihood of the model as the objective function (3.2) to optimize the model parameters (α, β, u, and v):

$$\begin{aligned} \mathcal{L}(\alpha, \beta, u, v | x, \sigma) =& \frac{1}{2\sigma^2} \sum_{n,p,t} (x_{pt}^{n} - u_n' v_{pt})^2 \\ +& \frac{1}{2\sigma_u^2} \sum_{n} \|u_n - \beta' f_n\|^2 + \frac{1}{2\sigma_v^2} \sum_{p,t} \|v_{pt} - \alpha_{c(p)t}\|^2 \\ +& \frac{1}{2\sigma_\alpha^2} \sum_{c,t} \|\alpha_{ct}\|^2 + \frac{1}{2\sigma_\beta^2} \sum_{k} \|\beta_k\|^2 \end{aligned} \tag{3.2}$$

The last two terms are added to reduce the generalization error with the following priors on the latent variables α and β:

$$\alpha_{ct} \sim \mathcal{N}(0, \sigma_\alpha I_K) \in \mathbb{R}^K,$$

and

$$\beta_k \sim \mathcal{N}(0, \sigma_\beta I_D) \in \mathbb{R}^D.$$

3.4.3 Learning Algorithm

In this section we introduce an efficient algorithms to optimize the latent factor model with large-scale data. In the proposed model, we have parameters in: 1) α, v for time-aware demand patterns for POI categories and POIs respectively; 2) u for latent demand preferences for individual regions; 3) finally β for regression coefficients between the region features and the demand preferences of each region. We iteratively update these parameters to optimize the objective function in 3.2.

Specifically, to optimize α with other parameters fixed, the problem is equivalent to minimize:

$$\frac{1}{2\sigma_v^2} \sum_{p,t} \|v_{p,t} - \alpha_{c(p),t}\|^2 + \frac{1}{2\sigma_\alpha^2} \sum_{c,t} \|\alpha_{c,t}\|^2 \tag{3.3}$$

Therefore, for each POI category c and time index t, to compute $\alpha_{c,t}$, we minimize:

$$\frac{1}{2\sigma_v^2} \sum_{p:c(p)=c} \|v_{p,t} - \alpha_{c,t}\|^2 + \frac{1}{2\sigma_\alpha^2} \|\alpha_{c,t}\|^2 \tag{3.4}$$

For this problem, we have closed form solution:

$$\alpha_{c,t} = \frac{\sum_{p:c(p)=c} v_{p,t}}{M_c + \sigma_v^2/\sigma_\alpha^2} \tag{3.5}$$

where $M_c = |\{p \mid c(p) = c\}|$ is the number of POIs in category c.

The problem to optimize β is the so-called ridge regression which is to minimize:

$$\frac{1}{2\sigma_u^2}\sum_n \|u_n - \beta' f_n\|^2 + \frac{1}{2\sigma_\beta^2}\sum_k \|\beta_k\|^2$$

Since $\mathbf{F}\mathbf{F}' + \sigma_u^2/\sigma_\beta^2 I_D$ is not singular, we also have closed form solution as:

$$\beta = (\mathbf{F}\mathbf{F}' + \sigma_u^2/\sigma_\beta^2 I_D)^{-1}\mathbf{F}u' \in \mathbb{R}^{D\times K}$$

where $\mathbf{F} \in \mathbb{R}^{D\times N}, u \in \mathbb{R}^{K\times N}$ and $I_D \in \mathbb{R}^{D\times D}$ is the identity matrix.

For updating u and v, we use gradient descent optimization. To this end, we have

$$\frac{\partial \mathcal{L}}{\partial u_n} = -\frac{1}{\sigma^2}\sum_{p,t}(x_{pt}^n - \langle u_n, v_{pt}\rangle)v_{pt} + \frac{1}{\sigma_u^2}(u_n - \beta' f_n)$$
$$\frac{\partial \mathcal{L}}{\partial v_{pt}} = -\frac{1}{\sigma^2}\sum_{n}(x_{pt}^n - \langle u_n, v_{pt}\rangle)u_n + \frac{1}{\sigma_v^2}(v_{pt} - \alpha_{c(p),t})$$

3.4.4 Variations and Extensions

There are several variations of our modeling process. For example, we can use the the following objective function to directly identify the POI demand of urban regions at the POI category level with region activity cube $\mathbf{Y}$:

$$\begin{aligned}\mathcal{L}(\alpha, \beta, u|x, \sigma) =& \frac{1}{2\sigma^2}\sum_{n,c,t}(y_{ct}^n - u_n'\alpha_{ct})^2 \\ &+ \frac{1}{2\sigma_u^2}\sum_n \|u_n - \beta' f_n\|^2 \\ &+ \frac{1}{2\sigma_\alpha^2}\sum_{c,t}\|\alpha_{ct}\|^2 + \frac{1}{2\sigma_\beta^2}\sum_k \|\beta_k\|^2\end{aligned} \tag{3.6}$$

In this way, we can estimate:

$$\hat{y}_{ct}^{n} \sim u_n' \alpha_{ct}.$$

This might be different with the simple aggregation of the results at the POI level, e.g.,

$$\hat{x}_{pt}^{n} \sim u_n' v_{pt},$$

$$\hat{y}_{ct}^{n} \sim \sum_{p \in c} \hat{x}_{pt}^{n},$$

where u and v are from Equation 3.2. We named the model variation in Equation 3.6 as RPDI_c, which will be investigated in our empirical studies in Section 3.5.

3.5 Experimental Results

In this section we first introduce the data and settings of our experiments. Then we evaluate the performances of the proposed region demand inference model. Finally we show the results of our model applying to POI demand ranking.

3.5.1 Experimental Data

All the experiments were performed on real-world datasets including one taxi trip dataset from the New York City (NYC)[2], one POI dataset collected from Google Map[3], one Location-based Social Network (LBSN) dataset collected from Foursquare[4], and one demographic dataset as introduced in Section 3.3.2.

The taxi trip dataset is generated by about 50,000 taxis in New York City from January to June, 2016, in total we have around 72 million trips collected. Each trip is

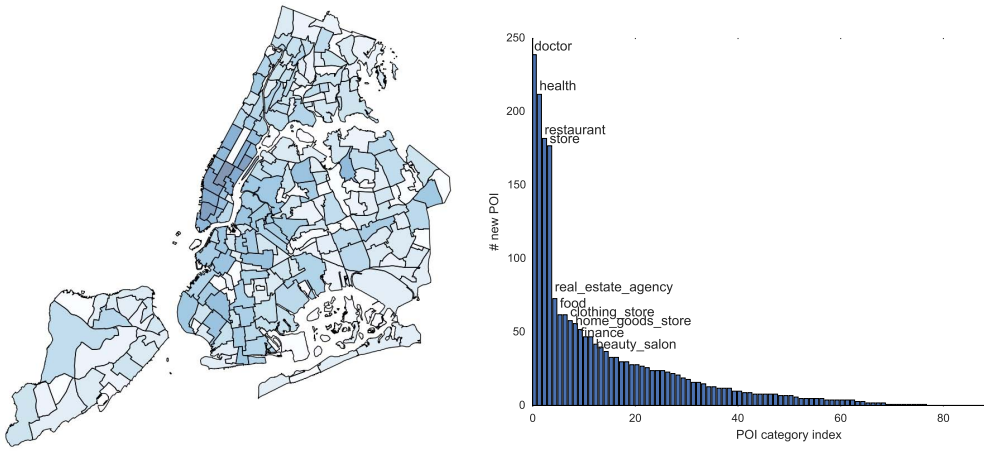
[2]http://www.nyc.gov/html/tlc/html/home/home.shtml
[3]https://developers.google.com/places/
[4]https://www.foursquare.com/

associated with pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. In this experiment we focus on the time, origin, and destination information related to taxi trips.

The POI dataset is collected from Google Place API and have a flat category structure, which contains 97 fine categories[5], such as restaurant, store, park, etc. Due to space limit, we cannot list all the categories here, please refer to Google Place API[6] for the whole list of categories. This dataset can be split into two sets, one contains 297,078 POIs created before June, 2016, which are employed to estimate the human activities. Another set contains 3,817 POIs, which are created after June, 2016, is used as our validation set of demand ranking. The distribution of newly created POIs are shown in Figure 3.7 (a) and (b) in terms of regions and categories, respectively. Here we assume that people are rational and the new POIs are created to meet the demands.

The Foursquare dataset includes 504,152 check-in observations for 55,717 POIs. Each check-in contains the user ID, check-in time, venue ID and the venue's geo-coordinates. Note the Foursquare dataset has 418 POI categories which is more detailed than Google POI, and we manually match the 418 POI categories to the Google POI categories to make them consistent.

[5]https://developers.google.com/places/supported_types
[6]https://developers.google.com/places



(a) Regions (b) Categories

Figure 3.7. The distribution of new POIs in NYC. Note that in (b) only top 10 most POI categories are labeled due to space.

3.5.2 Evaluation Metrics

In our experiments, we first evaluate the performances of our proposed latent factor model by comparing learned POI demands with observed demands. Then we apply our model for POI demand ranking, and the performances are presented.

Model Performance. We removed a uniform random subset of 10% of the entries in $\mathbf{X}$ (or $\mathbf{Y}$) as a test set and trained on the remaining 90%. We chose to remove random entries in $\mathbf{X}$ (or $\mathbf{Y}$) as opposed to random trips so as to avoid the obvious bias that regions will tend to revisit the same POIs. The goal of demand identification is to estimate the demands for a region which may not be observed. So, by limiting the test set to new region-POI pairs we are able to define an evaluation metric more inline with the problem we are solving. We learn the models and obtain the estimated POI demands for each region, then compare this estimation with testing data. The evaluation metric used for the comparison is Normalized Root-Mean-Square Error (NRMSE).

Normalized Root-Mean-Square Error (NRMSE). The root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. Normalizing the RMSE with respect to the standard deviation (or mean) facilitates the comparison between datasets or models with different scales.

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{x}_i - x_i)^2}{n}}, NRMSE = \frac{RMSE}{std(x)} \tag{3.7}$$

POI Demand Ranking. We rank our identified POI demands to see what's the most needed POI categories in a region, and which regions are with the most demand for certain POI category. Therefore, our experiments are conducted in two-folds: (1) Given a region, rank the demands for different POI categories; (2) Given a POI category, rank the demands of different regions.

For practical usage, we train a model for each time slot. To provide a unified region demand to users, the output of these models can be aggregated as

$$d_p^n = \sum_t w_t^n \cdot \hat{x}_{pt}^n,$$

where

$$w_t^n = \frac{|\{tr \in r_n, t\}|}{\sum_t |\{tr \in r_n, t\}|},$$

d_p^n stands for demand of region r_n for POI p. And further normalized by category

with supply information considered,

$$\bar{d}_c^n = \frac{1}{|\{p \in r_n, c\}| + 1} \cdot \sum_{p \in c} d_p^n.$$

Here $\bar{d}_c^n$ is the marginal demand for one POI in category c, which measures the potential demand can be delivered to a new POI. Our ranking result is given by ranking $\bar{d}_c^n$ in descending order. To evaluate the ranking list given, we use the newly created POIs after June, 2016 as our groundtruth of demand, which is ranked in descending order by the ratio of increased new POIs:

$$\frac{|\{q \in r_n, c\}|}{|\{p \in r_n, c\}| + 1},$$

where $q \in \mathcal{NP}, p \in \mathcal{P}$, and $\mathcal{NP}$ is the new POI set.

F-measure. F-measure combines precision and recall together with a harmonic mean, which is defined as

$$F_1@\,top\text{-}k = 2 \cdot \frac{Precision@k \times Recall@k}{Precision@k + Recall@k}. \tag{3.8}$$

Given a top-k ranking list S_{rank} sorted in a descending order based on the estimated demands, precision and recall can be obtained as follows:

$$Precision@\,top\text{-}k = \frac{S_{rank} \bigcap S_{new}}{k}, Recall@\,top\text{-}k = \frac{S_{rank} \bigcap S_{new}}{S_{new}},$$

where S_{new} are the POI categories newly created in the groundtruth data. The F-

measure for the entire city are computed by averaging all the F-measure values of all the regions (or categories).

Normalized Discounted Cumulative Gain (NDCG). Given a top-k ranking list sorted in a descending order of the estimated demands, NDCG (Järvelin & Kekäläinen, 2002) is defined as

$$NDCG@\,top\text{-}k = \frac{1}{IDCG} \times \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{log(i+1)}, \tag{3.9}$$

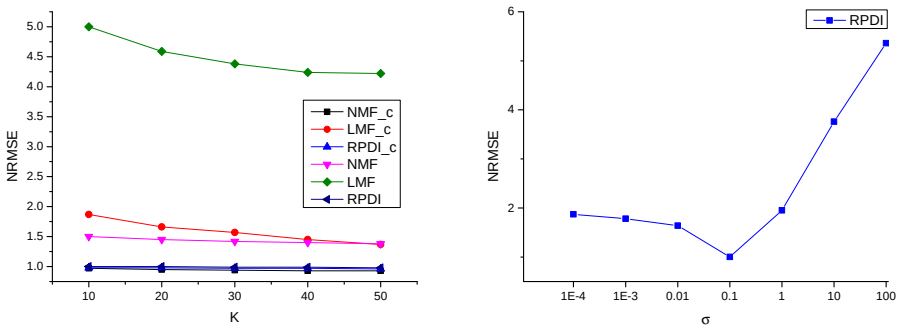
where IDCG is the maximum possible DCG for a given set of ranking list, and rel_i is 1 if the ranked POI category at position i is newly created and 0 otherwise. NDCG measures the ranking quality of the recommender system based on a graded relevance scale of recommendations. The NDCG for the entire city are computed by averaging all the NDCG values of all the regions (or categories).

3.5.3 Baselines

We compare the proposed method (RPDI) with five baselines, which are introduced as follows.

Non-Negative Matrix Factorization (NMF) is a matrix factorization model, which factorize a matrix into (usually) two matrices, with the property that all three matrices have no negative elements. This non-negativity makes the resulting matrices easier to inspect.

Logistic Matrix Factorization (LMF) (Johnson, 2014) is a factorization model for the implicit case in which it models the probability of a user choosing an item by a



(a) Dimension of latent factors (b) Priors

Figure 3.8. NRMSE of models

logistic function.

Moreover, NMF_c, LMF_c, RPDI_c are the methods we apply NMF, LMF, RPDI to the aggregated region activity cube $\mathbf{Y}$ at POI category level, respectively.

3.5.4 Results of Model Performances

We evaluate our proposed model and baseline models using the NRMSE evaluation metric for a differing number of latent factors K ranging from 10 to 50. As shown in Figure 3.8 (a), we can see RPDI achieves best performance among all the models, with relatively small NRMSE obtained. While LMF performs worst among all the methods. For the models applied to $\mathbf{Y}$, they are able to obtain relatively small NRMSE, because $\mathbf{Y}$ is much smaller than $\mathbf{X}$ and easier to be factorized. With the increasing of K, the NRMSE decreases slowly. We also find that increasing the number of the latent factors beyond 50 did not improve performance on our dataset. Moreover, the priors of latent variables $(\sigma, \sigma_\alpha, \sigma_\beta, \sigma_u, \sigma_v)$ are validated with values (0.0001, 0.001, 0.01, 0.1, 1, 10, 100), and we find 0.1 achieves the best performance (as shown in Figure 3.8 (b)).

3.5.5 Results of POI Demand Ranking

In the next, we investigate the performances of our methods on ranking region POI demand in two-folds: rank the POI demand for every region, and rank the region demand for every POI category.

First, given a region, we rank the demands for POI categories. And aggregate the results for all the regions as our final result. The performances in terms of F-measure and NDCG with respect to top-k categories are shown in Figure 3.9. In the figures, we can see that RPDI achieves better overall performances than the others, outperforming the second best model NMF by 8.5%. The performances of models using category visiting probabilities are not as good as the models using POI visiting probabilities, probably due to the aggregation of category cannot reveal people's choices of POIs when going out. Among the ranking lists of all the regions, the top 10 most needed POIs are as follows: restaurant, bar, lodging, health, doctor, dentist, school, clothing store, beauty salon, cafe. We can see most of these POIs are quite related to local businesses and can bring much convenience to local residents if demands can be satisfied. Moreover, to better illustrate the ranking results for regions, we show several examples of the top-10 identified demands of regions in Table 3.2.

