DBSCAN-Based No-Load Road Detection Algorithm

Jiandong Shang[1, a], Yunpeng Yang[1, b], Runjie Liu[1, c], Yaohuan Yang[1, d], Panle Li[1, e], Huihui Hao[2, f]

[1]*School of the Zhengzhou's Smarter City Zhengzhou University Henan 450000, China*
[2]*Institute of Industrial Technology of Zhengzhou University Henan 450000, China*
[a]*shangjiandong@zzu.edu.cn,* [b] *yu_yangyunpeng@163.com*
[c] *ierjliu@zzu.edu.cn@163.com,* [d]*15837196021@163.com*
[e]*185979505@qq.com,* [f] *400554841@qq.com*

Keywords: DBSCAN-Based; No-Load Road; Detection algorithm

Abstract: Taxi plays an important role in the urban passenger transport system. Good taxi service quality will help increase the happiness of urban residents. However, the issue of no-load taxis has always been one of the most concerned issues for urban managers. This paper uses the improved DBSCAN method to carry on the cluster analysis to the taxi trajectory, prompts the taxi driver to avoid the no-load area, and achieves the purpose that to raise the taxi utilization factor. The improved DBSCAN method introduces two distance metrics, segment feature distance and dynamic space-time distortion distance. The feature distances of the segment include the vertical distance, horizontal distance, and angular distance, which can effectively measure the distance between trajectory segments. The dynamic space-time distortion distance introduces spatial factors on the basis of the dynamic distortion distance, and successfully overcomes the similarity error problem caused by different trajectory lengths. At the same time, using the GPS data set of Nanjing taxi to carry out experiments, the results show that the algorithm has higher accuracy and stability than the algorithms in the literature.

1. Introduction

With the continuous increase of urban population and traffic flow, traffic problems have become a major issue in modern large cities. For example, during the rush hour in Beijing, buses and subways will be overloaded. Passengers need to wait at least 30 minutes before they can wait for rental. The car [1], but at other times, the vast majority of taxis are either empty or parked somewhere waiting for passengers. According to incomplete statistics, the idle rate of taxis in most cities is more than 50%, and some cities are even more than 80% [2]. This leads to a huge waste of transportation resources and how to effectively reduce the no-load ratio of taxis. Therefore, improving the utilization of taxis is one of the research hotspots of intelligent transportation.

Taxi no-load is a very important issue in intelligent transportation. The reduction of taxi no-load rate can effectively increase the utilization rate of taxis, increase the economic income of drivers, and reduce the waste of resources. By tapping the potential spatio-temporal characteristics of mass taxi trajectory data, taxi drivers can be prompted to avoid certain road sections in time, thereby improving the operating efficiency of taxis. Based on the spatio-temporal data collected by GPS,

this paper uses the clustering algorithm based on linear density to cluster the vehicle trajectory data to find the no-load section of the taxi in the urban road based on the characteristics of the continuity of the taxi trajectory. Using the taxi's real data set, systematic experiments and analysis were performed to verify the effectiveness of the algorithm.

2. Related Work

Taxi utilization is related to many factors such as population density, taxi strategy, and taxi driver experience. The literature [3] uses the method of classification to mine the vehicle trajectory, and establishes an abnormal vehicle trajectory model based on a smart vehicle. Literature [4] proposed a Hausdorff distance similarity measure based on line segments to detect anomalies in moving objects. The literature [5, 6, 7, 8] detects and analyzes hotspots of on-road and off-road passengers in road network space. On the one hand, it improves people's perception of urban spatial dynamics, and on the other hand, it provides guidance for the vehicle to find passengers as soon as possible. The literature [5] uses the L1-Norm SVM algorithm to extract the important feature patterns of the car hire-seeking behavior, and combines the triples of time, location, and strategy features to compare the tactics of cruising or in situ waiting. A large number of vehicle GPS data sets find efficient and inefficient passenger search strategies. This result provides an important theoretical basis for vehicle drivers to better find passengers. The paper [9] et al. used the time-sequence density clustering algorithm to propose a plot of road segmentation. Literature [10] believes that the utilization rate of taxis is closely related to time and space. Therefore, a collaborative filtering algorithm based on space-time context is proposed. This algorithm combines the idea of context filtering algorithm and improves the traditional collaborative filtering algorithm. Literature [11] based on the information bottleneck method clusters the location of empty taxis of taxis, thus guiding the vehicles to avoid idling and gathering areas, effectively improving the utilization rate of taxis.

