Recommendation for Ridesharing Groups Through Destination Prediction on Trajectory Data

Lei Tang, *Member, IEEE*, Zongtao Duan, Yishui Zhu, *Member, IEEE,* Junchi Ma, and Zihang Liu

***Abstract*— In this paper, we aim to provide an optimal passenger matching solution by recommending ridesharing groups of passengers from GPS trajectories. Existing algorithms for rider grouping usually rely on matching pre-selected origin-destination coordinates. Unfortunately, the semantics in the spatial layout (e.g., social interactions and properties of the locations) are ignored, leading to inaccuracies in discovering the ridesharing groups. Meanwhile, the destinations manually entered by users impact the accuracy of matching, as these addresses are usually not available in a road network or are not optimal for passenger pickup. This is particularly true when a passenger travels in a less familiar place. Given a set of passengers and the distribution of their destination, our approach is to compute the ridesharing matching between passengers. The raw GPS trajectories can be characterized by a combination of time constraints, traffic environments, and social activities. We first developed a PrefixSpan-prediction using a partial matching (P-PPM) destination-prediction algorithm to mine the frequent movement patterns from the trajectory data and determine the confidence of the movement rules. Our method uses the total travel time as the matching objective. Our approach is superior to the baseline methods in terms of accuracy (increased from 46% to 80%). We have also achieved significant improvements on other metrics, such as users' saved travel distance. We demonstrated that using our proposed method, a group of passengers could save over 19% of total travel miles, which shows that the ridesharing scheme could be effective.**

***Index Terms*— Ridesharing group, recommendation, destination prediction, trajectory.**

I. Introduction

AS AN alternative means of traveling short distances, ridesharing can help alleviate traffic congestion [1] and road wear and reduce air pollution, and energy consumption [2]–[4]. As a simple yet effective form of ridesharing, "slugging" allows the origin or destination of the passenger not to be on the way of a route of a driver [5]. Slugging is a unique form of ridesharing that has been around in the Northern Virginia and Washington, DC area since the 1970s, shortly after the high occupancy vehicle (HOV) lanes were opened for carpooling and vanpooling [6]. To use the HOV lane, a driver might be able to pick several passengers that have routes "covered" by his own. Passengers may be asked to meet at a specific location (e.g., points of interest (POIs) and bus stops) to take the ride and will be dropped at a particular place [7].

Considering multi-passenger slugging, people must share time and space resources with others in the same car simultaneously [8]. A major issue in ridesharing is passenger matching [9].Previous studies found that ridesharing only with friends or colleagues is restricted since the ridesharing is provided without considering a large number of potential participants with similar interests [10]. Zhang and Zhao [11] stated that optimal ridesharing can be achieved based on the interests of all individuals in a system. In real-world ridesharing, a set of passengers may be interested in pick-up points and destinations. For instance, commuters often would like to accompany others having the same routes [12]. Since they may prefer to travel together [13], it is possible to select common destinations to which passengers may agree to rideshare [14] and divide passengers into travel groups that fit in a car by capturing connections in the trajectory data. Khan et al. also showed that better ridesharing arrangements are possible and further enjoyable when riders agree to join a ridesharing group and reach a decision [15] in ridesharing plans, rather than going to their pre-chosen destinations by car randomly. A positive understanding of the role of common destinations in ridesharing is an important prerequisite for addressing passenger matching.

A typical form of slugging, called as origin-destination (OD)-slugging, is investigated herein. The passengers walk to the origins of the drivers, board at the departure time, debark at the drivers' destinations, and then walk to their destinations [16]. A case with one driver d, who is the owner of the car, and three riders r_1, r_2 and r_3 is considered. They want to go to Bell Tower, Xi'an, where parking fares are costly and there are many close-by locations to visit. They may not have a specific starting attraction in mind. Fig. 1 depicts the locations of riders and their origins. In the optimal solution, POIs p_1, p_2 and p_3 are presented as their destinations. d can pick-up r_1 and r_2 at his/her origin, q_1, and drop them off at d's destination, i.e., q_2; r_1 and r_2 then walk to p_1 and p_2, respectively. r_3 goes to p_3 by another car. If r_1 and r_2 go to their pre-selected destinations without first agreeing on common destinations and then joining a group, the total travel distance for all vehicles increases. Therefore, agreeing on destinations and grouping a set of riders related through spatial proximity is especially useful for producing further benefits for ridesharing [17].

Manuscript received August 3, 2018; revised January 7, 2019, May 16, 2019, July 30, 2019, and October 8, 2019; accepted December 16, 2019. Date of publication December 27, 2019; date of current version February 2, 2021. This work was supported in part by the Key Scientific and Technological Innovation Team of the Shaanxi Province, China, under Grant 2017KCT-29, and in part by the Key Research and Development Plan Project of the Shaanxi Province, China, under Grant 2019ZDLGY17-08 and Grant 2019ZDLGY03-09-01. The Associate Editor for this article was J. Sanchez-Medina. *(Corresponding authors: Lei Tang; Yishui Zhu.)*

The authors are with the School of Information Engineering, Chang'an University, Xi'an 710064, China (e-mail: tanglei24@chd.edu.cn; ztduana@chd.edu.cn; yszhu@chd.edu.cn; majunchi@chd.edu.cn; lzhhollyhang@163.com).

Digital Object Identifier 10.1109/TITS.2019.2961170

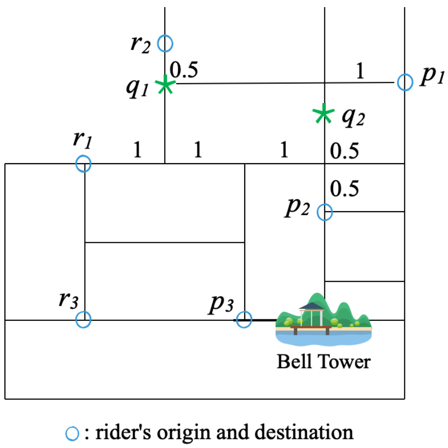


Fig. 1. Illustration showing that the total travel distance with ridesharing for all cars is 2+0.5=2.5 and without ridesharing is (1+1+2+1)+(0.5+1+2+0.5)=9.

In this paper, we develop a method to predict the destinations accurately, detect groups in trajectories, and divide the set of passengers into optimal groups that fit in a car. Existing methods for optimizing ridesharing usually rely on matching a driver and rider with a pre-selected OD and location data [18], [19]. Unfortunately, real group relationships may be missed. Hence, we focus on selecting groups based on trajectory-related information (e.g., spatial dispersion [20], temporal duration, and movement velocity) of individuals and the semantic properties of the space. In addition, extra work for the users to enter the full name of the destination would significantly degrade the passenger experience [21]. Knowing the behavioral factors that influence destination choices and identifying the destination to which the user wants to go will serve as a beneficial modeling tool for transit authorities [22] to extend existing shared mobility services. Unlike existing methods [23], [24] that attempt to suggest the nearest location, to which passenger might not be interested in going, from the actual destination, we take the sequences of each user's historical spaces as the candidate set for destination prediction. Thus, it is highly likely that the passengers will regard the predicted destination as an intended location [21], [24], accept the recommendation, and be interested in sharing any rides. Finally, an optimal group of riders whose proximity similarity is likely a manifestation of shared relationship is detected to minimize the total travel time. The contributions of the paper can be summarized as follows:

- We generalize OD-slugging by allowing riders to agree on destinations and form a group that fit in a car. The behaviors of participants are explicitly modeled.
- We define a new group discovery problem. We solve it by generating the semantic features of the spatial layout, learning the possible distribution of the intended destination, and imposing constraints on the extra walking distance and additional waiting time, thereby reducing the total travel time.
- Our extensive experiments demonstrated that our method is efficient and can be deployed in real-world applications.

The rest of this paper is organized as follows. In Section II, we briefly review the existing literature. In Section III, we present the group recommendation system and Section IV introduces the destination prediction system. We provide experiments that show the effectiveness of our proposed algorithms and end by discussing the limitations of the information used in the group recommendation as well as future directions to overcome these issues.

II. Literature Review

In contrast to new forms of shared-use mobility, slugging has been in existence for over 30 years and is entirely run informally by its passengers. Researchers have been fascinated by this phenomenon and have conducted studies in the past. Shaheen et al. analyzed the choice factors of slugging in San Francisco through 16 interviews with drivers and passengers [25]. They pointed out flexibility is an essential motivation for sluggers, and discussed the changes of behaviors, e.g., the distributions of departure and waiting times as well as the walking distance. In slugging, people must share time and space resources with others in the same car simultaneously. A person then becomes more dependent on his/her partners when ridesharing. Therefore, it is better for individuals to find a regular partner when sharing a ride [26]. Wang *et al.* [27] developed a social taxi-sharing system based on space-dependent preferences to improve the single driver-passenger match rates. Although previous researchers [28] investigated the idea that people may feel more comfortable when sharing rides with people who are friends, family, and colleagues, this kind of trust conscious ridesharing is either too restricted or too relaxed to be practical [10]. Wang *et al.* [29] also discussed that limiting ridesharing to friends while rejecting strangers also reduces ride choices and increases the detour cost; specifically, the spatial friendship distribution is sparse. Instead, riders further seek stranger car-mates that have similar preferences on travelling instead of being from the same household or friends.

