CrossMark

REGULAR PAPER

Intelligent bus routing with heterogeneous human mobility patterns

**Yanchi Liu[1] · Chuanren Liu[2] · Nicholas Jing Yuan[3] · Lian Duan[4] ·
Yanjie Fu[1] · Hui Xiong[1] · Songhua Xu[5] · Junjie Wu[6]**

Abstract 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.
In this paper, 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 trans-
portation 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 pro-
posed methods, extensive studies are performed on real-world data collected in Beijing which
contains 19 million taxi trips and 10 million bus trips.

✉ Hui Xiong
hxiong@rutgers.edu

[1] Management Science and Information Systems Department, Rutgers University, Newark,
NJ 07102, USA

[2] Drexel University, Philadelphia, PA, USA

[3] Microsoft Research, Beijing, China

[4] Hofstra University, Long Island, NY, USA

[5] New Jersey Institute of Technology, Newark, NJ, USA

[6] Beihang University, Beijing, China

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 [14]. 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 saves more than $9700 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 to understand people's mobility patterns and their choice among different transportation modes [4,21]. 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 paper. 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 paper, 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.

---

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

With integrated analysis on the mobility patterns, we aim to detect and replan flawed and less effective bus routes for attracting most number of potential bus riders. 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 suboptimal 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, nonlinear, nonconvex, and multi-objective NP-hard problem [13,27]. 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 paper 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 [14], 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 Sect. 2 and the preliminaries of this study in Sect. 3. Then, the transportation mode choice model is proposed in Sect. 4, followed by the flawed OD pairs detection in Sect. 5 and bus routing optimization in Sect. 6. Experimental results are presented in Sect. 7, and we discuss the results and give concluding remarks in Sect. 8.

2 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.

2.1 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 [20] and Liu et al. [28] suggested that human mobility patterns follow high degree of spatial and temporal regularity and thus predictable. Song et al. [34] suggested that human mobility has a predictability of 93%, and Montjoye et al. [15] showed that four spatial–temporal points are enough to uniquely identify 95% of the individuals. Utsunomiya et al. [37] 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 works on making use of human mobility patterns for different novel applications. Giannotti et al. [19,31] 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. [43] detected flawed designs in current road network with a frequent graph method on taxi GPS traces. Yuan et al. [40] 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. [18] 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 pickup 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. [26] mined Automated Fare Collection (AFC) data of London public transportation system with the aim to build more accurate travel route planners. Lathia and Capra [25] 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. [38] 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 problem except the following two [6,13]. Bastani et al. [6] 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 connecting multiple dense clusters was identified. This work aimed to maximize the sum of each connected trip cluster discarding other constraints such as time. Chen et al. [13] 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 [6], 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, Chen et al. [13] 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.

2.2 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 [9,14]. It is an intensive studied area in the urban planning and transportation field, known to be a complex, nonlinear, non-convex, multi-objective NP-hard problem [13,27]. Specifically, it focuses on the optimization of a number of objectives representing the efficiency of bus network under operational and resource constraints [11,16]. 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 [4,21]. 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 socioeconomic and attitude elements [9]. 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 [2,21]. 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 [21,22], such as linear programming, nonlinear programming, and heuristic algorithms [23]. Fan and Machemehl [16] formulated the bus route network design problem as a multi-objective nonlinear mixed integer model. Then Dijkstra's shortest path algorithm [1] and Yen's k-shortest path algorithm [39] 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 [10] considered travel time as a constraint and constructed routes which had minimized demand differences between them and shortest paths. More recently, Chakroborty and Dwivedi [12] 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.

3 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 Fig. 1. Later in the second (right) part of the framework, these mobility patterns will be modeled (in Sect. 4) 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 Sects. 5 and 6, respectively).

Unless otherwise stated, we use bold characters to represent nonscalar 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.

3.1 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 $\mathbf{S} = \{s_i | i = 1, \ldots, N\}$ 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 Fig. 2). Buses traveling north through Mingguang Bridge will stop at Mingguang Bridge
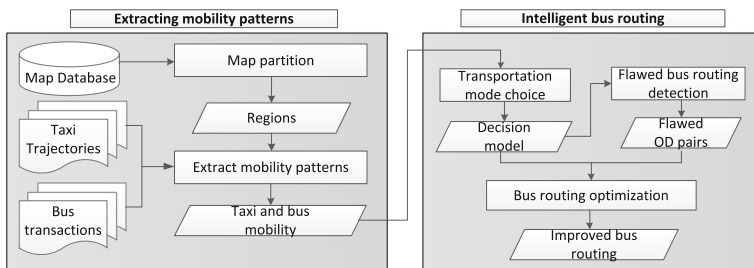

Fig. 1 Framework of our method

Fig. 2 Bus stop merging of Mingguang Bridge. Mingguang Bridge North and Mingguang Bridge South are merged together to represent 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 paper, we assume the stops in set **S** have already been merged.

After merging bus stops serving the same regions, we then partition the map using Voronoi diagram [5], 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 pickup/drop-off points for taxi trips are mapped to the regions located. Since we are focusing on the bus routing problem in this paper, 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 [36] and road-network-based partition [43]. In the following sections, we use **S** to represent both stops and their, respectively, associated regions. Please refer to Fig. 3a for the map segmentation in Beijing.

Now, we define the routing network $\mathbf{G} = (\mathbf{S}, \mathbf{E})$ with the bus stops **S** as nodes. The edges in **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 Fig. 3b for an example of the routing network with nodes plotted in red dots and edges in blue lines.

3.2 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 Fig. 4, there is a clear
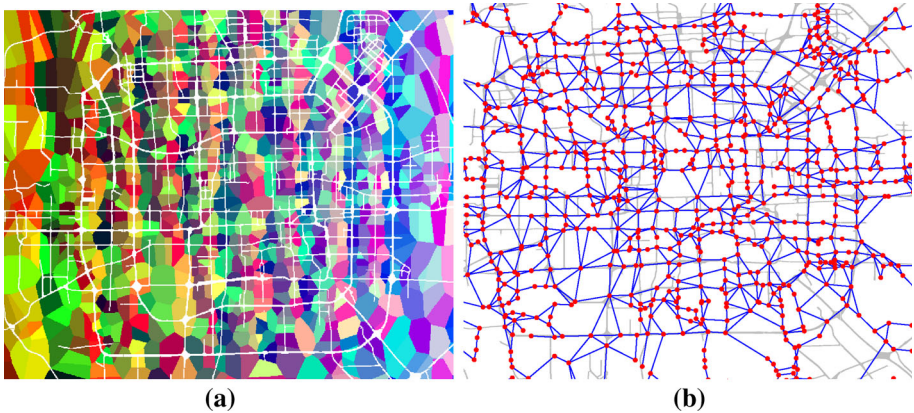
Fig. 3 Map of Beijing. **a** Map segmentation of Beijing. **b** Bus routing network of Beijing (color figure online)
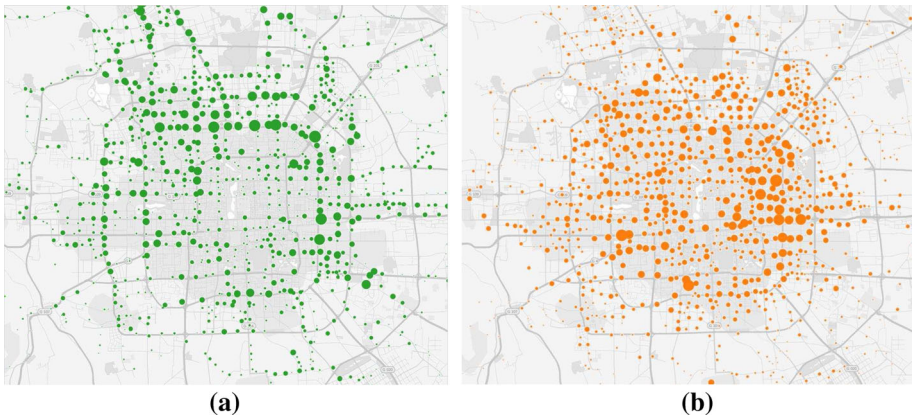

