Carpool Algorithm Based on Similarity Measure of Partition Trajectory Line

Cheng-De ZHANG\textsuperscript{1, a, *}, Geng DENG\textsuperscript{1, b}, Wu BO\textsuperscript{1, c}, Jun GUO\textsuperscript{2, d}, Yi-Peng SU\textsuperscript{1, e}, Sheng WANG\textsuperscript{1, f} and Yu-Xiao ZHANG\textsuperscript{1, g}

\textsuperscript{1}School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China
\textsuperscript{2}School of foreign language, Zhongnan University of Economics and Law, Wuhan 430073, China

\textsuperscript{a}chengdezhang@zuel.edu.cn, \textsuperscript{b}DG_Nikumata@163.com, \textsuperscript{c}15827403235@163.com, \textsuperscript{d}Guojun0029@163.com, \textsuperscript{e}suzi_liuyun@163.com, \textsuperscript{f}1819959027@qq.com, \textsuperscript{g}15827018523@163.com

*Corresponding author

Keywords: Partition trajectory line, Carpooling, Hausdorff distance.

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: (2) 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.

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 is a sharing of car journeys by persons with similar travel needs. Carpooling services on car-hailing platforms using GPS positioning technology connect drivers and passengers through APP. 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) Passengers pay less than taxis. Car-sharing services is helpful to improve the utilization of transport resources, 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 hot topic [2].

However, passengers and drivers often face two problems: First, passengers need to take a detour distance and waste addition time; second, drivers are difficult to select an optimal carrying solution for both passengers and drivers [3]. The more reasonable and effective type among them is carpooling model based on Hausdorff distance according to the result of the experiments [4]. 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 [5]. 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.

However, the 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. In response to these challenges, we propose a Carpool Algorithm Based on Similarity
Measure of partition trajectory line. We judge the fitting degree of the target trajectory, and select the best matching scheme according to the fitting degree rank.

**Similar Trajectory Matching Model**

**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.

\[
\Delta \delta = \left| \text{GPS}_{\text{Direction}}_{i} - \text{GPS}_{\text{Direction}}_{j} \right| \tag{1}
\]

where, \(\text{GPS}_{\text{Direction}}_{i}\) is the direction angle of point \(i\) in the GPS satellite position system. If \(75^\circ < \Delta \delta < 105^\circ\) or \(165^\circ < \Delta \delta < 195^\circ\), 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 \(\{L_i, K_i\}\).

**Trajectory Similarity Matching**

**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 \(A = \{a_1, a_2, \ldots, a_n\}\) and \(B = \{b_1, b_2, \ldots, b_n\}\), these two sets of Hausdorff distance can be defined as:

\[
H(A, B) = \max(h(A, B), h(B, A)) \tag{2}
\]

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:

\[
h(A, B) = \max_{a_i \in A} \left( \min_{b_j \in B} \|a_n - b_m\| \right) \tag{3}
\]

\[
h(B, A) = \max_{b_j \in B} \left( \min_{a_i \in A} \|b_m - a_n\| \right) \tag{4}
\]

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.

**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. This algorithm is designed to solve the problem of time constraints and matching route quality on the traditional carpool-matching algorithm, as shown in table 1:
**Algorithm: similar trajectory compatibility based on key line**

**Input:** the child point sequence of trajectory A as: \( T_{RA} = \{(L_{A_L}, K_{A_L}), (L_{A_R}, K_{A_R})\} \), the child point sequence of trajectory B as:
\( T_{RB} = \{(L_{B_L}, K_{B_L}), (L_{B_R}, K_{B_R})\} \)

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

1. **Initialization:** \( A_L = A_1 = A_2, B_L = B_1, B_R = B_2 \)
2. While \( \min\{|A_LA_R|, |B_LB_R|\} \neq 0 \) do
3. \( L = \min\{|A_LA_R|, |B_LB_R|\} \)
4. \( H(T_{RA}, T_{RB}) = \max(h(T_{RA}, T_{RB}), h(T_{RB}, T_{RA})) \)
5. If \( H(T_{RA}, T_{RB}) < h_{\infty} \) then
6. \( L_{sum} = L_1 + L_{sum} \)
7. Output \( L_{sum} \)
8. End if
9. \( A_L = A_R \), \( A_R = \{A \in \min\{A_RA_{Tnext}, B_RB_{Tnext}\}\} \)
10. \( B_L = B_R \), \( B_R = \{B \in \min\{A_RA_{Tnext}, B_RB_{Tnext}\}\} \)
11. **End while**
12. \( \xi_k = \frac{L_{sum}}{min[L^a_i|A_iA_{i+1}, L^b_j|B_jB_{j+1}]} - \sum f(\epsilon_i) \)
13. **Return** \( \xi_k \)

