A Performance Evaluation Model for Taxi Cruising Path Recommendation System

Huimin Lv, Fang Fang, Yishi Zhao, Yuanyuan Liu, and Zhongwen Luo(✉)

College of Information Engineering,
China University of Geosciences (Wuhan), Beijing, China
lvhuimin320@gmail.com, ffang1014@163.com,
{zhaoyishi,luozw}@cug.edu.cn,
yuanyuanliu@mails.ccnu.edu.cn

Abstract. Recommending an appropriate route to reduce taxi drivers' mileage spent without a fare is a long-standing challenge. The current solution has been to get the best route which has optimal performance, and the performance usually combined the conditional probability for getting a passenger and the cruising distance. However, the main reference has some limitation. To eliminate the limitation, a novel model is proposed to evaluate the candidate route performance. And based on this new model, a recommendation system is tested. Firstly, by mining the knowledge of the historical taxi trajectory, we extract the temporal probabilistic recommending points. Then based on it, the evaluation model is presented to estimate the performance of each candidate route. Finally, a route recommendation algorithm is used to get the optimal route for taxi drivers. And as the result, the experiment is performed on real-world taxi trajectories data set, and shows the effectiveness of the proposed model for evaluating the performance.

Keywords: Evaluation model · Mobile recommendation systems · Taxi drivers

1 Introduction

Nowadays, taxi service plays an important role in public transportation service in large cities. However, there are often a huge number of taxis cruising around the city with no passengers. The vacant taxis not only waste energy but also result in a traffic jam. So, a recommendation system to improve the performance of taxis is needed. And the advances of various technologies provide the possibility.

Indeed, most of the existent mobile recommendation systems are using the integration of the conditional probability and the cruising distance or others such as income to measure the performance of the route and then recommend the best one to the taxi [1–4]. However, we find that the living performance evaluation method is wrong in some cases. Using the existing methods will result in sending the taxi to a lower performance route. Hence, a new performance evaluation method is proposed in this paper. In addition, since taxi trajectories are big spatial-temporal data and how to extract the useful information like the mobility pattern of the passengers with consideration of the time factor is also challenging.

J. Kim et al. (Eds.): PAKDD 2017, Part II, LNAI 10235, pp. 156–167, 2017.
DOI: 10.1007/978-3-319-57529-2_13

To that end, in this paper, we propose a recommendation system based on historical trajectory taxi data. The key idea is that it utilizes a new route model to evaluate a candidate route and then provide an algorithm to find a potential passenger with the minimum cruising miles. Specifically, the contributions of this paper are as follows:

1. A novel model for evaluating the candidate route is proposed. The model computes the potential cruising distance along the route for picking up per passenger. The main difference of the new model from the traditional one is considering the passenger number of the route. A recommendation system for taxi drivers to minimize their cruising driving distance for taking per passenger is presented.
2. To verify the effectiveness of the model, we conduct extensive experiments on a real-world data set. And the result shows that the new model is more effective.

The remainder of this paper is organized as follows. Section 2 shows some related works. In Sect. 3, we formulate the problem of route recommendations for taxi drivers and introduce some preliminaries in the paper. Section 4 presents the generation of the temporal probabilistic recommending pick-up points. In Sect. 5, the recommending model is discussed in details. Section 6 shows some experimental results and the paper is concluded in Sect. 7.