In this paper, we study group recommendation from trajectories, aiming to identify groups of passengers [10] from trajectory data based on additional behaviorally driven markers of individual movement. The driver and all the passengers must agree on the costs and schedules, including the position and temporal elements. Furuhata *et al.* [30] classified the demands of ridesharing participants based on what information was used to form driver-passenger matches. The routing, OD-pairs, and departure or arrival times are primary criteria for seeking ridesharing participants. Most of the current methods are more focused on accurately matching such information predetermined by passengers. Bakkal *et al.* [31] proposed a novel method for ridesharing group recommendations. The Neo4j-based spatial-temporal tree was established using trajectory data to extract the varying time and spatial data. Passengers with similar travel times and locations were recommended to join a group. Rigby et al. considered vehicle accessibility as a ridesharing service. To improve the accuracy of pick-up, the proposed OppRide developed the network time prism with a road network to represent the service. A group of the passengers could subsequently be informed where they should be boarded and dropped [32]. SaRG [10] was presented to group passengers according to their social connections, i.e., each member of the ridesharing group should have acquaintance relationships (e.g., friends or colleagues). To meet the social comfort and trust in ridesharing, SaRG grouped the members whose trips were similar to that of the driver and were familiar to each other. However, there are a few different criteria, such as pick-up areas within a passenger-specified radius from specific addresses, predetermined meeting spots with a reasonable time window, and travel purpose, e.g., going home during school breaks [10]. Differences in the criteria selected may yield sub-optimal recommendations for a group. For example, OD-pairs and time cannot be applied well for a group recommendation in dynamic ridesharing due to the lack of knowledge of routing, leading to the inability to fully pick up the riders en route.

For existing matching methods, passengers manually input their requests and negotiate with driver(s) using additional communication channels (e.g., voice or text) to confirm an accurate pick-up location. It incurs extra work for the passengers to enter such information. The destination prediction is helpful to inform the passengers of the route and at which time they should be boarded. Thus, the passenger experience would be enhanced if the intended destination could be accurately predicted when a passenger submitted the request. Destination prediction mainly captures passenger preferences on movements and social interactions from trajectory data. Most existing approaches predict the destination according to the existing trip based on history trajectories. They typically use a Markov model to identify the transfer probability between two near neighbors and focus on the accuracy of provisioning [33]. For example, T-DesP model was proposed [24] to predict the destination by a Markov model and solved the problem of data sparsity by using a content-based tensor-decomposition method. Association rules were also introduced for destination prediction to detect the many spaces visited from a passenger's historical trip data. Rules with respect to movement patterns were then generated to conduct a match between the next location and common areas [34]. Lian and Xie [35] considered the correlation between a passenger's points of interest and his/her historical trip locations to discover the frequent subsequence of trajectories for prediction.

How to identify the destination to which a passenger wants to go is still a problem. Additional semantic information (e.g., activity and waiting time at specific locations) that is especially useful in capturing such preferences is ignored when employing a raw trajectory from GPS-enabled devices. Our work can accommodate various latent attributes embedded within the raw trajectory data and predict a destination list when a passenger requests the ridesharing. To achieve the benefits of ridesharing, it is possible to force riders to agree on destinations and to group a set of riders related through spatial-temporal movements. This helps to improve the willingness of the riders and the detour tolerance to share rides. For this, Wang *et al.* [27] employed a social network to achieve a taxi-sharing matching for higher travel satisfaction with persons of different affinities. Instead, we sought a group consisting of strangers with similar travelling preferences instead of being from the same household or friends or matching between a driver and single rider. Our study is most similar to work done by Ma and Wolfson [16], who also examined passengers' preferences in terms of travel time delays. However, we constrained the grouping of passengers by vehicle capacity, spatial proximity, and temporal duration, and measured the delay of the group by comparing those of all the members.

III. Group Recommendation System in OD-Slugging

We first introduce some notations. It is assumed that there is a driver u and passenger υ, each with travel plans.

Definition 1 (OD-Slugging): An OD-slugging is defined as a pair (u,V) with driver $u's$ path, i.e., P, where V denotes a set of passengers. The pick-up and drop-off location of $\upsilon \in V$ are along the path P, and each origin or destination of υ is not on the path.

Passengers and drivers must negotiate [36] in advance to determine all the details of the ride, including the ride time and the location to pick up and drop off the passengers. In general, the negotiation describes a 3-way handshake consisting of (1) the passenger making a request via an app, (2) the system matching and sending options to the passenger, and (3) the passenger receiving and accepting the arrangement. Our study on negotiation is most similar to work done by Dogru [37], who also examined the route matching by delivering the requests and offers of the system. However, as opposed to our work, the process of negotiation occurred between a single rider and a driver, not between the ridesharing groups and a set of drivers. Furthermore, there is no need for a passenger to create a route by entering a destination and desired stops in our model. The negotiation enables an automatic matching procedure between a group of passengers and the most suitable driver nearby.In slugging, a driver may pick-up and deliver multiple passengers to their destinations. Therefore, it is essential to assign a group to drivers optimally in a negotiation. Fig. 2 shows the negotiation model of the system in OD-slugging. The mobile connection and work flow between passengers and drivers are automatically completed through an existing ride service system, e.g., DiDi. A system receives and analyzes the requests of passengers, such as origin, departure time, acceptable waiting time and walking distance. The system subsequently analyzes the request according to the real-time supply responses of nearby drivers and other passengers' demands. A recommendation procedure between a group of passengers and the most suitable driver nearby is implemented accordingly, including destination prediction and optimizing travel time of a group. The matched driver and passengers will receive the information, including pick-up locations and associated departure time for the passengers and the drop-off locations for the driver.

Detailed information about the notation used in Section III and IV is provided in the appendix.

A. Group Recommendation Model

The ridesharing group recommendation system aims to divide riders into groups that fit in a car and minimize their

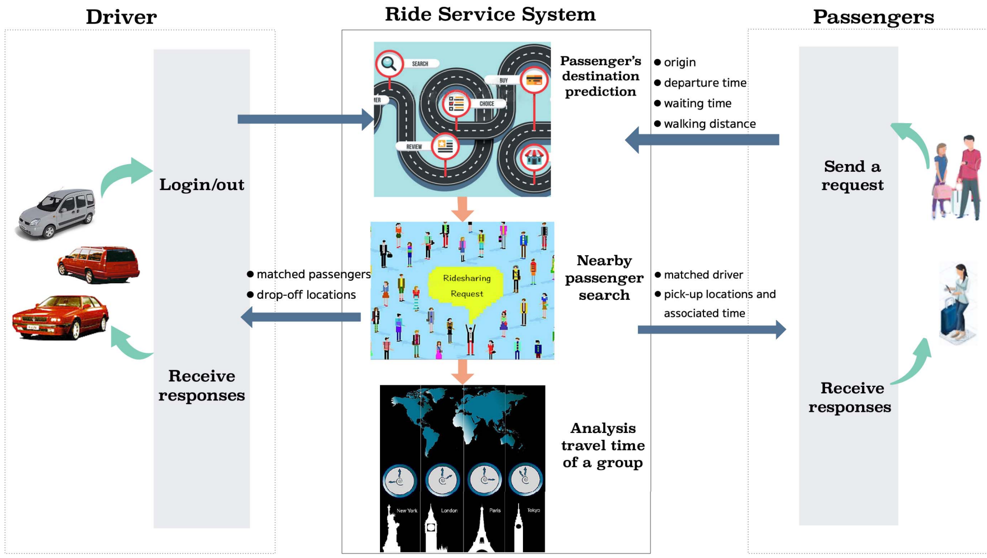


Fig. 2. Negotiation model in OD-slugging, including (a) passenger's destination prediction (b) nearby passenger search, and (c) analysis of travel time of a group.

travel time. The trajectories in a specific ridesharing group exhibit similarities in the movement-related features.

We considered the scenario of a single driver with multiple passengers, in which passenger-passenger pairs are discovered based on the commonly used criteria (i.e., location, destination, and ride time) and behaviorally-driven markers of individual movement. Due to the worsens traffic congestion, passengers may experience excess waiting times or even fail to obtain a ride. Thus, we extend the grouping criterion to consider the markers, i.e., waiting time from the presence of each passenger to the departure of a vehicle. The vehicle will pass by somewhere within walking distance of the passengers. We also introduce another marker, i.e., the walking time of a passenger between the location of origin/drop-off and pick-up/destination, into the grouping criterion.

For a given passenger $v_i \in V(i = 1, 2, \cdots, I)$ and driver u,the slugging recommendation is expressed as a matrix $M_u = (m_{ij})_{I \times I}$,

$$m_{ij} = \begin{cases} 1, & v_j \text{ is recommend to be a group with } v_i \\ 0, & otherwise \qquad\qquad\qquad 1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{I} \end{cases}$$

In this scenario, a driver picks up H passengers at most during his trip. This implies a necessary condition for the feasibility of a match between driver u and a ridesharing group: only if $\forall i, \sum_{j=1}^{I} m_{ij} \leq H, i \neq j$. Each passenger's requests on departure arise at their place of stay. The vehicle is assumed to be traveling on the shortest path between a set of locations for picking up and dropping off, where the travel time is independent of the flow [38]. For each passenger, whether they would accept the recommendation is based on their interests in saving travel time.

Definition 2 (Ridesharing Group): Given a passenger-passenger pair (v_i, v_j) and their OD pairs and time windows, a ridesharing group for slugging is defined as $G_u = \{(v_i, \cdots, v_H) | rank(M_u) = 1\}$, where the following is assumed:

1) The distance from the group to a driver is the maximum network distance from every group member's OD pairs to those of driver u, which is calculated as,

$$d^o(u, G_u) = \max_{v_i \in G_u} d^o(u, v_i), \;\; d^*(u, G_u) = \max_{v_i \in G_u} d^*(u, v_i) \quad (1)$$

Equation (1) yields the distance for each pair incrementally by Dijkstra's algorithm over a road network.