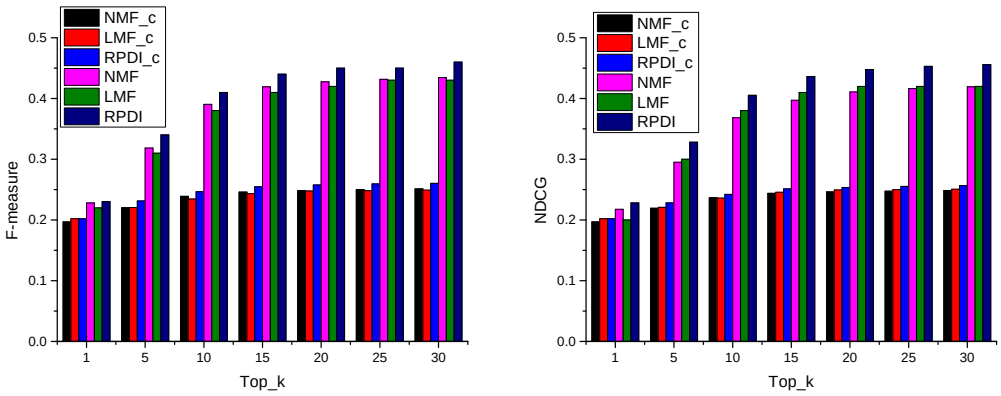
Then, given a POI category, we rank the POI demands of regions. And aggregate the results for all the POI categories as our final result. The performances in terms of F-measure and NDCG with respect to top-k regions are shown in Figure 3.10. In the figures, we can see that RPDI is still able to obtain better overall performances

Table 3.2. Identified POI demands for regions.

Region Name	**Identified Demands@top-10**	**Groundtruth**
Homecrest	'**restaurant**' 'bar' 'school' 'lodging' '**beauty salon**' '**store**' 'cafe' '**clothing store**' 'bakery' '**finance**'	'beauty salon' 'store' 'restaurant' 'clothing store' 'car repair' 'health' 'finance' 'jewelry store' 'doctor' 'hair care'
Central Harlem South	'**beauty salon**' 'dentist' '**store**' 'night club' '**health**' 'clothing store' 'electronics store' '**health**' 'general contractor' 'bank'	'school' 'beauty salon' 'store' 'gym' 'health' 'laundry' 'mosque' 'liquor store' 'doctor'
Clinton Hill	'**restaurant**' '**bar**' 'school' '**store**' '**beauty salon**' 'lodging' 'cafe' 'dentist' '**clothing store**' '**bakery**'	'beauty salon' 'store' 'bar' 'pharmacy' 'restaurant' 'bakery' 'grocery or supermarket' 'clothing store' 'car repair' 'liquor store'
Bedford	'**restaurant**' '**beauty salon**' '**school**' 'bar' 'lodging' 'cafe' '**clothing store**' 'night club' 'health' '**food**'	'school' 'beauty salon' 'food' 'store' 'restaurant' 'real estate agency' 'grocery or supermarket' 'clothing store' 'health' 'laundry'
Fordham North	'**beauty salon**' '**health**' '**clothing store**' '**restaurant**' 'moving company' 'car repair' 'finance' 'car wash' 'doctor' 'general contractor'	'beauty salon' 'restaurant' 'clothing store' 'health'

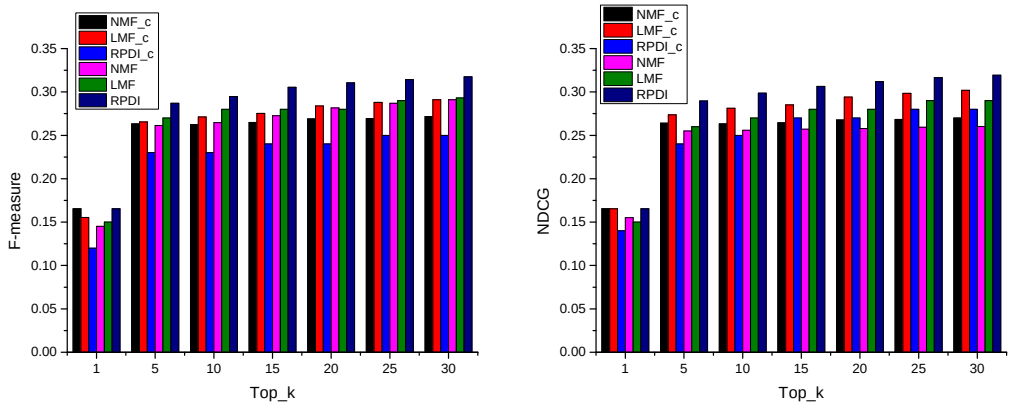
than the others. However, the ranking results are not as good as ranking for regions, since we have more regions than POI categories which makes it a harder problem. Similar to the ranking for regions, the performances of models using category visiting probabilities are not as good as the models using POI visiting probabilities.

To better illustrate the ranking results for POI categories, we show the estimated demands of two typical kinds of POIs, restaurants and health services. As shown in Figure 3.11, it is notable to see that the estimated demands of restaurants (shown



(a) F-measure (b) NDCG

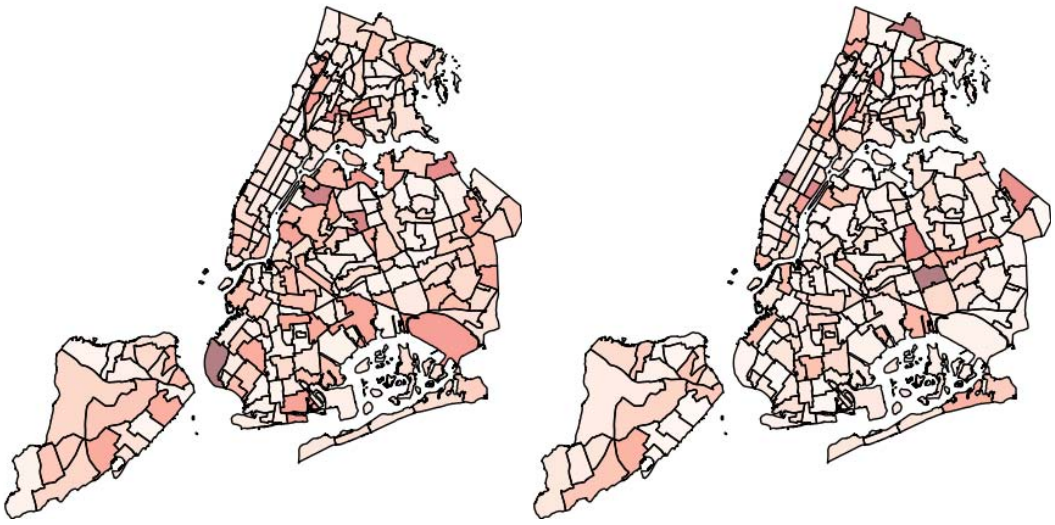
Figure 3.9. Rank categories for regions with different top-k



(a) F-measure (b) NDCG

Figure 3.10. Rank regions for categories with different top-k

in (a)) in Manhattan do not rank high among all the regions. Although Manhattan is the central area of NYC and the market of restaurants is huge, there is not much demand for new restaurants since the supply is also high. As for health services, there are higher demands in areas with lower household income in Brooklyn, Bronx, and Queens. From this point of view, the government should make efforts to allocate more health services in these areas.



(a) Restaurants (b) Health Services

Figure 3.11. Identified POI demands for categories

3.6 Related Work

In this section, we introduce relevant studies utilizing POI data and human mobility data: POI recommendation, site selection, and human mobility pattern mining.

POI Recommendation. POI recommendation, targeting at recommending the right POIs to the target users (Y. Liu et al., 2016; Lian et al., 2014; B. Liu & Xiong, 2013), can be seen as discovering POI demands for users. Unlike other recommender systems based on explicit user feedback, such as user ratings, POI recommendation is developed based on implicit user feedback, such as the check-in frequency. Recently, other implicit information, such as check-in locations, check-in time, and transition between POIs, have been exploited for POI recommendations.

Previous studies often used collaborative filtering (CF) algorithm to fuse the check-in information, e.g., user interest preferences, social influence, temporal influence and geographical influence (Q. Yuan et al., 2013; Gao et al., 2013). In (Ye et al.,

2011), Ye et al. considered the social influence under the framework of a user-based CF model, and modeled the geographical influence by a model-based method (a Bayesian CF algorithm). Moreover, Yuan et al. (Q. Yuan et al., 2013) and Gao et al. (Gao et al., 2013) introduced temporal preference to enhance the algorithm efficiency and effectiveness. The authors separated a day into different time slots and user preferences were learned for each slot, thus POIs can be recommended according to different times of a day. Cheng et al. (Cheng et al., 2012) considered more comprehensive information, such as the multi-center of user check-in patterns, and the skewed user check-in frequency. Moreover, Liu et al. (B. Liu et al., 2015) proposed a geographical probabilistic factor analysis framework to analyze the joint effects of multiple factors by considering user preference for locations as a multiplication of interest in the locations, location popularity and distance.

Different from the above works, we consider the POI demand problem at the region level. Moreover, we integrate the region supply information and demographic information into the proposed model to learn region demands more accurately.

Site Selection. Traditionally, the site selection problem, which is also called optimal placement problem, has been studied by researchers from land economy community using spatial interaction models, multiple regression discriminant analysis (Berman & Krass, 2002). In recent years, location-based services have been widely used to tackle this problem (Karamshuk et al., 2013). Karamshuk et al. selected optimal retail store location from a list of locations by using supervised learning with features mined from Foursquare check-in data. Li et al. (Li et al., 2015) studied ambulance stations site selection by using real traffic information so as to minimize the average travel time to

reach the emergency requests. Niu et al. (Niu et al., 2016) extracted discriminative features of gas stations from heterogeneous mobile data and then formalized a gas station ranking predictor to select gas station location. Xu et al. (Xu et al., 2016) proposed a framework to combine the spatial distribution of user demands with the popularity and economic attributes for optimal location selection.

Different from the above works, we consider a more general framework to identify region POI demands. In this framework, it learns demand of regions for all POIs simultaneously instead of focuses on specific POI categories like restaurant.

Human Mobility Pattern Mining. Understanding human mobility in urban environments is central to traffic forecasting, location-based services, and urban planning. A significant number of papers on human mobility analysis have been published in recent years thanks to the widely available mobility data, such as GPS data, cellular network data, and transportation data (Gonzalez, Hidalgo, & Barabasi, 2008b; Song et al., 2010).

To the best of our knowledge, we are the first to work on the problem of discovering region POI demand by leveraging human mobility patterns. Although there is no existing work on the exact application we are working on, there are many existing work on making use of human mobility patterns for different novel applications. Giannotti et al. (Giannotti, Nanni, Pinelli, & Pedreschi, 2007; Monreale, Pinelli, Trasarti, & Giannotti, 2009) developed trajectory pattern mining, and applied it to predict the next location at a certain level of accuracy by using GPS data. Zheng et al. (Y. Zheng, Liu, Yuan, & Xie, 2011) detected flawed designs in current road network with a frequent graph method on taxi GPS traces. Yuan et al. (N. J. Yuan et

al., 2015) proposed a topic-based inference model that discovers regions of different functions, such as educational areas and business districts, using both human mobility data and POIs.

3.7 Summary

In this chapter, we investigated how to exploit human mobility patterns, geographic data, and demographic data for identifying region POI demands. Along this line, we first proposed a framework, named Region POI Demand Identification (RPDI), to model POI demands with the daily needs identified from their large-scale mobility data. Specifically, in this framework, an urban space was first partitioned into spatially differentiated neighborhood regions formed by local communities. Then, the daily activity patterns of people traveling in the city were extracted from human mobility data. However, the trip activities, even aggregated, were sparse and insufficient to directly identify the POI demands, especially for underdeveloped regions. Therefore, with a proposed demand inference model considering POI preferences and supplies together with demographic features, we estimated the POI demands of all the regions simultaneously. As shown in the experimental results on real-world data, the proposed RPDI framework could provide effective POI demand identification for different regions.

CHAPTER 4

INTELLIGENT BUS ROUTING WITH HUMAN MOBILITY PATTERNS

4.1 Introduction

More and more people live in metropolitan areas with the rapid development of urbanization. One major side effect of urbanization is more frequent and intense traffic congestion due to more human activities within limited space, and consequently unnecessary energy consumption during traffic congestion. Public transportation (e.g., bus, subway) not only saves fuel and reduces congestion, but also offers a safe, affordable, and convenient way to travel (de Dios Ortuzar & Willumsen, 2011). According to American Public Transportation Association[1], Americans living in areas served by public transportation save 865 million hours of travel time and 450 million gallons of fuel annually in congestion reduction alone. A household that uses public transportation frequently save more than $9,700 every year. Although there are various and huge benefits by using public transportation, our current public transportation system is far from perfect and has much room for improvement. Better public transportation planning can significantly help to foster a more sustainable development and improve quality of life.

Traditional public transportation planning methods have relied on human surveys

[1] http://www.apta.com

to understand people's mobility patterns and their choice among different transportation modes (Aslam, Lim, Pan, & Rus, 2012) (Guihaire & Hao, 2008). Despite the substantial time and cost spent on the survey process, the macroscopic analysis based on surveys is too static to reflect the fast development of urban areas. Therefore, we need a more cost-effective and adaptive way to handle the classical transportation problem. In addition, we try to explore a new challenging problem of how to convert people who take private transportation (e.g. private car, and taxi) to take public transportation in this chapter. It is a more urgent research problem as most past research on public transportation planning focuses on how to design a system to satisfy the current need for public transportation instead of attracting more people to take public transportation. If transit agencies could have an effective tool to quantify travel demand and a choice model on how people choose public transportation and private transportation (e.g., private car, taxi), then recommendations on how to better design and optimize a given public transportation network could be proposed to attract more people to public transportation. As a result, cities would be able to better support people's travel demand through a regulated, efficient, more sustainable public transportation system.

Meanwhile, with the wide deployment of Automatic Fare Collection (AFC) systems on bus networks and Global Positioning System (GPS) devices on taxis, large amounts of bus transactions and taxi traces are collected. The availability of rich travel data and the emergence of big data technology enable the automatic analysis on human mobility patterns. We can detect the up-to-date patterns adaptively because travel data dynamically change with the development of urban areas. As

demonstrated in this chapter, this offers the possibility of optimizing public transportation by taking overall city traffic into account. By leveraging mobility patterns of public and private transportation, public transportation services can be designed in a way that accommodates different levels of demands and by doing so, attracts more potential riders and increases utilization efficiency of the public transportation system. For example, people may be more willing to choose bus over taxi when a better bus route with less travel time and stops is provided.

With integrated analysis on the mobility patterns, we aim to detect and re-plan flawed and less effective bus routes for attracting most number of potential bus riders (Y. Liu, Liu, Yuan, et al., 2017). There are two main challenges to achieve this goal: 1) modeling people's transportation mode choices for different Origin-Destination (OD) pairs; 2) optimizing bus routes with budget constraints to maximize converted bus rider number. For the first main challenge, it requires the understanding of human mobility pattern in a regional level. Since buses can only stop at bus stops and taxi can stop anywhere, it is hard to understanding human mobility patterns if their origins and destinations are not directly comparable. It requires us to integrate heterogeneous human mobility data together instead of focusing on single mode human mobility pattern. Through mapping their origins and destinations to their related regions, we are able to understand regional interactions by aggregating individual human mobility patterns. Such mapping can not only enable us work on unified features for both bus and taxi travel behaviors, but also reduce the computational complexity as the current base unit is regions instead of individuals. Therefore, the whole urban area needs to be properly partitioned into regions, and then regional travel patterns related to taxis

and buses can be modeled respectively. For the second main challenge, we only work on a sub-optimal problem of improving flawed bus routes instead of searching for a global optimized plan for some practical considerations. First, bus route optimization under different constraints has already been recognized as a complex, non-linear, non-convex, and multi-objective NP-hard problem (C. Chen et al., 2013) (C.-L. Liu, Pai, Chang, & Hsieh, 2001). Second, the global optimization is too intrusive as it changes many existing routes. Therefore, we identify flawed bus routes first, and then work on the optimization problem of improving the identified flawed bus routes.

The three main contributions of this chapter are as follows:

- **Transportation mode choice modeling**. We model transportation mode choices of different OD pairs separately using a spatio-functionally weighted regression method, providing the probabilities of taking bus and taxi for each OD pair. Note that we investigate mode choice as an aggregate problem (de Dios Ortuzar & Willumsen, 2011), which means we focus on people's group behavior of OD pairs other than individual behavior.

- **Bus routing optimization**. Given limited budgets for bus network restructuring, we propose a method to attract the maximum number of potential bus riders from private transportation.

- **Real evaluation**. We evaluate our method using a series of large-scale real GPS traces generated by 30,000 taxis and over 10 million bus transactions in Beijing from August to October in 2012. We also obtain data from the Beijing Bus Company, justifying the effectiveness of our method.

We begin by introducing related work in Section 4.7 and the preliminaries of this study in Section 4.2. Then the transportation mode choice model is proposed in Section 4.3, followed by the flawed OD pairs detection in Section 4.4 and bus routing optimization in Section 4.5. Experimental results are presented in Section 4.6, and we discuss the results and give concluding remarks in Section 4.8.

4.2 Preliminaries

We begin by introducing the routing network which provides the platform for bus routing optimization. The routing network contains bus stops and connections between them, with no bus route information included. We then generate the human mobility patterns between regions (nodes of the routing network) using taxi traces and bus transactions. These components are shown in the first (left) part of the framework in Figure 4.1. Later in the second (right) part of the framework, these mobility patterns will be modeled (in Section 4.3) to identify the factors affecting people's transportation choices. After that, we detect and optimize the flawed OD pairs with budget constraints to increase bus ridership (in Section 4.4 and Section 4.5, respectively).