Fig. 4 Trip origin distribution of Beijing. The size of dot is proportional to the related number of trips. **a** Bus trip origins. **b** Taxi trip origins

difference between the mobility patterns of these two transportation 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.

3.2.1 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**.
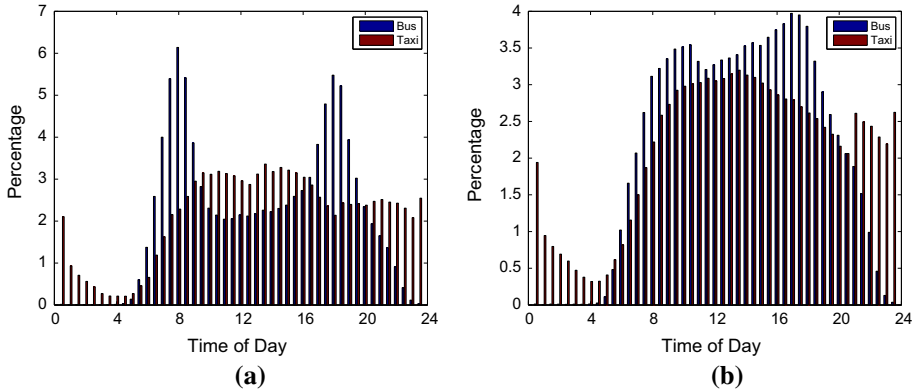
Fig. 5 Travel behavior in Beijing. The percentages of bus and taxi riders shown in *y* axis are calculated separately. **a** Weekday. **b** Weekend

Table 1 Temporal slots for weekday and weekend	Weekday		Weekend	
	Slot 1	5:00 a.m.–10:30 a.m.	Slot 5	5:00 a.m.–12:30 p.m.
	Slot 2	10:30 a.m.–4:00 p.m.	Slot 6	12:30 p.m.–7:30 p.m.
	Slot 3	4:00 p.m.–7:30 p.m.	Slot 7	7:30 p.m.–11:00 p.m.
	Slot 4	7:30 p.m.–11:00 p.m.		

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

3.2.2 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 5a 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 8 a.m. and 6 p.m., 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 1, which is derived according to the traffic and travel behaviors in different time of day [43]. Specifically, we first segment the time of day into 48 segments, each for half an hour. By comparing the number of bus and taxi riders in each segment to the total number of bus and taxi riders in a day (as shown in Fig. 5), and the speed in each segment to the average speed in a day (as shown in Fig. 6), these segments can be further merged into the temporal slots presented. In the same temporal slot, the semantic meaning of people's travel is similar. Figure 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
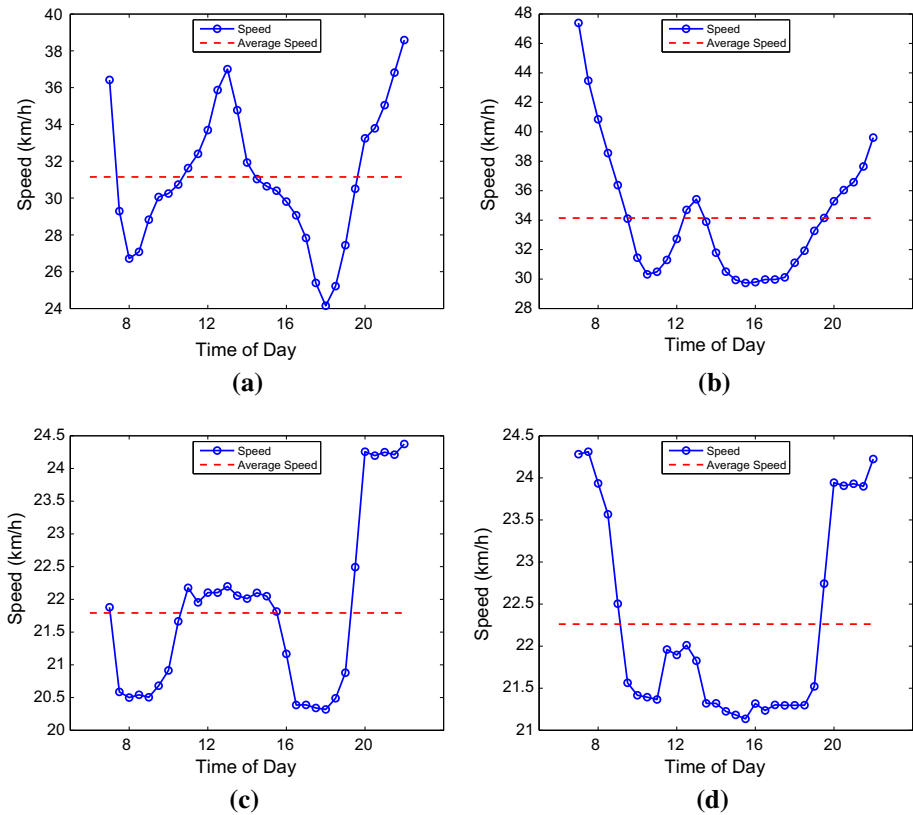
Fig. 6 Traffic conditions in Beijing. **a** Taxi on weekday. **b** Taxi on weekend. **c** Bus on weekday. **d** Bus on weekend

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 11 p.m. and 5 a.m. (as shown in Fig. 5), this paper focuses on the day bus lines, running from 5 a.m. to 11 p.m. We use $c = 1, \ldots, 7$ to represent the temporal slots, and each is associated with its time proportion in $STime^c$, for example, $STime^1 = 5 * 5.5$ h (5.5 h every day and 5 days every week).