2 Related Works

In the literature, a mass of research has been devoted to the recommendation system [5–9]. Based on the massive data of the taxis' trajectory, the route recommending system's main target is to provide the more efficient driving route for taxi drivers by finding the behavior pattern of the experienced taxi drivers, the potential flowing direction of the crowds, etc. Li et al. [10] pay attention to the prediction of the movement of human beings. They present an adaptive hot extraction algorithm to cluster the pick-up/drop-off events of the passengers. Awasthi et al. [11] propose a rule-based method to evaluate the fastest path in the city. In order to get the fastest route, they build a statistical model using the traffic log. Gonzalez et al. [12] develop an adaptive fastest path algorithm by considering the speed patterns mined from historical GPS trajectory data. Ge et al. [1] develop a mobile recommendation system to recommend a taxi driver with the shortest potential travel distance route for finding a passenger. Then some concern the carpool service [13, 14] to save energy and seek the balance of demand between the taxi drivers and passengers. In the T-Share system, users submit request of taking a taxi with the location of getting on and off, the number of passengers and the expected time to the destination through the phone. System maintains all states of the taxi in real-time in the back, and after receiving a request, search out the best cab which satisfies the conditions of the new user and the passengers already in the cab. In addition, other works care about the optimization of calculation. Statistics show that the time complexity of existing recommendation methods are usually exponentially [15]. Trestian et al. [16] use the orthogonal kd-tree. Yang et al. [4] propose a new kds-tree structure which is a binary tree and extended from kd-tree and ball-tree. In this article, we focus on the recommendation of the shortest potential cruising distance for taxi drivers. Different from the earlier studies, we propose a novel model to evaluate the performance of each candidate route and then based on it, get the optimal route for taxi drivers.

3 Problem Definition

Definition 1. Picking-up rate: picking-up rate is the probability of finding a passenger at one pick-up point.

Definition 2. Given a set of N potential pick-up points $C = \{C_1, C_2, \ldots, C_N\}$, a route R is a sequence of connected pick-up points, i.e., $R = (C_1 \rightarrow C_2 \rightarrow \ldots \rightarrow C_K)$ $(1 \leq K \leq N)$, The length of route R denotes as $|\vec{R}| = K$, the number of pick-up points. R_{set} is the set of R, which is generated from C. Independence probability set $P_R = (P_1, P_2, \ldots, P_K)$ denotes the set of picking-up rate of pick-up point and distance subset D_R represents the set of distance between each pair of pick-up points en route R.

Note that C_0 always denotes the current position of a taxi PoTaxi in this paper.

Definition 3. The taxi mobile routing recommendation problem is to recommend a profitable route to a taxi driver so that the potential cruising distance to a possible passenger is minimized.

As the calculating the potential cruising driving distance depends on the current position of the taxi, time period, route R, and the corresponding picking-up rate set P_R and the distance set D_R, the potential cruising distance function can be denoted as:

$$\mathrm{F}(\mathrm{PoTaxi}, \mathrm{T}, \mathrm{R}, P_R, D_R)$$

Note that, in this paper, we limit the length of route R to be K. This is because the calculating constraints and considering the practical applications.

So, the taxi mobile routing recommendation problem can be formulized as:

$$\min_{R \in R_{set}} \mathrm{F}(\mathrm{PoTaxi}, \mathrm{T}, \mathrm{R}, P_R, D_R)$$

Almost all the current existing researches of evaluating the performance of the route are using the integration of the conditional probability for getting a passenger and the cruising distance. However, this is wrong. Because it only takes the potential driving distance into account without considering the probability of picking up passengers. In other words, they do not consider the number of the passengers along the route. The potential cruising distance of finding a passenger will be the correct evaluation standard. So, in this paper, we put forth a novel model, the potential cruising distance function, which is not only considering the driving distance also the probability of picking up passengers, to evaluate the route from a taxi to a potential passenger, which will be discussed in detail later.

First of all, let's focus on a demonstration of recommending. Figure 1 shows an illustration example of two candidate paths.

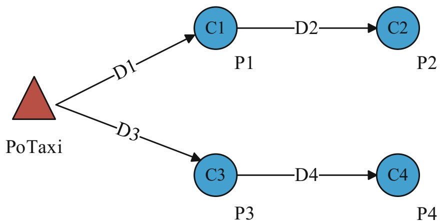


Fig. 1. An example of two candidate path

In this graph, node PoTaxi (C_0) represents the current position of an empty taxi at time period t, node C_i (i = 1, 2, 3, 4) denote the recommending pick-up point with the estimated picking-up rate P_i (i = 1, 2, 3, 4) respectively. D_i (i = 1, 2, 3, 4) indicate the distance between node C_{i-1} and node C_i. In addition, there are two candidate driving routes R_1 = {PoTaxi, C_1, C_2} and R_2 = {PoTaxi, C_3, C_4}. Note that the nodes in the path are sequential and assumed to be different from each other. This is because we do not allow taxi drivers to drive back and forth.

