CCF Transactions on Pervasive Computing and Interaction (2019) 1:240–249
https://doi.org/10.1007/s42486-019-00015-0

REGULAR PAPER

An efficient ride-sharing recommendation for maximizing acceptance on geo-social data

Lei Tang[1] · Meng Han[1] · Zongtao Duan[1] · Dandan Cai[1]

Received: 5 April 2019 / Accepted: 5 September 2019 / Published online: 24 September 2019

Abstract
Ride-sharing, which refers to assigning a set of riders for saving travel miles and alleviating traffic pressure, has drawn increasing attention. Existing works emphasize compatibility of potential riders on the basis of geographic proximity. They generally assume that no rejection would happen after the assignment is completed by the server. However, ignorance of psychological factors on ridesharing (e.g., trust on car mates) can lead to decrease rider acceptance. Thus, in this paper, we take the tendency of a rider to group with others into consideration and maximize riders' acceptance when sharing a trip. Specifically, we formally define the problem of maximizing riders' acceptance based on *people's* interests, social links, and employ social networking to facilitate finding a ridesharing group for the rider with the largest acceptance. We propose a new ride-sharing mode to recommend groups that travel together from geo-social data streams. To optimize the recommendation, we develop a heterogenous travel network, based on a proposed destination-prediction algorithm, to mine the similar spatial movements among a set of riders. Then, we measure the willingness of riders for joining in a group using social context. Finally, we progressively select the riders with high acceptance to be in the top-k results. We present the results of applying framework on real world social media data from the Twitter. Computational results show our method is able to significantly reduce the travel time when ridesharing, while keeping a high level of acceptance on real-world datasets.

Keywords Group recommendation · Ridesharing · Destination prediction · Social network · Acceptance

1 Introduction

Congestion, parking shortages, and frustration with existing dial-a-ride services are causing travelers to search for innovative technologies and services to address the mobility challenges. Ridesharing differs from traditional services in that requirements released by a rider arise dynamically over time at various locations, and drivers should physically travel to a meeting spot to perform the delivery task using their personal vehicles. With ridesharing, a group of riders can be recommended to a nearby driver and be shuttled to their destination. Instead, drivers in the vicinity of the riders would make a detour and make extra stops. For example, Tina and Tom want to travel to the airport, so they issue the ridesharing requests to the ride-matching server respectively, and then the server notifies Peter who happen to drive through the airport. Tina and Tom accept the recommendation, walk to the meeting spot(s) to form a ridesharing group. For this example, the ridesharing task has to be assigned to the people who are nearby and on the way to the airport. If Tom is a stranger (male) to Tina, the recommendation may get rejected as Tina dislikes sharing rides with whom she does not trust. And it is more proper to assign tasks considering the willingness of Tina. From our example, we can find the ridesharing recommendation is a critical issue in ridesharing.

To address these issues, we need information sources and methodologies for harvesting and evaluating a riders' interests. Most existing work adopt the routes specifying similar pick-up and drop-off locations and times (Furuhata et al. 2013) to form a ridesharing community. (Berlingerio et al. 2017a) claimed that it is possible to seek traveling buddies who have similar preferences on travelling instead of being from the same household or friends. They modelled *people's* interest in sharing a trip from a trajectory stream, to facilitate scalable and flexible companion discovery by maximizing the quality of ridesharing. However, existing work are based

✉ Lei Tang
tanglei24@chd.edu.cn

[1] School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China

on the assumption that no rejection would happen after the ridesharing recommendation and riders guarantee to share rides with others. In practice, a rejection would happen when the rider is no interested in the recommended ridesharing, which may result from extra trip time, long waiting time, distrusted car mates, and etc. In the typical application of ridesharing, DiDi, the riders are allowed to reject the recommendations only if the waiting time is too long, not the social discomfort when sharing a trip. The riders may be logged out of the application for low acceptance rate (Wang et al. 2019) and the system's matching success rate cannot be guaranteed.

Taking riders' rejections into concern, how to recommend the riders with a ridesharing group to maximize their acceptance is of great importance. In this paper, we formalize the problem of maximizing riders' acceptance in ridesharing recommendation. By incorporating the social interests for each rider-rider pair, this problem is to find potential partners to form a ridesharing group with drivers' delivery tasks. We propose a group search algorithm with a meta-path similarity strategy for mining spatial movement as well as a social network based algorithm for searching the trust links between each pair. We make the following contributions in this work.

- We formalize the problem of acceptance maximization in ridesharing. Then we introduce the framework of ridesharing recommendation is Sect. 3.
- We give the destination prediction method to learn the probability of interest for each rider-location pair. A heterogeneous travel network is then constructed for identifying the ridesharing relationship between riders and each driver. The latent semantic of relationship is represented and learned using a proposed meta-path.
- We propose a ridesharing group recommendation approaches in Sect. 4, including social data harvesting, interest modeling on topics and *top-1* search, to improve the ridesharing acceptance.

In addition to the contributions listed above, we further make some discussions of our framework in Sect. 5. We conducted experiments on real datasets to evaluate our approaches, and the results are shown in Sect. 6, and conclude the paper in Sect. 7.