2) Each passenger, v_i, arrives at the pick-up location in advance. The additional waiting time for traveling together is denoted as the time difference between all the v_i's arrivals and u's departure. The extra time cost associated with walking is $\frac{d^o(u,v_i)}{V}$, where V denotes the walking speed. $tw(,)$ indicates the time difference between two points.

$$TW(u, G_u) = \max_{v_i \in G_u} tw(u.t, v_i.t + \frac{d^o(u, v_i)}{V}) \quad (2)$$

Equation (2) allows the specification of a maximum passenger waiting time at the pick-up points.

The critical problem for the ridesharing group recommendation is to search similar movement patterns characterized by passengers' OD pairs and departure times. We define the issue of group recommendation in OD-slugging as an optimization for minimizing the travel times of passengers, which, given a driver considering the route and time, returns the top k pairs of passengers from the candidate sets that are highly similar to each other (ranked by their similarity values of OD pairs). Therefore, we divide the group recommendation model into two sub-models. One model predicts each passenger's destination, and the other model formulates an optimization problem to minimize the travel times of passengers. The relations between the two models are that the destination POIs are presented to find a set of riders whose origins and destinations are in a small distance range. We can subsequently

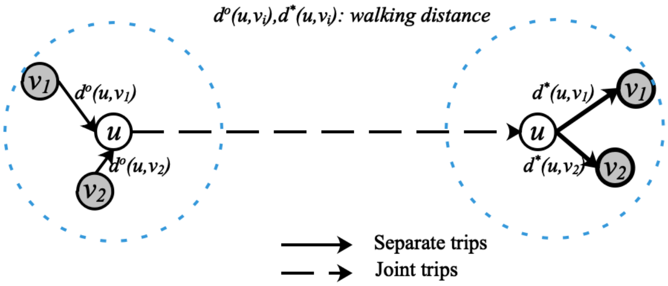


Fig. 3. Example of a single driver, ridesharing group arrangement.

group H riders by computing the group's total travel time. Finally, we identify the optimal group.

B. Optimization Model

In our system, multiple passengers can be recommended to travel together; thus, all these passengers contribute to the travel time. We set up the grouping criterion for two passengers by their extra walking distance and additional waiting times.

If a driver would like to share a ride with some passengers, the passengers will arrive at the pick-up location before the driver's departure. Thus, if two passengers v_1, and v_2 are grouped as G_u, then they can cover their walking distances (i.e. $d^o(u, v_i)(i = 1, 2)$) in the time interval of $TW(u, v_i)$. h_i indicates the constraint on walking; that is, $h_i = 1$ means v_1 and v_2 can travel together because both can arrive at the pick-up location in advance, and it is 0 otherwise. An OD-slugging trip includes the joint trip and the four separate trips. Equation (3) can be used to calculate the potential travel time of a ridesharing group of driver u. An illustration of Eq. (3) is shown in Fig. 3.

We seek to group passengers in the system in a way that minimizes the total travel time. v_1's and v_2's travel time on slugging trips is $TT(u, G_u)$, which includes the time interval from v_i's arrival to departure of driver u, i.e., $TW(u, G_u)$, and the time taken to walk towards v_i's destination. In particular, the travel time equals 0 if no one is recommended to share a ride for u, i.e., $h_i = 0$.

$$TT(u, G_u) = (TW(u, G_u) + \frac{V + d^*(u, G_u)}{V})^{h_i} - 1,$$
$$h_i = \begin{cases} 1, & \frac{d^o(u, G_u)}{V} \le \min_{v_i \in G_u} TW(u.t, v_i.t), \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

As mentioned above, there is a constraint in our model: $\sum_{i=1}^{I} h_i \le H$, which means that each driver can deliver at most H passengers at a time, leading to the following group recommendation problem, where the size of ridesharing group is H.

$$\begin{cases} \min_{h_i, H} TT(u, G_u), \\ s.t. \sum_{i=1}^{I} h_i \le H, \quad h_i \in \{0, 1\}. \end{cases} \tag{4}$$

In slugging, we provide optimal matches between potential passengers for a ridesharing group by predicting destinations of the passengers. First, we search the passengers with requests near the driver's origin in radius and time and subsequently predict the destinations of these passengers. Finally, a ridesharing group of H potential passengers with the minimum travel time is selected. Algorithm 1 describes the detailed procedure. A similarity search will obtains the driver-rider pairs with geographic proximity and time constraints (lines 2-6 in Algorithm 1). It then calculates the travel time of each pair to obtain all possible group subsets G_u for u after each enumeration of similarity search (lines 15 and 16). We return the recommendation M_u for each of them with key (u, G_u). The top-k v_i choices minimizing the TT consists of each group G_u. To support the group recommendation, the function $P - PPM(v_i)$ in line 14 from Algorithm 2 obtains the destination set P^* and identifies v_i's destination in line 15.

Algorithm 1 Ridesharing Group Discovery

Input: passenger set PS, passenger $v's$ request $T_v = (v.o, v.d)$, driver $u's$ offer $T_u = (u.o, u.d, u.t)$, departure time of a location lev_{time}, walking speed V, a set $SC_1 = \emptyset$, searching radius $\theta_d = d^o(u, G_u)$, time span for waiting $\theta_t = TW(u, G_u)$, $\|, \|$ denotes the road network distance between two spatial points, and destination set P^* for all passengers
Output: G_u

1: **for** each v in PS **do**
2:   finding $v's$ a stay point loc which is apart from $u.o$ less than θ_d
3:   $t_1 = \frac{\|loc, u.o\|}{V}$
4:   **if** tw$(u.t, t_1 + loc.lev_{time}) \le \theta_t$ **then**
5:     set loc to be $v.o$
6:     put v, $v.o$, t_1 in SC_1
7:   **else**
8:     delete v
9:   **end if**
10: **end for**
11: **while** $size(G_u) < H$ **do**
12:   **for** $i = 1$ to $size(SC_1) - 1$ **do**
13:     $P^* \leftarrow P\_PPM(v_i)$
14:     set loc which has maximal probability in P^* to be $v_i.d$
15:     calculate TT(u, v_i)
16:     put v_i in G_u
17:   **end for**
18: **end while**
19: find the H passengers who has minimal TT(u, G_u)
20: return G_u

IV. Destination Prediction

In this section, we describe our proposed destination-prediction model. By analyzing many users' trajectories, we discovered that the same passenger tends to go to a fixed set of location types. We supposed that the location type is an essential factor for predicting a passenger's intended destination. We modeled the probability distribution of a passenger's destination using our proposed P-PPM algorithm, in which the passenger's historical departure longitude, latitude and time are utilized.

Algorithm 2 P-PPM Destination Prediction

Input: trajectory model Tra_Loc, $minsup$, set of rules and their confidence R, set of frequent movement patterns FP, semantic trajectory model Tra_ST, set of the rules matching successfully R^*

Output: destination predicted D

$FP = PrefixSpan(<>, 0, Tra\_ST)$
for each Q^K in FP **do**
  determine R_{Q^K} and con_{S_K}
  put R_{Q^K}, con_{S_K} into R
end for
for $(j = h;\ j;\ j--)$ **do**
  match S_j from Tra_ST with R
  if R_{Q^K} in R includes the S_j as one of the front items **then**
    put R_{Q^K} in R^*
    break
  end if
end for
for each L_i in Tra_Loc **do**
  $p_i = PPM(N, Tra\_Loc)$
  match S_i of L_i with R^*
  $p'_i = \omega \cdot p_i + (1 - \omega) \cdot con_{S_i}$
end for
take L_i as the candidate destion D
return D

A. Model Description

1) Semantic Trajectory Model: A trajectory model that indicates the way a passenger υ traveled in the past sequentially over a 1-d period is defined as $Tra\_Loc_\upsilon = loc_0, \cdots, loc_i, \cdots, loc_h, (h > 1)$, where loc_i represents a location from the trajectory database. $loc_i = (L_i, arv_{time}, lev_{time})$, which is defined according to the latitude (lat) and longitude (lon) of a location L_i, the arrival time (arv_{time}) and departure time(lev_{time}). All m historical trajectories of passenger υ are collected in $Tra_\upsilon = (id, Tra\_Loc_\upsilon), id = 1, 2, \cdots, m$, where id distinguishes the trajectories.

Our previous study [39] suggested that the trajectory model can be characterized with semantic information using a combination of time constraints, traffic environments, and social activities. We have not generated a bag of words, similar to Wordnet [40]. However, similar to work done by Zheng *et al.* [41] and Yan *et al.* [42], we focus on the semantic information of a location, such as properties of the location service-visiting behaviors, generate the semantic features, and calculate the semantic similarity between trajectories using these features. Unlike Liu and Wang's work [20], which also extracted the semantic features from the trajectories to measure the stationary distribution of the object's residency probability on different semantic sites, we consider service-visiting behavior as a manifestation of social activities. Service-visiting behavior can tell us exactly what a passenger prefers and be recognized by associating the trajectory histories with the major POSs (points of service) (e.g., types of service in a mall).

TABLE I
EXAMPLE OF SEMANTIC TRAJECTORIES FROM A PASSENGER

	S_0	S_1	S_2	S_3
1st day	"Accommodation"	"FS"	"Medical"	
2nd day	"SAE"	"FS"	"Restaurant"	"Medical"
3rd day	"Accommodation"	"Medical"	"Restaurant"	"SAE"
4th day	"FS"	"Accommodation"	"Restaurant"	"Medical"

"SAE"="Science and education", "FS"="famous spots".

