

# Carpool Algorithm Based on Similarity Measure of Partition Trajectory Line

Chengde Zhang<sup>1\*</sup>, Geng Deng<sup>1</sup>, Wu Bo<sup>1</sup>, Jun Guo<sup>2</sup>, Yi-peng SU<sup>1</sup>, Sheng Wang<sup>1</sup>, Yu-xiao Zhang<sup>1</sup>

<sup>1</sup>School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

<sup>2</sup>School of foreign language, Zhongnan University of Economics and Law, Wuhan 430073, China

**Abstract:** Rapid increase of private cars leading to serious road congestion and air pollution problems, carpool has become an important choice for people to travel. In order to improve the service quality of carpool, trajectory matching is becoming a hot topic. However, traditional path matching algorithm based on Hausdorff distance has exposed two problems: ① Ignoring waiting time of general users; ② The direct calculation of the Hausdorff distance of the entire path cannot reflect the impact of special sections on the matching metric. To solve the above mentioned problems, this paper proposed an improved algorithm: ① Hausdorff Distance is combined with turning point; ② Turning point is taken as an effect way to segment trajectories. Compared to the classical algorithm MCT (Minimum completion time online mode scheduling algorithm), proposed algorithm get better performance.

**Keywords:** Partition Trajectory Line; Carpooling; Hausdorff Distance.

## 1 Introduction

With the rise of Chinese economy, there are more cars. This lead to road congestion and automobile exhaust pollution problems. Therefore car-sharing services on car-hailing platforms are becoming popular, which making full use of private car resources. Carpooling (Car-sharing) is the sharing of car journeys by persons with similar travel needs so that more than one person travels in a car. Carpooling services on car-hailing platforms using advanced GPS positioning technology connect drivers and passengers through smartphone application platforms so that the driver can carry more than one passenger at a time. Car-sharing service has many advantages. ① Carpooling services connect private car drivers and passengers so that passengers can travel more convenient and idle car resources can be used more efficiently. ② When paying for the car-sharing service, passengers just need to pay less than the cost of taking taxis which reduce each passenger's travel costs. Simultaneously, drivers can carry more than one passenger at a time that increase the drivers' income. Therefore, higher load factor rate, less empty loading rate and higher cars usage rate is helpful to alleviate urban traffic congestion. ③ As the legalization and standardization of carpooling, passengers can get effective legal protection when carpooling, which reduce passengers' travel security risks. In summary, Car-sharing services embodies a kind of travel mode in sharing economy, it is helpful to improve the utilization of transport resources and it has an important significance for relieving urban traffic congestion, reducing environment pollution and accelerating the urbanization process[1]. Therefore, Car-sharing services cause a wide social attention in China and make the carpooling problems becomes a new research hotspot [2].

However, passengers and drivers often face two problems: First, passengers need to take a detour distance and waste addition time. Even sometimes two passengers' destination are opposite. Second, when facing numerous carpooling requests, drivers often have difficulties to select an optimal carrying solution for both passengers and drivers [3]. Since the lack of accurate and effective car-sharing system, the whole process of carpooling can't run efficiently, which cause unnecessary loss for both passengers and drivers [4]. The more reasonable and effective type among them is carpooling model based on Hausdorff distance according to the result of the experiments [5]. It is found that the Hausdorff distance can calculate the distance between two points sets well, thus can be used to calculate the distance between two trajectories [6]. Therefore, according to three step: ① Extract path turning points, ② Divide path by trajectory line, ③ Building a Carpool Algorithm Based on Similarity Measure of Partition Trajectory Line, the key problem is Hausdorff Distance of ordered point set .

---

\*Corresponding author.

E-mail address: chengdezhang@zuel.edu.cn (Chengde Zhang).

This paper is organized as follows: Section 2 gives a brief overview of related work, Section 3 gives an introduction and analysis of the details of the proposed method, Section 4 of empirical study and its analysis. Section 5 concludes our work.