3.2.3 Statistical summarization

Now, we summarize the partitioned transitions $\mathbf{TR}^c_{ij} = \{\mathbf{tr} : \mathbf{tr}.o = i, \mathbf{tr}.d = j, \mathbf{tr}.lt \in c\}$ with statistics defined in Table 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 [7,33]. In this paper, we focus on improving bus routing to attract private transportation riders to public transportation, so we assume other perceived factors, such as comfort and safety, remain the same after the bus route change [7,33]. The definition of an OD pair is given as follows.

Table 2 Statistics of transition records for OD pair (i, j) in temporal slot c

Statistic	Definition
Volume	$BVol = |\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TVol = |\{\mathbf{tr} : \mathbf{tr}.m = 0\}|$
	$Vol = BVol + TVol$
Travel time	$BTime = \sum_{\mathbf{tr}:\mathbf{tr}.m=1}(\mathbf{tr}.at - \mathbf{tr}.lt)/|\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TTime = \sum_{\mathbf{tr}:\mathbf{tr}.m=0}(\mathbf{tr}.at - \mathbf{tr}.lt)/|\{\mathbf{tr} : \mathbf{tr}.m = 0\}|$
Travel distance	$BDist = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.td/|\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TDist = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.td/|\{\mathbf{tr} : \mathbf{tr}.m = 0\}|$
Velocity	$BVel = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.td/(\mathbf{tr}.at - \mathbf{tr}.lt)/|\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TVel = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.td/(\mathbf{tr}.at - \mathbf{tr}.lt)/|\{\mathbf{tr} : \mathbf{tr}.m = 0\}|$
Fare	$BFare = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.tf/|\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TFare = \sum_{\mathbf{tr}:\mathbf{tr}.m=0} \mathbf{tr}.tf/|\{\mathbf{tr} : \mathbf{tr}.m = 0\}|$
Stop number	$BStop = \sum_{\mathbf{tr}:\mathbf{tr}.m=1} \mathbf{tr}.sn/|\{\mathbf{tr} : \mathbf{tr}.m = 1\}|$
	$TStop = 0$ (no stops for taxi)

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 Sect. 4, 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 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: $v_{bus}^c = \lambda^c * v_{taxi}^c$, where λ^c is a constant for temporal slot c[13]. Different cities may have different λ, here we set $\lambda^c = <0.68, 0.67, 0.77, 0.65, 0.61, 0.68, 0.62>$, $c = 1, \ldots, 7$ for Beijing by comparing the difference between taxi and bus average speed in different temporal slots (as shown in Table 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 [13]. Since the bus speed has already taken the stop time into account when calculating λ, we use $t_0 = 0$ minutes in this paper. 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 routing network is considered fixed, $EDist^c$

Table 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

is invariant with respect to c, but $ETime^c$ varies along with the traffic situations in different temporal slots.

4 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.

4.1 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 [7,9]. 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 that other factors such as 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:

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

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 paper, 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 Fig. 7a, with the increasing in distance of OD pairs, the percentage of people taking a bus is also increased. 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 Fig. 7b, with the increase in *BDist*/*TDist*, the percentage of people taking bus is decreased.

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
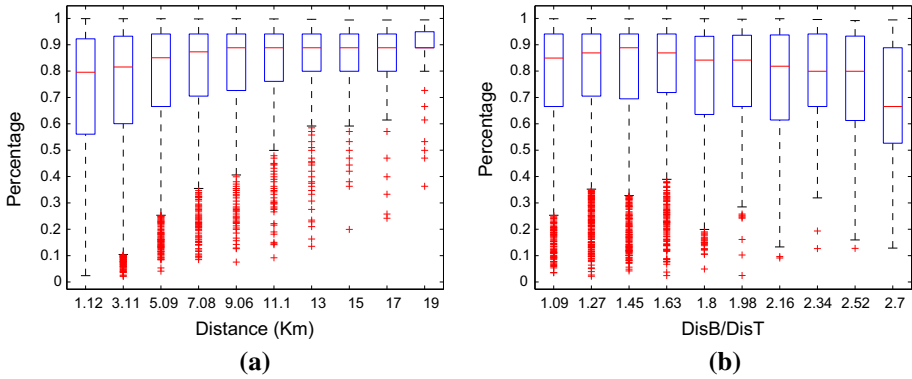
Fig. 7 Trip distribution wrt. distance. **a** Travel distance of taxi. **b** Difference of travel distance
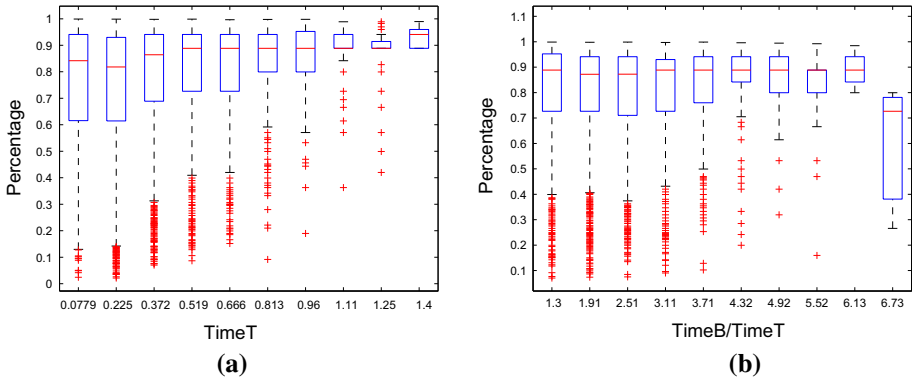

Fig. 8 Trip distribution wrt. time. **a** Travel time of taxi. **b** Difference of travel time

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-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 Fig. 8a, with the increase in 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 Fig. 8b, with the increase in *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 Fig. 9a, with the increase in fare of OD pairs, the percentage of people taking bus increases. That is 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 Fig. 9b 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 paper, we use the bus stop number per kilometer *BStop/TDist* to evaluate whether it affects
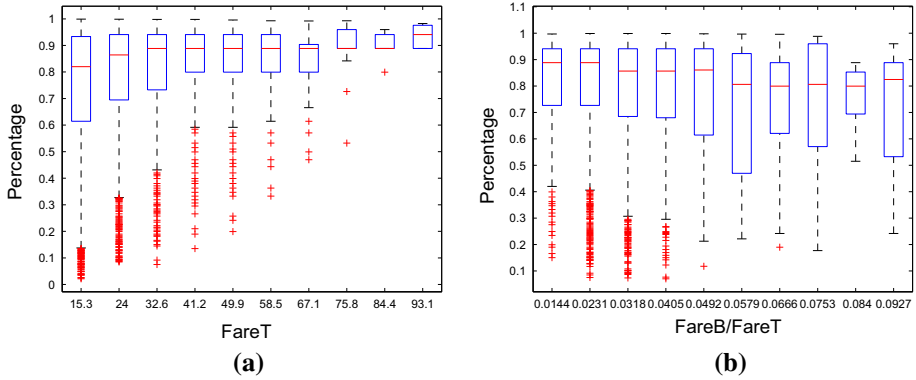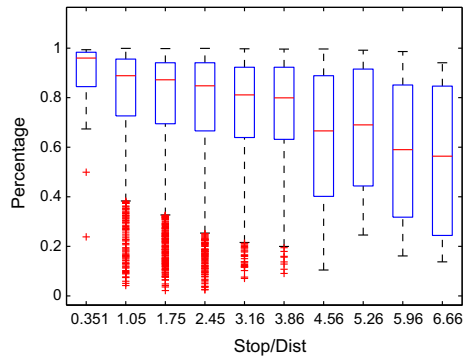
Fig. 9 Trip distribution wrt. fare. **a** Taxi fare. **b** Difference of fare