Definition 3 (Semantic Trajectory): A semantic trajectory is defined as a sequence of trajectories $Tra\_ST_\upsilon = S_i$, $(1 \le i \le h)$, with a type mapping function $\tau : L \to S$, where each $S \in Tra\_ST_\upsilon$ is one semantic type associated with L_i.

For example,$\tau : L_1 \to$ "$Science\ and\ education''$ implies that the semantic type of the location is science and education. Similarly, $S\_Tra_\upsilon = (id, Tra\_ST_\upsilon), id = 1, 2, \cdots, m$ is a set of historical trajectories with semantic information of passenger υ. Table I displays the semantic trajectories of a passenger over four days.

2) Prediction Using Trajectory Dataset: We employed the PPM model [43] [44] for predicting the intended destination.

Definition 4 (Contextual Sequence): For each $L_i = loc_{n+1}$ in Tra_Loc_υ, a contextual sequence, loc_n^N, is a sequence of length N where loc_n is the destination, formalized as $loc_n^N = loc_{n-(N-1)}, \ldots, loc_{n-1}, loc_n$.

Thus, by looking for the segments of trajectories with the same loc_n^N in Tra_Loc_υ, we can predict the probability distribution of L_i from the number of trajectories observed based on the PPM model. We calculate the probability p_i of passenger's next destinations based on the contextual sequence. $f(x_i|loc_n^N)$ represents the number of loc_n^N where x_i is the destination, $S(loc_n^N)$ denotes the total number of loc_n^N of different destinations, and $A(loc_n^N)$ describes the set of destinations whose the contextual sequences are same with loc_n^N.

$$p_i = p(L_i|loc_n^N) = \frac{f(L_i|loc_n^N)}{S(loc_n^N)},$$
$$S(loc_n^N) = \sum\nolimits_{i=1}^{|A(loc_n^N)|} f(x_i|loc_n^N), x_i \in A(loc_n^N), \quad (5)$$

Equation(5) shows that, given an L_i in L, if there are a sequence of trajectories being the same as loc_n^N, the probability p_i that indicates L_i would be a destination is determined and returned by $p(L_i|loc_n^N)$. Otherwise, it begins to decrease N by 1, updates the contextual sequence to be loc_n^{N-1} for predicting using $p(esc|loc_n^N)$. The "esc" in Eq.(6) indicates the escape code that is provided in the PPM model to terminate the procedure. $p(esc|loc_n^N)$ indicates the prediction probability of the escape code of loc_n^N, $f(esc|loc_n^N)$ denotes the frequency of escape code of loc_n^N, and $f(esc|loc_n^N) = |A(loc_n^N)|$.

$$p(esc|loc_n^N) = \frac{f(esc|loc_n^N)}{S(loc_n^N) + f(esc|loc_n^N)}, \quad (6)$$

We then decrease $f(esc|loc_n^N)$ by 1, identify the number of loc_n^{N-1} in Tra_Loc_υ and search the sequence of trajectories being the same as loc_n^{N-1} in Tra_Loc_υ. If there is no such sequence, we continue to decrease the frequency of the

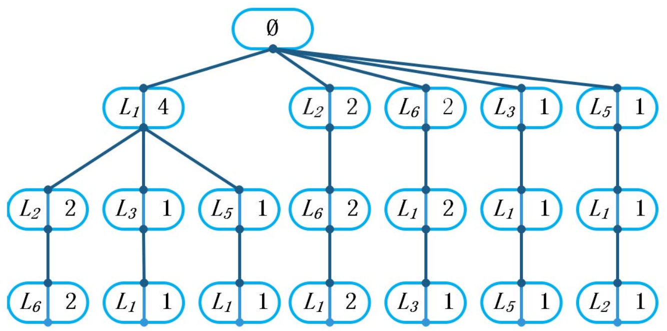


Fig. 4. Illustration of PPM model, where each location in Tra_Loc_{υ} and its two followers form a branch (i.e., $N = 3$). Each node includes the location and associated frequency.

escape code until finding the contextual sequence loc_n^{N-r} in the $(r+1)^{th}$ round. The probability is then calculated as:

$$p_i = \left[\prod_{j=k+1}^{N} p(esc|loc_n^j)\right] \cdot p(L_i|loc_n^k), \quad k = N-r,$$

$$p(L_i|loc_n^k) = \frac{f(L_i|loc_n^k)}{S(loc_n^k) + f(esc|loc_n^k)}, \tag{7}$$

If no trajectory is found while the frequency of the escape code equals 1, we assign the p_i to be 0. Thus, we can determine the predicted destination with the maximum probability and predict the destinations of a ridesharing group by traversing from loc_n^N to loc_n^1 in each member's historical trajectories.

Example: Given $Tra\_Loc_{\upsilon} = L_1, L_2, L_6, L_1, L_3, L_1, L_5, L_1, L_2, L_6, L_1$, we can structure the PPM model based upon a tree. Fig.4 illustrates that there are three length-1 contextual sequences of L_1's, that is, $\langle L_3 \rangle, \langle L_5 \rangle$ and $\langle L_6 \rangle$, where the number of L_6 is 2 in Tra_Loc_u. The probability for the destination L_3 is calculated when its length-2 contextual sequence is $\langle L_6, L_1 \rangle$. Since there are no contextual sequences of L_2, PP calculates the probability of escape code and obtains the prediction probability.

$$p_3 = \frac{f(L_3|L_6L_1)}{S(L_6L_1) + f(esc|L_6L_1)} = \frac{1}{1+1} = \frac{1}{2},$$
$$p_2 = p(esc|L_6L_1) \times p(L_2|L_1) = \frac{1}{1+1} \times \frac{3}{4+3} = \frac{3}{14}$$

3) Mining Frequent Movements: We train the PPM model to calculate the probability of passengers' historical destinations based on departure longitude, latitude and departure time. It then provides a list of the predicted destinations ranked by the probability. To further optimize the prediction efficiency, we consider the semantic information of the trajectories, extract the frequent movement patterns, and analyze the possible service-visiting behaviors at a certain location. The association rules indicating the relations between different frequent movement patterns are used to improve the accuracy of prediction.

Given a trajectory Tra_ST_{υ}, Q^k is the sequence of K terms, S_j, that is $Q^k = S_1, \cdots, S_j, \cdots, S_K$. Given the number, q, of trajectories including Q^k in S_Tra_{υ}, the support rate refers to the ratio of the number of trajectories including Q^k to the total number of trajectories (assumed to be m) in S_Tra_u; that is, $sup(Q^k) = \frac{q}{m}$. When $sup(Q^k) > minsup$, Q^k is considered to be a frequent movement pattern. $minsup$ is a predefined minimal support rate. In this paper, we employ the PrefixSpan algorithm [45] to extract the frequent movement patterns form S_Tra_u.

TABLE II

EXAMPLE OF FREQUENT MOVEMENT PATTERNS WHERE $minsup$=0.4

Frequent movement pattern	support rate
$<$ "Accommodation" $>$	3/4=0.75
$<$ "FS" $>$	3/4=0.75
$<$ "Medical" $>$	4/4=1
$<$ "SAE" $>$	2/4=0.5
$<$ "Restaurant" $>$	3/4=0.75
$<$ "Accommodation","Medical" $>$	3/4=0.75
$<$ "Accommodation","Restaurant" $>$	2/4=0.5
$<$ "FS","Medical" $>$	3/4=0.75
$<$ "FS""Restaurant" $>$	2/4=0.5
$<$ "Restaurant","Medical" $>$	2/4=0.5
$<$ "FS","Restaurant","Medical" $>$	2/4=0.5

Example: By extracting from the semantic trajectories given in Table I, we construct the frequent movement patterns and associated support rate. In Table II, $Q^2 = S_1, S_2$ is the generated frequent movement pattern using the PrefixSpan algorithm, where S_1 and S_2 denotes the semantic types of "Accommodation," and "Medical", respectively. There are three movement patterns including Q^2 in S_Tra_u; thus, the support rate of Q^2 is 0.75, based on four sequences in total.

Definition 5 (Movement Rule): Given a frequent movement pattern $Q^k = S_j, (1 \leq j \leq K)$, the movement rule R_{Q^K} is defined as $\langle S_1, \cdots, S_j, \cdots, S_{K-1} \rangle \Rightarrow \langle S_K \rangle$, where $\langle S_1, \cdots, S_j, \cdots, S_{K-1} \rangle$ and $\langle S_K \rangle$ are the antecedent and consequent of R_{Q^K}, respectively.

Therefore, the confidence of the movement rule indicates the possibility of its consequent according to its antecedent, expressed as follows, where $sup(S_1, \cdots, S_K)$ is the support rate of given sequence $(S_1, \cdots, S_K)$.

$$con_{S_K} = con(R_{Q^k}) = \frac{sup(S_1, \cdots, S_K)}{sup(S_1, \cdots, S_{K-1})}, \tag{8}$$

Therefore, by detecting the frequent movement patterns from a passenger's historical semantic trajectories, it is possible to predict the semantic type of destination.

B. P-PPM Destination Prediction

We propose a P-PPM algorithm that employs the PPM model for destination prediction, extracts the frequent movement patterns to define several rules, and determines the confidence of rule for improving the prediction accuracy. Fig. 5 illustrates the destination prediction.