Unless otherwise stated, we use bold characters to represent non-scalar variables, e.g., vectors, sequences, sets, and graphs. We use a comma in brackets to concatenate row vectors or stack column vectors horizontally, and a semicolon in brackets to concatenate column vectors or stack row vectors vertically. We use $\langle \cdot, \cdot \rangle$ to represent the inner product of two vectors.

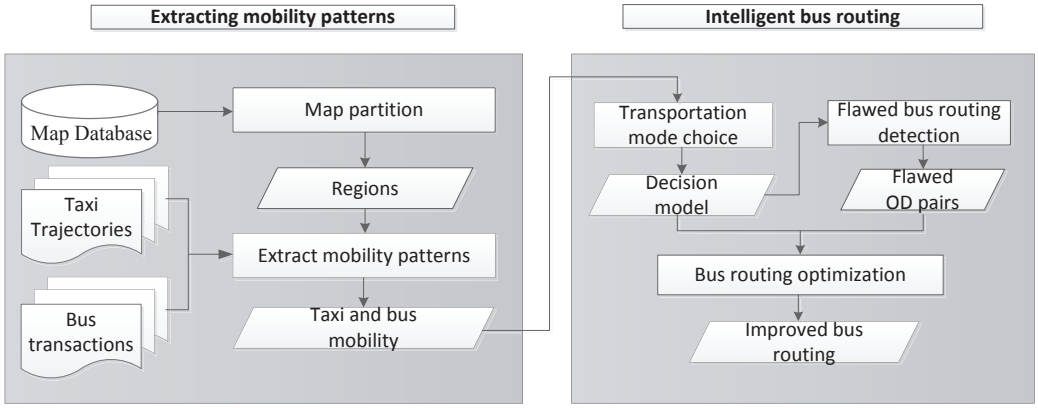


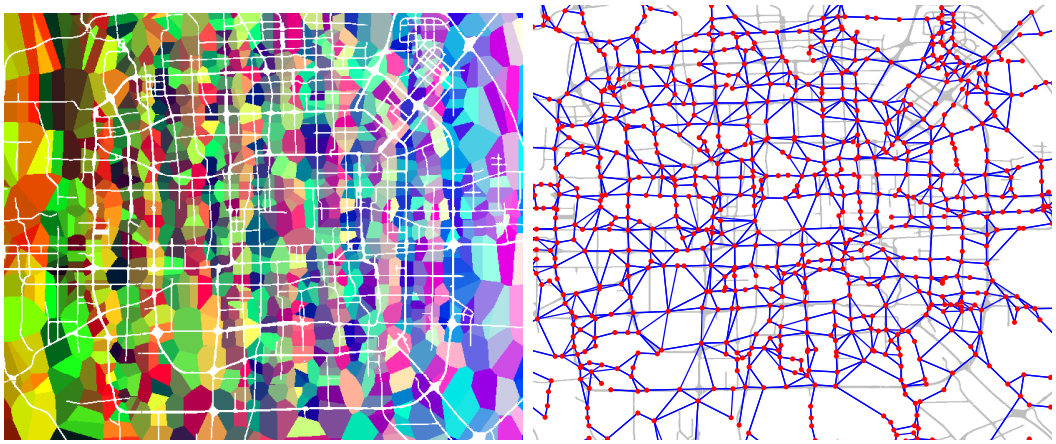
Figure 4.1. Framework of our method

4.2.1 Routing Network

As buses can only stop at bus stops and taxis can stop anywhere, we need to construct a common routing network for both buses and taxis. First, we partition the urban area into disjoint regions served by buses and taxis. Through disjoint regions, we can modify bus routes to attract the corresponding taxi passengers. To this end, we partition the urban area using bus stops $\mathbf{S} = \{s_i | i = 1, \cdots, N\}$ to align service regions for both buses and taxis. Considering duplicated bus stops on different sides of the same street, we have merged stops with same names, or stops with different names but actually share the same place. For instance, for each of the bridges (also called overpasses) there are usually two or four stops around it, e.g., Mingguang Bridge North and Mingguang Bridge South at Xueyuan Road (as shown in Figure 4.2). Buses traveling north through Mingguang Bridge will stop at Mingguang Bridge North, but not Mingguang Bridge South. So we can merge these two stops into one, which represents Mingguang Bridge. In the rest of this chapter, we assume the stops in set $\mathbf{S}$ have already been merged.

Figure 4.2. Bus stop merging of Mingguang Bridge. Mingguang Bridge North and Mingguang Bridge South are merged together to represent Mingguang Bridge.

After merging bus stops serving the same regions, we then partition the map using Voronoi diagram (Aurenhammer, 1991), which is a partitioning of a plane into regions based on distance to points in a specific subset of the plane. In our map partition problem, we treat the whole city as the plane, and bus stops as the points. With Voronoi diagram applied, the city can be partitioned into regions based on distance to bus stops. As a result, there is one region formed for each bus stop, and pick-up/drop-off points for taxi trips are mapped to the the regions located. Since we are focusing on the bus routing problem in this chapter, we assume bus stops are reasonably designed and distributed in the city. So if a person taking taxi wants to take bus instead, then the nearest bus stop will be his/her first choice. This partition method effectively describes the travel demand around bus stops comparing to other partition methods, such as grid-based partition (Sun, Yuan, Wang, Si, & Shan, 2011), road-network-based partition (Y. Zheng et al., 2011). In the following sections, we



(a) Map segmentation of Beijing (b) Bus routing network of Beijing

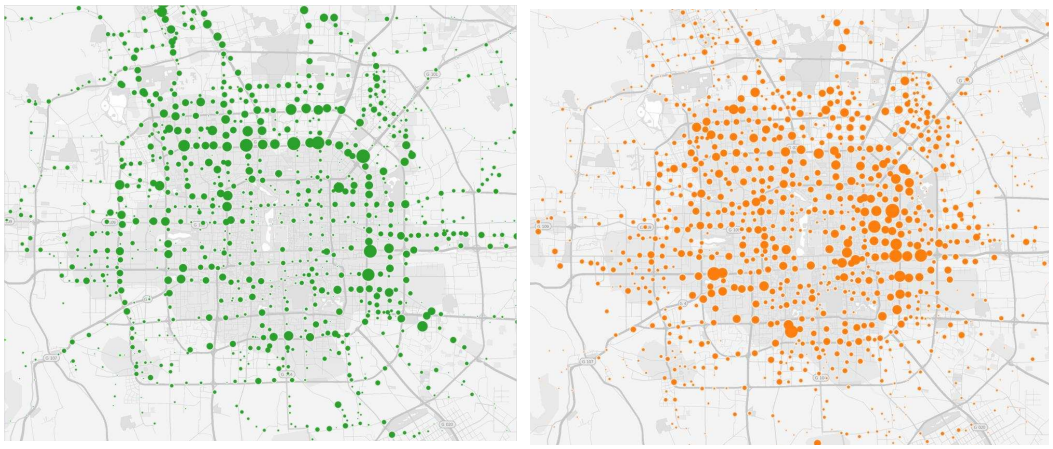
Figure 4.3. Map of Beijing

use $\mathbf{S}$ to represent both stops and their respectively associated regions. Please refer to Figure 4.3 (a) for the map segmentation in Beijing.

Now we define the routing network $\mathbf{G} = (\mathbf{S}, \mathbf{E})$ with the bus stops $\mathbf{S}$ as nodes. The edges in $\mathbf{E}$ are direct connections of neighbor bus stops which means there is a route existing from one stop to another without transiting other stops. Specifically, we have edge $\mathbf{e} = (s_i, s_j) \in \mathbf{E}$, if there is a direct road connection between the head stop s_i and tail stop s_j without traveling through other regions, where $s_i, s_j \in \mathbf{S}$. The edges are generated from existing road segments. Please refer to Figure 4.3 (b) for an example of the routing network with nodes plotted in red dots and edges in blue lines.

4.2.2 Human Mobility Pattern

The human mobility patterns contain travel information for both bus and taxi riders, representing public and private transportation respectively. As shown in Figure 4.4, there is a clear difference between the mobility patterns of these two transportation



(a) Bus trip origins (b) Taxi trip origins

Figure 4.4. Trip origin distribution of Beijing. The size of dot is proportional to the related number of trips.

modes. We retrieve these information by constructing transition records from the taxi traces and bus transactions, and then we summarize these information with a comprehensive set of statistics. Also, we have observed that, people's behaviors and thus their mobility patterns vary significantly over different days and different time periods of a day. Therefore, we apply temporal partition on the transition records before summarizing the statistics. We give details of these three steps as follows.

Transition Construction

We construct the transition records with the following definition:

Definition 1 *A transition* **tr** *contains the following attributes: origin o, destination d, transportation mode m (0 and 1 stand for taxi and bus respectively), leaving time lt, arriving time at, travel distance td, travel fare tf and number of stops sn. The set of all the transitions is notated as* **TR**.

Specifically, we project each bus and taxi trip to the nodes of the routing network

Table 4.1. Temporal slots for weekday and weekend.

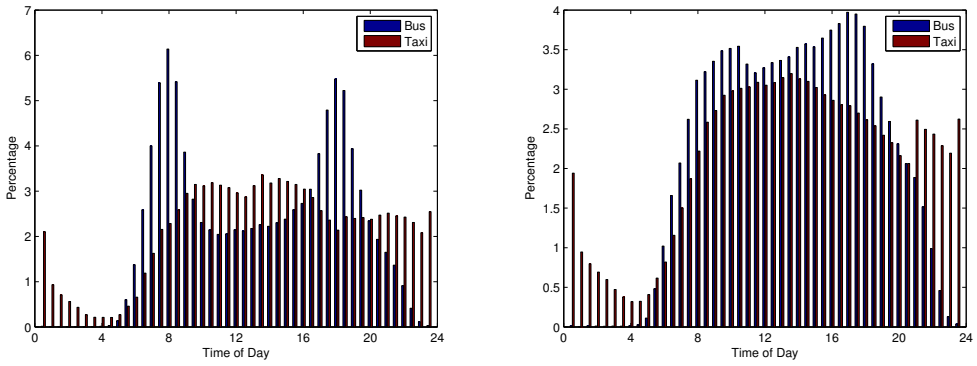
Weekday		**Weekend**	
Slot 1	5:00am-10:30am	Slot 5	5:00am-12:30pm
Slot 2	10:30am-4:00pm	Slot 6	12:30pm-7:30pm
Slot 3	4:00pm-7:30pm	Slot 7	7:30pm-11:00pm
Slot 4	7:30pm-11:00pm		

G, turning a trip into a transition. The travel distance of a taxi trip is calculated using the sum of the road distance of all consecutive GPS points in the trace, and the travel distance of a bus trip is calculated using the sum of the road distance of all consecutive bus stops traveled through.

Temporal Partition

People go to different places on weekends (including public holidays in China) in comparison with weekdays. Also, people's preferences among different transportation modes vary over different time periods of the day. For example, people prefer public transportation to commute, which usually happens during the morning and evening rush hours. Figure 4.5 (a) shows the distribution of bus and taxi riders during the day on weekdays. We can see there are two high peaks of bus riders around 8am and 6pm, which are the morning and evening rush hours. In contrast, people often prefer private transportation for business transit during the day.

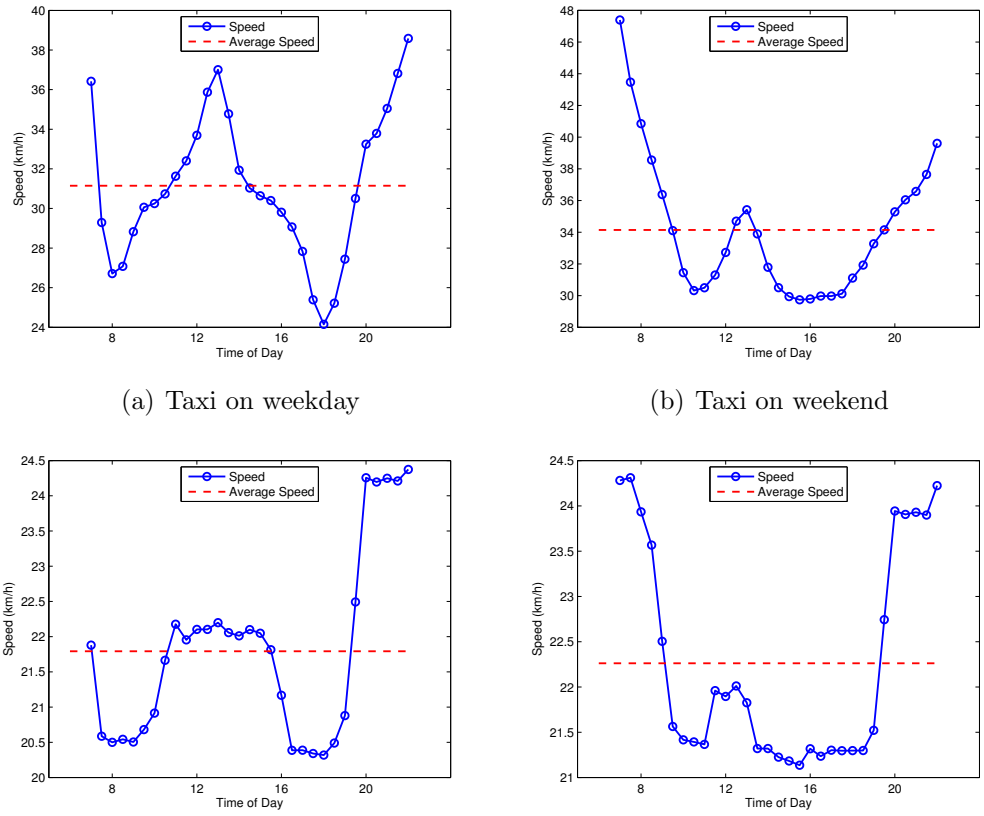
To incorporate these facts, we segment the transitions **TR** based on the leaving time lt to the temporal slots in Table 4.1, which is derived according to the traffic and travel behavior in different time of day (Y. Zheng et al., 2011). Specifically, we first segment the time of day into 48 segments, each for half an hour. By comparing the



(a) Weekday (b) Weekend

Figure 4.5. Travel behavior in Beijing. The percentages of bus and taxi riders shown in y-axis are calculated separately.

number of bus and taxi riders in each segment to the total number of bus and taxi riders in a day (as shown in Figure 4.5), and the speed in each segment to the average speed in a day (as shown in Figure 4.6), these segments can be further merged into the temporal slots presented. In the same temporal slot, the semantic meaning of people's travel are similar. Figure 4.5 shows the travel behavior of riders on weekdays and weekends, from which we can see the travel behaviors of bus and taxi differ in different time slots. For example, slot 1 corresponds to people going to work and slot 3 corresponds to people leaving from work, the number of people taking bus is much higher than other slots since bus is a major commuting transportation method. Since few people take buses between 11pm and 5am (as shown in Figure 4.5), this chapter focuses on the day bus lines, running from 5am to 11pm. We use $c = 1, \cdots, 7$ to represent the temporal slots, and each is associated with its time proportion in $STime^c$, for example $STime^1 = 5 * 5.5$ hours (5.5 hours every day and 5 days every week).



(a) Taxi on weekday

(b) Taxi on weekend

(c) Bus on weekday

(d) Bus on weekend

Figure 4.6. Traffic conditions in Beijing

Statistical Summarization

Now we summarize the partitioned transitions $\mathbf{TR}_{ij}^{c} = \{\mathbf{tr} : \mathbf{tr}.o = i, \mathbf{tr}.d = j, \mathbf{tr}.lt \in c\}$ with statistics defined in Table 4.2 for each OD pair (i, j), temporal slot c, and transportation mode *bus/taxi*, respectively. With these six statistics, which are volume, travel time, travel distance, velocity, fare, and stop number, we well depict the transportation modes and travel demands of OD pairs (Beirão & Cabral, 2007) (Redman, Friman, Gärling, & Hartig, 2013). In this chapter, we focus on improving bus routing to attract private transportation riders to public transportation, so we assume other perceived factors, such as comfort, safety, remain the same after the bus route change (Beirão & Cabral, 2007) (Redman et al., 2013). The definition of an OD pair is given as follows.

Definition 2 *An OD (Origin-Destination) pair* (o, d) *is a pair of regions with origin* $o = s_i$*, destination* $d = s_j$*, where* $s_i, s_j \in \mathbf{S}$*. We write it as* (i, j) *for short.*

Specifically, for each OD pair (i, j) and temporal slot c, we compute *BVol*, *BTime*, *BDist*, *BVel*, *BFare*, *BStop* for bus and *TVol*, *TTime*, *TDist*, *TVel*, *TFare*, *TStop* for taxi. For example, $BTime_{ij}^{c}$ is the average bus travel time of all the bus trips from origin i to destination j during the temporal slot c. In Section 4.3, we will further leverage these statistics to extract features and build the transportation mode choice models. As we have contended earlier, the mobility patterns are significantly different across different temporal slots, and for that reason, we have partitioned the records into different temporal slots. Thus here, the aforementioned statistics are summarized for each temporal slot respectively. As a result, we will build the transportation mode

Table 4.2. Statistics of transition records for OD pair (i, j) in temporal slot c.