2 Related work

The financial benefits of sharing trip-related expenses may motivate people to participate in ridesharing. People usually share time and space resource with other strangers in the same car simultaneously. Therefore, it is better for individuals to find a regular partner when sharing a ride (Elbery et al. 2013). Although de Abreu e Silva (2013) investigated the idea that people may feel more comfortable when riding with his/her friends or a regular partner, going from an acquaintanceship-based social group (a household or a company), this kind of trust conscious ridesharing is either too restricted or too relaxed to be practical.

Forming a ridesharing group in which each *rider's* trip is similar to that of the driver (Li et al. 2017) is necessary. Driver and all the riders must agree on the costs and schedules, including position and temporal elements. Furuhata et al. (2013) classified the demands of ridesharing participants according to what is the information used to form driver-rider matches. A detailed routing, OD-pairs, departure and arrival time are common for ridesharing participants. Most of the current methods are more focused on accurately matching such information predetermined by users. Bakkal et al. (2017) proposed a novel method for ridesharing group recommendations. The Neo4j-based spatial-temporal tree was established by using the trajectory data to extract the travel time and semantic type of a destination. Users with similar travel time and location were recommended to join a group. Rigby et al. considered the vehicle accessibility as a ridesharing service. In order to improve the accuracy of pickup, the proposed OppRide developed the network time prism with road network to represent the service. A group of riders can be then informed where should be boarded and dropped (Rigby et al. 2016b). SaRG (Li et al. 2015) was presented to group riders according to their social connections, e.g., check-in locations. In order to meet the social comfort and trust in ridesharing, the members' trip is similar to that of the driver, and is familiar with each other.

For existing ridesharing recommendation methods, riders must manually input their travel demands when they submit their requests and negotiate with driver(s) using additional communication channels (e.g., voice or text) to confirm an accurate pick-up location. This incurs extra work for the users to enter such information, especially the full name of the destination. Successful ridesharing requires to increase the interest of a person to participate in ridesharing, especially once the ridesharing-booking APP is started. The destination prediction is helpful to inform the users which route, and at which time, she or he should be boarded, without bringing attention to enter the full name of the destination. Thus, the acceptance can be improved if the intended destination can be accurately predicted when a user submits the request. Destination prediction mainly captures user preferences on movements and social interactions from trajectory data. Most existing approaches always predict destination according to existing trip based on history trajectories. They typically use Markov model to identify the transfer probability between two near neighbors and focus on the accuracy of provisioning (Nadembega et al. 2015). For example, T-DesP (destination prediction based on big trajectory data) model was proposed (Li et al. 2016) to predict the destination by a Markov model and solved the problem of data sparsity by using the content-based tensor-decomposition method. Association rules is also introduced for destination prediction, the idea of which is to detect the frequent spaces visited from a user's historical trip data. The rules w.r.t. movement patterns are then generated to conduct a match between the next location and frequent spaces (Karamshuk et al. 2011).

How to recommend a ridesharing group that rider wants to join in is still a problem. A recommendation, with consideration for riders' social interests, will be acceptable. Due to personal safety or social considerations (Agatz et al. 2012), we need to translate such social data into features that are relevant for a format of ridesharing that riders and drivers would accept. Therefore, building trust between unacquainted ridesharing pairs and determining whether there are other rides in which people would prefer to travel together are essential attributes to building trust in ridesharing (Cici et al. 2014). Our study is most similar to work done by Berlingerio et al. (2017a), who also examined user's preferences in terms of trajectories and social interest. However, as opposed to our work, we constrain grouping riders by vehicle capacity and spatial proximity, and measure the travel time of a group by comparing that of all members' and drivers. Moreover, in order to reach the best compromise between user satisfactory and recommendation quality when travelling with strangers, we further recommend a group by predicting a user's intended destination based on his historical trajectories data.

3 Problem definition

3.1 Ridesharing group

Assume driver set $D\{d_n\}$ and rider set $R\{r_i\}$, each driver will be assigned a group of riders which have the similar requests, including departure time, origin, and destination. The rider may make a rejection for not being interested in any of the assignment. Regarding this rejection, we extend thus the grouping criterion to consider the *people's* willingness to share rides. In addition, the vehicle will pass by somewhere in the vicinity of the riders for delivery task. We also introduce the walking time of a rider between his or her origin to meeting point into the grouping criterion.

Each riders' demands on departure and destination arise at their place of stay. The vehicle is assumed to be traveling on the shortest path between a set of locations of picking up and dropping off, where the ride time are independent of traffic flow (Daganzo and Sheffi 1977).

Definition 1 (*Ridesharing group*) Given a rider-rider pair (r_i, r_j) and their origins, time windows and trust links between them, the ridesharing groups subsets for a driver d_n is defined a set $G_{d_n} = \{G^q_{d_n}\}, G^q_{d_n} = \{r_i | i \neq j, m_{ij} = 1\}$ from a candidate group set $G = \{G_t | t = 1, 2, \dots, Q\}$, where the following is assumed:

1. The trust link between riders is expressed as,

$$m_{ij} = \begin{cases} 1, & r_j \text{ is grouped with } r_i \\ 0, & otherwise\ . \end{cases} \tag{1}$$

2. The distance from the group to driver is the maximum network distance from every group member's OD pairs to the pick-up and drop-off locations ($i.e.$,$d_n.u$, $d_n.d$), and calculated as, respectively,

$$\begin{aligned} D^u_n(q) &= \max_{r_i \in G^q_{d_n}} distance(d_n.u, r_i.o), \\ D^d_n(q) &= \max_{r_i \in G^q_{d_n}} distance(d_n.d, r_i.d). \end{aligned} \tag{2}$$

Equation (2) obtains the distance for each driver-rider pair incrementally by *Dijkstra's* algorithm over a road network.