Algorithm 2 gives the process of the P-PPM algorithm.

Given the Tra_Loc_{υ}, with respect to a passenger υ, there is a location $loc_{n+1} = L_i$, a length-N contextual sequence loc_n^N of loc_{n+1}, and the semantic trajectory model $Tra\_ST_n^N = S_{n-(N-1)}, \cdots, S_n$.

We first employ the PrefixSpan algorithm to extract the set of frequent movement patterns, FP, satisfying the constraints

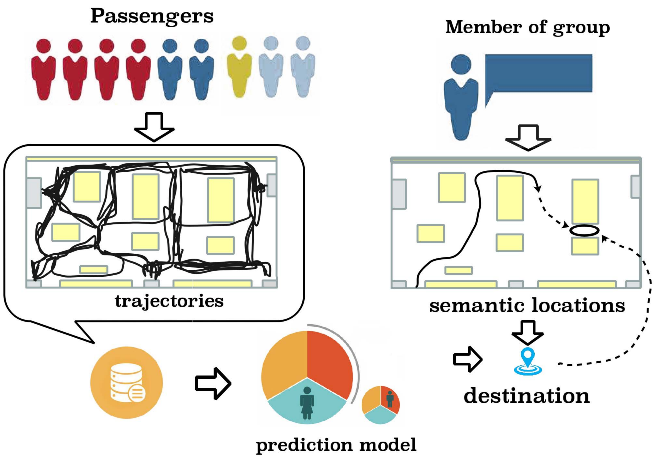


Fig. 5. Destination prediction for a ridesharing group, where the sequences of each passenger's historical travel destinations are taken as the candidate set.

on $minsup$ in $Tra\_ST_n^N$ by deleting the patterns with a length of 1. $FP = Q^k(k > 1)$ (lines 1 in Algorithm 2). For each pattern in FP, a set of movement rules, R, is then identified, where $R = R_{Q^K}$, and the confidence of each rule is determined. Next, we search all movement rules that regard the items of $Tra\_ST_n^N$ as their antecedent in R. If nothing is found, $Tra\_ST_n^{N-1}$ is used as the antecedent of the rule; then, the search is executed repeatedly until all the rules have been compared with the antecedent selected. A set R^* of the rules matching successfully is also generated (lines 7-13). We subsequently build a destination-prediction model to predict the probability p_i of each candidate L_i in Tra_Loc_u. Finally, we consider the S_i of each L_i as the consequence of a rule to compare all rules in R^* and compute the rule confidences of all candidates to obtain the confidence set $Con^* = cons_{S_i}(1 < i \leq h)$. If the match does not work, the confidence returns zero.

Through the confidence, the probability p_i of each candidate L_i can be refined as follows:

$$p'_i = \max_{1<i\leq h} (\omega \cdot p_i + (1-\omega) \cdot cons_{S_i}) \tag{9}$$

$\omega \in [0, 1]$ is the trade-off factor for PPM-based prediction and correction components. When $\omega = 1$, only a PPM-based prediction is used; when $\omega = 0$, only the frequent movement patterns are used.

V. Experiment

In this section, we evaluate the proposed group recommendation and destination prediction model. For the destination prediction model, we compare our proposed method with PPM and Markov methods using the Geolife dataset [46] for Beijing, China. The dataset was collected in (Microsoft Research Asia) by 182 users over three years (from April 2007 to August 2012). A GPS trajectory of this dataset was represented by a sequence of time-stamped points, each of which contained the information of latitude, longitude, and altitude. The dataset includes 17,621 trajectories with a total distance of about 1.2 million kilometers and an entire duration of

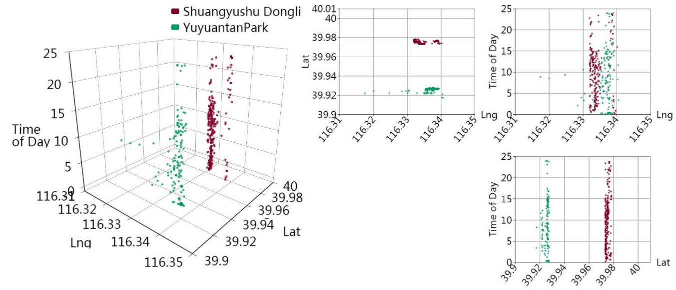


Fig. 6. Departure time, longitude and latitude distributions Shuangyushu Dongli and Yuyuantan Park (places in Beijing).

48,000+ hours. Furthermore, the set of points of interest (POI) covering 60% of Beijing was also used to identify the types of each location and define the semantic trajectory model. To solve data sparsity problem [47] and improve the performance of prediction, we first divided the area of the city center into 50 × 50 cells, each with a side of 500 m, and obtain 2,500 grid centroids. We referred to Han et al.'s work [48] and calibrated the raw trajectories by employing the grid centroids as the anchor points. We skimmed over the technical part of trajectory calibration since it is not the focus of this paper. Finally, we extracted 3891 stay points and obtained the dataset consisting of 528 semantic trajectories related to the above points. For the group recommendation model, we compare our model with two other (PPM- and Markov-based) methods using multiple evaluation criteria.

Fig. 6 shows the three-dimensional distributions of the departure time, longitude and latitude(lev_{time}, lon, lat) for the two different destinations Shuangyushu Dongli and Yuyuantan Park of this user. We can see that the three-dimensional distributions can easily distinguish the two different destinations.

Two sets of experiments were conducted for validate destination-predicting algorithm performance and evaluating the effectiveness of the group recommendation. For drivers' benefits, parameter H was set to 2 for rides with three people. The group recommendation system affects many core metrics. Among these metrics, we chose several important ones [49], which are listed in Table IV.

A. Destination Prediction Experiments

1) Parameter Study: We randomly split the Geolife dataset into training and test datasets using a ratio 7:3, i.e., 70% of the data were used for training and the remaining 30% for testing. It has been suggested that the best prediction performance can be achieved when the order N of the Markov model is set to 3 [50]. When the order was increased, more considerable computing resources are required, leading to poor prediction accuracy. Owing to the PPM model being developed from the Markov model, the order of N was set to 3 for predictions using the P-PPM model.

We study the hyper-parameters ω, which is the trade-off term for combing PPM-based prediction and correction components. The result is shown in Fig.7. The prediction accuracy decreased when the value of parameter ω increased from 0.2 to 0.55. The accuracy may not vary if we further increase ω.

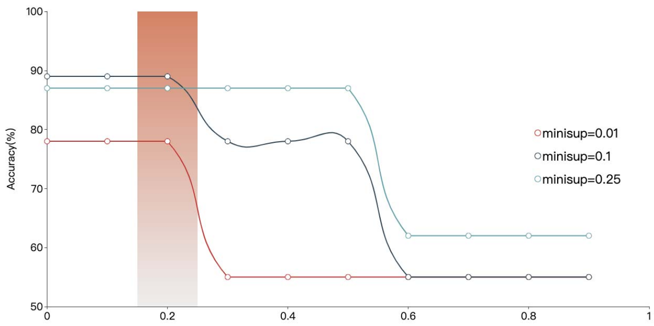


Fig. 7. Parameter study of P-PPM when varying ω.

Based on the curves, setting ω to 0.2 was reasonable because both objectives were combined most appropriately.

2) Performance Analysis: We compare our proposed model with the following three baseline methods.

- Traditional Markov Model. The traditional Markov Model obtains the transition matrix. The location that has the largest cumulative transition probability from the others is regarded as the predicted destination.
- PPM model. PPM identifies the destination by calculating the number of contextual sequences from the history trajectories.
- Sub-Trajectory Synthesis (SubSyn) [47] method. SubSyn aims to address the data sparsity problem. It decomposes the history trajectories into sub trajectories and combines them into different new trajectories to improve the coverage of trajectories.

We compared three algorithms in terms of prediction accuracy in Fig. 8. The figure shows that the PPM and P-PPM algorithms had the highest accuracy when N was 3, while the accuracy decreases as N increased. Thus, an increase in the lengths of contextual sequences led to complexity in searching the trajectories with the same lengths. Due to the data sparsity problem, it may not find such trajectories for matching, which resulted in a declining accuracy. This also illustrated that early historical trajectory has little effect on the destination prediction, as proposed in [50]. The accuracy of the P-PPM model can reach 80%, while those of the Markov and PPM models were only approximately 46% and 66%, respectively. The P-PPM algorithm had evident advantages in accuracy. The improvement results illustrated that allowing an analysis of a passenger's interests in the movement patterns in a spatial layout could lead to the ability to fully predict the destination to which the passenger wants to go.

The SubSyn(g) that is chosen for the comparison was limited to the order-3 P-PPM model, denoted as PPM(3). Table III illustrates the effectiveness of our model by returning the top-k reference destination based on the sorting of their probabilities. SubSyn(40) and SubSyn(50) indicate the method using grids with granularity value $\boldsymbol{g} = \mathbf{40}$ and **50**, respectively. This shows that when k is low, e.g., if k is 1, both hit rates of these methods were less than 55%. As k was increased, the hit rates also increased gradually. This also proved that our model can achieve a higher hit rate than the others. The bold number shows the optimal value among the results of these algorithms.