Statistic	**Definition**
Volume	$BVol = \lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TVol = \lvert\{\mathbf{tr} : \mathbf{tr}.m = 0\}\rvert$
	$Vol = BVol + TVol$
Travel Time	$BTime = \sum_{\mathbf{tr}:\mathbf{tr}.m=1}(\mathbf{tr}.at - \mathbf{tr}.lt)/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TTime = \sum_{\mathbf{tr}:\mathbf{tr}.m=0}(\mathbf{tr}.at - \mathbf{tr}.lt)/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 0\}\rvert$
Travel Distance	$BDist = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.td/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TDist = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.td/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 0\}\rvert$
Velocity	$BVel = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.td/(\mathbf{tr}.at - \mathbf{tr}.lt)/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TVel = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.td/(\mathbf{tr}.at - \mathbf{tr}.lt)/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 0\}\rvert$
Fare	$BFare = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.tf/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TFare = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.tf/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 0\}\rvert$
Stop Number	$BStop = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.sn/\lvert\{\mathbf{tr} : \mathbf{tr}.m = 1\}\rvert$
	$TStop = 0$ (No stops for taxi)

choice model for each slot respectively.

In addition, using the transition records, we also compute some statistics of the routing network, e.g., edge distance, edge travel time, which will be used later for the bus routing optimization. Specifically, for each direct connection edge $\mathbf{e} \in \mathbf{E}$, we compute its travel distance d and travel time t for bus along the connection edge $\mathbf{e}$. We obtain d by projecting the head and tail stops of $\mathbf{e}$ to the map and calculate the shortest travel distance on the road map. To obtain the bus travel time t, we consider the travel speed v on each edge $\mathbf{e}$ obtained by using taxis as flowed sensors. Due to the speed difference between taxi and bus in different time slots, we estimate the bus speed as follows: $v^c_{bus} = \lambda^c * v^c_{taxi}$, where λ^c is a constant for temporal slot c(C. Chen et al., 2013). Different cities may have different λ, here we set $\lambda^c =< 0.68, 0.67, 0.77, 0.65, 0.61, 0.68, 0.62 >, c = 1, \cdots, 7$ for Beijing by comparing the difference between taxi and bus average speed in different temporal slots (as shown in Table 4.3). By using bus speed divided by taxi speed, we get λ in different temporal slots. It follows that $t_{bus} = t_0 + \frac{1}{\lambda} * t_{taxi} = t_0 + \frac{1}{\lambda} * d/v_{taxi}$, where t_0 is a constant indicating the time for a bus stop (C. Chen et al., 2013). Since the bus speed has already taken the stop time into account when calculating λ, we use $t_0 = 0$ minutes in this chapter. We represent all the edge travel distances in a vector $EDist^c \in \mathbb{R}^{|E|}$, and all the edge travel time in $ETime^c \in \mathbb{R}^{|E|}$, where c signifies the temporal slot when computing the statistics. As noted, these statistics can be specific for each temporal slot. Indeed, when the network is considered fixed, $EDist^c$ is invariant with respect to c; but $ETime^c$ varies along with the traffic situations in different temporal slots.

Table 4.3. Average speed (km/h) in different temporal slots.

Slot	1	2	3	4	5	6	7
Bus	21.66	21.97	20.75	24.25	22.79	21.43	24.04
Taxi	31.95	32.65	27.08	37.34	37.25	31.54	38.65
λ	0.68	0.67	0.77	0.65	0.61	0.68	0.62

4.3 Transportation Mode Choice Model

In this section, we learn a transportation mode choice model to estimate the probabilities of people taking bus given origin, destination (the OD pair) and departing time (the temporal slot). To achieve this goal, we first extract features that contribute to the decision process of choosing transportation mode. Then a spatio-functionally weighted regression model is proposed to estimate the probability of taking bus p given these features.

4.3.1 Feature Extraction

Understanding travel behavior and the reasons for choosing one transportation mode over another is an essential issue. However, travel behavior is complex. The choice of transportation mode is influenced by various factors, such as travel time, monetary cost, accessibility and reliability (Beirão & Cabral, 2007) (Ceder, 2007). Each transportation mode has its advantages and disadvantages. In general, people choose taxis because of their shorter travel distance and time, and choose buses for their lower cost. Here we focus on factors related to bus routing and consider other factors like accessibility and reliability remain unchanged.

Given the statistical summarization of an OD pair (i, j) in a temporal slot c, we

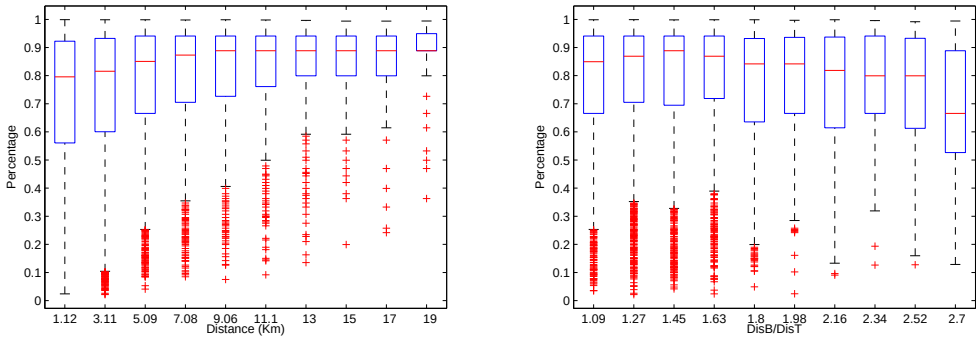
extract the features $\mathbf{X}_{ij}^c$ to better describe the OD pair and compare the difference between the transportation modes:

$$\mathbf{X}_{ij}^c = (TDist_{ij}^c, \frac{BDist_{ij}^c}{TDist_{ij}^c}, TTime_{ij}^c, \frac{BTime_{ij}^c}{TTime_{ij}^c}, \\ TFare_{ij}^c, \frac{BFare_{ij}^c}{TFare_{ij}^c}, \frac{BStop_{ij}^c}{TDist_{ij}^c}),$$

where details and our motivations are given as follows.

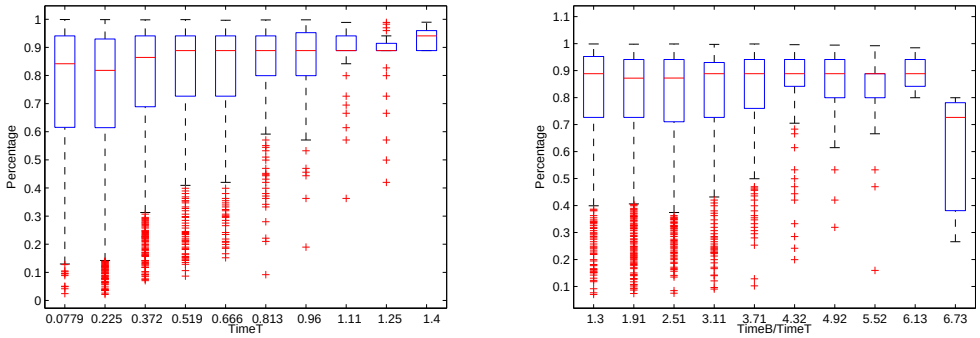
Distance related features. Distance influences people's choice in an intuitive way. It is usually the first factor that comes to mind when traveling, e.g., how far is the destination from the origin. In this chapter, distance related features include two parts: shortest road distance and distance ratio of buses and taxis. Here we use *TDist* to represent the shortest road distance of the OD pair, since it stands for the choice of experienced drivers which is usually the best in real. As shown in Figure 4.7 (a), with the increasing of distance of OD pairs, the percentage of people taking a bus is also increase. On the other hand, the ratio of the travel distance of buses and taxis $BDist/TDist$ describes the difference between these two. A larger $BDist/TDist$, which is larger than 1, indicates a longer travel distance by bus than taxi. As shown in Figure 4.7 (b), with the increase of $BDist/TDist$, the percentage of people taking bus is decrease.

Time related features. After the distance is determined, people consider time constraints. Usually one travels in a limited time, for which he/she has to choose a proper transportation mode that satisfies their time constraints. For example, if he/she is in a hurry, he/she will probably choose taxi over bus. Similar to distance



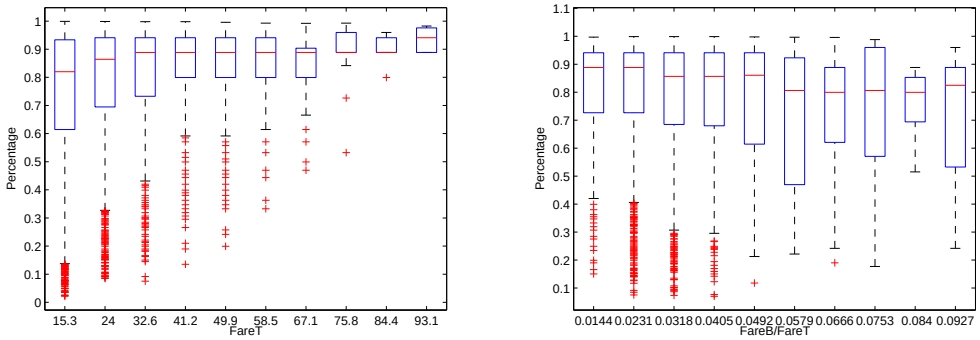
(a) Travel distance of taxi (b) Difference of travel distance

Figure 4.7. Trip distribution wrt. distance



(a) Travel time of taxi (b) Difference of travel time

Figure 4.8. Trip distribution wrt. time



(a) Taxi fare (b) Difference of fare

Figure 4.9. Trip distribution wrt. fare

related features, the time related features include two parts: the travel time of taxi and travel time ratio of buses and taxis. As shown in Figure 4.8 (a), with the increase of travel time of OD pairs, the percentage of people taking bus also increases. Here we use $TTime$ as a baseline for the travel time of OD pair, and the travel time ratio of bus and taxi $BTime/TTime$ describes the difference of these two. A larger $BTime/TTime$, which is larger than 1, indicates a longer travel time by bus than taxi. As shown in Figure 4.8 (b), with the increase of $BTime/TTime$ of OD pair, the percentage of people taking bus decreases.

Fare related features. Monetary cost is another factor people need to consider. As shown in Figure 4.9 (a), with the increase in fare of OD pairs, the percentage of people taking bus increases. That's because for long distances the taxi fare is much higher than bus. When the taxi fare is fixed, with the fare ratio of bus and taxi $BFare/TFare$ increasing, we can see from Figure 4.9 (b) that the number of people taking bus decreases.

Stop number related features. Too many stops will affect the riding experience of a trip, not only is the stop a waste of time, but waiting is also an unpleasant process. One main advantage of a taxi is that it has no stop in the middle of a trip, while a bus has many stops. In this chapter, we use the bus stop number per kilometer $BStop/TDist$ to evaluate whether it affects people's decisions to choose the bus. As shown in Figure 4.10, with an increase in $BStop/TDist^T$, the percentage of people taking bus drops quickly.

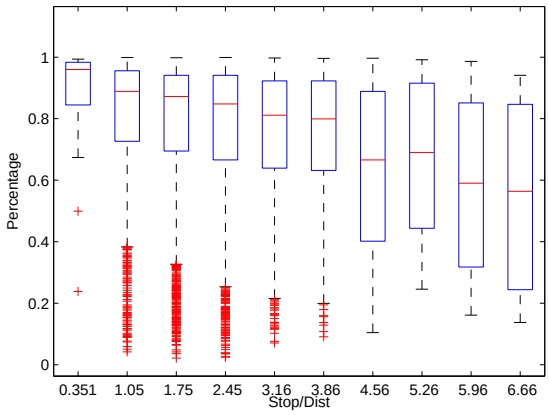


Figure 4.10. Trip distribution wrt. stop number

4.3.2 Spatio-functionally Weighted Regression

Given the features $\{\mathbf{X}_{ij}^c\}$, and the historical trip numbers of buses and taxis, we propose a spatio-functionally weighted logistic regression model (SFWLoR) to connect the features and people's transportation mode choices. First, for a given temporal slot c and an OD pair (o, d), we build a regression model between the probability of taking bus and the features as $\hat{p}_{od}^c = f(\langle \mathbf{X}_{od}^c, \mathbf{W}_{od}^c \rangle)$, where $\mathbf{W}_{od}^c$ is the model coefficient vector to be estimated. Since we want to estimate a probability distribution, we use the prediction function $f(z) = \frac{1}{1+\exp(-z)}$, which leads our model to logistic regression. Then, the regression model is locally fitted with all the observations $\{(\mathbf{X}_{ij}^c, p_{ij}^c) : s_i, s_j \in \mathbf{S}\}$, where p_{ij}^c is the observed probability of taking bus from origin s_i to destination s_j in temporal slot c, estimated with historical transition records. By fitting the model we obtain $\mathbf{W}_{od}^c$ which minimize the model error. After the coefficients $\mathbf{W}_{od}^c$ have been obtained, we can use the fitted model to predict the probability of taking bus from s_o to s_d with given route in the future. Finally, we repeat the above steps to learn $\mathbf{W}_{od}^c$ for each OD pair (o, d) in each temporal slot c,

where $s_o, s_d \in \mathbf{S}$.

The motivation of our proposed SFWLoR is as follows. We note that transportation mode preferences vary over different temporal slots as well as different OD pairs, due to differences in trip purpose and lifestyle. Indeed, different regions have different functions (J. Yuan, Zheng, & Xie, 2012), and the preferences of people from residential areas to commercial areas may differ from that of people from commercial areas to residential areas. On the other hand, travel preferences are more likely to be the same if two region pairs are near each other, sharing similar functions and lifestyles. As shown in Figure 4.11, we have three OD pairs (o_1, d_1), (o_2, d_2), and (o_3, d_3). When learning $\mathbf{W}_{o_1 d_1}$, we use the observations from (o_2, d_2), and (o_3, d_3). However, d_1 and d_2 both locate in university areas, while d_3 locates in bar area. The traveling purposes of (o_1, d_1) would probably more similar to (o_2, d_2) than (o_3, d_3). In order to better learn the traveling behavior of (o_1, d_1), we should assign more weight on observation of (o_2, d_2) than (o_3, d_3). Other than SFWLoR, a spatio-functionally weighted linear regression model (SFWLiR) which adopts linear regression instead of logistic regression is proposed for more efficient computation.

In these weighted models, we learn $\mathbf{W}^c_{od}$ specifically for each OD pair (o, d), with all the observations $\{(\mathbf{X}^c_{ij}, p^c_{ij}) : s_i, s_j \in \mathbf{S}\}$. However, we have different weights $\omega^{(ij)}_{od}$ for each observation (i, j) when estimating $\mathbf{W}^c_{od}$ which minimizes the total loss $\sum_{ij} \omega^{(ij)}_{od} Loss(p^c_{ij}, f(\langle \mathbf{X}^c_{ij}, \mathbf{W}^c_{od} \rangle))$ (Strutz, n.d.). $Loss(\cdot, \cdot)$ is the loss function of regression for each observation.

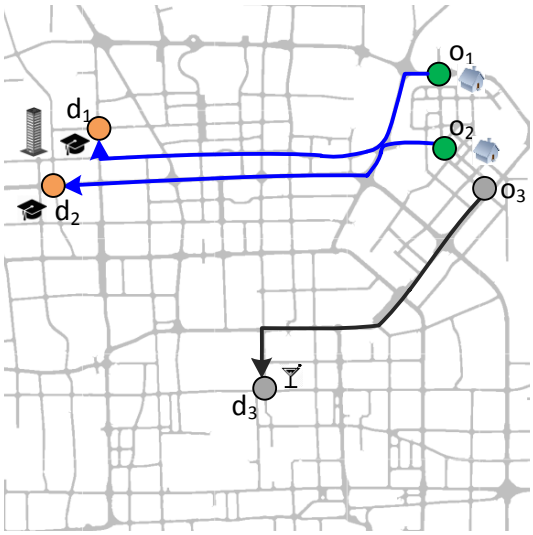


Figure 4.11. Weighted regression example. Among the three OD pairs, (o_1, d_1) and (o_2, d_2) are more similar than with (o_3, d_3). When modeling (o_1, d_1), a higher weight should be assigned to (o_2, d_2) than (o_3, d_3).

The *observation weight* of (i, j) for target OD pair (o, d) is defined as

$$\omega_{od}^{(ij)} = \exp(-\frac{\alpha_{od}^{(ij)}}{2h_\alpha}) \cdot \exp(-\frac{\beta_{od}^{(ij)}}{2h_\beta}) = \exp(-\frac{\alpha_{od}^{(ij)}}{2h_\alpha} - \frac{\beta_{od}^{(ij)}}{2h_\beta}), \tag{4.1}$$

where h_α, h_β are parameters that control the scaling at which the weights are computed, $\alpha_{od}^{(ij)}$ is the spatial distance of (i, j) and (o, d), and $\beta_{od}^{(ij)}$ is the functional distance of (i, j) and (o, d). With higher distance between (i, j) and (o, d), (i, j) will have lower weight when fitting the model. These two distances are calculated as follows.