Nowadays, almost all the methods of calculating the potential cruising distance from the taxi to a potential passenger are integrating the conditional probability with the cruising distance. For example, in Fig. 1, the potential cruising distance of route R_1 could be $P_1D_1 + (1 - P_1)P_2(D_1 + D_2)$, for route R_2, it will be $P_3D_3 + (1 - P_3)P_4(D_3 + D_4)$. In some cases, it makes sense, just like the probability is almost similar but the distance is very different. In other cases, however, this is not really applicable. As Fig. 2 shows, under this circumstance, the method is wrong.

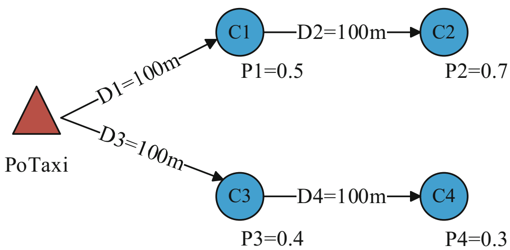


Fig. 2. A concrete example of routes

According to the previous method, the calculated potential cruising distance respectively is 120 m and 76 m, so the better route will be R_2 = {PoTaxi, C_3, C_4}. However, it is inconsistent with the facts, and clearly that we should choose route R_1 rather than R_2 in any cases unless you do not want to make a profit. The potential cruising distance from the taxi to a potential passenger cannot be simply integrating the conditional probability with the cruising distance. Furthermore, we should also take the probability of picking up passengers into account.

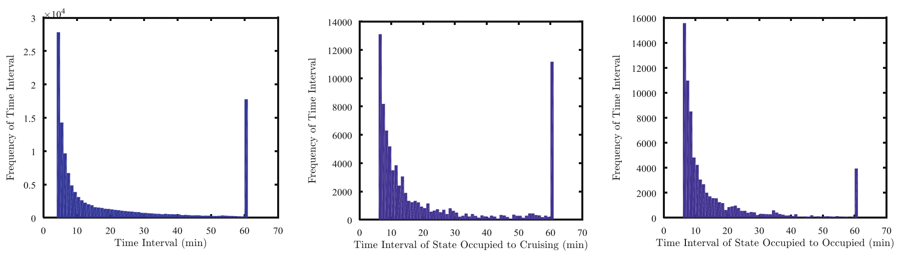
4 Temporal Probabilistic Recommending Point Generation

In this section, we show how to generate the temporal probabilistic recommending points. There are two main steps: clustering based upon the pick-up points of the experienced drivers, and calculate the probability of each recommended point.

4.1 Clustering Based on the Pick-up Points of the Experienced Drivers

To generate the recommended points, firstly, the experienced drivers are extracted from a large number of taxi track data. Then we can get their pick-up and drop-off points at different time period. Secondly, calculating the pair wise driving distance of these pick-up points of different time period using the Google Maps Distance Matrix API. Finally, clustering based on the calculated driving distance.

The driving time and the driving occupancy rates are the main factors to extract the experienced drivers, while the state of the driver is important to calculate the driving time and the driving occupancy rates. We consider drivers with plenty of driving time and high driving occupancy rates to be experienced. In general, there is three status of the driver's driving state: occupied, cruising and out-of-service state. Driver's driving time is the time when the state is not out-of-service, and the occupancy rate is the ratio of driving time of occupied to total driving time. Assume that there are two continuous GPS points of a driver, the state of the two points are occupied, but the time interval is greater than an hour, can we expect this time interval as occupied driving time? Figure 3(a) shows the distributions of the time interval of two continuous GPS points of more than 500 drivers in San Francisco over a period of about 30 days. Figure 3(b) and (c) show the distributions of the time interval of two continuous GPS points when the state changes from occupied to cruising and occupied to occupied of these drivers. From the figures, it's clear that some intervals are greater than one hour. Based on this observation, we conduct lots of experiments to get the best threshold for calculating the driving time and the driving time with passengers. Then, we extract the experienced drivers with their pick-up points at different time period. Figure 4 shows the distribution of pick-up points of experienced drivers.