3. Each rider,r_i , arrives at the meeting point in advance. The additional waiting time is denoted as a time window (TW) between the meeting time of the group, ($i.e.$,$r_i.t$), and the departure time of the driver ($i.e.$,$d_n.t$), is defined. $distance(d_n.u, r_i.o)$ represents the extra walking time, V denotes a walking speed, and then the time cost is deduced as $\frac{diatance(d_n.u, r_i.o)}{V}$ due to the extra walking.

$$TW_n(q) = \max_{r_i \in G^q_{d_n}} TW\left(d_n.t, r_i.t + \frac{diatance(d_n.u, r_i.o)}{V}\right) \tag{3}$$

Equation (3) allows the specification of a maximum rider waiting times at the meeting points. Equations (2) and (3) affect the riders' possibility to share rides.

We set up the grouping criterion for each rider-rider pair by their extra walking distance and additional waiting time. Assume that there are a group $G^q_{d_n}$ to be assigned to driver d_n , the travel time of group $G^q_{d_n}$ due to the walking extra miles and waiting additional minutes is calculated in Eq. (3), and $h_n = [h^i_n]$ indicates that a set of riders can be grouped if and only if they can arrive the pickup location before $d_n's$ departure. Therefore, Eq. (4) demonstrates that the travel time equals to 0 if no one is recommended to share a ride for d_n.

$$\begin{aligned} TT_n(q) &= \left(TW_n(q) + \frac{D^d_n(q)}{V}\right)^{h^i_n(q)} - 1 \\ h^i_n(q) &= \begin{cases} 1, & D^u_n(q)/V \leq \min_{r_i \in G^q_{d_n}} TW\left(d_n.t, r_i.t\right) \\ 0, & otherwise. \end{cases} \end{aligned} \tag{4}$$

4. For each rider-rider pair (r_i, r_j), we define the trust links as a probability-of-interest measure $P_{ij} \in (0, 1)$, which is the probability that r_i would be interested in sharing rides with r_j. P_{ij} is affected by social interactions, and can be mined from riders' social network.

3.2 Acceptance maximization in ridesharing recommendation (AM-R) problem

Due to the rejection from riders, we need to propose algorithms to assigning each driver d_n a candidate group set G for maximizing the acceptance. Since r_i makes a rejection on a recommendation $M = \{(d_n, G^q_{d_n})$ if and only if he or she is not interested in sharing rides with r_j in a G_{d_n} recommended. The acceptance of M can be computed as follows, where N is the number of riders.

$$E(M) = \sum_{\substack{\leq i \leq N \\ i+1 \leq j \leq N}} m_{ij} \times p_{ij} \tag{5}$$

To ensure fairness, let d_n picks up K riders at most in each task. So the system has two constraints when doing the recommendation

$$\begin{aligned} &\forall n \neq m,\ G_{d_n} \cap G_{d_m} = \emptyset \\ &\forall n, \| h_n \| \leq k. \end{aligned} \tag{6}$$

We define the problem of ridesharing recommendation as an optimization for maximizing the acceptance when traveling in G_{d_n} with a smallest travel time. It returns the top q groups of riders from the candidate sets G. We now define a score function for the recommendation M, $S(M)$, and denote the optimal recommendation M as M_{opt}, then

$$M_{opt} = \arg\max_M S(M) = \arg\max_M \sum_{\substack{1 \leq i \leq N \\ i+1 \leq j \leq N}} m_{ij} \times p_{ij} \tag{7}$$

subject to constraints (6).

Definition 2 (*AM-R*) Given a driver d_n , a ridesharing group set of size Q, grouping between two riders m_{ij}, and probability of interest p_{ij} for each rider-rider pair in each $G^q_{d_n}$, the problem of maximizing acceptance in ridesharing is to find a recommendation instance M such that the score $S(M)$ (7) is maximized, subjecting to constraints (6).

4 Exact solutions

A ridesharing recommendation is based on a supervised setting, as a result, we need to extract features for the links between r_i and r_j. In this section, we first present a destination prediction algorithm which exactly enumerates all possible ridesharing groups by linking the riders with similar spatial movements, and then propose a trust link-search algorithm which adopts social network to find the optimal recommendation.

4.1 Group discovery from predicted semantic destination

The semantic destinations visited are especially useful in capturing latent relationships among group members (Shaheen et al. 2016). A semantic trajectory model describes the spatial-temporal movements for riders that includes all $S(S > 1)$ locations visited by r_i over a 1 day period as $Seqloc_{ri} = \{loc_s\}$,where loc_s represents a spot from the trajectory database, $loc_s = (lat, lon, arv_{time}, lev_{time}, L_s, loc_{category_s})$, which is defined according to the latitude (lat) and longitude (lon), the visiting time (arv_{time}, lev_{time}), the name (L_s) of loc_s , and the semantic types $loc_{category_s}$. All m distinguishing trajectories of r_i are collected in Seq_{ri}.