## 2 Related Work

The key problem of carpooling is the matching of multiple passengers' routes. The static carpooling problem is typically modled as VRPPDTW problem (vehicle routing problem with pickup and deliveries and time windows) using heuristic methods to allocate [7]. Due to the lack of effective information technology and communication technology support, the research of automatically generating request matching list met the bottleneck. However, with the development of intelligent transportation system ITS, geographic information system GIS, global positioning system GPS and mobile communication technology, irregular dynamic personal carpooling requests can be respond in time [8]. The dispatch center can get driver's position and movement track through the ITS technology, and can obtain the position and the demand of mobile users in real time [9]. In reference [10], a multi-agent system based on path planning and unknown environment optimization technology is proposed, which is successful in learning path planning. Reference [11] solve the problem using dynamic taxi pooling scheduling algorithm based on Sufferage, this method only considers the minimum execution time MET (Minimum Execution Time) but does not take the fitting degree between the multiple starting point and the multi terminal ("many-to-many") path into account. In order to solve the Multi-start to multi-end problem, Shao Zengzhen [12] proposed matching degree cluster to adjust the vehicle route to divide more passengers into a car, and then use the prior cluster to arrange the order of passengers' getting on and off the car, finally using move in and out strategy to further optimize the carpooling program. Because Hausdorff distance can be used to calculate the distance between two point sets, it is also used to calculate the distance between to tracks. Xiao Qiang et al. [13] use Hausdorff distance to quantify the matching degree of two tracks in reference [2]. Atev et al. [14] proposed a dual spectral clustering algorithm based on Hausdorff distance and successfully applied it to the vehicle trajectory analysis. Reference [15] use the matching degree of clustering method to get the optimal matching of carpooling demand which proposed in the distribution algorithm based on improved Hausdorff distance [14]. Reference [16] proposed a time constrained Hausdorff distance based on the space and time trajectory similarity measure, using the sliding window to improve the algorithm in the reference [15].

These researches all used the Hausdorff distance algorithm to solve the path matching problem. However, there are still the following challenges in combination with the Hausdorff distance and carpooling problem:

①The turning points defined by Hausdorff distance is disordered, but the turning point sequence of passenger's route is orderly. How to order the turning point sequence needs further research. ②The time trajectory and space trajectory of passenger route are corresponding to each other, so how to define and analyze the time constraint of Hausdorff distance is a difficult point; ③Currently, the widely used Hausdorff distance carpooling algorithms have large enough error and need to be further improved.

In response to these challenges, we propose a Carpool Algorithm Based on Similarity Measure of partition trajectory line. The Hausdorff Distance of sub-trajectory between two paths can be analyzed dynamically by means of turning points and trajectory line. On this account, we judge the fitting degree of the target trajectory, and select the best matching scheme according to the fitting degree rank.

## 3 Similar Trajectory Matching Model

### 3.1 Pretreatment

Due to the high occupancy vehicle track, the complex and large amount of calculation process is not conducive to carpool track matching analysis and calculation. So we utilize turning point thought whose decision is related to satellite positioning system for vehicle location direction angle. When their direction angle changes, the vehicle is calculated adjacent points in the direction of the angle difference.

$$\Delta\delta = \left| GPS_{Direction_i} - GPS_{Direction_j} \right| \quad (1)$$

where,  $GPS_{Direction_i}$  is the direction angle of point  $i$  in the GPS satellite position system. If  $75^\circ < \Delta\delta < 105^\circ$  or  $165^\circ < \Delta\delta <$

195°, this point is regarded as turning point. By extracting all the pre-given track and the initial point and the end, according to the time ascending form turning point sequence, thus to extract, refactoring, simplifying the vehicle track, compressed them into turning point sequence with time node  $\{(L_i, K_i)\}$ .

### 3.2 Trajectory Similarity Matching

#### 3.2.1 Hausdorff distance

Traditional Hausdorff distance is used to measure the similarity between the two unordered collections; it also can be used as a form of definition of the distance between the two sets. Suppose there are two sets of collections  $A = \{a_1, a_2, \dots, a_n\}$  and  $B = \{b_1, b_2, \dots, b_n\}$ , these two sets of Hausdorff distance can be defined as:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (1)$$

where,  $\forall a_i, a_i \in A, \forall b_j, b_j \in B$  makes that  $\text{dist}(a_i, b_j) < H(A, B)$ , as:

$$h(A, B) = \max_{a_i \in A} (\min_{b_j \in B} \|a_i - b_m\|) \quad (2)$$

$$h(B, A) = \max_{b_j \in B} (\min_{a_i \in A} \|b_m - a_n\|) \quad (3)$$

where,  $\|b_m - a_n\|$  is the distance paradigm between the point set A and B, that is each point in collection  $b_m$  to distance from point A set of nearest  $a_n$  distance  $|b_m - a_n|$ , A maximum distance as  $h(B, A)$ . Bidirectional Hausdorff distance is the bigger one between  $h(A, B)$  and  $h(B, A)$ , which measures the different maximum mismatch degree between two point sets.