TABLE III
HIT RATE

	Hit_1	Hit_2	Hit_3
P-PPM(2)	33.8	**69**	**84.5**
SubSyn(40)	46.6	55.7	60
SubSyn(50)	**51.4**	58.6	62.8

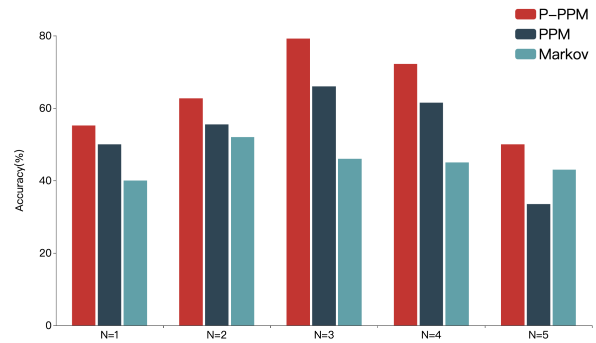


Fig. 8. Comparisons of prediction accuracy at different values of N.

We examine the offine learning procedure, where the prediction horizons were in the range of 3 d to 1 month when $\omega = 0.2$(see Table V). This demonstrated that when more trajectories were observed, the accuracies of the destination prediction tasks increased as well. The increment rate was the largest when horizon changes from 15 d to 1 month. Since the data was relatively sparse in 15 d, the performance decreased.

B. Group Recommendation System Experiments

1) Experimental Setting: We selected data from 62 trajectories in the Geolife dataset from October 2008 to July 2009. The trajectories sampled from October 2008 to April 2009 were taken as the training dataset, and those sampled from May 2009 to July 2009 were used as the test dataset. We first trained a destination prediction model and investigated the solution quality of the three algorithms. The parameter V was set to 1.2 m/s [14], the search radius and time span for waiting were set to 0.1 km and 15 min [25], respectively.

2) Experimental Results and Analysis: We have manually labeled the number of people sharing a ride with all passengers in the testing dataset based on the temporal duration $\theta_t = 30min$. If the result was greater than, or equal to 1, it was considered to be positive samples (i.e., each passenger-pair can be classified as being a group or being not a group), while the remaining ones were negative.

We first used the reduction rate of total travel time to measure the effectiveness of the aggregation level. Referring to a previous report [9], we used the reduction rate of the total trip in terms of distance to measure the effectiveness with various settings of time span and search radius, as shown in Fig. 9. This illustrated the changes in the reduction rate if the users share rides. This enabled us to understand the level of temporal and spatial aggregation that should be adopted to achieve the most active group ride strategy. The values of the reduction rate changed with increases in the radius and time, i.e., more additional passengers could be incorporated into some groups with a broader search region and time span. For example, there

TABLE IV
PERFORMANCE INDICES

Model	Indicator	Description
	RT	Number of trajectories in the testing set
Destination Prediction	CT	Number of trajectories whose destinations were correctly predicted
	$\frac{CT}{RT}$	Accuracy or Acc.
	CT@k	Number of accurate predictions of top-k predicted destinations
	$Hit_k = \frac{CT@k}{RT}$	Hit rate of destination prediction with the condition of top-k
	$\frac{\sum_{m=H.N+1}^{M} \|\|v_m.o, v_m.d\|\| - \sum_{n=0(d^0(u_n,G_{u_n})+d^*(u_n,G_{u_n})+\|\|u_n.o,u_n.d\|\|)}^{N}}{\sum_{mH.N+1\|\|v_m.o,v_m.d\|\|}^{M}}$	Reduction rate
Destination Prediction	TP	Number of passenger pairs correctly classified as being a group
	FP	Number of passenger pairs mistakenly classified as being a group
	FN	Number of passenger pairs mistakenly classified as not being a group
	Precision	$TP/(TP + FP)$
	Recall	$TP/(TP + FN)$

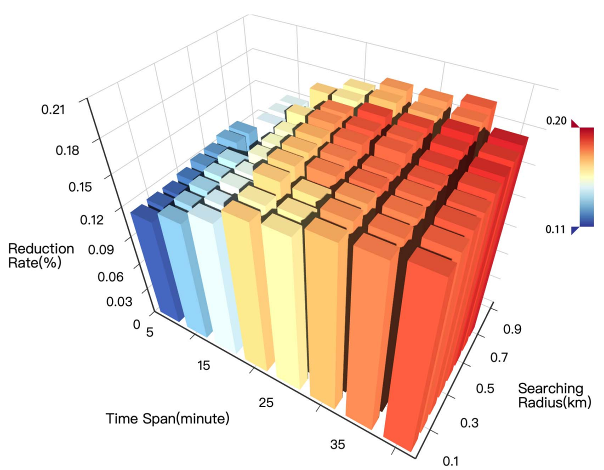


Fig. 9. Effectiveness under various settings of time span and search radius where rides with four-passengers. Different colors represent different reduction rate of the total trip.

was an improvement of 7% in the reduction rate when the radius was increased from 0.3 to 0.4 km and time span was increased from 5 to 10 min. Additional improvements in the reduction rate were not more than 0.3% with a higher level of aggregation. The reduction rate could be increased if we further increased the level of aggregation. The results also indicate that a ridesharing group can help to save over 11.5% of total travel miles where rides with four-passengers (i.e., $H = 4$) if it was applied in the real world with a search radius of 100 m and time span of 5 min, and the group recommendation scheme may be sufficient.

Fig. 10(a) shows the precision of group recommendations based on the P-PPM, Markov, and PPM models for the order N of the prediction model. The results showed that our model was significantly better than the other two for different values of N. A trend was evident for the change of precision of all of the models with respect to increments in N. When N was set to 2, the precision reached the highest accuracy. After,

TABLE V
THE PREDICTION ACCURACY OF DIFFERENT HORIZONS

prediction horizons (days)	3	7	15	30
N	Acc.	Acc.	Acc.	Acc.
1	0.38	0.51	0.46	0.58
2	0.36	0.4	0.33	0.42
3	0.33	0.36	0.31	0.4

with increasing N, the prediction decreased. However, this indicated that the precision computed by the P-PPM was on average about 10% higher than those of the other two methods.

We next studied the recall of the algorithm based on P-PPM with the other two algorithms in Fig. 10(b). The recall of the P-PPM model for destination prediction was better. The higher the recall was, the greater the number of potential ridesharing passengers that were identified correctly was and the fewer cases there were in which the potential ridesharing passengers could not be detected. Additionally, too large of a value of N could lead to a decrease in the precision and recall of the three models, resulting in a reduction in the effect of discovering potential passengers for ridesharing.

We also discuss how the new groups for ridesharing differed when selecting a single origin versus trying more extended areas of influence. Fig.11(a) illustrates the regions of influence on the grouping ($H = 2$) based on two above setting. When the regions were extended, the number of groups increased significantly since the areas of further away passengers would be intersected into the group. However, the areas of influence were too short with values of 0, that is, if a single origin was selected for grouping, the passengers might be more difficult to group. When the region was beyond a specific limit (i.e., 800 m in Fig.11), the value stops changing and stays steady since no more new candidates could be detected. We also compare the influence of region for group recommendation, under different hyperparameter H. Similarly, there was an increased tendency for precision, and an oscillation in a steady

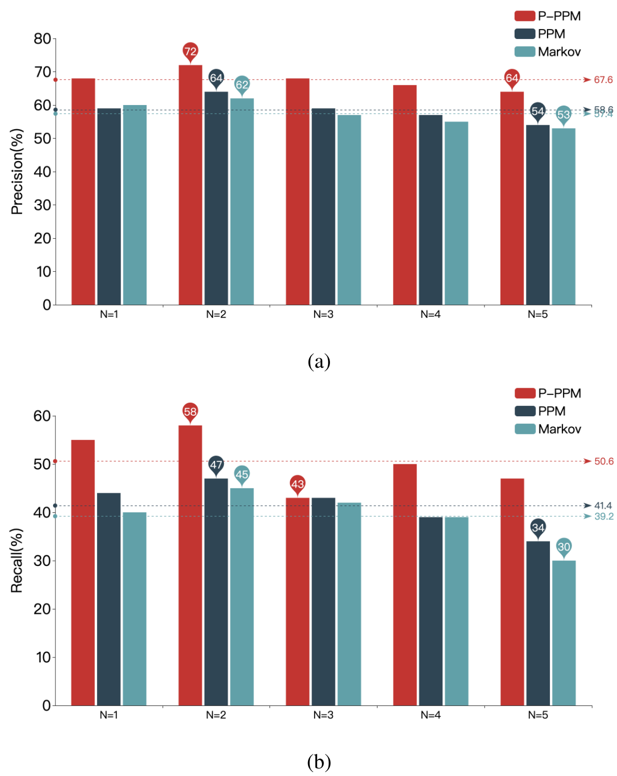


Fig. 10. Comparison of group recommendation.

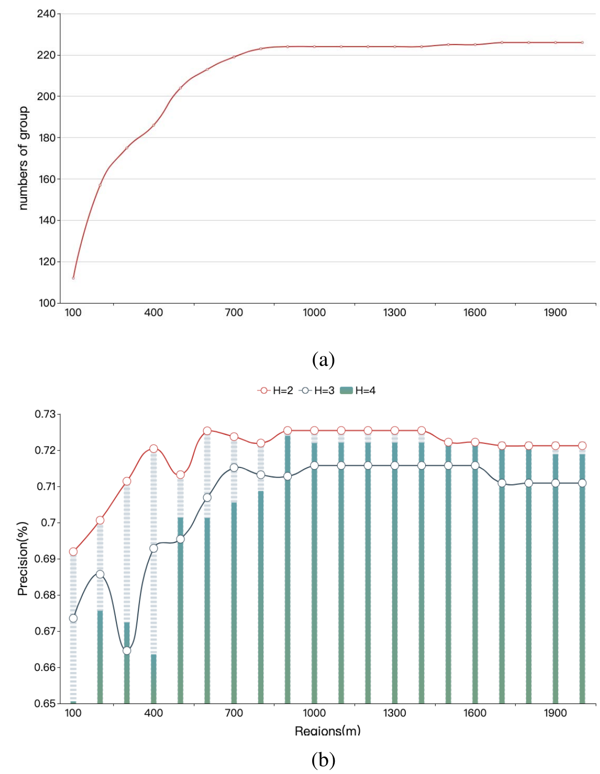


Fig. 11. The influence of regions for grouping.