We first employ a PPM (prediction by partial matching)-based prediction (PP) algorithm as shown in Algorithm1. In the PP algorithm, a type of destinations with maximal probability by enumeration procedure is conducted. Suppose that Nth-*order* contexts of $loc_{s+1} \in Seqloc_{ri}$ are ordered in a sequence $loc^N_s[loc_{s-(N-1)}, \ldots, loc_{s-1}, loc_s]$, PP enumerates all the possible subsequences of length j for in seq_{ri} from $j = N$ to $j = 1$ (lines 8–10 in Algorithm 1). That is, for the last locations, loc_{s+1}, PP calculates its probability distribution as being the next destination. If there are not such sequence when $j = N$, PP simply looks up the $loc^j_n$to computer the probability after each enumeration for $j = N - 1$. Finally, in Algorithm 1, PP obtains the riders' destination sets (line 18).

Algorithm 1 PP

Input: $r_i's$ historical trajectories Seq_{ri},current spot loc_s,probability $P(L_i|loc_s)$
Output: Destination set P^*

Initialzation :$P^* \leftarrow \emptyset$
$L_i \leftarrow loc_{s+1}$
for $L_i \in Seq_{r_i}$ **do**
  while $N! = 0$ **do**
    if L_i is neighbor of loc_s^N **then**
      calculate $p(L_i|loc_s)$
    else
      while (no loc_s^j for L_i) **do**
        predict the probability of escape code
        $j \leftarrow N - 1$;
      end while
    end if
    calculate $p(L_i|loc_s)$
  end while
end for
put L_i and $p(L_i|loc_s)$ into P^*
rank top_k L_i using $p(L_i|loc_s)$
return p^*

Example 1 Suppose $Seqloc_{ri} = \{L_1, L_2, L_6, L_1, L_3, L_1, L_5, L_1, L_2, L_6, L_1\}$, and is given as a 2th Tree shown in Fig. 1. Considering L_1 as loc_s, PP first enumerates all possible context sequence of L_3 for $j = 2$: $[L_6, L_1]$, and achieves the prediction probability $p(L_3|L_1) = \frac{1}{2}$. There are not the context sequences of L_2 , PP calculates the probability of escape code and obtains the prediction probability $p(L_2|L_1) = \frac{3}{7}$ when $j = 1$.

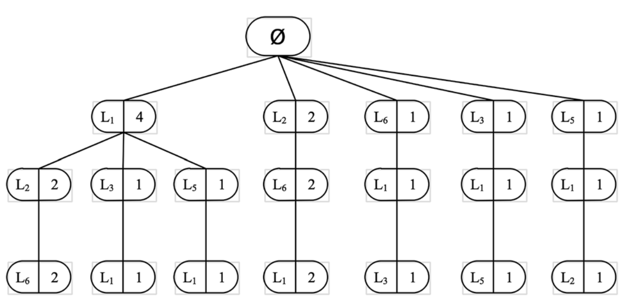

Fig. 1 A tree with probability distribution where the length of contextual sequence is 2. Each node includes the location and associated frequency count

Given any destination set $P^\star$ for R, *Find_Group* is proposed to cluster the riders based on a heterogeneous travel network developed in our previous work (Tang et al. 2018). The ridesharing relationship can be described using a meta-path '$U \xrightarrow{depart} T$' , or short as UT shown in Table 1. Hence the similarity search based on UT will obtain the driver-rider pairs with geographic proximity (lines 2, 3 in Algorithm 2). *Find_Group* calculates the travel time of each driver-rider pair to get all possible group subsets G_{d_n} for d_n after each enumeration of similarity search (lines 13–15). We store the recommendation M for each of them in a Dictionary with key $(d_n, G^q_{d_n})$. It is obvious that the top-k r_i making the TT_n minimal consists each group $G^q_{d_n}$. To support the discovery, the function $PP(r_i)$ used in line 11 in Algorithm 2 returns all the *size_j* subsets of set $P^\star$.

Table 1 Definition of meta-path based on ridesharing relationship

Meta-path	Semantics of meta-path	Similarity travel relationship based on meta-path
U→T	Users travel pass at a certain area within a certain time period	If there is U→T→U, the two users depart a time span apart, then the ridesharing relationship between them is similar

Algorithm 2 $Find\_group$

Input: riders set R , r_i′origin $r_i.o$, d_n′ pickup $d_n.u$, a set $SC_1 = \emptyset$, searching redius θ_d, time span for waiting θ_t **Output:** G_{dn}

for each r_i in R **do**
  if $dis\tan ce(r_{i.}.o, d_n.u) \leq \theta_d)$ **then**
    $pathsim(r_i, d_n, \theta_t)$
    put r_i , $r_i.o$ in SC_1
  else
    delete r_i
  end if
end for
while $size(G^q_{d_n}) < k$ **do**
  for $i = 1$ to $size(SC_1) - 1$ **do**
    $P* \leftarrow PP(r_i)$
    set loc_s which has maximal probability in $P*$ to be $r_i.d$
    calculate $TT_n(q)$
    if $TT_n(q) - \delta < 0$ **then**
      put r_i in $G^q_{d_n}$
    end if
  end for
end while
$Dictionary(d_n, G^q_{d_n})$
G_{dn}

4.2 Group recommendation based on social network

Chaube et al. (2010) pointed out that people have significantly higher willingness and detour tolerance to share rides with whom they know. That is because generally people have higher trust in whom they have the same social interests. Therefore, we search the rider-rider pair with same social interests to identify the trust links.