Fig. 10 Trip distribution wrt. stop number

people's decisions to choose the bus. As shown in Fig. 10, with an increase in $BStop/TDist^T$, the percentage of people taking bus drops quickly.

4.2 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 $\hat{p}_{od}^c = f(\langle \mathbf{X}_{od}^c, \mathbf{\Theta}_{od}^c \rangle)$, where $\mathbf{\Theta}_{od}^c$ 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 $\mathbf{\Theta}_{od}^c$ which minimize the model error. After the coefficients $\mathbf{\Theta}_{od}^c$ 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 $\mathbf{\Theta}_{od}^c$ 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 [40], and the
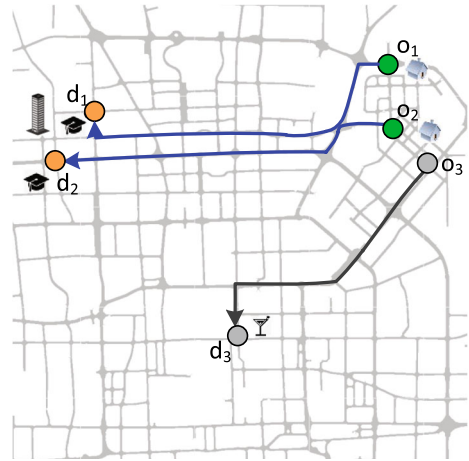
preferences of people from residential areas to commercial areas may differ from that of people from commercial areas to residential areas. On the other hand, travel preferences are more likely to be the same if two region pairs are near each other, sharing similar functions and lifestyles. As shown in Fig. 11, we have three OD pairs (o_1, d_1), (o_2, d_2), and (o_3, d_3). When learning $\Theta_{o_1 d_1}$, 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 Θ_{od}^c specifically for each OD pair (o, d), with all the observations $\{(\mathbf{X}_{ij}^c, p_{ij}^c) : s_i, s_j \in \mathbf{S}\}$. However, we have different weights $\omega_{od}^{(ij)}$ for each observation (i, j) when estimating Θ_{od}^c which minimizes the total loss $\sum_{ij} \omega_{od}^{(ij)} Loss(p_{ij}^c, f(\langle \mathbf{X}_{ij}^c, \Theta_{od}^c \rangle))$ [35]. $Loss(\cdot, \cdot)$ is the loss function of regression for each observation.

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

$$\omega_{od}^{(ij)} = \exp\left(-\frac{\alpha_{od}^{(ij)}}{2h_\alpha}\right) \cdot \exp\left(-\frac{\beta_{od}^{(ij)}}{2h_\beta}\right) = \exp\left(-\frac{\alpha_{od}^{(ij)}}{2h_\alpha} - \frac{\beta_{od}^{(ij)}}{2h_\beta}\right), \tag{1}$$

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 distances 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).

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

where $dist(s_i, s_j)$ is the Euclidean distance between the bus stops in regions 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.

$$\beta_{od}^{(ij)} = \frac{\mathrm{dcos}(s_i, s_o) + \mathrm{dcos}(s_j, s_d)}{2}, \tag{3}$$

where $\mathrm{dcos}(s_i, s_j)$ is the cosine distance calculated by $\mathrm{dcos}(s_i, s_j) = 1 - \frac{\mathbf{n}_i \cdot \mathbf{n}_j}{\|\mathbf{n}_i\| \cdot \|\mathbf{n}_j\|}$. The vector $\mathbf{n}_i = <n_1, n_2, \ldots, n_k>$ contains the POI distribution of the ith region, and k is the number of POI category. More details about POI information are given in Sect. 7.1.

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.

5 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 are no bus routes traveling 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.

5.1 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 Sect. 4, *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 [8].

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 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 12a 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 12b shows an example of searching the skyline. OD pair 1

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

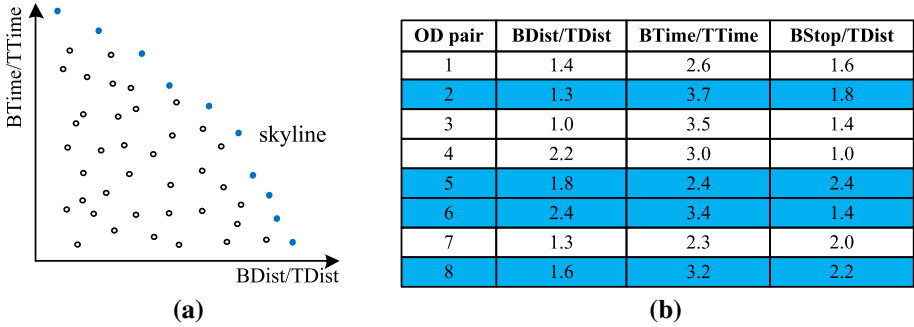**(a)**   **(b)**

Fig. 12 An example of skyline detection. **a** Two-dimensional skyline. **b** Searching skyline (color figure online)

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, points 5, 6, and 8 are detected as the skyline, while points 3, 4, and 7 are dominated by the skyline.

Note we want to find the region pairs with most 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.

5.2 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 [41]. 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 $R_{T,ij}^c$ in temporal slot c, we can derive the features $\mathbf{X}_{ij}^c$ of R_{ij}^c. 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:

$$\Delta BVol_{ij} = \sum_c Vol_{ij}^c \times \left(f_{ij}^c \left(R_{T,ij}^c \right) - p_{ij}^c \right), \tag{4}$$

and top K flawed OD pairs will be selected as candidates for bus routing optimization. Moreover, we compute $f_{ij}^c(R_{T,ij}^c) = f(\langle \mathbf{X}_{T,ij}^c, \mathbf{\Theta}_{ij}^c \rangle)$ as proposed in Sect. 4. $\mathbf{\Theta}_{ij}^c$ is the learned coefficient vector, and later we will show how to derive the features $\mathbf{X}_{T,ij}^c$ with the route $R_{T,ij}^c$.