(a) (b) (c)

Fig. 3. Distributions of time interval of two continuous GPS points: (a) all status; (b) occupied to cruising; (c) occupied to occupied

As Fig. 4 shows, different time periods have various numbers of pick-up events. In other words, there is different pick-up probability. And we can find that the trend is in accordance with our common sense, where the picking-up events happen less in the midnight (02:00–04:00) and higher during the night (18:00–22:00). After obtaining the historical pick-up points at different time periods, we use driving distance to cluster these points into N clusters for different time periods. Using driving distance rather than simply the spherical distance or Euclidean distance can make more accurate recommended results. Furthermore, we use the Cluto [17] for clustering by using vcluster clustering programs with parameters "−clmethod = direct". Eventually, the center of each cluster is the recommending points we needed.

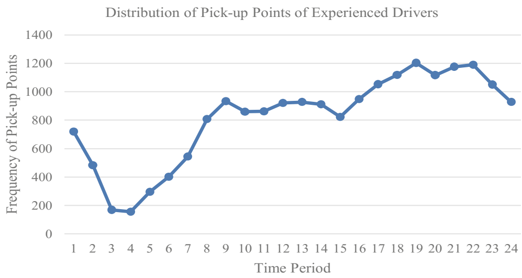


Fig. 4. Distribution of pick-up points of experienced drivers. The size of the timeslot is one hour, where 1 stands for 00:00–00:59, 2 stands for 01:00–02:59, etc.

4.2 Calculation of Probability of Recommended Points

To generate the probability of each recommended point, we measure the number of taxis, which pick up passengers when passing by the cluster while unoccupied. After getting the clusters, we should obtain the temporal-spatial coverage of each cluster. For each point in each cluster, we get the distance to the center of the cluster, then obtain the average distance for each cluster. The temporal-spatial coverage defines as a circle with radius of the average distance.

Definition 4. The probability of finding a passenger for each cluster c at time period t can be estimated as:

$$\mathrm{P}(\mathrm{c}, \mathrm{t}) = \frac{|states(cruising \rightarrow occupied)|}{|states(cruising)|}$$

where |status(cruising)| denotes the number of cruising taxis which passed by cluster c at time period t, and |status(cruising→occupied)| is the number of these cruising taxis which passed by cluster c at time period t and changed their state from cruising to occupied.

Since the probability of picking-up is very sensitive to time, for time period t, we divide it into several small ones. Then we calculate |status(cruising→occupied)| and

|status(cruising)| for each cluster respectively, and finally get the probability of finding a passenger for each cluster c at time period t.

5 Optimal Route Recommendation

In this section, we introduce the technical details for searching a route with the shortest potential cruising distance, which is we consider the optimal route. Firstly, we will show the model for measuring the potential cruising distance of each candidate route. Next, a recommending algorithm to get the optimal route will be presented.

5.1 The Potential Cruising Distance

Assume that there is an empty taxi, now we recommend it to the next place C_1. The distance between the taxi and C_1 is D_1. The probability of picking-up at C_1 is P_1. So, the potential cruising distance is D_1/P_1. The potential cruising distance from a taxi to a potential passenger is calculated based on the probability of the recommended points.