In this section, we will accumulate information for interests between riders at first. Actually, we notice users can publish online posts at the locations. And we accumulate the text and timestamps information of the online posts checked in at a certain location as the auxiliary information for trust link with riders. From a statistical point of view, information from posts published, including both functions and text contents, can reveal some properties of interactions between users. For example, posts published when ridesharing can contain some phrases depicting the scenes in terms of safety, speed, and convenience etc. Moreover, the '*Like*' and '*Share*' functions indicate approval for posts between friends. These feedbacks also contribute to gauging opinions on a range of topics. So we can know more about the similarity between riders from the information accumulated from online posts.

We extract the topics of interest (Guidotti and Berlingerio 2016) with rider r_j in each $G^q_{d_n}$ to determine $p_{ij}$from a set of tweets,$Tweet(r_i)$. A latent dirichlet allocation (LDA) (Lee et al. 2017) is then applied to learn the latent topics of $Tweet(r_i)$ through word splitting, stop-word filtering, and part of speech. A vector t_{r_i} is then established for the intended topics corresponding to r_j. For each rider-rider pair (r_i, r_j), the probability-of-interest of sharing rides, p_{ij} is expressed by the following Eq. (8):

$$p_{ij} = \frac{t_{r_i} \cdot t_{r_j}}{\parallel t_{r_i} \parallel \cdot \parallel t_{r_j} \parallel}. \tag{8}$$

We look up the *Dictionary* to get the optimal recommendation after each enumeration of recommendation for d_n. When there is only the last left d_n, it is obvious that the top $G^q_{d_n}$ with largest $S(M)$ is the optimal assignments.

Example 2 Suppose $D = \{d_1, d_2\}$, $R = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$, and $q, k = 3$ is given as below ,All possible groups subsets for d_1 are first enumerated:$\{(d_1, (r_1, r_2)), (d_1, (r_3, r_4, r_6))\}$, and achieves optimal assignments $d_1 - (r_1, r_2)$. Then the algorithm traces back to d_2 , the possible group sets are $\{(d_2, r_5), (d_2, (r_1, r_3)), (d_2, (r_7, r_8))\}$. There are three options, i.e., $d_2 - r_5, d_2 - (r_1, r_3), d_2 - (r_7, r_8)$ It then looks up the recommendation from the Dictionary: $d_1 - (r_1, r_2)$ for case $d_2 - r_5$ and $d_2 - (r_7, r_8)$, and get the optimal solution: $d_2 - (r_7, r_8)$.

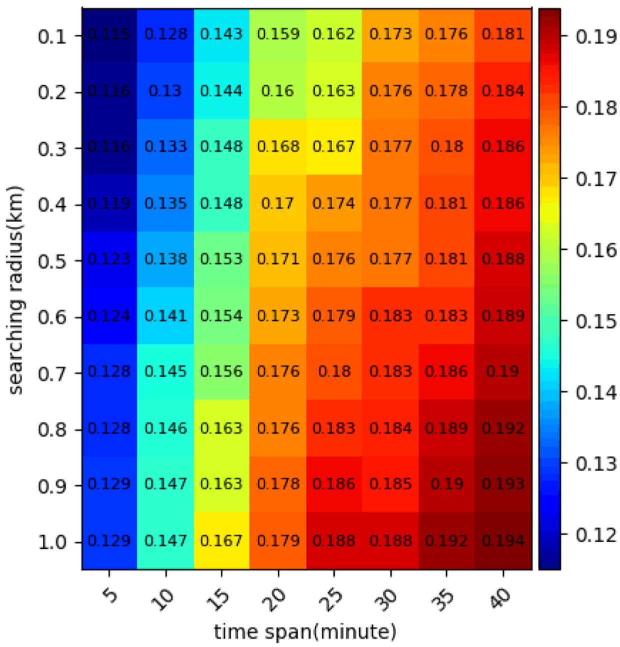

Fig. 2 Effectiveness under various settings of time span and search radius where rides with four-passengers. Each color represents different reduction rate of total trip

p_{ij}	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
r_1	1	0.8	0.9	0.8	0.3	0.5	0.8	0.8
r_2	0.8	1	0.3	0.3	0.9	0.9	0.2	0
r_3	0.9	0.3	1	0.4	0	0.6	0.5	0.5
r_4	0.8	0.3	0.4	1	0.2	0.3	0.9	0.6
r_5	0.3	0.9	0	0.2	1	0	0	0.4
r_6	0.5	0.9	0.6	0.3	0	1	0.2	0.1
r_7	0.8	0.2	0.5	0.9	0	0.2	1	0.9
r_8	0.8	0	0.5	0.6	0.4	0.1	0.9	1