**Turning point trajectory sequences are extracted based on time constraints.** Turning point sequence of A can be described as \( T_{RA} = \{(L_i, K_i)\} \) and B’s turning point sequence is described as \( T_{RB} = \{(L_i, K_i)\} \). 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 \( T_{RA} \) and \( T_{RB} \), we draw a latitude direction perpendicular about this point, and the perpendicular is the trajectory line.

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.

\( h_{\infty} \) represents the critical threshold of the improved Hausdorff distance similarity measurement, which is used to measure the maximum fault tolerance for track similarity compatibility.

\( A \in \min\{A_RA_{Tnext}, B_RB_{Tnext}\} \) represents the terminal point of shorter track between \( A_RA_{Tnext} \) and \( B_RB_{Tnext} \), \( A_{Tnext} \) 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:

$$\xi = \frac{L_{\text{sum}}}{\min\{\sum_{i=1}^{n-1}|A_i|, \sum_{j=1}^{m-1}|B_j|\}} - \sum_{i} f(\epsilon_i)$$  \hspace{1cm} (5)

where $L_{\text{sum}}$ represents total length of similar sub-trajectory, $\min\{\sum_{i=1}^{n-1}|A_i|, \sum_{j=1}^{m-1}|B_j|\}$ represents minimum length of track A and track B, $\sum_{i} f(\epsilon_i)$ 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.

**Experiment**

**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. 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.

**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 fig. 2, 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.

In fig 3, 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.

In fig 4, 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 Improved Hausdorff. At the condition of $h_{\epsilon}=3$, Success Rate reach 100%, with 30% improvement.
Figure 3. Passengers’ average waiting time. Figure 4. Passengers’ average waiting time.

Table 2. Experiments.

<table>
<thead>
<tr>
<th>Similarity_degree</th>
<th>$h_\alpha=1$</th>
<th>$h_\alpha=1.2$</th>
<th>$h_\alpha=1.4$</th>
<th>$h_\alpha=1.6$</th>
<th>$h_\alpha=1.8$</th>
<th>$h_\alpha=2$</th>
<th>$h_\alpha=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequec_e_B</td>
<td>total $L_{p}$</td>
<td>$L_{p}$ Success_rate</td>
<td>total $L_{p}$</td>
<td>$L_{p}$ Success_rate</td>
<td>total $L_{p}$</td>
<td>$L_{p}$ Success_rate</td>
<td>total $L_{p}$</td>
</tr>
<tr>
<td>Improve</td>
<td>0.1</td>
<td>89.9</td>
<td>0.1</td>
<td>94.8</td>
<td>0.1</td>
<td>97.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Hausdorff</td>
<td>0.1</td>
<td>353</td>
<td>0.1</td>
<td>375</td>
<td>0.1</td>
<td>395</td>
<td>0.1</td>
</tr>
<tr>
<td>MCT_algorithm</td>
<td>0.1</td>
<td>21h</td>
<td>2%</td>
<td>245</td>
<td>4%</td>
<td>278</td>
<td>4%</td>
</tr>
</tbody>
</table>

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

$$\text{total}_{prT} = \frac{\text{sum}(\text{timeDifference})}{\text{sum}(\text{Direction_sequence})}$$  \hspace{1cm} (6)

where \text{timeDifference} is the wait time for passengers to carpool. \text{sum}(\text{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:

$$\text{Success}_{rate} = \frac{\text{sum}(\text{number})}{\text{sum}(\text{Direction_sequenceB}) \times \text{sum}(\text{Direction_sequenceA})}$$  \hspace{1cm} (7)

where \text{sum}(\text{number}) is the number of carpooling routes which meet the critical threshold of detour similarity measure. \text{sum}(\text{Direction_sequenceB}) \times \text{sum}(\text{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. 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. Compared with MCT, the success rate of improved Hausdorff carpool matching algorithm increases 14.29%.

**Summary**

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.

**Acknowledgment**

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

**Reference**