6 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
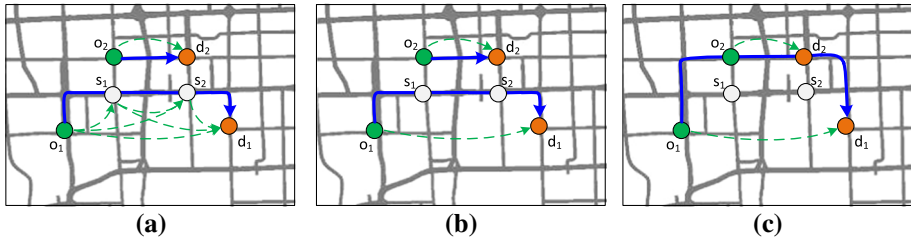
Fig. 13 Routing optimization comparison, **a** general bus routing; **b** bus routing for network renewal; **c** bus routing with constraints. *Blue solid line* stands for bus route, while *green dashed line* stands for bus travel demand of OD pair (color figure online)

[32]. In this section, we start by formulating a bus routing optimization problem, in light of the transportation mode choice model in Sect. 4. Following is our proposed solution to this problem.

6.1 Problem formulation

A general problem formulation One main goal of bus routing optimization is to accommodate bus travel demand [32]. In this paper, 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 $(\ldots, s_i, \ldots)$ 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 $R_{od} = (s_o, \ldots, s_i, \ldots, s_d)$ maximizes the bus riders of all stops traveled. In other words, R_{od} is the solution maximizing the objective function:

$$\mathcal{F}(R_{od}) = \sum_c \sum_{\substack{(s_i, s_j) \in R_{od} \\ s_i \prec s_j}} Vol_{ij}^c \times f_{ij}^c(R_{ij}), \tag{5}$$

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, $R_{ij} = (s_i, \ldots, s_j)$ is the subroute of R_{od} from stop s_i to s_j. Taking OD pair (o_1, d_1) in Fig. 13a 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 $f_{ij}^c(R_{ij}) = f(\langle \mathbf{X}_{ij}^c, \mathbf{\Theta}_{ij}^c \rangle)$ as proposed in Sect. 4. 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 paper 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 Fig. 13b, 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

$$\mathcal{F}(R_{od}) = \sum_c \left(Vol_{od}^c \times f_{od}^c(R_{od}) - BVol_{od}^c \right)$$
$$= \sum_c Vol_{od}^c \times f_{od}^c(R_{od}) - \sum_c BVol_{od}^c.$$

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

$$\mathcal{F}(R_{od}) = \sum_c Vol^c_{od} \times f^c_{od}(R_{od}). \tag{6}$$

Bus routing optimization with constraints Furthermore, in a real application of bus routing optimization, multiple flawed OD pairs need to be considered simultaneously 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 Fig. 13b that show the routing result when optimizing the two OD pairs independently, leading to two routes which exceed the budget constraints; Fig. 13c 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^c_k from Sect. 4 (i.e., f^c_k parameterized by $\Theta^c_k = \Theta^c_{o_k d_k}$), for each flawed OD pair (o_k, d_k), $k = 1, \ldots, K$, we optimize the total bus ridership under budget constraints. Supposing the optimal bus routes are $\mathbf{R} = \{R_k : k = 1, \ldots, K\}$, where R_k has an origin o_k and a destination d_k, our objective function is as follows,

$$\mathcal{F}(\mathbf{R}) = \sum_c \sum_k Vol^c_k \times f^c_k(R_k). \tag{7}$$

where $Vol^c_k = Vol^c_{o_k d_k}$. As stated previously, $f^c_k(R_k) = f(\langle \mathbf{X}^c_k, \Theta^c_k \rangle)$ and we will show how to derive the features $\mathbf{X}^c_k$ of R_k in Sect. 6.2.

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

$$\text{cost}(\mathbf{R}) \leq \mathbf{C}, \tag{8}$$

where the function *cost* is calculated with all the bus routes in $\mathbf{R}$, 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.

6.2 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, $k = 1, \ldots, K$, 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 that belong to R_k, while the taxi-related features remain the same and 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 Sect. 3), 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

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

$$B_k^c = \left(TDist_k^c; \ 0; \ TTime_k^c; \ 0; \ TFare_k^c; \ \frac{BFare_k^c}{TFare_k^c}; \ 0 \right),$$

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 $WTime_k^c = WTime^c, \forall k = 1, \ldots, K$, this cost can be written as

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

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

$$\text{cost}(\mathbf{R}) = \sum_c \frac{STime^c}{WTime^c} \sum_e \sum_k R_{ke} ECost_e^c$$

$$= \sum_c \sum_k \frac{STime^c}{WTime_k^c} BCost_{R_k},$$

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 min as the waiting time for all bus routes in this paper. In sum, our constraints in Eq. 8 can be linear with respect to the decision variables in R. However, the objective in Eq. 7 is nonlinear with the prediction function $f(z) = \frac{1}{1+\exp(-z)}$, the consequent optimization problem is nonconvex, 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

Fig. 14 A sample bus network. The edge distance and time are shown as $e = (EDist, ETime)$
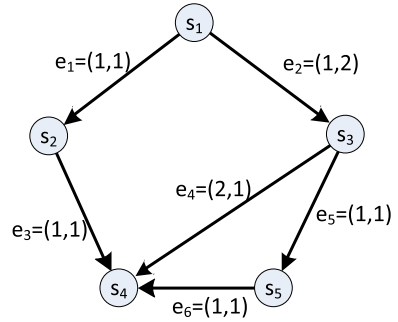
show results of the routing optimization with both nonlinear and linear prediction functions, and it can be seen that the relaxed linear approach can approximate optimal routing effectively.

An example To help understand the optimization process, we take the following sample network (shown in Fig. 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 $\sum BDist \le 3$ and $\sum BTime \le 3$. 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

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

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

given $(TDist_k, TTime_k, TFare_k, BFare_k) = (1, 1, 10, 1)$ which are constants. With Θ_k trained from our SFWLoR model, we have $z' = \langle \mathbf{X}'_k, \Theta_k \rangle$, 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 $f(R'_k) > f(R''_k)$.

6.3 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 [24] 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 unrelaxed 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:

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

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, we have:

$$
\begin{aligned}
A &= (A_1; \cdots ; A_K), \\
B &= \sum_c \sum_k Vol_k^c \langle B_k^c, \Theta_k^c \rangle, \\
P &= (P_1; \cdots ; P_K), \\
Q &= (Q_1; \cdots ; Q_K), \\
L &= \text{diag}(\mathcal{L}, \ldots, \mathcal{L}), \\
r &= (r_1; \cdots ; r_K).
\end{aligned}
$$

Here, $A_k = \sum_c Vol_k^c (A_k^c)' \Theta_k^c$, $P_k = \sum_c \frac{STime^c}{WTime_k^c} EDist^c$, $Q_k = \sum_c \frac{STime^c}{WTime_k^c} ETime^c$, and $A_k, P_k, Q_k \in \mathbb{R}^{|\mathbf{E}|}$. The matrix $\mathcal{L}$ represents the graph $\mathbf{G}$ (defined in Sect. 6) 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 $k \ne i, j$. 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 $\mathbf{0} \le R \le \mathbf{1}$. 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:

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

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

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

$$H(\mathcal{F}) = \sum_c \text{diag}\left(H_1^c, \ldots, H_K^c\right),$$

where

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

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

7 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.