5 Discussions

In some extreme case, i.e.,$\forall_q, |G^q_{d_n}| = 1$, all the possible groups include a rider due to the distributed demands, e.g., (d_2, r_5) in Example 2. We can increase the searching radius θ_d and time span for waiting θ_t. Generally, conducting questionnaires among riders for these parameters is an effective approach. In this work, we perform the maximum likelihood estimation on the reduction rate of travel time under the constraint, that is, the total walk time cannot exceed the total time in a ride-share trip for a rider (Stiglic et al. 2015).

$$RR\left(G^q_{d_n}\right) = 1 - \frac{\sum_i^k TT_n(q)}{\sum_i^k TW\left(t_i.o, t_i.d\right)} \tag{9}$$

The reduction rate of total trip with various settings of time span and search radius is shown in Fig. 2. The values of reduction rate increase with respect to increments in radius and time; that is, more additional riders can be incorporated into candidate groups with a larger search region and time span. It gains an improvement of 7% in reduction rate by increasing radius from 0.3 km to 0.4 km, and time span from 5 to 10 min, additional improvements reduce to only 0.3% and 0.2% respectively with higher level of aggregation. Therefore, the proposed algorithms can still be adopted with a simple parameter setting.

In this work, we specially focus on the acceptance problem in ridesharing recommendation, which has not studied in existing works. The optimal matching studied in existing works are addressed with different settings, e.g., a detailed routing, OD-pairs, departure and arrival time are common for ridesharing participants (Furuhata et al. 2013). Most of the current methods are more focused on accurately matching such information predetermined by riders. OppRide developed the network time prism with road network to represent the service. A set of rider can be then informed where should be boarded and dropped (Rigby et al. 2016a). However, it is possible to combine these methods with ours to achieve a more desirable solution. For example, one way to make existing works acceptable is to adopt this work as a subsequent step. The output of OppRide is a set of driver-rider pairs. There pairs can be further to restrict so that only high trust pairs are taken into account for recommendation.

6 Experiments

We use expected acceptance to evaluate the performance of the proposed methods on Twitter database under different settings. All our experiments are run on an Intel Xeon E5-2620 CPU @2.10 GHZ with 64 GB RAM.

6.1 Experiments on real data

In this section, we further investigate the performance of PP, ridesharing recommendation on real data. There are

Table 2 Performance indices of PP

Indices	Description
Pre(R)	Destinations predicted for all riders in R
$T(R^*)$	Destinations visited for riders in test set
$A@N$	Accuracy at N, $A@N = \sum_{n=1}^{N} \frac{Pre(r_n)\cap T(r_n)}{Pre(r_n)}$ during a day
$R@N$	Recall at N, $R@N = \sum_{n=1}^{N} \frac{Pre(r_n)\cap T(r_n)}{T(r_n)}$ during a day

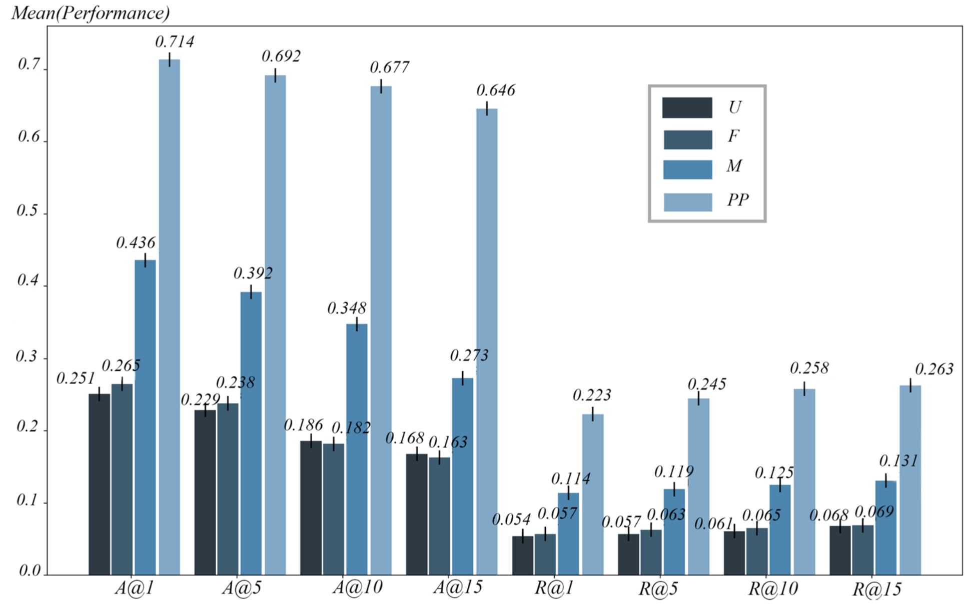

Fig. 3 Effect of N on accuracy and recall

two datasets used here, one for destination prediction and the other one for social interest accumulation. In Twitter, check-ins and tweets of users are collected over the period of October 2016–January 2017. The check-ins and tweets are treated as our trajectories and topics respectively.