#### 3.2.2 Similar trajectory matching based on turning points

In view of the traditional carpool matching problem, we propose a similar trajectory matching algorithm which is based on turning point line with smaller complexity of trajectory, fewer detours time and higher similar trajectory compatibility scheme. This algorithm is designed to solve the problem of time constraints and matching route quality on the traditional carpool-matching algorithm, and the specific algorithm is shown in table 1:

**Table 1. Algorithm**

---

Algorithm: similar trajectory compatibility based on key line

---

Input: the child point sequence of trajectory A as:  $Tr_A = \{(L_{A_L}, K_{A_L}), (L_{A_R}, K_{A_R})\}$ , the child point sequence of trajectory B as:

$Tr_B = \{(L_{B_L}, K_{B_L}), (L_{B_R}, K_{B_R})\}$ , trajectory Line interval length L

Output: The total length of similar sub-trajectory  $L_{sum}$ , the compatibility between two sub-trajectories  $\xi_k$

---

(1): Initialization:  $A_L = A_1, A_R = A_2, B_L = B_1, B_R = B_2$ ;

(2): **While**  $(\min\{|A_L A_R|, |B_L B_R|\}) \neq 0$  **do**

(3):  $L = \min\{|A_L A_R|, |B_L B_R|\}$ ;

(4):  $H(Tr_A, Tr_B) \max(h(Tr_A, Tr_B), h(Tr_B, Tr_A))$ ;

(5): **if**  $H(Tr_A, Tr_B) < h_\alpha$  **then**

(6):  $L_{sum} = L_1 + L_{sum}$ ;

(7): **Output**  $L_{sum}$ ;

(8): **End if**

(9):  $A_L = A_R, A_R = \{A \in \min\{A_R A_{Tnext}, B_R B_{Tnext}\}\}$ ;

(10):  $B_L = B_R, B_R = \{B \in \min\{A_R A_{Tnext}, B_R B_{Tnext}\}\}$ ;

(11): **End while**

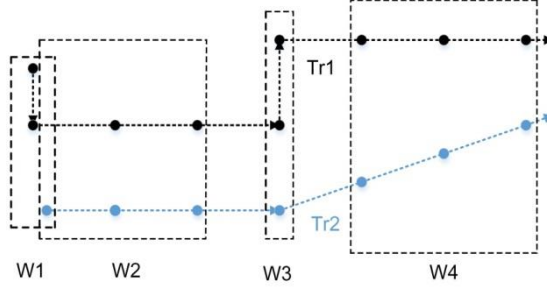
(12):  $\xi_k = \frac{L_{sum}}{\min\{\sum_{i=1}^{n-1} |A_i A_{i+1}|, \sum_{j=1}^{m-1} |B_j B_{j+1}|\}} - \sum_i^k f(\epsilon_i)$ ;

(13): **Return**  $\xi_k$

---

Turning point trajectory sequences are extracted based on time constraints. Turning point sequence of A can be described as

$Tr_A \{(L_i, K_i)\}$  and B's turning point sequence is described as  $Tr_B \{(L_m, K_m)\}$ . So the trajectory is compressed to the turning point sequence by the time sequence. Both infinitesimal method and turning point trajectory are used to divide the turning points' sequence, to improve the accuracy of Hausdorff distance measurement. As for the choice about key points, we select several points between in the start-stop turning point, which are interval  $\beta m$  and include start-stop turning point, to avoid the measurement error caused by the too much distance between in the turning points of Hausdorff distance, As for every key point of turning point sequence  $Tr_A$  and  $Tr_B$ , we draw a latitude direction perpendicular about this point, and the perpendicular is the trajectory line.



**Figure 1. Line molecular path diagram**

The segment between the trajectory lines is approximate to a point, when the direction along longitude turned to along latitude, which apparently mismatch with a track, therefore proves the reasonability of dividing the trajectory using lines along the latitude.

$h_\alpha$  represents the critical threshold of the improved Hausdorff distance similarity measurement, which is used to measure the maximum fault tolerance for track similarity compatibility.

$A \in \min\{A_R A_{Tnext}, B_R B_{Tnext}\}$  represents the terminal point of shorter track between  $A_R A_{Tnext}$  and  $B_R B_{Tnext}$ ,  $A_{Tnext}$  and  $B_{Tnext}$  represent the next turning point of track A and track B respectively.