Definition 5. If the current position of a taxi is PoTaxi, and follow the route R = {PoTaxi (C_0), C_1, C_2,…, C_n} at time period t. It may pick up passengers at C_1 with the probability P(C_1), or at C_2 with the probability $(1 - P(C_1))P(C_2)$. For each pick-up point C_i, the picking-up rate is following as:

$$\mathrm{P}(C_i|\mathrm{R},\mathrm{t}) = \begin{cases} P(C_i,t); & i = 1 \\ P(C_i,t)\prod_{j=1}^{i-1}\left(1 - P\left(C_j,t\right)\right). & i > 1 \end{cases}$$

In addition, we use D(C_j, C_{j+1}) to represent the driving distance between pick-up point C_j and pick-up point C_{j+1}. Thus,

Definition 6. The potential cruising distance function F can be defined as:

$$\mathrm{F} = \frac{\sum_{i=1}^{n-1}\left(P(C_i|R,t)\sum_{j=0}^{i-1}D\left(C_j,C_{j+1}\right)\right) + P(C_n|R,t)/\mathrm{p}(C_n,t)\sum_{j=0}^{n-1}D\left(C_j,C_{j+1}\right)}{\sum_{i=1}^{n}P(C_i|R,t)}$$

By observing the form of the potential cruising distance function, we can simplify this formula and re-write as:

$$\mathrm{F} = \frac{\sum_{i=1}^{n}\frac{P(C_i|R,t)D(C_{i-1},C_i)}{P(C_i,t)}}{\sum_{i=1}^{n}P(C_i|R,t)}$$

To clearly explain our potential cruising distance function, we illustrate it via an example. Figure 5 shows an example of a recommended cruising route PoTaxi → C_1 → C_2 with the corresponding probability {P_1, P_2}, driving distance {D_1, D_2}

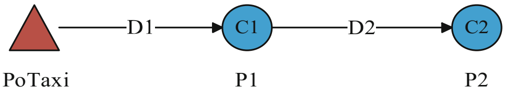


Fig. 5. An example of a recommended cruising route

respectively. The potential driving distance could be $(P_1D_1 + (1 - P_1)(D_1 + D_2))$ while it may have passengers of $(P_1 + (1 - P_1)P_2)$. Therefore, the potential cruising distance to a potential passenger will be:

$$\frac{P_1D_1 + (1 - P_1)(D_1 + D_2)}{P_1 + (1 - P_1)P_2}$$

5.2 Optimal Route Recommendation

In this subsection, we introduce the method for recommending the route with the shortest potential cruising distance to target taxi. Once the capacities of all the road are obtained based on the evaluation model proposed above, we can recommend a trajectory to a taxi given its current location and time.

Figure 6 shows the pseudo-code of the recommending algorithm. Given the current location (PoTaxi) and time (T) of the taxi; first of all, we can obtain the set of the recommended cluster nodes of current time (C_{set}), the set of the probability of the cluster nodes of the current time (P_{set}) which is corresponding to the Cset, and the driving distance matrix of the cluster nodes (D_{set}). Then, based on the above mentioned and the length of the route (k), all the candidate route can be gotten by the function GetCandidateRouteSet (). Next, for all candidate routes, the potential cruising distance

```
Algorithm GetRecommendingRoute (PoTaxi, T, k, Cset, Pset, Dset) {
    ShortestCruisingDistance = + ∞;
    CandidateRouteSet = GetCandidateRouteSet (k);
    for each route in CandidateRouteSet {
        PotentialCruisingDistance = GetPotentialCruisingDistance (PoTaxi,
T, Cset, Pset, Dset);
        if ShortestCruisingDistance > PotentialCruisingDistance
            ShortestCruisingDistance = PotentialCruisingDistance;
        end if
    The route with the minimum ShortestCruisingDistance is the optimal
recommending route;
    end for
}
```

Fig. 6. The recommending algorithm

should be calculated by the evaluation model we proposed, which is encapsulated in the function GetPotentialCruisingDistance (). Finally, the route with the minimum potential cruising distance will be the optimal recommending route.