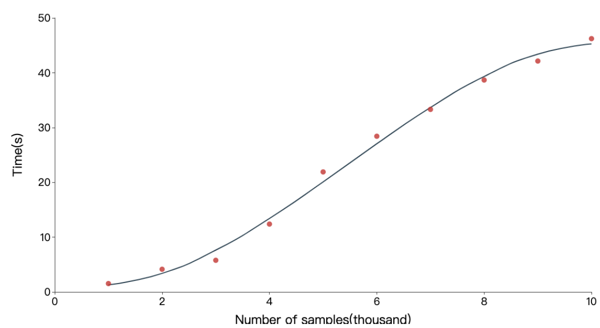


Fig. 12. Times of recommendation versus the number of samples used.

value follows. The results also indicate that rides with two-passengers may contribute the most to group recommendation.

To demonstrate the timeliness of grouping riders, which would be critical to the performance of any mobility system, we have demonstrated the efficiency of our method in Fig.12. All our experiments were conducted in a desktop with eight core i7-6700 CPU and 16G memory. The experiments with the group recommendation could be completed in about 1 min, demonstrating that our method may be welcome for an online system due to its simplicity and higher prediction efficiency.

VI. Discussion

A. Group Detection for Slugging

Joining a group and reducing unwanted operations on system leads to a better arrangement for slugging. This issue is of significance for policymakers to understand people's opinions about slugging and improve the experience to motivate people to continue to use it. Moreover, grouping multiple passengers into a vehicle results in a great reduction in the number of vehicles, vehicle miles traveled, vehicle hours traveled, and negative externalities associated with car travel [51], such as emissions and congestion. Thus, with the high levels of group recommendation adoption, slugging services would have a noticeable impact on urban traffic conditions. Compared to the social recommendation scenarios [52], group detection for slugging scenario is different from several perspectives. Different passengers make requests dynamically over time at various locations, usually resulting in a more widely distributed departure location and time. Furthermore, different drivers pass an area with different speeds, which increases the difficulties in determining the meeting time, and further grouping more passengers. These observations indicate that spatial information should be combined with other cues (such as velocity information and individual preference) to produce a more robust similarity measurement. The weighted combination of various measures is usually leveraged for similarity search. However, how to tune this weight is a challenge because it would lead to limited flexibility in processing real data. Although the online survey on potential target users enables us to assign the weights to different measurements. Such a study still must to incorporate a larger number of features into their user interfaces, at a higher cost for implementation time and code.

B. Automatically Predicting Destination

Although there is a lack of information about intended destinations, we can exploit the knowledge that was been learned from the training data in a batch learning fashion. We take

TABLE VI
SYMBOLS

Symble	Description
u, v_i	driver, the i-th passenger
$u.t, v_i.t$	departure time of driver and a passenger
m_{ij}	indicator for whether v_i, v_j can be grouped
M_u	$M_u = (m_{ij})_{I\times I}$, slugging recommendation among all I passengers
H	number of passengers for delivery in a vehicle during u's trip
G_u	a ridesharing group of u
$d^o(u, G_u), d^*(u, G_u)$	extra walking distance of group from origin/drop-off to pick-up/destination of each passenger
$TW(u, G_u)$	additional waiting time for departure as a group
h_i	Indicator for whether v_is will arrive the pickup location before u depart
$TT(u, G_u)$	total travel time
Tra_Loc_u, Tra_ST_u	trajectory model and semantic trajectory model of u
Tra_u, S_Tra_u	set of trajectories and set of semantic trajectories of u
loc_i	information of the location including coordinate, arrival time and departure time.
loc_n^N	contextual sequence of loc_{n+1}, and the length is N
$minsup$	minimal support rate
Q^K	a frequent movement pattern, and $sup(Q^k) > minsup$
R_{Q^K}	a movement rule associated with Q^k
Con_{S_i}	confidence of R_{Q^K}
FP	set of frequent movement patterns, $FP = Q^k$
R	a set of movement rule for each pattern in FP and associated confidence, $R = R_{Q^K}$
p_i, p_i'	prediction possibility and refined possibility
R^*	set of rules used for matching successfully

the sequences of each passenger's historical travel destinations as the candidate set for intended destination prediction. The system can capture the real-time information (from the forecast or manual destination entry) once the ridesharing-booking app is initiated and issue an online recommendation. It enables passengers to change their requests and matches the drivers and potential co-passengers.

We believe the model can be further improved by utilizing online learning techniques. Currently and in the near future, new types of data (e.g., social media, landmarks, or suggestions from other people) can be introduced to improve an existing destination prediction system. In this scenario, training data with new features will be added to the prediction system, while old features are still retained [53]. Furthermore, different passengers may have different opinions on the definition of destination, and thus, a universal model might not always be optimal for every group. Online learning in group recommendation can be potentially more challenging than the existing work [21], [51], as the model should be trained on the new data arriving sequentially and transfer useful knowledge from the universal model to personalize the recommendation for every individual in an online learning manner.

VII. Conclusion

In this paper, we proposed a novel group recommendation system for OD-slugging. The proposed method combines additional semantic information with raw trajectory data and constructs a prediction model based on passengers' historical data. As a result, an optimal group of passengers whose movement similarities are a manifestation of shared relationship is detected to minimize the total travel time. Experimental results from the Geolife dataset show that the P-PPM model exhibited better prediction accuracy than the other three models. Moreover, our proposed method is the best in terms of saved travel time, which is the most important metric in slugging, and a ridesharing group may help to save over 19% of total travel miles if applied in the real world.

In future work, we plan to investigate the following interesting problems further:

1) Other scenarios of ridesharing, for example, detour ridesharing, must satisfy the request of the pick-up and/or drop-off locations and the departure and arrival time of the driver. In such scenarios, searching for riders with similar routes will be beneficial, and we may extend our method for the more general scenarios.

2) The trajectory data from GPS–enabled devices are widely used in our group recommendation system. We plan to incorporate socially aware information further to discover the latent relationships among group members, meeting the personalized needs of different passengers in ridesharing.

Appendix

See Table VI.

References

[1] S. Shaheen *et al.*, "Shared mobility: Current practices and guiding principles," Federal Highway Admin., Washington, DC, USA, Tech. Rep., 2016.

[2] N. D. Chan and S. A. Shaheen, "Ridesharing in North America: Past, present, and future," *Transp. Rev.*, vol. 32, no. 1, pp. 93–112, 2012.

[3] D. Zhang, T. He, Y. Liu, S. Lin, and J. A. Stankovic, "A carpooling recommendation system for taxicab services," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 254–266, Sep. 2014.

[4] G. Auffret *et al.*, "Multimedia access and retrieval (panel session): The state of the art and future directions," in *Proc. 7th ACM Int. Conf. Multimedia*, 1999.

[5] M. Burris *et al.*, "Casual carpooling scan report," Office Transp. Manage., Federal Highway Admin., Washington, DC, USA, Tech. Rep. FHWA-HRT-12-053, 2012.

[6] D. E. LeBlanc, *Slugging: The Commuting Alternative*. Washington, DC, USA: Forel Publishing, 1999.

[7] D. E. LeBlanc. (2018). Information About Slugging and Slug Lines. Forel Publishing Company, LLC. Woodbridge, VA, USA. [Online]. Available: http://www.slug-lines.com/Slugging/Map.asp

[8] G. Laporte, F. Meunier, and R. W. Calvo, "Shared mobility systems," *4OR*, vol. 13, no. 4, pp. 341–360, 2015.

[9] X. Qian, W. Zhang, S. V. Ukkusuri, and C. Yang, "Optimal assignment and incentive design in the taxi group ride problem," *Transp. Res. B, Methodol.*, vol. 103, pp. 208–226, Sep. 2017.

[10] Y. Li, R. Chen, L. Chen, and J. Xu, "Towards social-aware ridesharing group query services," *IEEE Trans. Services Comput.*, vol. 10, no. 4, pp. 646–659, Jul./Aug. 2017.

[11] H. Zhang and J. Zhao, "Mobility sharing as a preference matching problem," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2584–2592, Jul. 2019.

[12] H. Dong, L. Ma, and J. Broach, "Promoting sustainable travel modes for commute tours: A comparison of the effects of home and work locations and employer-provided incentives," *Int. J. Sustain. Transp.*, vol. 10, no. 6, pp. 485–494, 2016.

[13] Y. Artan, O. Bulan, R. P. Loce, and P. Paul, "Passenger compartment violation detection in HOV/HOT lanes," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 395–405, Feb. 2016.

[14] M. Berlingerio, B. Ghaddar, R. Guidotti, A. Pascale, and A. Sassi, "The GRAAL of carpooling: GReen And sociAL optimization from crowd-sourced data," *Transp. Res. C, Emerg. Technol.*, vol. 80, pp. 20–36, Jul. 2017.

[15] A. Khan, O. Correa, E. Tanin, L. Kulik, and K. Ramamohanarao, "Ride-sharing is About Agreeing on a Destination," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2017, p. 6.