3. Feature-based Trail Avoidance Section Model

With the development of the urban economy, the road network is becoming more and more complicated, and the characteristic trajectories on different roads are often independent of each other $DSTW(Tr^i, Tr^j) = \infty$. If the conventional method is used for cluster analysis of the trajectory set, the data volume will be very large and affected by the feature trajectories on adjacent roads, and the clustering effect will be poor. For this reason, this paper categorizes feature trajectories and then constructs road-based data sets.

3.1 Trajectory Mapping and Classification

Trajectory mapping is to map the trajectory data to the actual road network and lay the foundation for trajectory classification. In order to improve the accuracy of the mapping, this article uses the API provided by the Baidu development platform to batch map latitude and longitude data to the actual road network. Baidu Maps provides a web map search service covering almost all cities in the country, and Baidu Maps also provides a series of related application programming interfaces through which developers can easily interact with Baidu Maps. The mapped float car data format is as follows:

Table1 Mapping Dataset

ID	Longitude	Latitude	Status	Date	Time	Road name
806401925795	118.8042	32.088	0	2010/9/2	18:52:20	Lake Road
806401925801	118.8042	32.08812	0	2010/9/2	19:05:20	Lake Road
…	…	…	…	…	…	…

Using a serial method to classify taxi data, the time complexity will be very high. Assuming that the trajectory of the entire city is classified, the time it takes cannot be tolerated. Therefore, this paper uses hadoop to classify the feature trajectories. Hadoop's parallel execution mechanism and built-in classification algorithm can greatly shorten the implementation time of classification algorithms. By designing reasonable (key, value) key-value pairs, you can achieve arbitrary classification. This article uses key=ID+ road name. The data after classification is as follows:

Table 2 Class

key		value				
ID	Lake Road	Longitude	Latitude	Status	Date	Time
806401925795	Lake Road	118.8042	32.088	0	2010/9/2	18:52:20
806401925801	Lake Road	118.8042	32.08812	0	2010/9/2	19:05:20
…	…	…	…	…	…	…

3.2 Extracting Feature Trajectories and Cluster Analysis

The set of road network is $R = \{r_1, r_2, \cdots r_k\}$, r_i ($i = 1, 2, \cdots k$) indicates the i road. In this paper, 24 hours are divided into m periods, for example: $N = \{n_1, n_2, \cdots n_m\}$, where n_i represents the i period of time. With the help of the previous trajectory classification, it is easy to obtain the inner n_i trajectory r_j dataset S_{ij}. Connect the same feature points ID in time series to S_{ij} form a feature track. The feature trajectory has positive significance for guiding the taxi driver to plan the route. The higher the no-load rate of a certain section of road is, the less likely the taxi is to be carried to the guest on this section of the road. Therefore, it is not easy for an empty car to cruise on this section. From the point of view of passengers, the easier it is for passengers to find a taxi on this section of the road, helping passengers to make reasonable arrangements for their own trip.

Traditional trajectory clustering algorithms often use metrics such as Manhattan distance, norm, and cosine to measure trajectory similarity. These metrics do not take into account the inconsistent track segments. Obviously, the situation of inconsistent track segments is very common. If this problem cannot be solved well, the accuracy of the clustering algorithm will be greatly affected. Therefore, this paper divides the trajectory similarity measure into two parts: segment feature distance and dynamic space distortion distance. The fragment feature distance integrates three distances such as vertical distance, angle distance, and horizontal distance, and proposes a new feature distance. Based on the feature distance of the segment, the feature distance matrix between the trajectories is constructed, and then the similarity between the trajectories is measured by the dynamic twist distance. The dynamic twist distance comprehensively considers both time and space factors, and can well measure the situation of trajectory spatiotemporal misalignment. After obtaining the trajectory similarity matrix, the DBSCAN algorithm is used to cluster the trajectories.