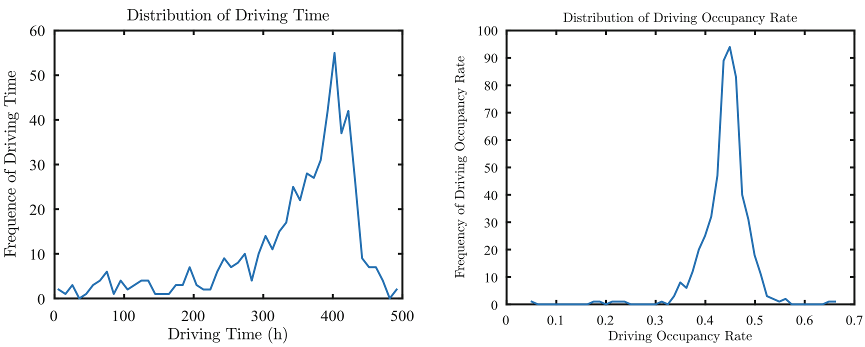
6 Experiments

In this section, to demonstrate the effectiveness of the proposed evaluation model and evaluate the performance of the proposed recommendation system, we have done extensive experiments on real-world data sets.

6.1 Experiment Data

In this paper, we train our system using the real-world data sets collected in the San Francisco Bay Area in 30 days, which provided by the Exploratorium-the museum of science, art and human perception through the cabspotting project. The mobility traces are the records of more than 500 taxis' driving states in consecutive time. Each record can be expressed as a tuple: (unique taxi ID, latitude, longitude, status, time stamp).

In the experiments, we obtain the experienced drivers by exploring the important properties of the drivers: driving time and driving occupancy rate. Figure 7 shows the distributions of the driving time and the driving occupancy rate. From Fig. 4 we can see that the picking-up events occur most frequently during the time period 18:00–19:00, and during 14:00–15:00, the gradient has a sharp change. So, we will focus on this two time period in the experiment. In total, 1203 pick-up points of experience drivers and 561573 points of all taxis are obtained during 18:00–19:00, and 822 pick-up points of experience drivers and 509362 points of all taxis are obtained during 14:00–15:00. All potential pick-up points are clustered into 10 clusters. Table 1 shows the information of the 10 temporal probabilistic pick-up points during 18:00–19:00. And Table 2 shows the information of the 10 temporal probabilistic pick-up points during 14:00–15:00. Note that the latitude and longitude represent the corresponding centroid of the cluster, the $P(C_i)$ represents the picking-up rate in the cluster C_i.



(a) (b)

Fig. 7. Distribution of: (a) Driving time; (b) Driving occupancy rate

Table 1. Description of the 10 clusters during 18:00–19:00

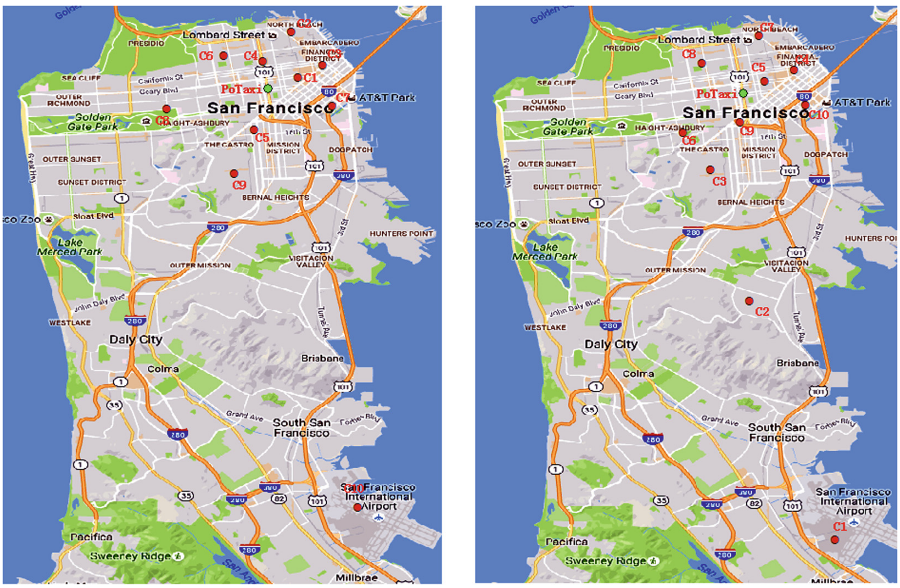
No.	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Latitude	37.78647	.80349	.79091	.79240	.76676	.79456	.77573	.77438	.75038	.43327
Longitude	−122.40942	.41193	.40027	.42260	.42574	.43721	.39663	.45873	.43327	.38711
P(Ci)	0.8795	0.7039	0.8888	0.8713	0.7856	0.7383	0.5831	0.6935	0.8377	0.4419