We evaluate the spatial distance of (i, j) and (o, d) by comparing the travel distances of origin regions i and o, destination regions j and d, separately. Then use the

average of these two distances as the spatial distance of (i, j) and (o, d).

$$\alpha_{od}^{(ij)} = \frac{\text{dist}(s_i, s_o) + \text{dist}(s_j, s_d)}{2}, \tag{4.2}$$

where $\text{dist}(s_i, s_j)$ is the Euclidean distance between the bus stops in s_i and s_j.

Since each POI serves certain function, thus region function is highly related to the POI distributed in this region. Here we measure the functional distance of two regions by comparing difference of POI distributions in these regions.

$$\beta_{od}^{(ij)} = \frac{\text{dcos}(s_i, s_o) + \text{dcos}(s_j, s_d)}{2}, \tag{4.3}$$

where $\text{dcos}(s_i, s_j)$ is the cosine distance calculated by

$$\text{dcos}(s_i, s_j) = 1 - \frac{\mathbf{n}_i \cdot \mathbf{n}_j}{\|\mathbf{n}_i\| \cdot \|\mathbf{n}_j\|}.$$

The vector $\mathbf{n}_i =< n_1, n_2, ..., n_k >$ contains the POI distribution of the i-th region, and k is the number of POI category. More details about POI information are given in Section 4.6.1.

Note that the observation weight can also be extended by adding other distances of regions if found to be impacting the choice of transportation mode.

4.4 Flawed OD Pair Identification

In this section we detect flawed OD pairs with which bus routing is problematically designed. People may have to take a long detour traveling with the bus routing

or even there is no bus routes travel through two regions with high travel demand. People would like to take taxi other than take bus in these bus routes because bus is so inconvenient. Here we first detect the flawed OD pairs with problematical bus routing and further improve them in the next section.

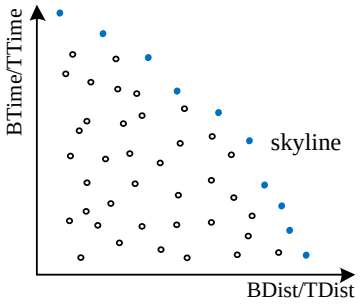
4.4.1 Skyline Patterns

Skyline detection method is used here to find the flawed OD pairs for every time slot separately. Then they are combined together as the flawed OD pair set.

As stated in Section 4.3, $BDist/TDist$, $BTime/TTime$, $BStop/TDist$ will model the connectivity and the accessibility between two regions through bus comparing to taxi, and $BFare/TFare$ will model the monetary cost between them. Specifically, $BDist/TDist$, $BTime/TTime$, and $BStop/TDist$ capture the property of the connection between an OD pair. A region pair with a big $BDist/TDist$ or $BTime/TTime$ means people have to take a long detour traveling from one region to the other, or they have to travel through congested road segments. A big $BStop/TDist$ means people have to stop many times during the trip which very likely will degrade the rider experience. In this step, we aim to retrieve the OD pairs with a big $BDist/TDist$, a big $BTime/TTime$, and a big $BStop/TDist$ which indicate problematic bus routing.

We first select the region pairs having the number of transitions above the average. Then, we find the skyline set from these selected region pairs according to above features, using skyline operator (Borzsony, Kossmann, & Stocker, 2001).

Definition 3 *The skyline is defined as those points which are not dominated by any other point. A point dominates another point if it is as good or better in all dimensions*



(a) 2-dimensional skyline

OD pair	BDist/TDist	BTime/TTime	BStop/TDist
1	1.4	2.6	1.6
2	1.3	3.7	1.8
3	1.0	3.5	1.4
4	2.2	3.0	1.0
5	1.8	2.4	2.4
6	2.4	3.4	1.4
7	1.3	2.3	2.0
8	1.6	3.2	2.2

(b) Searching skyline

Figure 4.12. An example of skyline detection.

and better in at least one dimension.

Specifically in our problem, each OD pair (i, j) is not dominated by others, in terms of $BDist/TDist$, $BTime/TTime$, and $BStop/TDist$. That is, there is no OD pair having a bigger $BDist/TDist$, $BTime/TTime$, and $BStop/TDist$ than (i, j). Figure 4.12 (a) depicts an example of the skyline set in a two-dimensional axis where a point denotes a OD pair. Clearly, no blank points simultaneously have a bigger $BTime/TTime$ and a bigger $BDist/TDist$ than the skyline points in blue. Figure 4.12 (b) shows an example of searching the skyline. OD pair 1 is not considered as skyline because it is dominated by OD pair 8. However, point 2 is not dominated by point 8 as point 2 has a bigger $BTime/TTime$ than point 8. Likewise, point 5, 6 and 8 are detected as the skyline while point 3, 4, and 7 are dominated by the skyline.

Note we want to find the region pairs with urgent needs to improve the bus service rather than all the flawed region pairs. Seeking the skyline from the region pairs with a large volume of trips, we guarantee the detected skyline is related to many people's travel and each statistic is calculated based on a large number of observations.

4.4.2 Candidate Selection

With all the flawed OD pairs detected, we further select top K OD pairs which can attract most riders as the candidates to be optimized. People traveling between these OD pairs have a relatively low probability of taking bus and can be improved dramatically after the bus routing rework.

Routes traveled by taxi usually indicate the practically best driving directions (J. Yuan, Zheng, Xie, & Sun, 2013). It is reasonable for us to use the travel route of taxi for each flawed OD pair (i, j) as the upper bound of the bus route. Then, with the travel route of taxi $R^c_{T,ij}$ in temporal slot c, we can derive the features $\mathbf{X}^c_{ij}$ of R^c_{ij}. Finally, with the above information and the transportation mode choice model, we are able to calculate the upper bound of probability of taking bus for every flawed OD pair.

For all the flawed OD pairs, we rank them in descending order according to the potentially increased bus rider number, which is calculated as follows:

$$\Delta BVol_{ij} = \sum_c Vol^c_{ij} \times (f^c_{ij}(R^c_{T,ij}) - p^c_{ij}), \tag{4.4}$$

and top K flawed OD pairs will be selected as candidates for bus routing optimization. Moreover, we compute $f^c_{ij}(R^c_{T,ij}) = f(\langle \mathbf{X}^c_{T,ij}, \mathbf{W}^c_{ij} \rangle)$ as proposed in Section 4.3. $\mathbf{W}^c_{ij}$ is the learned coefficient vector, and later we will show how to derive the features $\mathbf{X}^c_{T,ij}$ with the route $R^c_{T,ij}$.

4.5 Bus Routing Optimization

Routing refers to the specifics of bus service alignment based on certain objective functions and a set of constraints, both as individual routes and as a system of routes working together (Pratt, Evans, et al., 2004). In this section we start by formulating a bus routing optimization problem, in light of the transportation mode choice model in Section 4.3. Following is our proposed solution to this problem.

4.5.1 Problem Formulation

A general problem formulation. One main goal of bus routing optimization is to accommodate bus travel demand (Pratt et al., 2004). In this chapter with the transportation mode choice model, we can estimate bus demand dynamically for different routing results, which further allows us to both accommodate and maximize bus travel demand.

Specifically, we denote a bus route by a sequence of bus stops $(\cdots, s_i, \cdots)$ and we search for the optimal bus routes which maximize the total number of bus riders. Given OD pair (o, d) and the transportation mode choice model, one optimized routing $R_{od} = (s_o, \cdots, s_i, \cdots, s_d)$ maximizes the bus riders of all stops traveled. In other words, R_{od} is the solution maximizing the objective function:

$$\mathcal{F}(R_{od}) = \sum_{c} \sum_{\substack{(s_i, s_j) \in R_{od} \\ s_i \prec s_j}} Vol_{ij}^{c} \times f_{ij}^{c}(R_{ij}), \tag{4.5}$$

where $(s_i, s_j) \in R_{od}$ and $s_i \prec s_j$ indicate R_{od} passes s_i earlier and s_j later, $R_{ij} = (s_i, \cdots, s_j)$ is the sub-route of R_{od} from stop s_i to s_j. Take OD pair (o_1, d_1) in Figure

4.13 (a) for example, to get the route (o_1, s_1, s_2, d_1) as the optimal route, we need to maximize the riders taking bus for the following six OD pairs: (o_1, s_1), (o_1, s_2), (o_1, d_1), (s_1, s_2), (s_1, d_1), and (s_2, d_1) (drawn in green dashed lines). Moreover, we compute $f_{ij}^c(R_{ij}) = f(\langle \mathbf{X}_{ij}^c, \mathbf{W}_{ij}^c \rangle)$ as proposed in Section 4.3. And later we will show how to derive the features $\mathbf{X}_{ij}^c$ with the route R_{ij}.

Bus routing for network renewal. This problem can be well fitted into new bus route design, where there previously were no bus routes. However, in this chapter we aim to rework the existing bus routing, in which case, it is unnecessary to change well-designed bus routes but only flawed ones. As shown in Figure 4.13 (b), to find an optimal route for OD pair (o_1, d_1) we now only need to consider the bus travel demand between (o_1, d_1). Hence, the objective function for optimizing a flawed OD pair (o, d) is to maximize the converted bus rider number of (o, d), which is

$$\begin{aligned}\mathcal{F}(R_{od}) &= \sum_c (\, Vol_{od}^c \times f_{od}^c(R_{od}) - BVol_{od}^c) \\ &= \sum_c Vol_{od}^c \times f_{od}^c(R_{od}) - \sum_c BVol_{od}^c.\end{aligned}$$

Note that $\sum_c BVol_{od}^c$ stands for the current bus rider number which is a constant. Therefore our objective function is equal to

$$\mathcal{F}(R_{od}) = \sum_c Vol_{od}^c \times f_{od}^c(R_{od}). \tag{4.6}$$

Bus routing optimization with constraints. Furthermore, in a real application of bus routing optimization, multiple flawed OD pairs need to be considered simulta-

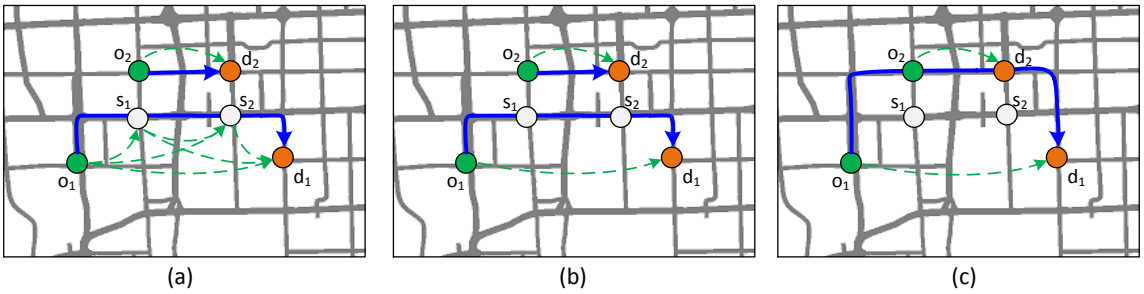


Figure 4.13. Routing optimization comparison, (a) general bus routing; (b) bus routing for network renewal; (c) bus routing with constraints. Blue solid line stands for bus route, while green dashed line stands for bus travel demand of OD pair.

neously due to various constraints. For instance, the bus company (or government) is constrained by a limited budget which does not always allow for implementation of the identified optimal transit solution and service design. As an example, we have two flawed OD pairs as shown in Figure 4.13: (b) shows the routing result when optimize the two OD pairs independently, leading to two routes which exceed the budget constraints; (c) shows the routing results using a multiple optimization method, leading to one route under the budget constraint.

Following this line, the bus routing optimization problem is formulated as follows. Given choice models f_k^c from Section 4.3 (i.e., f_k^c parameterized by $\mathbf{W}_k^c = \mathbf{W}_{o_k d_k}^c$), for each flawed OD pair (o_k, d_k), $k = 1, \cdots, K$, we optimize the total bus ridership under budget constraints. Supposing the optimal bus routes are $\mathbb{R} = \{R_k : k = 1, \cdots, K\}$, where R_k has an origin o_k and a destination d_k, our objective function is as follows,

$$\mathcal{F}(\mathbf{R}) = \sum_c \sum_k Vol_k^c \times f_k^c(R_k). \tag{4.7}$$

where $Vol_k^c = Vol_{o_k d_k}^c$. As stated previously, $f_k^c(R_k) = f(\langle \mathbf{X}_k^c, \mathbf{W}_k^c \rangle)$ and we will show

how to derive the features $\mathbf{X}_k^c$ of R_k in Section 6.2.

We consider multiple budgets (e.g., total route length, total service time) under the following constraints:

$$\text{cost}(\mathbb{R}) \leq \mathbf{C}, \tag{4.8}$$

where the function *cost* is calculated with all the bus routes in $\mathbb{R}$, and the budgets allowed to stay within are defined in vector $\mathbf{C}$. Note that when there is no budget constrain or the budgets are large enough, the above problem becomes an independent routing problem for each OD pair.

4.5.2 Problem Solution

To find the optimal route, we consider the routing network $\mathbf{G} = (\mathbf{S}, \mathbf{E})$. For each edge $e = (i, j) \in \mathbf{E}$ connecting the bus stops s_i and s_j, we define $R_k \in \mathbb{R}^{|\mathbf{E}|}$, where $R_{ke} = 1$ if and only if route R_k passes edge e, and $R_{ke} = 0$ otherwise. Also, for each bus stop $s \in \mathbf{S}$, we define $in(s) = \{(s', s) \in \mathbf{E}\}$ and $out(s) = \{(s, s') \in \mathbf{E}\}$ as the incoming edges and outgoing edges of s. To ensure the route has and only has one origin and one destination, also no loop exists, for route $R_k, k = 1, \cdots, K$, we have

$$\sum_{e \in out(o_k)} R_{ke} = \sum_{e \in in(d_k)} R_{ke} = 1,$$

$$\sum_{e \in in(o_k)} R_{ke} = \sum_{e \in out(d_k)} R_{ke} = 0,$$

$$\sum_{e \in out(s)} R_{ke} = \sum_{e \in in(s)} R_{ke}, \forall s \neq o_k, d_k.$$

Should be noted that, R_k passes bus stop s if and only if $\sum_{e \in out(s)} R_{ke} = \sum_{e \in in(s)} R_{ke} = 1$ for $s \neq o_k, d_k$.

Given route R_k, we need to get the features $\mathbf{X}_k^c$ of it to calculate the probability of people taking bus using our mode choice model. The bus related features can be aggregated from all the edges belong to R_k, while the taxi related features remain the same thus can be obtained from historical data. Therefore, to derive the features $\mathbf{X}_k^c$ for route R_k at temporal slot c, we have

$$BDist_{R_k}^c = \sum_e R_{ke} \times EDist_e^c = \langle EDist^c, R_k \rangle$$
$$BTime_{R_k}^c = \sum_e R_{ke} \times ETime_e^c = \langle ETime^c, R_k \rangle$$
$$BStop_{R_k}^c = \sum_s \sum_{e \in out(s)} R_{ke} = \sum_e R_{ke} = \langle \mathbf{1}, R_k \rangle$$

where $EDist^c, ETime^c \in \mathbb{R}^{|\mathbf{E}|}$ are travel distance and time on edges (introduced in Section 4.2), and $\mathbf{1} \in \mathbb{R}^{|\mathbf{E}|}$ is a row vector of ones. We will also use $\mathbf{0} \in \mathbb{R}^{|\mathbf{E}|}$ as a row vector of zeros.

By letting

$$A_k^c = (\mathbf{0}; \frac{1}{TDist_k^c} EDist^c; \mathbf{0}; \frac{1}{TTime_k^c} ETime^c; \mathbf{0}; \mathbf{0}; \frac{1}{TDist_k^c} \mathbf{1}),$$
$$B_k^c = (TDist_k^c; 0; TTime_k^c; 0; TFare_k^c; \frac{BFare_k^c}{TFare_k^c}; 0),$$

we obtain all the features of R_k as $\mathbf{X}_k^c = A_k^c R_k + B_k^c$.