6.2 Performance of PP algorithm

To prepare for the evaluation, we split all check-ins dataset into a training and test set by the ratio 6:4. We consider four algorithms for the accuracy and recall of prediction: user-based CF(U), Friend-based CF(F), Meta-path based Recommendation(M) and PP. Table 2 shows the indicators used.

The evaluation results of all algorithms are reported in Fig. 3. It can be observed that the accuracy of all algorithms goes down with the increase of $N(N = 1, 5, 10, 15)$, while the recall goes up. This is because a larger N means more choices for destinations, which results in larger recall. More specifically, in terms of both accuracy and recall, U and F, without considering the current location of users, obtain similar results and rank bottom, throughout all values of N, following M and PP. M searches the similar users by taking the semantic relationships into account, and thus attains a better performance.However, there is a large gap between M and PP, which concluded that the similarities between users without considering temporal information is not effective for destination prediction.

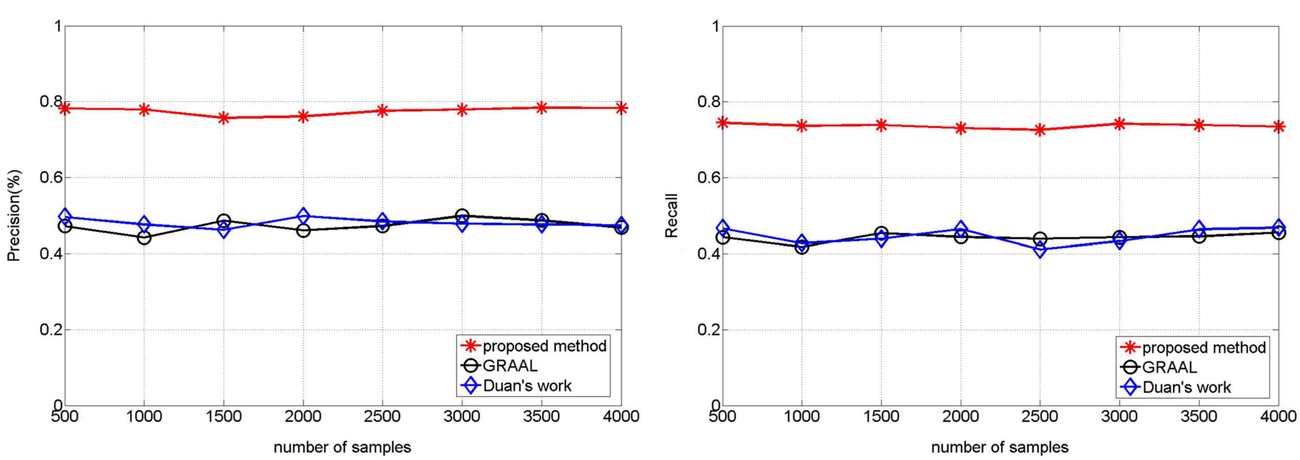

Fig. 4 Evaluate ridesharing recommendation methods on real data

6.3 Performances of ridesharing recommendation

We adopt HDP (Teh et al. 2006) method to obtain the number of topics in the tweets. The experiment has been repeated 1000 times, and the average number of topics (i.e., 38.3) is reported. We set the dimension of t_{ri} to 38.

With the candidate groups on hand, we analyze the precision and recall of our proposed algorithm, GRAAL (Berlingerio et al. 2017b), and Duan et al. (2018) work. With the ground truth for similarity of any two users in Twitter, it is possible to validate the performance with a similarity threshold θ. We vary the number of samples, with default values of $\theta(0.5)$, as shown in Fig. 4. Our method still performs well on this real dataset, with precision and recall ranking top and around 78%, because it represents the latent social interests better and incorporates them into the similarity measures. For GRAAL and Duan et. al.'s work, it shows that Duan et.al.'s work attains better precision in most cases because the user's interest in particular places is associated with the semantic types. However, the precision of Duan et.al.'s work reduces when number of samples is 1500, 3000, 3500 respectively, that is difficult to compare the two algorithms..

7 Conclusion

In this paper, we formulate the problem of maximizing acceptance in ridesharing recommendation, which aims to maximize riders' acceptance by assigning proper ridesharing groups to riders. The problem is addressed by measuring the willingness of riders for joining in a group using social context. Our proposed method leverages the spatiotemporal trajectories and underlying trust links between riders, to reach the best compromise between user satisfactory and recommendation quality when ridesharing with strangers. Experiments have been conducted to evaluate the proposed method on real data, that is, the method is the top choice with its better performance w.r.t. acceptance and efficiency.