The turning point sequences, A and B, are divided into several turning sub-trajectory. Through the turning point sub-trajectory formed among each line to process the similarity matching, thus all turning points of the two trajectories involved in the matching traversal calculation. Gaining the total length of all similar sub-trajectory, by comparing the proportion of the shorter matching trajectory length with the critical threshold of the improved Hausdorff distance similarity measurement to determine whether the similarity between the two tracks.

Calculate the sub-trajectory length meeting the conditions, and the compatibility between similar tracks as:

$$\xi = \frac{L_{sum}}{\min\{\sum_{i=1}^{n-1} |A_i A_{i+1}|, \sum_{j=1}^{m-1} |B_j B_{j+1}|\}} - \sum_{i=1}^k f(\epsilon_i) \quad (4)$$

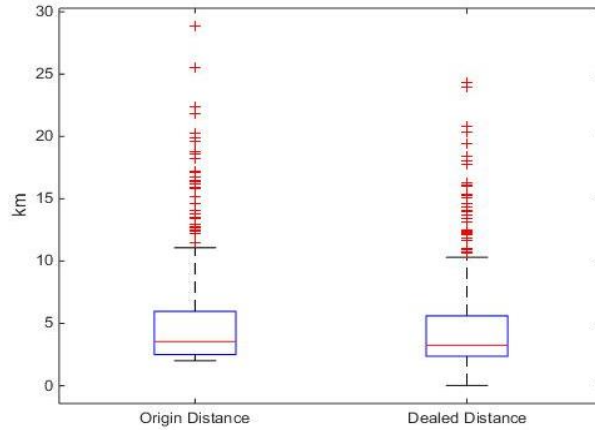
where  $L_{sum}$  represents total length of similar sub-trajectory,  $\min\{\sum_{i=1}^{n-1} |A_i A_{i+1}|, \sum_{j=1}^{m-1} |B_j B_{j+1}|\}$  represents minimum length of track A and track B,  $\sum_{i=1}^k f(\epsilon_i)$  represents influence of Hausdorff distance between each track. According to sorting of track similarity compatibility  $\xi$  of all carpooling schemes,  $\max\{\xi_k\}$  corresponds the scheme k is the real-time optimum carpooling scheme.

## 4 Experiment

### 4.1 Data Source

By randomly selecting 670 traffic routes of National Road, The Mountain Road, South Lake Road and Xiongchu Road in administrative road map of Wuhan City, Hubei Province, China, we simulate carpooling demand of passengers in the same search matching cycle. Then we get optimal carpool scheme through track similarity matching and track processing. In order to verify the reasonability and necessity of the turning point, the tracks are compared in Figure 2.

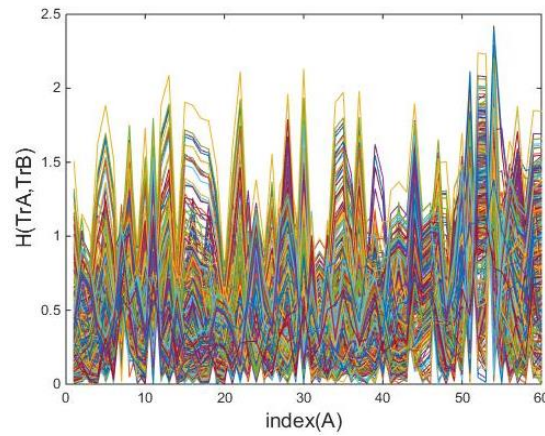
Although the calculation of the turning point will pay a price, the overall running time shortened greatly ensure the timeliness of carpooling matching scheduling and reduce the overhead and burden of server computing.



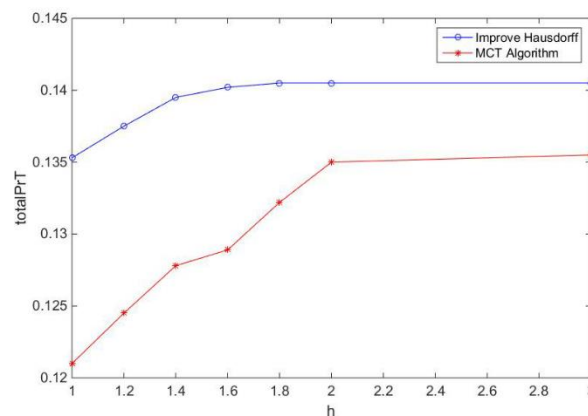
**Figure 2. The boxplot of Original distance and processed distance**

#### 4.2 Trajectory Matching

After a preliminary screening, we matched the track of the rest of routes. Put the travelling turning contrail sequences of passengers who need carpool into set B and put the travelling turning contrail sequences of passengers who are on board into set A. Match every route in B with every route in A, and then calculate the improved Hausdorff distance. In figure 3, a route set was labeled on the Y-axis, and the Hausdorff distance from a certain route in B to a certain route in A was labeled on X-axis. Every curve stands for a B route.