For the constraints, we limit the service route length and driving time introduced

per unit time by the overall routing R on all traveled edges. By letting service waiting time $WTime_k^c$ be the time interval between two consecutive buses of route R_k at temporal slot c, if $WTime_k^c = WTime^c$, $\forall k = 1, \cdots, K$, this cost can be written as

$$\text{cost}(\mathbb{R}) = \sum_c \frac{STime^c}{WTime^c} \sum_e [\sum_k R_{ke} > 0] ECost_e^c,$$

where $ECost_e^c = (EDist_e^c; ETime_e^c)$ is a two dimensional column vector encoding both the travel distance and time on edge $\mathbf{e}$. A relaxed calculation which avoids the boolean test operator ($[\cdot]$) can be formulated as

$$\begin{aligned} \text{cost}(\mathbb{R}) &= \sum_c \frac{STime^c}{WTime^c} \sum_e \sum_k R_{ke} ECost_e^c \\ &= \sum_c \sum_k \frac{STime^c}{WTime_k^c} BCost_{R_k}, \end{aligned}$$

where $BCost_{R_k} = (BDist_{R_k}^c; BTime_{R_k}^c)$ encodes the route travel distance and time of R_k. As noted, this also allows us to calculate different waiting time for different bus routes. Since we do not focus on the scheduling of bus, we use 15 minutes as the waiting time for all bus routes in this chapter. In sum, our constraints in Equation 4.8 can be linear with respect to the decision variables in R. However, the objective in Equation 4.7 is non-linear with the prediction function $f(z) = \frac{1}{1+\exp(-z)}$, and the consequent optimization problem is non-convex and the gradient directed searching will result only a local optimal. We also exploit the choice model with a linear prediction function $\tilde{f}(z) = z$, which leads to a constrained linear programming problem. In experiments, we will show results of the routing optimization with both

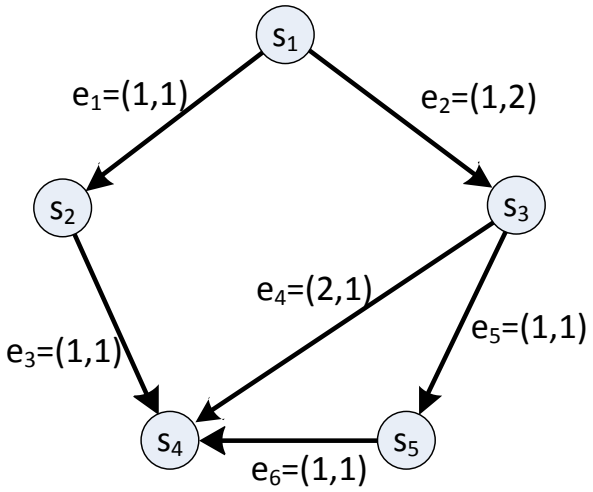


Figure 4.14. A sample bus network. The edge distance and time are shown as $e = (EDist, ETime)$.

non-linear and linear prediction functions, and it can be seen that the relaxed linear approach can approximate optimal routing effectively.

An example. To help understand the optimization process, we take the following sample network (shown in Figure 4.14) as an example. In this network, it has five bus stops $\mathbf{S} = \{s_1, s_2, s_3, s_4, s_5\}$, and six edges $\mathbf{E} = \{e_1 = (1, 2), e_2 = (1, 3),$ $e_3 = (2, 4), e_4 = (3, 4), e_5 = (3, 5), e_6 = (5, 4)\}$ between them. Now we want to find an optimized bus route R_{14} for (s_1, s_4) (which can be shorten as R_k given $(o_k = s_1, d_k = s_4)$) in one temporal slot, with constraints as $\sum BDist \leq 3$ and $\sum BTime \leq 3$. Since we only have one OD pair in one temporal slot to optimize, the objective function becomes $\mathcal{F}(\mathbf{R}) = Vol_k \times f(R_k)$. Vol_k is a constant, so we only need to find a route from s_1 to s_4 with highest $f(R_k)$.

We have three candidate routes for this kth OD pair (s_1, s_4), which are $R'_k = (e_1 = 1; e_2 = 0; e_3 = 1; e_4 = 0; e_5 = 0; e_6 = 0)$, $R''_k = (e_1 = 0; e_2 = 1; e_3 = 0; e_4 = 1; e_5 =$

$0; e_6 = 0)$, and $R'''_k = (e_1 = 0; e_2 = 1; e_3 = 0; e_4 = 0; e_5 = 1; e_6 = 1)$. For candidate route R'_k, we have

$$\mathbf{X}'_k = A_k R'_k + B_k$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/1 & 1/1 & 1/1 & 2/1 & 1/1 & 1/1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/1 & 2/1 & 1/1 & 1/1 & 1/1 & 1/1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/1 & 1/1 & 1/1 & 1/1 & 1/1 & 1/1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 10 \\ 1/10 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 10 \\ 1/10 \\ 2 \end{bmatrix}$$

given $(TDist_k, TTime_k, TFare_k, BFare_k) = (1, 1, 10, 1)$ which are constants. With $\mathbf{W}_k$ trained from our SFWLoR model, we have $z' = \langle \mathbf{X}'_k, \mathbf{W}_k \rangle$, and $f(R'_k) = \frac{1}{1+\exp(-z')}$. Similarly, for candidate route R''_k we can also get $f(R''_k) = \frac{1}{1+\exp(-z'')}$, and candidate route R'''_k has travel time 4 which exceeds our constraint and will be excluded. Finally we choose R'_k as R_k if $f(R'_k) > f(R''_k)$.

4.5.3 Computation Details

In general, the resultant integer programming is NP-complete. However, since we optimize only the most flawed OD pairs instead of the overall bus routing, the problem is of a reasonable scale and it turns out that the branch-and-bound algorithm (Land & Doig, 1960) can solve the problem efficiently for flawed bus routing in Beijing. In the more general cases, we can also relax the binary requirements to $R_{ke} \in [0, 1]$, which can be interpreted as the probability of route R_k passing edge $\mathbf{e}$. A solution of the relaxed problem signifies how we should route the bus from origin to destination, so that the maximum transportation needs are satisfied by the bus service. To recover the solution for the un-relaxed problem, we can iteratively remove the edge with the smallest probability, until there is a unique route for an OD pair.

More details of the solution are provided as follows. With the prediction function $\tilde{f}(z) = z$, our bus routing optimization problem can be written as:

$$\begin{aligned} \max \quad & \langle A, R \rangle + B \\ \text{s.t.} \quad & \mathbf{0} \leq R \leq \mathbf{1} \\ & \langle P, R \rangle \leq p \\ & \langle Q, R \rangle \leq q \\ & LR = r \end{aligned}$$

Here, $R = (R_1; \cdots; R_K)$ is the vector of all routes to be optimized. A, B, P, Q, L, r are constant matrices constructed with the observed data. p, q are user-specified

parameters on the budget constraints, where p is the maximum of service distance while q is the maximum of service time, per unit time respectively. To be specific,

$$A = (A_1; \cdots ; A_K),$$

$$B = \sum_c \sum_k Vol_k^c \langle B_k^c, \mathbf{W}_k^c \rangle,$$

$$P = (P_1; \cdots ; P_K),$$

$$Q = (Q_1; \cdots ; Q_K),$$

$$L = \text{diag}(\mathcal{L}, \cdots , \mathcal{L}),$$

$$r = (r_1; \cdots ; r_K).$$

Here, $A_k = \sum_c Vol_k^c (A_k^c)' \mathbf{W}_k^c$, $P_k = \sum_c \frac{STime^c}{WTime_k^c} EDist^c$, $Q_k = \sum_c \frac{STime^c}{WTime_k^c} ETime^c$, and $A_k, P_k, Q_k \in \mathbb{R}^{|\mathbf{E}|}$. The matrix $\mathcal{L}$ represents the graph $\mathbf{G}$ (defined in Section 4.5) with rows corresponding to nodes (bus stops) and columns corresponding to edges: for $\mathbf{e} = (i, j)$, we let $\mathcal{L}_{ie} = -1$, $\mathcal{L}_{je} = 1$, and $\mathcal{L}_{ke} = 0$ for $k \neq i, j$. r_k is a vector of all 0's except of 1 at o_k and of -1 at d_k.

With these notations, the problem can be solved by calling the MATLAB function:

$$linprog(-A, [P'; Q'], [p; q], L, r, \mathbf{0}, \mathbf{1}).$$

This procedure relaxes the binary constraints on R to be $\mathbf{0} \leq R \leq \mathbf{1}$. To solve the problem without relaxation, one can run:

$$bintprog(-A, [P'; Q'], [p; q], L, r).$$

As for the prediction function $p(z) = f(z) = \frac{1}{1+\exp(-z)}$ which leads the transportation mode choice model to spatio-functionally weighted logistic regression, the objective function for bus routing optimization is:

$$\mathcal{F}(\mathbb{R}) = \sum_c \sum_k Vol_k^c \frac{1}{1+\exp(-\langle \mathbf{W}_k^c, A_k^c R_k + B_k^c \rangle)}.$$

This can be solved by gradient directed searching, such as the function $fmincon$ in MATLAB. Specifically, we have the gradients and hessian as follows:

$$\frac{\partial \mathcal{F}(\mathbb{R})}{\partial R_k} = \sum_c Vol_k^c \times f(z_k^c)(1-f(z_k^c))\langle \mathbf{W}_k^c, A_k^c \rangle,$$

$$H(\mathcal{F}) = \sum_c \mathrm{diag}(H_1^c, \cdots, H_K^c),$$

where

$$z_k^c = \langle \mathbf{W}_k^c, A_k^c R_k + B_k^c \rangle,$$

$$H_k^c = Vol_k^c \times f(z_k^c)(1-f(z_k^c))(1-2f(z_k^c))(A_k^c)' \mathbf{W}_k^c (\mathbf{W}_k^c)' A_k^c.$$

4.6 Experimental Results

In this section we first introduce the data and settings of our experiments. Then we evaluate the results of the proposed transportation mode choice model, followed by the evaluation of flawed OD pairs. Finally we show the results of our bus routing optimization model.

4.6.1 Data and Settings

Bus transactions. Bus transactions are generated by BMAC smart card system[2] installed on all the buses in Beijing. We select the data from the same time span as the taxi data, from August, 2012 to November, 2012. This dataset contains the following information: card id, bus route number, boarding and alighting, time, fare (N. Yuan, Wang, Zhang, Xie, & Sun, 2013). Note that a random sampling method is used to recover bus trips to match taxi trips, where the ratio of bus trips to taxi trips is about 3.5:1[3].

Taxi GPS traces. These taxi GPS traces are generated by about 30,000 taxis in Beijing from August to November, 2012. Each GPS point is associated with a label indicating if the taxi is occupied or not. Here we only focus on the occupied points which form taxi trips of riders, from pick-up points to drop-off points. Table 4.4 shows some statistics of the two trip datasets.

Bus routes and road map. 1) We have the bus route data, which contains 2,427 stops and 1,058 routes in the urban area of Beijing. After we merge the redundant stops, we obtain 1,250 stops and we partition the urban area into 1,250 regions accordingly. We use the stops/regions as nodes of our routing network. 2) We have the road map data containing 196,307 road segments and their locations. We use this data to construct the connection edges of our routing network. For the 1,250 routing nodes, we have 3,855 connection edges.

[2]http://www.bmac.com.cn/
[3]http://www.bjjtw.gov.cn/

Table 4.4. Statistics of the datasets.

Datasets	**Properties**	**Statistics**
Taxi GPS Traces	Number of taxis	29,964
	Effective days	114 (77 weekdays, 37 weekends)
	Time period	Aug. 2012 - Nov. 2012
	Number of occupied GPS points	333M
	Number of occupied trips	19M
	Total trip distance(km)	156M
Bus Transactions	Number of bus stops	7,810
	Time Period	Aug. 2012 to Nov. 2012
	Number of car holders	701,250
	Number of trips	10M

POI data. A Beijing POI dataset in the year 2012 is employed to compute the functional observation weights. The number of POIs $\mathbf{n}_i = \langle n_1, \cdots, n_{10} \rangle$ in region s_i is counted following the categories shown in Table 4.5.

4.6.2 Transportation Mode Choice Model

Baselines. To the best of our knowledge, there is no existing work specifically on the modeling of transportation mode choice with a data-driven method. We evaluate the effectiveness of our spatio-functionally weighted regression (SFWLoR, SFWLiR) with a set of widely used methods and their extensions, including unweighted logistic regression (LoR), temporal logistic regression (TLoR) and temporal linear support vector machine (TLiSVM).

- A Logistic Regression model (LoR) on the data before segmented to temporal slots. That means we treat the whole day as one temporal slot and it evaluates

Table 4.5. Category of POIs.

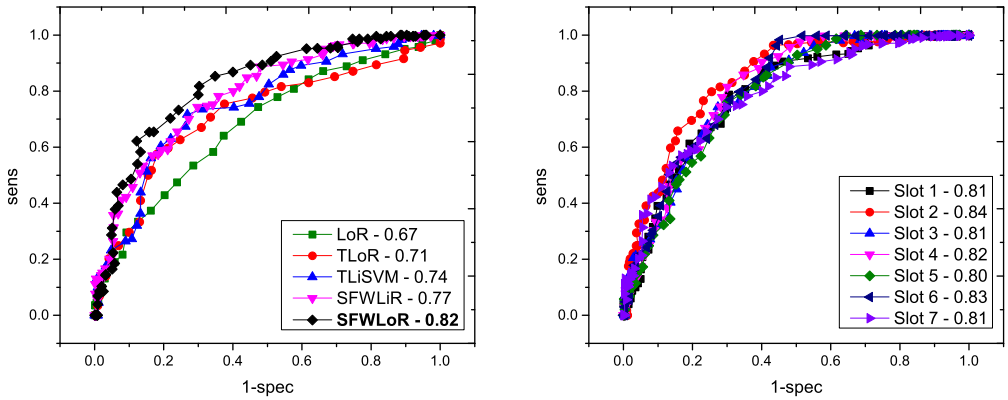
	Category	**Sub-categories**	**Number**
1	Home	Apartment building	29,246
2	Work	Government & office building	71,915
3	Education	School, training center	15,489
4	Food	Restaurant	36,723
5	Shopping	Shop, mall, outlet	56,520
6	Entertainment	Museum, theater, club	7,897
7	Outdoor	Park, sports field	2,211
8	Transportation	Airport, railway & bus station	15,287
9	Health care	Hospital, medical center, pharmacy	9,768
10	Car service	Car sale, repair, gas station	10,781

if the preference changes through the day.

- A Temporal Logistic Regression model (TLoR), which estimates people's choices in different temporal slots.

- A Temporal Linear Support Vector Machine (TLiSVM), which estimates people's choices in different temporal slots.

We use the receiver operating characteristics (ROC) curve and the area under ROC (AUC) (Fawcett, 2006) to evaluate the performance of the transportation mode choice models. The ROC curve is obtained by drawing pairs of sensitivity and false positive rate (1-specificity) at different cutoff points, i.e., every 0.01 from 0 to 1 in our experiments. The sensitivity (*sens*) is defined as the proportion of true positives as compared to the total positive class, whereas specificity (*spec*) comprises the



(a) Different methods (b) SFWLoR on different slots

Figure 4.15. Results of all the OD pairs. The number listed is the AUC score of each ROC curve.

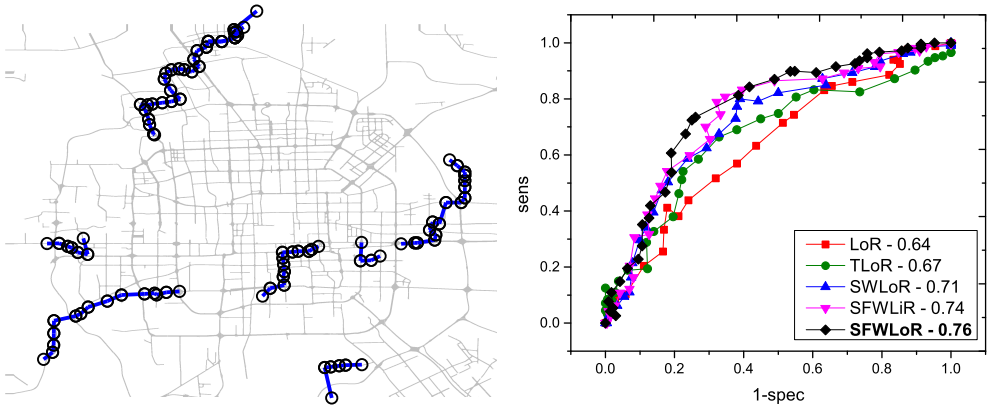
proportion of true negatives in relation to the total negative class.

$$sens = tp/(tp + fn), \tag{4.9}$$

$$spec = tn/(tn + fp), \tag{4.10}$$

where tp, fp, tn and fn are true positives, false positives, true negatives and false negatives, respectively.

Results. We evaluate the models with 10-fold cross-validation in each temporal slot separately, and then use the average of different temporal slots as the final result. Figure 4.15 (a) shows the overall performance of each method, and SFWLoR on each temporal slot. From the figure we can see SFWLoR outperforms other methods. The models perform better on weekdays (Slot 1,2,3,4) than on weekends (Slot 5,6,7), because there is a lot of variation occurs on weekend trips as compared to weekday trips and it increases the difficulty of modeling(Ashish, 2004).



(a) Bus routes changed (b) ROC curve

Figure 4.16. Results of OD pairs with route changed.

Other than the experiments with an overall evaluation on all OD pairs, we notice that routes of 8 bus lines (shown in Figure 4.16 (a)) changed in Beijing urban area started from Sep. 21, 2012[4], which is in the middle of our dataset, from August to November 2012. This gives us a chance to further test the effectiveness of our model by using data before Sep. 21, 2012 as training data, and the data after as testing data. Specifically, we summarize the statistics of OD pairs for these two periods separately, and train the mode choice model with training data and then test it on the testing data. With an analysis of the changed routes, we select 86 OD pairs which were affected by the route change. The ROC curves on the 86 OD pairs are shown in Figure 4.16 (b).

As shown in Figure 4.16, the ROC curves exhibit a consistent trend with the previous results in Figure 4.15. We can see our method demonstrates an advantage compared to other methods.