Table 2. Description of the 10 clusters during 14:00–15:00

No.	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Latitude	37.61475	.70423	.75358	.79087	.78653	.76744	.80369	.79360	.77130	.77758
Longitude	−122.38618	.41852	.43296	.40164	.41253	.44328	.41476	.43607	.42185	.39720
P(Ci)	0.4955	0.2967	0.8182	0.7764	0.7974	0.6861	0.6739	0.6852	0.5736	0.5866

6.2 Effectiveness

In this section, we compare the proposed model in the paper with the PTD function [1]. Here, we show the optimal driving routes recommended by the PTD function and our new model. Figure 8 shows the potential recommending points (the red points) within the time period 18:00–19:00 (a) and 14:00–15:00 (b) and the assumed position of the empty taxi to be recommended (the green point). Tables 3 and 4 shows the results of the recommendation during the time period 18:00–19:00 and 14:00–15:00 respectively.



(a) (b)

Fig. 8. Route recommendation. The red points denote the potential recommending points, and the green point denotes the target taxi (Color figure online)

Table 3. The results of the recommendation during the time period 18:00–19:00

Routes	Our Method	PTD
k = 3	PoTaxi→C_1→C_3→C_4	PoTaxi→C_1→C_3→C_7
k = 4	PoTaxi→C_1→C_3→C_4→C_6	PoTaxi→C_1→C_3→C_4→C_6
k = 5	PoTaxi→C_1→C_3→C_4→C_6→C_2	PoTaxi→C_1→C_3→C_4→C_6→C_2

Table 4. The results of the recommendation during the time period 14:00–15:00

Routes	Our Method	PTD
k = 3	PoTaxi→C_5→C_4→C_7	PoTaxi→C_5→C_4→C_{10}
k = 4	PoTaxi→C_5→C_4→C_7→C_8	PoTaxi→C_5→C_4→C_{10}→C_2
k = 5	PoTaxi→C_5→C_4→C_7→C_8→C_6	PoTaxi→C_5→C_4→C_7→C_8→C_2

In the experiment, during the time period 18:00–19:00, the driving distance between C3 and C4 is 2355 m, and the driving distance between C3 and C7 is 2643 m. As it can be seen from Table 3, when k = 3, the optimal route is PoTaxi → C1 → C3→ C7 generated by the PTD method while PoTaxi → C1 → C3 → C4 is generated by our proposed method during the time period 18:00–19:00. Obviously, our method works much better than the PTD. Because of the picking-up rate of C4 is higher than C7 and the distance between C3 and C4 is smaller than the distance between C3 and C7. And the potential driving distance of our method is about 1336 m while the potential driving distance of PTD is about 1350 m.

7 Conclusion and Future Work

In this paper, a novel model for evaluating the candidate route is proposed. Based on it, we design a recommendation system for taxi drivers to minimize their cruising driving distance before taking passengers regarding the time and location of the taxi. To be specific, we first put forward the temporal probabilistic recommending pick-up points by exploring the historical trajectory data of taxi drivers. Then we introduce the novel evaluation model, and based on it, we provide an algorithm to get the optimal route of different time and location for taxi drivers. As a result, we can use the model to rank each candidate route and get the optimal route for recommending.