4. System Experiments and Analysis

This section uses the real data set of Nanjing City to evaluate the effectiveness of the algorithm. At the same time, it compares the efficiency of the algorithm in this paper with the algorithm in [9]. The core part of the algorithm is implemented in C++. The experimental platform is an I5 processor, 8G memory, and the operating system is a win7 system.

The evaluation of clustering effect adopts two methods compared with the algorithm in the literature [9] and CDbw validity index clustering. The higher CDbw effectiveness evaluation index value indicates better clustering effect.

4.1 Cluster Effect Test

In this paper, the clustering effect tests are performed on Data1, Data2, and Data3. The algorithm directly reads the preprocessed results into memory and then performs cluster analysis. This algorithm can find clusters of arbitrary shape and identify outliers. The clustering effect is shown in the figure. Red is the outlier and the circle is the boundary point. Figure 1 shows the clustering effect of the algorithm on Data1. The parameter is set to Eps=1773.1 and MinPts=4. Figure 3 For the clustering effect on Data2, the parameters are set to Eps=882.103, MinPts=5. From the clustering results, it can be found that there are more empty vehicles in those sections. Figure 5 shows the clustering effect of the algorithm on Data3. The parameters are set to Eps = 300.268 and MinPts = 4. The implementation results show that the algorithm can identify clusters and isolated points better. In addition, the algorithm of this paper is compared with the algorithm in [9]. Figure 2, Figure 4, and Figure 6 are the clustering effects of the algorithms in Data1, Data2, and Data3 in [9]. Compared with the results, the algorithm in this paper and the algorithm in [9] can get better clustering effect under the same parameters.
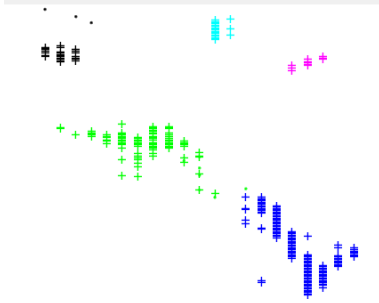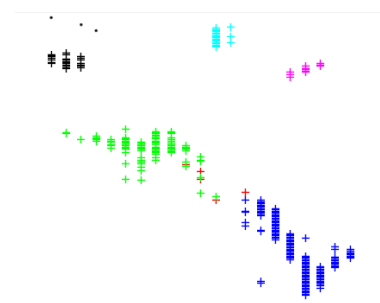

Figure 1 Algorithm Clustering Results　　　　Figure 2 Literature [9] Operational Results
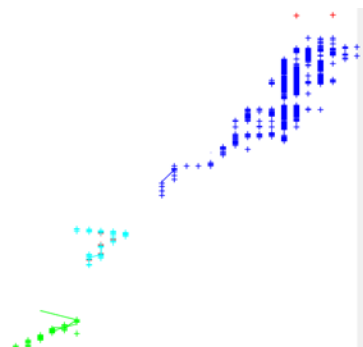
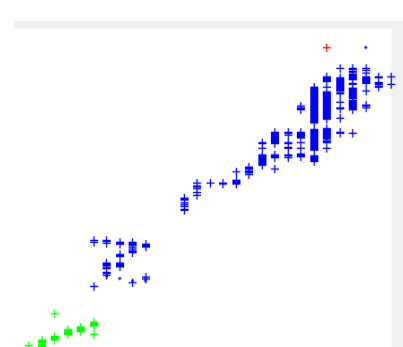

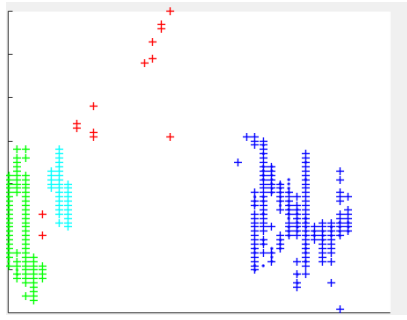Figure 3 Algorithm Clustering Results　　　Figure 4 Literature [9] Operational Results