[4]http://www.bjjtw.gov.cn/

4.6.3 Flawed OD Pairs

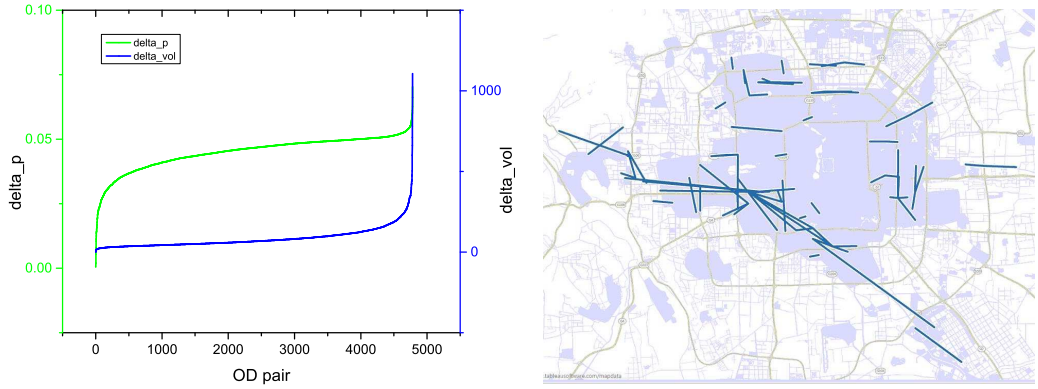
Using skyline detection, totally 651 flawed OD pairs are detected, with each time slot about 100 flawed OD pairs. More experimental results of flawed OD pair identification are shown in Figure 4.17: (a) shows the changes of probability (green line) and volume (blue line) of taking bus after using taxi routes as the upper bound of bus routes; (b) shows us top 100 flawed OD pairs.

From Figure 4.17 (a) we can see with the improvement of bus routing, an average of 5 percent increase of probability taking bus is expected for all OD pairs. Moreover, we find that the bus volume increase follows Zipf's law (Manning & Schütze, 1999), which means most of the volume increase happens among a few OD pairs. This further validate our method which focuses on these flawed OD pairs instead of all.

Figure 4.17 (b) shows us the distribution of the top 100 flawed OD pairs. By comparing the flawed OD pairs to the trip distributions of buses and taxis in Figure 4.4, we can see the OD pairs selected well reflect the travel demand of buses in the south western area of Beijing. There are many taxi trips but few bus trips are found, indicating the possibility of attracting riders from taxis by improving bus service.

4.6.4 Routing Optimization

Given the top K flawed OD pairs with a descending rank of potential increases of bus riders, we evaluate our objective function (Maximum Converted Rider, MCR) on different K. Two different solutions for MCR, MCR-LiR and MCR-LoR, are presented, using linear and logistic regression choice models respectively. We use shortest distance (SD), shortest time (ST), and maximum rider with taxi demand



(a) Change of prob. and vol. of bus (b) Top 100 flawed OD pairs

Figure 4.17. Results of flawed OD pair identification.

(MRT) (Guihaire & Hao, 2008) (C. Chen et al., 2013), which are the most widely used routing methods in practice, as baselines of our method. Accordingly, the objective functions of these three baselines in our experiment are as follows,

$$\mathcal{F}_{\mathcal{SD}}(\mathbb{R}) = \sum_{c}\sum_{k} BDist_{R_k}, \tag{4.11}$$

$$\mathcal{F}_{\mathcal{ST}}(\mathbb{R}) = \sum_{c}\sum_{k} BTime_{R_k}, \tag{4.12}$$

$$\mathcal{F}_{\mathcal{MRT}}(\mathbb{R}) = \sum_{c}\sum_{k} TVol_{R_k}. \tag{4.13}$$

What's more, according to Equation 4.7 the objective function of MCR is

$$\mathcal{F}_{\mathcal{MCR}}(\mathbb{R}) = \sum_{c} Vol_k^c \times f_k^c(R_k) = \sum_{c}\sum_{k} B\hat{V}ol_{R_k}. \tag{4.14}$$

From which we can see, $Vol_k^c \times f_k^c(R_k)$ means the predicted bus rider number $B\hat{V}ol_k^c$ after the route network renewal. So our method is trying to routing based on future bus travel demand not the current one. This makes our method not only can

Table 4.6. Results of bus routing on top K flawed OD pairs.

K	**Methods**	**BDist** (km)	**BTime** (hour)	**BStop** (#)	**BFare** (CNY)	**Prob.**	**ΔBVol**
100	SD	3.65	0.26	3.72	0.44	0.83	450.2
	ST	3.83	0.24	4.01	0.45	0.83	469.6
	MRT	3.72	0.27	3.98	0.45	0.83	511.3
	MCR-LiR	3.67	0.26	3.83	0.44	0.84	669.8
	MCR-LoR	3.43	0.25	3.66	0.44	0.86	732.5
200	SD	4.25	0.31	4.55	0.48	0.81	248.3
	ST	4.31	0.29	4.53	0.49	0.81	303.2
	MRT	3.72	0.31	4.38	0.49	0.82	354.6
	MCR-LiR	4.40	0.33	4.35	0.49	0.85	428.1
	MCR-LoR	4.54	0.34	4.42	0.49	0.86	483.7
500	SD	4.33	0.32	4.77	0.48	0.79	191.1
	ST	4.52	0.30	4.76	0.49	0.81	231.8
	MRT	3.72	0.32	4.73	0.49	0.82	278.6
	MCR-LiR	4.55	0.34	4.66	0.49	0.84	321.3
	MCR-LoR	4.59	0.34	4.73	0.49	0.85	350.5

accommodate bus travel demand but also able to maximize it based on the prediction.

Results of top 100, 200, and 500 flawed OD pairs are shown in Table 4.6, where the columns show average values of statistics of each OD pair. Specifically, we first use these methods to find the best routes $\hat{\mathbb{R}}$ for our identified flawed OD pairs. For every OD pair in different temporal slots, the transportation mode model is used to predict the probability of taking bus $\hat{p} = f(\hat{R})$. Together with the total travel demand of each pair, the bus rider number can be obtained. By comparing this to historical bus rider numbers, we then get the change of bus rider number $\Delta BVol$. Please note that here we only use taxis to represent private transportation, and the real effect of

buses can be enlarged when other private transportation modes (e.g., private car) are considered.

As shown in Table 4.6, we can see MCR-LoR, MCR-LiR provide routes that lead to highest probabilities of people taking bus, because them successfully measures the trade-off between different factors and lead to a maximum convert number from taxi riders to bus riders. While MRT obtains third best routing results, it focuses on maximizing the taxi riders on each route. However, not all the taxi riders willing to convert to bus and they will stick to taxi no matter there is a bus line exists or not. Especially in commercial areas, the taxi riders are very high but the conversion rate to bus is low. On the other hand, we see ST performs better than SD, which indicates people consider time a more important factor than distance. Although some of the routes found by our method are the same as results found by either SD, ST or MRT, we can still provide suggestions on the selection of them. From this point of view, our transportation mode choice model can serve as a criteria for choosing candidate bus routes.

A real example of bus routes found for flawed OD pairs is shown in Figure 4.18, where includes two flawed OD pairs (*Xiaohongmen*, *Qianmen*) and (*Shazikou*, *Qianmen*). The routes generated by SD, ST, MRT and MCR are shown in green, red, black, and blue lines respectively. From the figure we can see SD and ST both generate two routes, which are similar to each other, while MRT and MCR generate a single route traveled through these two OD pairs. Moreover, we found that this route share same subroutes with bus line 93 which is newly added by the Beijing Bus Company from March, 2013.

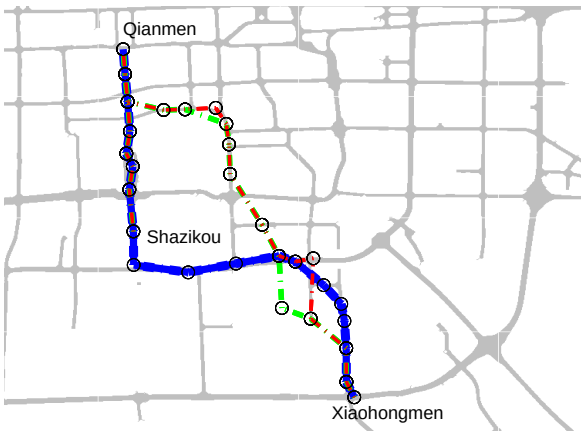


Figure 4.18. An example of routes generated.

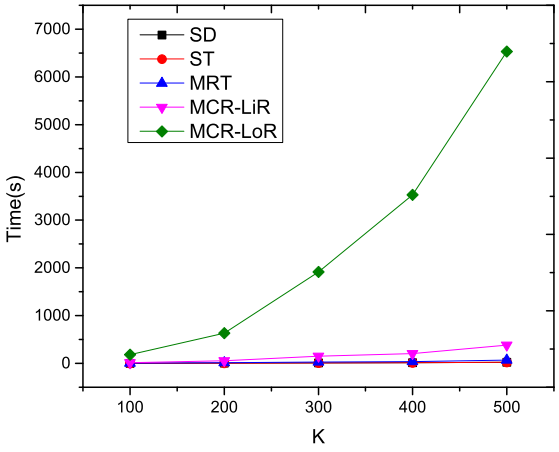


Figure 4.19. Running time of bus routing.

Efficient Study. Figure 4.19 presents the efficiency of the four methods for different K. From this figure we can see ST, SD and MRT are the fastest among these four, since they don't involve the bus travel demand prediction phase. While MCR-LoR costs the most time for computing results. We note that MCR-LiR is much faster than MCR-LoR but the performance is not much worse. In real applications, MCR-LiR would be recommended for large-scale bus routing. Since this application usually works in an offline manner, MCR-LoR would also be used for better planning results.

4.7 Related Work

Our work is related to two research areas, the first one is human mobility pattern mining and the second one is bus route network optimization.

4.7.1 Human Mobility Pattern Mining

Understanding human mobility in urban environments is central to traffic forecasting, location-based services, and urban reconstruction. A significant number of papers on human mobility analysis have been published in recent years thanks to the widely available mobility data, such as GPS data, cellular network data, and transportation data. Gonzalez (Gonzalez et al., 2008b) and Liu et al. (L. Liu, Hou, Biderman, Ratti, & Chen, 2009) suggested that human mobility patterns follow high degree of spatial and temporal regularity and thus predictable. Song et al. (Song et al., 2010) suggested that human mobility has a predictability of 93% and Montjoye et al. (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013) showed that four spatial-temporal points are enough to uniquely identify 95% of the individuals. Utsunomiya et al. (Utsunomiya, Attanucci, & Wilson, 2006) also reported the consistency of daily travel patterns with public transportation transaction data.

To the best of our knowledge, we are the first to work on the problem of improving bus routing to attract taxi riders by leveraging human mobility patterns. Although there is no existing work on the exact application we are working on, there are many existing work on making use of human mobility patterns for different novel applications. Giannotti et al. (Giannotti et al., 2007; Monreale et al., 2009) developed trajectory pattern mining, and applied it to predict the next location at a certain

level of accuracy by using GPS data. Zheng et al. (Y. Zheng et al., 2011) detected flawed designs in current road network with a frequent graph method on taxi GPS traces. Yuan et al. (J. Yuan et al., 2012) proposed a topic-based inference model that discovers regions of different functions, such as educational areas and business districts, in a city using both human mobility data and points of interests (POIs). Ge et al. (Ge et al., 2010) developed a mobile recommender system to maximize the probability of business success and reduce energy consumption, which has the ability in recommending a sequence of pick-up points for taxi drivers or a sequence of potential parking positions.

Specifically, there are some but not many efforts have been made to understand and improve public transportation leveraging human mobility patterns. Lathia et al. (Lathia, Froehlich, & Capra, 2010) mined Automated Fare Collection (AFC) data of London public transportation system with the aim to build more accurate travel route planners. Lathia and Capra (Lathia & Capra, 2011) also analyzed Oyster card data of London to estimate future travel habits. By analyzing historical travel traces, they have been able to extract features about when, where, and how often individual travels, that can then be predicted with a high level of accuracy. Watkins et al. (Watkins, Ferris, Borning, Rutherford, & Layton, 2011) conducted a study on the impact of providing real-time bus arrival information directly on riders' mobile phones, and found it reduced not only the perceived wait time of those already at a bus stop, but also the actual wait time experienced by customers who plan their journey using such information.

However, these papers are not dedicated to the bus route network planning prob-

lem except the following two (Bastani, Huang, Xie, & Powell, 2011; C. Chen et al., 2013). Bastani et al. (Bastani et al., 2011) leveraged historical taxi GPS trips to suggest a flexible bus route. The authors first grouped taxi trips into different clusters with similar starting time, duration, origin, and destination. Then a route connects multiple dense clusters was identified. This work aimed to maximized the sum of each connected trip cluster discarding other constraints like time. Chen et al. (C. Chen et al., 2013) investigated night bus route planning using large scale taxi GPS traces, which aimed to find a bus route with a fixed frequency, maximizing the number of passengers expected along the route subject to the total travel time constraint. Similar to (Bastani et al., 2011), this work first clustered "hot" areas with dense passenger trips and treated these clusters as candidate bus stops. Then several rules were derived to build bus routing graph and finally generated candidate bus routes to maximize the number of bus riders. However, (C. Chen et al., 2013) focused on night bus routing planning, and this paper mainly focus on day bus routing planning which is more complicated with more routes and riders considered. Moreover, these two papers only considered one single transportation mode, i.e., taxi, and considered the taxi travel demand as new bus demand. They assumed all the taxi riders were willing to take bus if there was one. However, different taxi riders may have different probabilities converting to bus based on the bus routes provided. From this point of view, we propose a mode choice model using heterogenous human mobilities to learn the probabilities of people choosing between taxi and bus. As a result, our proposed routing optimization approach can predict the number of bus riders by integrating the mode choice model, and is able to both accommodate and maximize future bus

travel demand.

Motivated by the above novel applications, we aim at the problem of optimizing bus routing by comparing the difference of human mobility patterns between taxi and bus riders. Unlike the above mentioned work focusing on single mode human mobility pattern, we integrate heterogeneous human mobility data together to better represent the mobility of a city. Thus we are able to analyze the difference and relation between different human mobility patterns and make planning in the city level considering the dynamic transitions between different transportation modes.

4.7.2 Bus Route Network Optimization

Bus route network optimization addresses the problems of how to design a new bus route network or how to redesign bus routes in an existing network (Ceder, 2007; de Dios Ortuzar & Willumsen, 2011). It is an intensive studied area in the urban planning and transportation field, known to be a complex, non-linear, non-convex, multi-objective NP-hard problem (C. Chen et al., 2013; C.-L. Liu et al., 2001). Specifically, it focuses on the optimization of a number of objectives representing the efficiency of bus network under operational and resource constraints (Chakroborty, 2003; Fan & Machemehl, 2006). The outcome is a set of routes that cover the required OD pairs in the network, and on which user demands can be fulfilled. The general process of bus network optimization is as follows: (1) user travel demand generation; (2) bus stops and routing network construction; (3) formulation of optimization model with objectives and constraints; (4) construction of candidate bus routes; and (5) calculation of final solutions.

Traditional bus network design primarily considered passenger flows and user requirements gathered from census and household travel surveys (Aslam et al., 2012) (Guihaire & Hao, 2008). In a general survey, multi-type information is obtained, such as Origin-Destination, transportation mode and distance, trip purpose, routes selected on a trip, fare paid, type of payment, frequency of use by time of day, and socio-economic and attitude elements (Ceder, 2007). As related surveys normally cost several millions dollars each time for a metropolitan area, a common practice is to conduct such surveys once every several years.

Some widely used objectives include shortest distance, shortest travel time, maximum passenger flow, and maximum area coverage while constraints include travel time, route length, capacity and so on (Asadi Bagloee & Ceder, 2011) (Guihaire & Hao, 2008). However, these objectives may be generated from different perspectives, i.e., the operator and the riders, and need to be considered simultaneously. Therefore, there are no clear-cut criteria for evaluating the "goodness" of a bus network and a trade-off rather than optimal solution is often achieved due to the conflicting objectives.

With user demands and objectives determined, a variety of approaches have been proposed in formulating and solving the bus route network optimization problem (Guihaire & Hao, 2008) (Kepaptsoglou & Karlaftis, 2009), such as linear programming, non-linear programming, and heuristic algorithms (Kim, Shekhar, & Min, 2008). Fan and Machemehl (Fan & Machemehl, 2006) formulated the bus route network design problem as a multi-objective nonlinear mixed integer model. Then Dijkstra's shortest path algorithm (Ahuja, 1993) and Yen's k-shortest path algorithm

(Yen, 1971) were combined to generate all candidate routes. At last a genetic algorithm procedure was used to select an optimal set of routes. Ceder and Wilson (Ceder & Wilson, 1986) considered travel time as a constraint and constructed routes which had minimized demand differences between them and shortest paths. More recently, Chakroborty and Dwivedi (Chakroborty & Wivedi, 2002) used a demand driven node-addition approach. They estimated the incoming and outgoing passengers of each network node, and used the demand to guide the construction of routes. In the meanwhile, they used the connectivity of nodes, route length and number of nodes per route as constraints.