[16] S. Ma and O. Wolfson, "Analysis and evaluation of the slugging form of ridesharing," in *Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.* New York, NY, USA: ACM, 2013, pp. 64–73.

[17] T. Hilgert, M. Kagerbauer, T. Schuster, and C. Becker, "Optimization of individual travel behavior through customized mobility services and their effects on travel demand and transportation systems," *Transp. Res. Procedia*, vol. 19, no. 3, pp. 58–69, 2016.

[18] R. Uddin, M. N. Rice, C. V. Ravishankar, and V. J. Tsotras, "Assembly queries: Planning and discovering assemblies of moving objects using partial information," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2017, p. 24.

[19] J. Guo, Y. Zhu, A. Li, Q. Wang, and W. Han, "A social influence approach for group user modeling in group recommendation systems," *IEEE Intell. Syst.*, vol. 31, no. 5, pp. 40–48, Sep./Oct. 2016.

[20] S. Liu and S. Wang, "Trajectory community discovery and recommendation by multi-source diffusion modeling," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 898–911, Apr. 2017.

[21] L. Zhang *et al.*, "A taxi order dispatch model based on combinatorial optimization," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 2151–2159.

[22] L. Tang, Z. Duan, and Y. Zhao, "Toward using social media to support ridesharing services: Challenges and opportunities," *Transp. Planning Technol.*, vol. 42, no. 4, pp. 355–379, 2019.

[23] W. Huang, Y. Zhang, Z. Shang, and J. X. Yu, "To meet or not to meet: Finding the shortest paths in road networks," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 4, pp. 772–785, Apr. 2018.

[24] X. Li, M. Li, Y.-J. Gong, X.-L. Zhang, and J. Yin, "T-DesP: Destination prediction based on big trajectory data," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2344–2354, Aug. 2016.

[25] S. A. Shaheen, N. D. Chan, and T. Gaynor, "Casual carpooling in the San Francisco Bay Area: Understanding user characteristics, behaviors, and motivations," *Transp. Policy*, vol. 51, pp. 165–173, Oct. 2016.

[26] A. Elbery, M. Elnainay, F. Chen, C.-T. Lu, and J. Kendall, "A carpooling recommendation system based on social VANET and geo-social data," in *Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2013, pp. 556–559.

[27] Y. Wang, R. Kutadinata, and S. Winter, "The evolutionary interaction between taxi-sharing behaviours and social networks," *Transp. Res. A, Policy Pract.*, vol. 119, pp. 170–180, Jan. 2019.

[28] G. H. de Almeida Correia, J. de Abreu e Silva, and J. M. Viegas, "Using latent attitudinal variables estimated through a structural equations model for understanding carpooling propensity," *Transp. Planning Technol.*, vol. 36, no. 6, pp. 499–519, 2013.

[29] Y. Wang, S. Winter, and N. Ronald, "How much is trust: The cost and benefit of ridesharing with friends," *Comput., Environ. Urban Syst.*, vol. 65, pp. 103–112, Sep. 2017.

[30] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, and S. Koenig, "Ridesharing: The state-of-the-art and future directions," *Transp. Res. B, Methodol.*, vol. 57, pp. 28–46, Nov. 2013.

[31] F. Bakkal, S. Eken, N. S. Savaç, and A. Sayar, "Modeling and querying trajectories using Neo4j spatial and TimeTree for carpool matching," in *Proc. IEEE Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Jul. 2017, pp. 219–222.

[32] M. Rigby, S. Winter, and A. Krüger, "A continuous representation of ad hoc ridesharing potential," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2832–2842, Oct. 2016.

[33] A. Nadembega, A. Hafid, and T. Taleb, "A destination and mobility path prediction scheme for mobile networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2577–2590, Jun. 2015.

[34] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 157–165, Dec. 2011.

[35] D. Lian and X. Xie, "Mining check-in history for personalized location naming," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, p. 32, 2014.

[36] B. Barabino, S. Salis, and A. Assorgia, "Application of mobility management: A Web structure for the optimisation of the mobility of working staff of big companies," *IET Intell. Transp. Syst.*, vol. 6, no. 1, pp. 87–95, 2012.

[37] M. Dogru, "Car pooling with GIS map server and Web services," Ph.D. dissertation, Dept. Inf. Technol., Univ. Zurich, Zürich, Switzerland, Aug. 2004, pp. 1–83.

[38] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transp. Sci.*, vol. 11, no. 3, pp. 253–274, Aug. 1977.

[39] Z. Duan, L. Tang, X. Gong, and Y. Zhu, "Personalized service recommendations for travel using trajectory pattern discovery," *Int. J. Distrib. Sensor Netw.*, 14(3):1550147718767845, 2018.

[40] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[41] B. Zheng, N. J. Yuan, K. Zheng, X. Xie, S. Sadiq, and X. Zhou, "Approximate keyword search in semantic trajectory database," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Sep. 2015, pp. 975–986.

[42] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semantic trajectories: Mobility data computation and annotation," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, p. 49, 2013.

[43] T.-Y. Wong, J. C.-S. Lui, and M. H. Wong, "Markov chain modelling of the probabilistic packet marking algorithm," *Int. J. Netw. Secur.*, vol. 5, no. 1, pp. 32–40, 2007.

[44] V. S. Tiwari, A. Arya, and S. Chaturvedi, "Scalable prediction by partial match (PPM) and its application to route prediction," *Appl. Informat.*, vol. 5, no. 1, p. 4, 2018.

[45] J. Pei *et al.*, "PrefixSpan,: Mining sequential patterns efficiently by prefix-projected pattern growth," in *Proc. 17th Int. Conf. Data Eng.*, Apr. 2001, pp. 215–224.

[46] Y. Zheng *et al.*, "Geolife GPS trajectory dataset-user guide," Microsoft Res., Beijing, China, Tech. Rep., 2011. [Online]. Available: https://www.microsoft.com/enus/research/publication/geolife-gps-trajectory-dataset-user-guide

[47] A. Y. Xue, J. Qi, X. Xie, R. Zhang, J. Huang, and Y. Li, "Solving the data sparsity problem in destination prediction," *VLDB J.*, vol. 24, no. 2, pp. 219–243, 2015.

[48] H. Su, K. Zheng, J. Huang, H. Wang, and X. Zhou, "Calibrating trajectory data for spatio-temporal similarity analysis," *VLDB J.*, vol. 24, no. 1, pp. 93–116, Feb. 2015.

[49] I. Avazpour, T. Pitakrat, L. Grunske, and J. Grundy, "Dimensions and metrics for evaluating recommendation systems," in *Recommendation Systems in Software Engineering*. Berlin, Germany: Springer, 2014, pp. 245–273.

[50] H. Abu-Ghazaleh and A. S. Alfa, "Application of mobility prediction in wireless networks using Markov Renewal Theory," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 788–802, Feb. 2010.

[51] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "Enhancing urban mobility: Integrating ride-sharing and public transit," *Comput. Oper. Res.*, vol. 90, pp. 12–21, Feb. 2018.

[52] M. Ye, X. Liu, and W.-C. Lee, "Exploring social influence for recommendation: A generative model approach," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2012, pp. 671–680.

[53] P. Zhao, S. C. H. Hoi, J. Wang, and B. Li, "Online transfer learning," *Artif. Intell.*, vol. 216, pp. 76–102, Nov. 2014.

Lei Tang (M'14) received the Ph.D. degree in computer science and technology in 2012.

She was a Visiting Researcher at the Chair of Information Systems, Mannheim University, Germany, from 2009 to 2010. She is currently working with the School of Information Engineering, Changan University, China. Her research interests include the area of intelligent transportation system and pervasive computing. She is a member of ACM and China Computer Federation (CCF). She received the Distinguished Young Scholars Award from the Universities of Shaanxi.

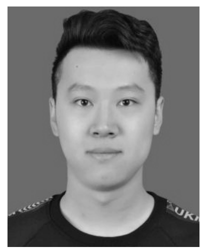
Zongtao Duan received the Ph.D. degree in computer science from Northwestern Polytechnical University, China, in 2006.

He is currently a Professor with the School of Information Engineering, Chang'an University, China. He was a Post-Doctoral Research Fellow with the University of North Carolina at Chapel Hill, USA, from 2009 to 2010. His research interest includes context-aware computing in transportation.

Dr. Duan is a member of CCF, CCF High Performance Computing, and Pervasive Computing Technical Committee.

Yishui Zhu (M'15) received the Ph.D. degree in human sciences from Waseda University, Saitama, Japan, in 2012. She is currently an Associate Professor with the Department of Software Engineering, School of Information Engineering, Chang'an University, Xi'an, China. She has been interested extensively in research works in the fields of service computing, network information systems, human informatics, and traffic services systems. She seeks to exploit the rich interdependence between theory and practice in her work with interdisciplinary and integrated approaches. Her recent research interests cover human–centric ubiquitous computing, traffic services, block-chain, behavior and cognitive informatics, big data, personal study behavior analytics, and individual modeling. She is a member of the IEEE CS, ACM USA, CCF and YOCSEF China.

Junchi Ma received the Ph.D. degree in computer science and engineering from Southeast University, China, in 2017. He is currently working with the School of Information Engineering, Chang'an University, China. His research interests include software reliability and intelligent transportation systems.



Zihang Liu is currently pursuing the degree with the School of Computer Science, Chang'an University, China. His research interests include data mining and network embedding. He is a Student Member of CCF.