Since the model is more complicated and loses some good properties like monotonic, future research will focus on the improvement of the efficiency of the recommendation algorithm. Moreover, choosing routes is like game, if all the taxis are recommended to the same route at the same time, the system is inefficient and fewer taxis will be a winner. So more efforts will be studying the taxi game strategy.

References

1. Ge, Y., Xiong, H., Tuzhilin, A., et al.: An energy-efficient mobile recommender system. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 899–908. ACM (2010)

2. Hu, H., Wu, Z., Mao, B., Zhuang, Y., Cao, J., Pan, J.: Pick-Up tree based route recommendation from taxi trajectories. In: Gao, H., Lim, L., Wang, W., Li, C., Chen, L. (eds.) WAIM 2012. LNCS, vol. 7418, pp. 471–483. Springer, Heidelberg (2012). doi:10.1007/978-3-642-32281-5_45
3. Hwang, R.H., Hsueh, Y.L., Chen, Y.T.: An effective taxi recommender system based on a spatio-temporal factor analysis model. Inf. Sci. **314**, 28–40 (2015)
4. Yang, W., Wang, X., Rahimi, S.M., Luo, J.: Recommending profitable taxi travel routes based on big taxi trajectories data. In: Cao, T., Lim, E.-P., Zhou, Z.-H., Ho, T.-B., Cheung, D., Motoda, H. (eds.) PAKDD 2015. LNCS (LNAI), vol. 9078, pp. 370–382. Springer, Cham (2015). doi:10.1007/978-3-319-18032-8_29
5. Letchner, J., Krumm, J., Horvitz, E.: Trip router with individualized preferences (trip): incorporating personalization into route planning. In: Proceedings of the National Conference on Artificial Intelligence, Menlo Park, CA, vol. 21, no. 2, p. 1795. MIT Press, Cambridge, AAAI Press, London, 1999, 2006
6. Malviya, N., Madden, S., Bhattacharya, A.: A continuous query system for dynamic route planning. In: 2011 IEEE 27th International Conference on Data Engineering, pp. 792–803. IEEE (2011)
7. Li, B., Zhang, D., Sun, L., et al.: Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 63–68. IEEE (2011)
8. Yuan, J., Zheng, Y., Zhang, L., et al.: Where to find my next passenger. In: Proceedings of the 13th International Conference on Ubiquitous Computing, pp. 109–118. ACM (2011)
9. Qu, M., Zhu, H., Liu, J., et al.: A cost-effective recommender system for taxi drivers. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 45–54. ACM (2014)
10. Li, X., Pan, G., Wu, Z., et al.: Prediction of urban human mobility using large-scale taxi traces and its applications. Front. Comput. Sci. **6**(1), 111–121 (2012)
11. Awasthi, A., Lechevallier, Y., Parent, M., et al.: Rule based prediction of fastest paths on urban networks. In: Proceedings of 2005 IEEE Intelligent Transportation Systems 2005, pp. 978–983. IEEE (2005)
12. Gonzalez, H., Han, J., Li, X., et al.: Adaptive fastest path computation on a road network: a traffic mining approach. In: Proceedings of the 33rd International Conference on Very Large Data Bases, pp. 794–805. VLDB Endowment (2007)
13. Ma, S., Zheng, Y., Wolfson, O.: T-share: a large-scale dynamic taxi ridesharing service. In: 2013 IEEE 29th International Conference on Data Engineering (ICDE), pp. 410–421. IEEE (2013)
14. Ma, S., Zheng, Y., Wolfson, O.: Real-time city-scale taxi ridesharing. IEEE Trans. Knowl. Data Eng. **27**(7), 1782–1795 (2015)
15. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)
16. Trestian, I., Huguenin, K., Su, L., et al.: Understanding human movement semantics: a point of interest based approach. In: Proceedings of the 21st International Conference on World Wide Web, pp. 619–620. ACM (2012)
17. Karypis, G.: CLUTO. http://glaros.dtc.umn.edu/gkhome/views/cluto