The above work assumed that travel demands were statically determined by user survey or population estimation. Different from that, we integrate the bus route network design problem with real mobility demand, also the travel demands are dynamically estimated with different routing results. This enables us to plan bus routes to maximize number of bus riders by converting from taxi, which cannot be fulfilled by existing methods.

4.8 Summary

In this chapter, we focused on the identification and optimal planning of the flawed bus routes to improve the utilization efficiency of the public transportation service, according to the transportation mode choice model built on real data. First, we partitioned the urban area into disjoint regions on which an integrated analysis of the taxi traces and the bus transactions is conducted. Second, based on the integrated analysis, we proposed a localized transportation mode choice model, with which we

can dynamically predict the bus travel demand for different bus routing. Then, we leveraged this model to optimize the bus routes by maximizing the bus ridership with budget constraints. At last, we provided a solution for the identified most flawed region pairs in the urban area. Extensive studies, which validated the effectiveness of our methods, were performed on real world data collected in Beijing which contains 19 million taxi trips and 10 million bus trips.

The work reported in this chapter showed how to optimize bus routing to attract more bus riders from taxi. Improvements can be made through several different directions. First, we can further take bus stop location selection into account. In this way, we can optimize bus routing and bus stop location simultaneously to meet people's travel demands. Second, more transportation modes can be considered, for example, bus network optimization can be conducted together with subway system and city bike system. This can help to model the whole city travel demand as a whole and better serve our goal to make public transportation more attractive to riders.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this dissertation, we identified several unique challenges for smart living in mobile environments, and then introduced how we use advanced data mining techniques to address these challenges.

First, we investigated how to exploit both user interests and their evolving sequential preferences for recommending POIs for a specific time period. Along this line, we first proposed a unified framework to integrate user interests and their sequential preferences. Specifically, the distributions of the temporal intervals between dependent POIs were studied for measuring users evolving sequential preferences based on their historical check-in sequences. Here, to address the challenge of estimating the distributions with only sparse observations, we developed a bi-weighted low-rank graph construction model to identify a set of bi-weighted graph bases, which in turn can be leveraged for learning user interests and their sequential preferences in a coherent way. Furthermore, we provided a mobile regularization method to effectively incorporating human mobility patterns captured by user mobility data to improve the performances of POI recommendations. As shown in the experimental results on real-world data, unlike existing POI recommender systems, the proposed WWO recommender system can provide effective POI recommendations for a future specified time period by capturing user evolving sequential preferences.

Second, we investigated how to exploit human mobility patterns, geographic data, and demographic data for identifying region POI demands. Along this line, we first proposed a framework, named Region POI Demand Identification (RPDI), to model POI demands with the daily needs identified from their large-scale mobility data. Specifically, in this framework, an urban space was first partitioned into spatially differentiated neighborhood regions formed by local communities. Then, the daily activity patterns of people traveling in the city were extracted from human mobility data. However, the trip activities, even aggregated, were sparse and insufficient to directly identify the POI demands, especially for underdeveloped regions. Therefore, with a proposed demand inference model considering POI preferences and supplies together with demographic features, we estimated the POI demands of all the regions simultaneously. As shown in the experimental results on real-world data, the proposed RPDI framework could provide effective POI demand identification for different regions.

Finally, we focused on the identification and optimal planning of the flawed bus routes to improve the utilization efficiency of the public transportation service, according to the transportation mode choice model built on real data. First, we partitioned the urban area into disjoint regions on which an integrated analysis of the taxi traces and the bus transactions is conducted. Second, based on the integrated analysis, we proposed a localized transportation mode choice model, with which we can dynamically predict the bus travel demand for different bus routing. Then, we leveraged this model to optimize the bus routes by maximizing the bus ridership with budget constraints. At last, we provided a solution for the identified most flawed

region pairs in the urban area. Extensive studies, which validated the effectiveness of our methods, were performed on real world data collected in Beijing which contains 19 million taxi trips and 10 million bus trips.

Future Work

Scientific discovery and business solutions are demanded to be more data-driven, therefore data analytics will be increasingly critical for scientific, industrial, and business problems. Also the data sources are becoming more diverse, which calls for more sophisticated modeling and learning methods. Following are some future directions to explore towards smart living: First, it is interesting to extend our previous research from macro-level POI demand modeling to micro-level POI popularity prediction which provides guidelines for business site selection, supply chain management, and urban planning. Second, it would be interesting to enhance the research of mobile recommender systems by jointly considering Who (personalized preference), What (item characteristics), When (temporal dependency), Where (geographic influence), as well as other context information. Finally, it is interesting to start from the perspectives of smart living, mobile analytics, and personalized techniques, to combine urban data with geographical, mobile, and personalized dimensions, to make informed decisions, to reduce managerial risk, and finally, to create livable environments where people and businesses can thrive. Specifically, it is interesting to investigate the following three topics: (i) geographic ranking systems for location analysis and business site selection, (ii) mobile recommender systems for mobile user targeting, and (iii) mobile data analytics for smart transportation and planning.

BIBLIOGRAPHY

Ahuja, R. K. (1993). *Network flows* (Unpublished doctoral dissertation). TECHNISCHE HOCHSCHULE DARMSTADT.

Asadi Bagloee, S., & Ceder, A. A. (2011). Transit-network design methodology for actual-size road networks. *Transportation Research Part B: Methodological*, *45*(10), 1787–1804.

Ashish, A. (2004). *A comparison of weekend and weekday travel behavior characteristics in urban areas* (Unpublished doctoral dissertation). USF.

Aslam, J., Lim, S., Pan, X., & Rus, D. (2012). City-scale traffic estimation from a roving sensor network. In *Proceedings of the 10th acm conference on embedded network sensor systems* (pp. 141–154).

Aurenhammer, F. (1991, September). Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, *23*(3), 345–405. doi: 10.1145/116873.116880

Bastani, F., Huang, Y., Xie, X., & Powell, J. W. (2011). A greener transportation mode: flexible routes discovery from gps trajectory data. In *Gis* (pp. 405–408).

Beirão, G., & Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport policy*, *14*(6), 478–489.

Berman, O., & Krass, D. (2002). The generalized maximal covering location problem. *Computers & Operations Research*, *29*(6), 563–581.

Blondel, V. D., Gajardo, A., Heymans, M., Senellart, P., & Van Dooren, P. (2004). A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review*, *46*(4).

Borzsony, S., Kossmann, D., & Stocker, K. (2001). The skyline operator. In *Data engineering, 2001. proceedings. 17th international conference on* (pp. 421–430).

Canny, J. (2004). Gap: a factor model for discrete data. In *Sigir* (pp. 122–129).

Ceder, A. (2007). *Public transit planning and operation: theory, modeling and practice*. Elsevier, Butterworth-Heinemann.

Ceder, A., & Wilson, N. H. (1986). Bus network design. *Transportation Research Part B: Methodological*, *20*(4), 331–344.

Chakroborty, P. (2003). Genetic algorithms for optimal urban transit network design. *Computer-Aided Civil and Infrastructure Engineering*, *18*(3), 184–200.

Chakroborty, P., & Wivedi, T. (2002). Optimal route network design for transit systems using genetic algorithms. *Engineering Optimization*, *34*(1), 83–100.

Chen, C., Zhang, D., Zhou, Z.-H., Li, N., Atmaca, T., & Li, S. (2013). B-planner: Night bus route planning using large-scale taxi gps traces. In *Pervasive computing and communications (percom), 2013 ieee international conference on* (pp. 225–233).

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. In *Icde* (pp. 900–911).

Cheng, C., Yang, H., King, I., & Lyu, M. R. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. In *Aaai*.

Cheng, C., Yang, H., Lyu, M. R., & King, I. (2013). Where you like to go next: Successive point-of-interest recommendation. In *Ijcai* (pp. 2605–2611).

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In *Kdd* (pp. 1082–1090).

de Dios Ortuzar, J., & Willumsen, L. G. (2011). *Modelling transport*.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, *3*.

Fan, W., & Machemehl, R. B. (2006). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. *Journal of transportation engineering*, *132*(1), 40–51.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, *27*(8), 861–874.

Feng, S., Li, X., Zeng, Y., Cong, G., Chee, Y. M., & Yuan, Q. (2015). Personalized ranking metric embedding for next new poi recommendation. In *Ijcai* (pp. 2069–2075).

Fu, Y., Xiong, H., Ge, Y., Zheng, Y., Yao, Z., & Zhou, Z.-H. (2016). Modeling of geographic dependencies for real estate ranking. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *11*(1), 11.

Gao, H., Tang, J., Hu, X., & Liu, H. (2013). Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of recsys* (pp. 93–100).

Ge, Y., Liu, Q., Xiong, H., Tuzhilin, A., & Chen, J. (2011). Cost-aware travel tour recommendation. In *Kdd* (pp. 983–991).

Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., & Pazzani, M. (2010). An energy-efficient mobile recommender system. In *Proceedings of kdd* (pp. 899–908).

Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007). Trajectory pattern mining. In *Kdd* (pp. 330–339).

Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, *43*(2), 103–114.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008a). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008b). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM TOIS*, *20*(4), 422–446.

Jeh, G., & Widom, J. (2002). Simrank: a measure of structural-context similarity. In *Proceedings of the eighth acm sigkdd international conference on knowledge discovery and data mining* (pp. 538–543).

Johnson, C. C. (2014). Logistic matrix factorization for implicit feedback data. *NIPS*, *27*.

Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-spotting: mining online location-based services for optimal retail store placement. In *Proceedings of kdd* (pp. 793–801).

Kepaptsoglou, K., & Karlaftis, M. (2009). Transit route network design problem: review. *Journal of transportation engineering*, *135*(8), 491–505.

Kim, S., Shekhar, S., & Min, M. (2008). Contraflow transportation network reconfiguration for evacuation route planning. *Knowledge and Data Engineering, IEEE Transactions on*, *20*(8), 1115–1129.

Kumar, R., Mahdian, M., Pang, B., Tomkins, A., & Vassilvitskii, S. (2015). Driven by food: Modeling geographic choice. In *Proceedings of wsdm* (pp. 213–222).

Land, A. H., & Doig, A. G. (1960). An automatic method of solving discrete programming problems. *Econometrica*, 497–520.

Lathia, N., & Capra, L. (2011). Mining mobility data to minimise travellers' spending on public transport. In *Proceedings of the 17th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1181–1189).

Lathia, N., Froehlich, J., & Capra, L. (2010). Mining public transport usage for personalised intelligent transport systems. In *Data mining (icdm), 2010 ieee 10th international conference on* (pp. 887–892).

Li, Y., Zheng, Y., Ji, S., Wang, W., Gong, Z., et al. (2015). Location selection for ambulance stations: a data-driven approach. In *Proceedings of gis* (pp. 85–94).

Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014). Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Kdd* (pp. 831–840).

Liu, B., Fu, Y., Yao, Z., & Xiong, H. (2013). Learning geographical preferences for point-of-interest recommendation. In *Kdd* (pp. 1043–1051).

Liu, B., & Xiong, H. (2013). Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of sdm* (Vol. 13, pp. 396–404).

Liu, B., Xiong, H., Papadimitriou, S., Fu, Y., & Yao, Z. (2015). A general geographical probabilistic factor model for point of interest recommendation. *TKDE*, *27*(5), 1167–1179.

Liu, C.-L., Pai, T.-W., Chang, C.-T., & Hsieh, C.-M. (2001). Path-planning algorithms for public transportation systems. In *Intelligent transportation systems, 2001. proceedings. 2001 ieee* (pp. 1061–1066).

Liu, L., Hou, A., Biderman, A., Ratti, C., & Chen, J. (2009). Understanding individual and collective mobility patterns from smart card records: A case study in shenzhen. In *Intelligent transportation systems, 2009. itsc'09. 12th international ieee conference on* (pp. 1–6).

Liu, Q., Chen, E., Xiong, H., Ge, Y., Li, Z., & Wu, X. (2014). A cocktail approach for travel package recommendation. *TKDE*, *26*(2), 278–293.

Liu, Q., Ge, Y., Li, Z., Chen, E., & Xiong, H. (2011). Personalized travel package recommendation. In *Icdm* (pp. 407–416).

Liu, Y., Liu, C., Liu, B., Qu, M., & Xiong, H. (2016). Unified point-of-interest recommendation with temporal interval assessment. In *Proceedings of kdd* (pp. 1015–1024).

Liu, Y., Liu, C., Lu, X., Teng, M., Zhu, H., & Xiong, H. (2017). Point-of-interest demand modeling with human mobility patterns. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 947–955).

Liu, Y., Liu, C., Yuan, N. J., Duan, L., Fu, Y., Xiong, H., ... Wu, J. (2017). Intelligent bus routing with heterogeneous human mobility patterns. *Knowledge and Information Systems*, *50*(2), 383–415.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Mnih, A., & Salakhutdinov, R. (2007). Probabilistic matrix factorization. In *Advances in neural information processing systems* (pp. 1257–1264).

Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of kdd* (pp. 637–646).

Niu, H., Liu, J., Fu, Y., Liu, Y., & Lang, B. (2016). Exploiting human mobility patterns for gas station site selection. In *Proceedings of dasfaa* (pp. 242–257).

Pilinkienė, V. (2008a). Market demand forecasting models and their elements in the context of competitive market. *Engineering Economics*, *60*(5).

Pilinkienė, V. (2008b). Selection of market demand forecast methods: Criteria and application. *Engineering Economics*, *58*(3).

Pratt, R. H., Evans, I., et al. (2004). *Traveler response to transportation system changes. chapter 10- bus routing and coverage.*

Redman, L., Friman, M., Gärling, T., & Hartig, T. (2013). Quality attributes of public transport that attract car users: A research review. *Transport Policy*, *25*, 119–127.

Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2010). Factorizing personalized markov chains for next-basket recommendation. In *Www* (pp. 811–820).

Richtárik, P., & Takáč, M. (2014). Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Math. Prog.*, *144*(1-2), 1–38.

Sang, J., Mei, T., Sun, J.-T., Xu, C., & Li, S. (2012). Probabilistic sequential pois recommendation via check-in data. In *Gis* (pp. 402–405).

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021.

Strutz, T. (n.d.). Data fitting and uncertainty.

Sun, J., Yuan, J., Wang, Y., Si, H., & Shan, X. (2011). Exploring space–time structure of human mobility in urban space. *Physica A: Statistical Mechanics and its Applications*, *390*(5), 929–942.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, 416–424.

Utsunomiya, M., Attanucci, J., & Wilson, N. (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. *Transportation Research Record: Journal of the Transportation Research Board*, *1971*(1), 119–126.

Wang, J., & Zhang, Y. (2013). Opportunity model for e-commerce recommendation: right product; right time. In *Sigir* (pp. 303–312).

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where is my bus? impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, *45*(8), 839–848.

Xu, M., Wang, T., Wu, Z., Zhou, J., Li, J., & Wu, H. (2016). Demand driven store site selection via multiple spatial-temporal data. In *Proceedings of gis* (pp. 40–49).

Yap, G.-E., Li, X.-L., & Philip, S. Y. (2012). Effective next-items recommendation via personalized sequential pattern mining. In *Database systems for advanced applications* (pp. 48–64).

Ye, M., Yin, P., & Lee, W.-C. (2010). Location recommendation for location-based social networks. In *Gis.*

Ye, M., Yin, P., Lee, W.-C., & Lee, D.-L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of sigir* (pp. 325–334).

Yen, J. Y. (1971). Finding the k shortest loopless paths in a network. *management Science*, *17*(11), 712–716.

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining* (pp. 186–194).

Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2013). T-drive: Enhancing driving directions with taxi drivers' intelligence. *Knowledge and Data Engineering, IEEE Transactions on*, *25*(1), 220–232.

Yuan, N., Wang, Y., Zhang, F., Xie, X., & Sun, G. (2013, Dec). Reconstructing individual mobility from smart card transactions: A space alignment approach. In *Icdm* (p. 877-886). doi: 10.1109/ICDM.2013.37

Yuan, N. J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., & Xiong, H. (2015). Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *27*(3), 712–725.

Yuan, Q., Cong, G., Ma, Z., Sun, A., & Thalmann, N. M. (2013). Time-aware point-of-interest recommendation. In *Sigir* (pp. 363–372).

Zhang, J.-D., & Chow, C.-Y. (2015). Spatiotemporal sequential influence modeling for location recommendations: A gravity-based approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *7*(1), 11.

Zhang, J.-D., Chow, C.-Y., & Li, Y. (2014). Lore: Exploiting sequential influence for location recommendations. In *Gis*.

Zhao, G., Lee, M. L., Hsu, W., & Chen, W. (2012). Increasing temporal diversity with purchase intervals. In *Sigir* (pp. 165–174).

Zheng, V. W., Cao, B., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative filtering meets mobile recommendation: A user-centered approach. In *Aaai*.

Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., & Chang, E. (2014). Diagnosing new york city's noises with ubiquitous data. In *Proceedings of ubicomp* (pp. 715–725).

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011). Urban computing with taxicabs. In *Proceedings of ubicomp* (pp. 89–98).