7.1 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 to November, 2012. This dataset contains the following information: card id, bus route number, boarding and alighting, time, fare [42]. 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 pickup points to drop-off points. Table 4 shows some statistics of the two trip datasets.

Bus routes and road map (1) We have the bus route data, which contains 2427 stops and 1058 routes in the urban area of Beijing. After we merge the redundant stops, we obtain 1250 stops and we partition the urban area into 1250 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 1250 routing nodes, we have 3855 connection edges.

POI data A Beijing POI dataset in the year 2012 is employed to compute the functional observation weights. The number of POIs $\mathbf{n}_i = \langle n_1, \ldots, n_{10}\rangle$ in region s_i is counted following the categories shown in Table 5.

Platform The algorithms are implemented in MATLAB 2013b and C# on Visual Studio 2012. All the experiments are conducted on a 64-bit machine with 3.40 GHz Intel Core i7 CPU and 16GB memory.

---

[2] http://www.bmac.com.cn/.

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

Table 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	333 M
	Number of occupied trips	19 M
	Total trip distance (km)	156 M
Bus transactions	Number of bus stops	7810
	Time period	Aug. 2012–Nov. 2012
	Number of car holders	701,250
	Number of trips	10 M

Table 5 Category of POIs

	Category	Subcategories	Number
1	Home	Apartment building	29,246
2	Work	Government and 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	7897
7	Outdoor	Park, sports field	2211
8	Transportation	Airport, railway and bus station	15,287
9	Health care	Hospital, medical center, pharmacy	9768
10	Car service	Car sale, repair, gas station	10,781

7.2 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 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) [17] to evaluate the performance of the transportation mode choice models. The ROC curve
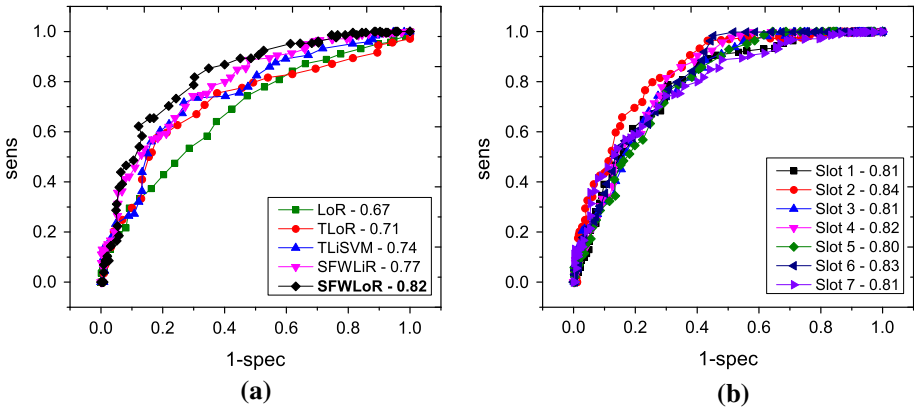
Fig. 15 Results of all the OD pairs. The number listed is the AUC score of each ROC curve. **a** Different methods. **b** SFWLoR on different slots
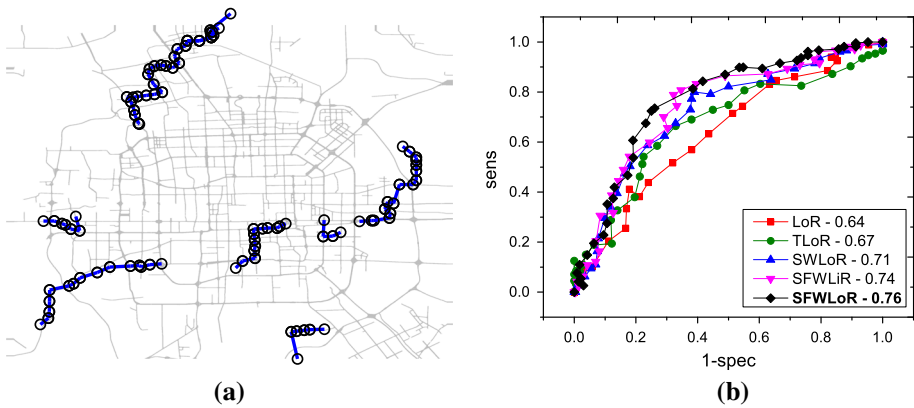

Fig. 16 Results of OD pairs with route changed. **a** Bus routes changed. **b** ROC curve

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

$$sens = tp/(tp + fn), \qquad (9)$$

$$spec = tn/(tn + fp), \qquad (10)$$

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

Results We evaluate the models with tenfold cross-validation in each temporal slot separately and then use the average of different temporal slots as the final result. Figure 15a 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 1–4) than on weekends (Slot 5–7), because there is a lot of variation occurring on weekend trips as compared to weekday trips and it increases the difficulty of modeling [3].
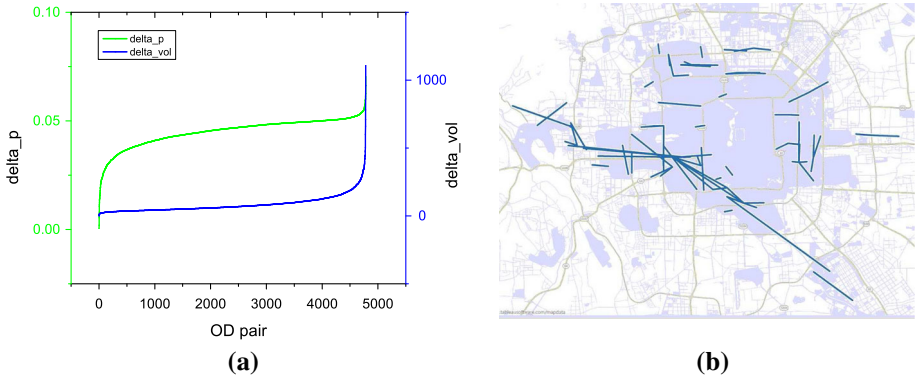
Fig. 17 Results of flawed OD pair identification. **a** Change of prob. and vol. of bus. **b** Top 100 flawed OD pairs (color figure online)

Other than the experiments with an overall evaluation on all OD pairs, we notice that routes of 8 bus lines (shown in Fig. 16a) 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 Fig. 16b.

As shown in Fig. 16, the ROC curves exhibit a consistent trend with the previous results in Fig. 15. We can see our method demonstrates an advantage compared to other methods.