References

Agatz, N., Erera, A., Savelsbergh, M., Wang, X.: Optimization for dynamic ride-sharing: a review. Eur. J. Oper. Res. **223**(2), 295–303 (2012)

Bakkal, F., Eken, S., Savaş, N.S., Sayar, A.: Modeling and querying trajectories using NEO4J spatial and timetree for carpool matching. In: 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 219–222. IEEE (2017)

Berlingerio, M., Ghaddar, B., Guidotti, R., Pascale, A., Sassi, A.: The graal of carpooling: green and social optimization from crowd-sourced data. Transp. Res. Part C: Emerg. Technol. **80**, 20–36 (2017a)

Berlingerio, M., Ghaddar, B., Guidotti, R., Pascale, A., Sassi, A.: The graal of carpooling: green and social optimization from crowd-sourced data. Transp. Res. Part C Emerg. Technol. **80**, 20–36 (2017b)

Chaube, V., Kavanaugh, A.L., Perez-Quinones, M.A.: Leveraging social networks to embed trust in rideshare programs. In: 2010 43rd Hawaii International Conference on System Sciences, pp. 1–8. IEEE (2010)

Cici, B., Markopoulou, A., Frias-Martinez, E., Laoutaris, N.: Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 201–211. ACM (2014)

Daganzo, C.F., Sheffi, Y.: On stochastic models of traffic assignment. Transp. Sci. **11**(3), 253–274 (1977)

de Abreu e Silva, J., et al.: Using latent attitudinal variables estimated through a structural equations model for understanding carpooling propensity. Transp. Plan. Technol. **36**(6), 499–519 (2013)

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

Elbery, A., ElNainay, M., Chen, F., Lu, C.-T., Kendallm J.: A carpooling recommendation system based on social VANET and geo-social data. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 556–559. ACM (2013)

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, Xiaoqing, Koenig, Sven: Ridesharing: the state-of-the-art and future directions. Transp. Res. Part B Methodol. **57**, 28–46 (2013)

Guidotti, R., Berlingerio, M.: Where is my next friend? recommending enjoyable profiles in location based services. In: Complex Networks VII, pp. 65–78. Springer (2016)

Karamshuk, D., Boldrini, C., Conti, M., Passarella, A.: Human mobility models for opportunistic networks. IEEE Commun. Mag. **49**(12), 157–165 (2011)

Lee, L.C., Liong, C.-Y., Jemain, A.A.: Q-mode versus R-mode principal component analysis for linear discriminant analysis (LDA). In: AIP Conference Proceedings, vol. 1842, p. 030024. AIP Publishing (2017)

Li, Y., Rui, C., Lei, C., Xu, J.: Towards social-aware ridesharing group query services. IEEE Trans. Serv. Comput. **10**, 646–659 (2015)

Li, X., Li, M., Gong, Y.-J., Zhang, X.-L., Yin, J.: T-desp: destination prediction based on big trajectory data. IEEE Trans. Intell. Transp. Syst. **17**(8), 2344–2354 (2016)

Li, Y., Chen, R., Chen, L., Jianliang, X.: Towards social-aware ridesharing group query services. IEEE Trans. Serv. Comput. **10**(4), 646–659 (2017)

Nadembega, A., Hafid, A., Taleb, T.: A destination and mobility path prediction scheme for mobile networks. IEEE Trans. Veh. Technol. **64**(6), 2577–2590 (2015)

Rigby, M., Winter, S., Kruger, A.: A continuous representation of ad hoc ridesharing potential. IEEE Trans. Intell. Transp. Syst. **17**, 2832–42 (2016a)

Rigby, M., Winter, S., Krüger, A.: A continuous representation of ad hoc ridesharing potential. IEEE Trans. Intell. Transp. Syst. **17**(10), 2832–2842 (2016b)

Shaheen, S.A., Chan, N.D., Gaynor, T.: Casual carpooling in the San Francisco bay area: understanding user characteristics, behaviors, and motivations. Transp. Policy **51**, 165–173 (2016)

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M.: The benefits of meeting points in ride-sharing systems. Transp. Res. Part B Methodol. **82**, 36–53 (2015)

Tang, L., Zhao, Y., Duan, Z., Chen, J.: Efficient similarity search for travel behavior. IEEE Access **6**, 68760–68772 (2018)

Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. Publ. Am. Stat. Assoc. **101**(476), 1566–1581 (2006)

Wang, Y., Kutadinata, R., Winter, S.: The evolutionary interaction between taxi-sharing behaviours and social networks. Transp. Res. Part A Policy Pract. **119**, 170–180 (2019)

Lei Tang (Sichuan, China, 1983) received the PhD degree in computer science and technology in 2012. She is currently work in the school of information engineering, Chang'an University, P. R. China. She has been a visiting researcher at the chair of information systems, Mannheim University, Germany in 2009–2010. Her research interests include the area of Heterogeneous network, intelligent transportation system, and pervasive computing. Dr. Tang is a member of ACM, IEEE, and a member of CCF (China Computer Federation).

Zongtao Duan (ShaanXi, China, 1977) is currently a professor at the school of information engineering, Chang'an University, P. R. China. He received his Ph.D. in computer science from Northwestern Polytechnical University, P. R. China in 2006. He has been a postdoctoral research fellow in the University of North Carolina, American in 2009-2010. His research interests include context-aware computing in transportation. Dr. Duan is a member of CCF, CCF High Performance Computing, and-Pervasive Computing Technical Committee.

Meng Han (Shandong, China, 1996) is currently study in the school of software engineering, Chang'an University, China. She is a student member of CCF. Her research interest is network embedding.

Dandan Cai (Xi'an, China,1995) received the master degree in the school of software engineering, Chang'an University, China. Her research interest is social recommendation.