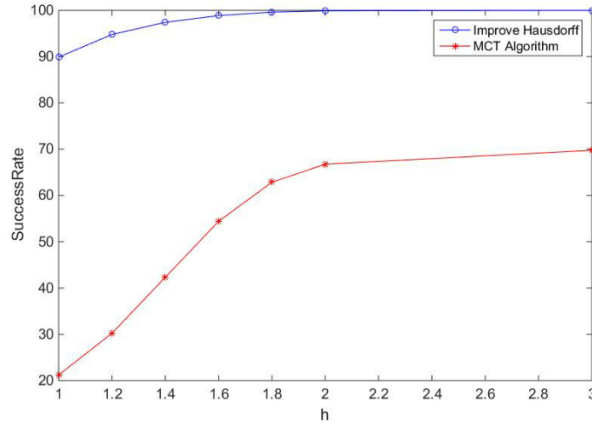


**Figure 3. Processed Hausdorff's track matching**



**Figure 4. Passengers' average waiting time**

In figure 4, h-axis is a threshold for similarity measurement, while Y-axis represents the average of passengers' waiting time. We can infer that the red curve is lower than the blue one overall. It indicates that the waiting time for user of Improve Hausdorff is higher than MCT.



**Figure 5. Carpool success rate**

In figure 5, Y-axis represents the percentage of the successfully matched number of A routes and B routes. Thus it indicates that the number of successfully matched routes of MCT algorithm is lower than that in Improve Hausdorff. At the condition of  $h_{\alpha}=3$ , Success Rate reach 100%, with 30% improvement for MCT.

**Table 2. Experiments**

Similarity degree	$h_{\alpha}=1$		$h_{\alpha}=1.2$		$h_{\alpha}=1.4$		$h_{\alpha}=1.6$		$h_{\alpha}=1.8$		$h_{\alpha}=2$		$h_{\alpha}=3$	
Sequence_B	total_prT	Success_rate	total_prT	Success_rate	total_prT	Success_rate	total_prT	Success_rate	total_prT	Success_rate	total_prT	Success_rate	total_prT	Success_rate
Improve Hausdorff	0.1353h	89.91%	0.1375h	94.80%	0.1395h	97.42%	0.1402h	98.88%	0.1405h	99.63%	0.1405h	99.90%	0.1405h	99.99%
MCT_algorithm	0.121h	21.22%	0.1245h	30.24%	0.1278%	42.34%	0.1289h	54.43%	0.1322h	62.87%	0.1355h	69.77%	0.135h	69.77%

More details about experiments are shown in Table 2, the average waiting time is calculated as follows:

$$total_{prT} = \frac{\sum(timeDifference)}{\sum(Direction\_sequence)} \quad (5)$$

where *timeDifference* is the wait time for passengers to carpool. *sum(timeDifference)* is the total number of carpooling passengers. "Success rate" is the matching success rate of a route which the passengers in demand have in track B and all the routes in track A. it is also the ratio of quantity that the car-pooling route satisfy the critical threshold of similarity measurement to total quantity. The calculation formula is as follows:

$$Success_{rate} = \frac{\sum(number)}{\sum(Direction\_sequenceB) * \sum(Direction\_sequenceA)} \quad (6)$$

where *sum(number)* is the number of carpooling routes which meet the critical threshold of detour similarity measure. *sum(Direction\\_sequenceB) \* sum(Direction\\_sequenceA)* is the product of the number of A routes and B routes. As every B route should match all A routes. There are 610 B routes in all. So the denominator is the product of these two.  $h_{\alpha}$  stands for threshold of similarity measure, which is used to evaluate the performance of the matching routes. The larger  $h_{\alpha}$  is, the worse the matching routes performs. The smaller  $h_{\alpha}$  is, the better the matching routes performs and the less time passengers wait. If  $h_{\alpha}=1$ , the passengers' average waiting time of which time bound is reduced by 67.8% compared with carpooling scheme without time

constraints. With restriction of  $h_\alpha$  broadening, success rate of matching routes with time bound is higher and higher. If  $h_\alpha=3$ , the overall success rate run up to 84.06%. Due to the relaxation of the critical similarity measurement threshold, the average waiting time has increased. It means that passengers' waiting time for carpooling is greatly shortened after adding time constraints. And the success rate is still maintained at a high level. At the same time, compared with MCT, if  $h_\alpha=3$  the success rate of improved Hausdorff carpool matching algorithm increases 14.29% compared with the MCT, shortening 0.0405h in time, which is superior to MCT.