7.3 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 Fig. 17a that shows the changes in probability (green line) and volume (blue line) of taking bus after using taxi routes as the upper bound of bus routes; Fig. 17b shows us top 100 flawed OD pairs.

From Fig. 17a we can see with the improvement of bus routing, an average of 5 % increase in probability taking bus is expected for all OD pairs. Moreover, we find that the bus volume increase follows Zipf's law [30], which means most of the volume increase happens among a few OD pairs. This further validates our method which focuses on these flawed OD pairs instead of all of them.

Figure 17b 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 Fig. 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 http://www.bjjtw.gov.cn/.

Table 6 Results of bus routing on top K flawed OD pairs

K	Methods	BDist (km)	BTime (h)	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

7.4 Routing optimization

Given the top K flawed OD pairs with a descending rank of potential increases in 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 (MRT) [13,21], 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,

$$\mathcal{F}_{\mathcal{SD}}(\mathbf{R}) = \sum_c \sum_k BDist_{R_k}, \tag{11}$$

$$\mathcal{F}_{\mathcal{ST}}(\mathbf{R}) = \sum_c \sum_k BTime_{R_k}, \tag{12}$$

$$\mathcal{F}_{\mathcal{MRT}}(\mathbf{R}) = \sum_c \sum_k TVol_{R_k}. \tag{13}$$

What's more, according to Eq. 7 the objective function of MCR is

$$\mathcal{F}_{\mathcal{MCR}}(\mathbf{R}) = \sum_c Vol_k^c \times f_k^c(R_k) = \sum_c \sum_k B\hat{V}ol_{R_k}. \tag{14}$$

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 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 6, where the columns show average values of statistics of each OD pair. Specifically, we first use these methods

**Fig. 18** An example of routes
generated (color figure online)
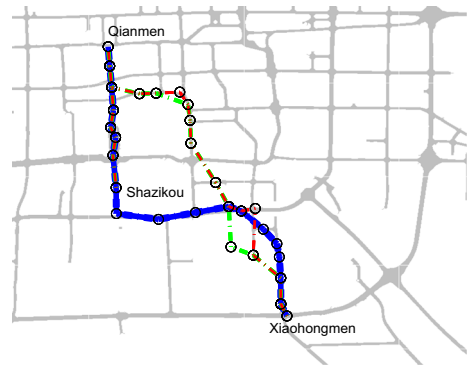
to find the best routes $\hat{\mathbf{R}}$ for our identified flawed OD pairs. For every OD pair in different temporal slots, the transportation mode model is used to predict the probability of taking bus $\hat{p} = f(\hat{R})$. Together with the total travel demand of each pair, the bus rider number can be obtained. By comparing this to historical bus rider numbers, we then get the change in 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 6, we can see MCR-LoR and MCR-LiR provide routes that lead to highest probabilities of people taking bus because they successfully measure 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 Fig. 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.

Efficient study Figure 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 do not 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 off-line manner, MCR-LoR would also be used for better planning results.

Fig. 19 Running time of bus routing
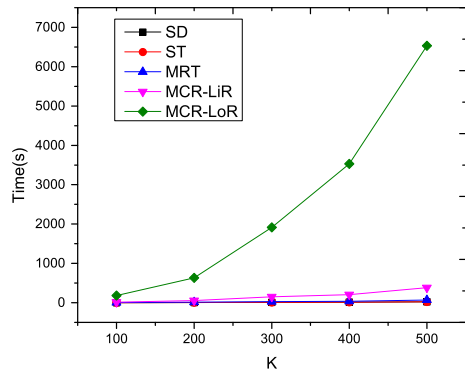
8 Conclusion

In this paper, 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 paper 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.

References

1. Ahuja RK (1993) Network flows. PhD Thesis, Technische Hochschule Darmstadt
2. Bagloee SA, Ceder AA (2011) Transit-network design methodology for actual-size road networks. Trans Res Part B Methodol 45(10):1787–1804
3. Agarwal A (2004) A comparison of weekend and weekday travel behavior characteristics in urban areas. PhD Thesis, USF
4. 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. ACM, pp 141–154

5. Aurenhammer F (1991) Voronoi diagrams: a survey of a fundamental geometric data structure. ACM Comput Surv 23(3):345–405

6. Bastani F, Huang Y, Xie X, Powell JW (2011) A greener transportation mode: flexible routes discovery from GPS trajectory data. In: GIS. ACM, pp 405–408

7. Beirão G, Cabral JAS (2007) Understanding attitudes towards public transport and private car: a qualitative study. Transp Policy 14(6):478–489

8. Borzsony S, Kossmann D, Stocker K (2001) The skyline operator. In: 17th international conference on data engineering, 2001. Proceedings. IEEE, pp 421–430

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

10. Ceder A, Wilson NHM (1986) Bus network design. Transp Res Part B Methodol 20(4):331–344

11. Chakroborty P (2003) Genetic algorithms for optimal urban transit network design. Comput Aided Civ Infrastruct Eng 18(3):184–200

12. Chakroborty P, Wivedi T (2002) Optimal route network design for transit systems using genetic algorithms. Eng Optim 34(1):83–100

13. 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: PerCom. IEEE, pp 225–233

14. de Dios Ortuzar J, Willumsen LG (2011) Modelling transport, Wiley press

15. de Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: the privacy bounds of human mobility. Sci Rep 3(1376):1–5

16. Fan W, Machemehl RB (2006) Optimal transit route network design problem with variable transit demand: genetic algorithm approach. J Transp Eng 132(1):40–51

17. Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27(8):861–874

18. Ge Y, Xiong H, Tuzhilin A, Xiao K, Gruteser M, Pazzani M (2010) An energy-efficient mobile recommender system. In: KDD, pp 899–908

19. Giannotti F, Nanni M, Pinelli F, Pedreschi D (2007) Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 330–339

20. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782

21. Guihaire V, Hao J-K (2008) Transit network design and scheduling: a global review. Transp Res Part A Policy Pract 42(10):1251–1273

22. Kepaptsoglou K, Karlaftis M (2009) Transit route network design problem: review. J Transp Eng 135(8):491–505

23. Kim S, Shekhar S, Min M (2008) Contraflow transportation network reconfiguration for evacuation route planning. IEEE Trans Knowl Data Eng 20(8):1115–1129

24. Land AH, Doig AG (1960) An automatic method of solving discrete programming problems. Econometrica 28(3):497–520

25. Lathia N, Capra L (2011) Mining mobility data to minimise travellers' spending on public transport. In: KDD. ACM, pp 1181–1189

26. Lathia N, Froehlich J, Capra L (2010) Mining public transport usage for personalised intelligent transport systems. In: ICDM, pp 887–892

27. Liu C-L, Pai T-W, Chang C-T, Hsieh C-M (2001) Path-planning algorithms for public transportation systems. In: Proceedings. 2001 IEEE intelligent transportation systems. IEEE, pp 1061–1066

28. 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: ITSC. IEEE, pp 1–6

29. Liu Y, Liu C, Yuan NJ, Duan L, Fu Y, Xiong H, Xu S, Wu J (2014) Exploiting heterogeneous human mobility patterns for intelligent bus routing. In: ICDM. IEEE, pp 360–369

30. Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge

31. Monreale A, Pinelli F, Trasarti R, Giannotti F (2009) Wherenext: a location predictor on trajectory pattern mining. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 637–646

32. Pratt RH, Evans IV et al (2004) Traveler response to transportation system changes. Chapter 10, Bus routing and coverage

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

34. Song C, Zehui Q, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021

35. Tilo S (2010) Data fitting and uncertainty: a practical introduction to weighted least squares and beyond. Vieweg, Teubner, Wiesbaden. ISBN:3834810223

36. Sun JB, Yuan J, Wang Y, Si HB, Shan XM (2011) Exploring space–time structure of human mobility in urban space. Phys A 390(5):929–942
37. Utsunomiya M, Attanucci J, Wilson N (2006) Potential uses of transit smart card registration and transaction data to improve transit planning. Trans Res Rec 1971(1):119–126
38. Watkins KE, Ferris B, Borning A, Rutherford GS, Layton D (2011) Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Trans Res Part A Policy Pract 45(8):839–848
39. Yen JY (1971) Finding the k shortest loopless paths in a network. Manag Sci 17(11):712–716
40. Yuan J, Zheng Y, Xie X (2012) Discovering regions of different functions in a city using human mobility and pois. In: KDD. ACM, pp 186–194
41. Yuan J, Zheng Y, Xie X, Sun G (2013) T-drive: enhancing driving directions with taxi drivers' intelligence. IEEE Trans Knowl Data Eng 25(1):220–232
42. Yuan NJ, Wang Y, Zhang F, Xie X, Sun G (2013) Reconstructing individual mobility from smart card transactions: a space alignment approach. In: ICDM, pp 877–886
43. Zheng Y, Liu Y, Yuan J, Xie X (2011) Urban computing with taxicabs. In: Proceedings of the 13th international conference on Ubiquitous computing. ACM, pp 89–98

Yanchi Liu is currently a Ph.D. student in the Management Science and Information Systems Department at Rutgers, the State University of New Jersey. He received the B.E. degree from the Civil Aviation University of China in 2005, and the Ph.D. degree from the University of Science and Technology Beijing in 2012. His main research interests include data mining, mobile computing, and urban computing.
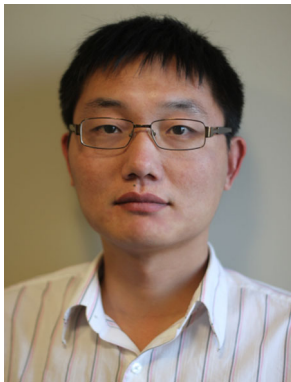
Chuanren Liu received the B.S. degree from the University of Science and Technology of China (USTC), the M.S. degree from the Beijing University of Aeronautics and Astronautics (BUAA), and the Ph.D. degree from Rutgers, the State University of New Jersey. He is currently an assistant professor in the Decision Science and Management Information Systems Department, Drexel University. His research interests include data mining and knowledge discovery, and their applications in business analytics.

Nicholas Jing Yuan is a researcher in Microsoft Research Asia. He received the Ph.D. degree in Computer Science in 2012 and the B.S. degree in Mathematics in 2007, both from University of Science and Technology of China. Over the past few years, he has published a series of papers in top conferences/journals such as KDD, TKDE, Ubicomp, WWW, and ICDE as a leading author. He also regularly served as program committee members among top conferences in the data mining community. He was awarded Microsoft Fellowship in 2011. Currently, his research interests include spatial–temporal data mining, behavioral mining, and computational social science. He is a member of ACM and IEEE.

Lian Duan is currently an assistant professor in the Department of Information Systems and Business Analytics at Hofstra University. He received the Ph.D. in Management Sciences at the University of Iowa in 2012, and the Ph.D. in Computer Sciences at Chinese Academy of Sciences in 2007. His main research interests include correlation analysis, community detection, business analytics, and geospatial analysis.

Yanjie Fu received the B.E. degree from the University of Science and Technology of China, Hefei, China, 2008, the M.E. degree from the Chinese Academy of Sciences, Beijing, China, 2011. He is currently a Ph.D. candidate in the Management Science and Information Systems Department at Rutgers University. His research interests include data mining, urban computing, and mobile analytics.

Hui Xiong is currently a Full Professor and Vice Chair of the Management Science and Information Systems Department, and the Director of Rutgers Center for Information Assurance at the Rutgers, the State University of New Jersey, where he received a 2-year early promotion/tenure (2009), the Rutgers University Board of Trustees Research Fellowship for Scholarly Excellence (2009), and the ICDM-2011 Best Research Paper Award (2011). He received the B.E. degree from the University of Science and Technology of China (USTC), China, the M.S. degree from the National University of Singapore (NUS), Singapore, and the Ph.D. degree from the University of Minnesota (UMN), USA. His general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. He has published prolifically in refereed journals and conference proceedings (3 books, 60+ journal papers, and 90+ conference papers). He is a co-Editor-in-Chief of Encyclopedia of GIS, an Associate Editor of IEEE Transactions on Data and Knowledge Engineering (TKDE), IEEE Transactions on Big Data (TBD), ACM Transactions on Knowledge Discovery from Data (TKDD), ACM Transactions on Management Information Systems (TMIS), and Knowledge and Information Systems (KAIS). He has served regularly on the organization and program committees of numerous conferences, including as a Program Co-Chair of the Industrial and Government Track for the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), a Program Co-Chair for the IEEE 2013 International Conference on Data Mining (ICDM), and a General Co-Chair for the IEEE 2015 International Conference on Data Mining (ICDM). He is an ACM Distinguished Scientist and a senior member of the IEEE.

Songhua Xu is an assistant professor at Information Systems Department at New Jersey Institute of Technology. He received his Ph.D., M.Phil., M.S., all in computer science from Yale University. His primary research interests include information retrieval and management, web search and data mining, innovative applications of AI, and intelligent systems for biomedical applications. He also has secondary research interests spanning across visual computing, multimedia, human–computer interaction, computer graphics, visualization, digital arts and design. He is particularly passionate about building human-centered applications that benefit people and society through advanced computing techniques.

Junjie Wu received his Ph.D. degree in Management Science and Engineering from Tsinghua University in 2008. He is currently a Professor in Information Systems Department of Beihang University, the director of Research Center for Social Computing, and the vice director of Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations. His general area of research is data mining and complex network. He has published prolifically in many leading conferences and journals such as KDD and TKDE. He is the recipient of NSFC Excellent Youth Award (2013) and National Excellent Doctoral Dissertation Award (2010).