## 5 Summary

In carpool problem, carpool matching algorithm is the key point which face both speed and quality at the same time. Multiple paths are transformed into turning point sequence. Improved Hausdorff distance has more advantages than the traditional carpool matching algorithm MCT on the speed of dealing with data and the success ratio of matching. In the future, we will focus on GPS and road description.

## 6 Acknowledgment

This work was supported by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 16YJC860026).

## 7 Reference

- [1] Xiaoting Wang, Christopher Leckie, Hairuo Xie, Tharshan Vaithianathan. Discovering the Impact of Urban Traffic Interventions Using Contrast Mining on Vehicle Trajectory Data [J]. Lecture Notes in Computer Science. April 2015 Volume 9077 pp 486-497
- [2] Álvaro Ribeiro, Daniel Castro Silva, Pedro Henriques Abreu . MoCaS: Mobile Carpooling System[J] Advances in Intelligent Systems and Computing. 2015 Volume 353 pp 913-922
- [3] HUANG Y, POWELL J W. Detecting regions of disequilibrium in taxi services under uncertainty[C] //SIGSPATIAL'12: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. New York: ACM, 2012:139-148.
- [4] Jia Yao, Anthony Chen, Seungkyu Ryu, Feng Shi A general unconstrained optimization formulation for the combined distribution and assignment problem[J]. Transportation Research Part B . 2014
- [5] Cao Yanyan, Cui Zhiming, Wu Jian, Sun Yong, et al. An improved Hausdorff distance and spectral clustering vehicle trajectory pattern learning approach [J]. Computer Applications and Software. 2012
- [6] Zhang, X., Ouyang, M. & Zhang, X. Small scale crowd behavior classification by Euclidean distance variation-weighted network [J]. Multimedia Tools and Applications, October 2016, Volume 75, Issue 19, pp 11945–11960
- [7] Bulent catay. A new saving-based ant algorithm for the Vehicle Routing Problem with Simultaneous Pickup and Delivery[J]. Expert Systems with Applications, 2010, 37:6809-6817.
- [8] TANG L L, ZHENG W B, WANG Z Q, et al. Space time analysis on the pick-up and drop off of taxi passengers based on GPS big data [J]. Journal of Geo- Information Science, 2015, 17 (10) : 1179-1186.
- [9] FENG Tian. Dynamic Taxipooling Scheduling Algorithm Based on Sufferage [J], Computer Knowledge and Technology, 2011, 28(7):7019-7023
- [10] Zhang, X., Chen, G., Han, Y. et al. Modeling and analysis of bus weighted complex network in Qingdao city based on dynamic travel time [J]. Multimedia Tools and Applications, December 2016, Volume 75, Issue 24, pp 17553–17572
- [11] Zafar, K. & Baig, A.R. Optimization of route planning and exploration using multi agent system [J]. Multimedia Tools and Applications, January 2012, Volume 56, Issue 2, pp 245–265
- [12] SHAO Zeng-Zhen, WANG Hong-Guo, LIU Hong, et al. Single Carpooling Problem Based on Matching Degree Clustering Algorithm [J]. Journal of Software, 2012, 23(2):204-212.
- [13] XIAO Q, HE R C, ZHANG W, et al. Algorithm research of taxi carpooling based on fuzzy clustering and fuzzy recognition[J]. Journal of Transportation Systems Engineering and Information Technology, 2014, 14 (5): 119-125.
- [14] Atev S, Miller G, Papanikolopoulos N P. Clustering of vehicle trajectories[J]. IEEE Trans on Intelligent Transportation Systems, 2010, 11(3): 647-657.
- [15] LIU Chun, TAN Mengxi, SHAO Xiongkai, et al. Carpooling algorithm research based on location data mining [J]. Computer Engineering and

d Applications, 2015-12-31