Figure 5 Algorithm Clustering Results   Figure 6 Literature [9] Operational Results
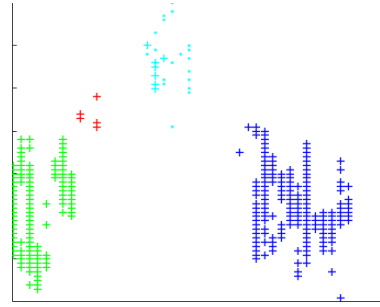
Table 3 Parameter settings

Dataset	Eps		MinPts		CDbw	
	Normal	promotion	Normal	promotion	Normal	Promotion
DataE1	300.596	1300.596	5	5	3167.92	7568.2
	882.103	882.103	4	4	5994.55	7898.56
DataE2	300.268	300.268	5	5	3513.51	4602.41
DataE3	300.596	1300.596	5	5	3167.92	7568.2

5. Conclusion

In view of the low utilization rate of urban taxis, this paper proposes a cluster analysis of urban empty car trajectories using the DBSCAN method based on on-board GPS data. In this paper, the vehicle trajectory is regarded as a sequence of points with spatiotemporal characteristics. By connecting these point sequences, the vehicle feature trajectory is formed. In order to improve the clustering effect of DBSCAN on the trajectory, this paper introduces the segment feature distance and dynamic spatiotemporal distortion distance. The feature distances of the segment include the vertical distance, horizontal distance, and angular distance, which can effectively measure the distance between trajectory segments. The dynamic space-time distortion distance introduces spatial factors on the basis of the dynamic distortion distance, and successfully overcomes the similarity error problem caused by different trajectory lengths. At the same time, the experiment was performed on the real GPS data set of the taxi in Nanjing. The results show that the algorithm has higher accuracy and stability than the algorithms in the literature. The algorithm can accurately identify the city's no-load aggregation area, which helps taxi drivers to avoid these empty areas in time and increase the utilization rate of city taxis.

References

[1] Jing W, Hu L, Shu L, et al. RPR: recommendation for passengers by roads based on cloud computing and taxis traces data[J]. Personal and Ubiquitous Computing, 2016, 20(3):337-347.

[2] Lee J. Traveling Pattern Analysis for the Design of Location-Dependent Contents Based on the Taxi Telematics System[J]. 2008.

[3] Song Yang. (2013). Vehicle abnormal track model. (Doctoral dissertation, Chang'an University).

[4] Jinyang Chen. (2012). Data mining of moving object trajectories. (Doctoral dissertation, Ningbo University).

[5] Li B, Zhang D, Sun L, et al. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset[C]// IEEE International Conference on Pervasive Computing and Communications Workshops. IEEE, 2011:63-68.

[6] Shen Y, Zhao L, Fan J. Analysis and Visualization for Hot Spot Based Route Recommendation Using Short-Dated Taxi GPS Traces[J]. Information, 2015, 6(2):134-151.

[7] Lee J, Shin I, Park G L. Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation[C]//

International Conference on Networked Computing and Advanced Information Management. IEEE, 2008:199-204.

[8] Tang Luliang, Zheng Wenbin, Wang Zhiqiang, et al. Spatio-temporal distribution of GPS trajectories from urban taxis. [J]. Journal of Geoinformatics, 2015, 17(10): 1179-1186.

[9] Jing Cheng, Jiajun Liu, Yong Gao. Temporal and Spatial Characteristics of Beijing Taxi Trips Based on Time Series Clustering Method[J]. Journal of Geoinformatics, 2016, 18(9):1227-1239.

[10] Ding Y, Liu S, Pu J, et al. HUNTS: A Trajectory Recommendation System for Effective and Efficient Hunting of Taxi Passengers[C]// IEEE, International Conference on Mobile Data Management. IEEE Computer Society, 2013:107-116.

[11] Yuan J, Zheng Y, Zhang C, et al. An Interactive-Voting Based Map Matching Algorithm[C]// Eleventh International Conference on Mobile Data Management. IEEE Computer Society, 2010:43-52.

[12] Ge Y, Xiong H, Tuzhilin A, et al. An energy-efficient mobile recommender system[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, Dc, Usa, July. DBLP, 2010:899-908.

[13] Yuan J, Zheng Y, Xie X, et al. Driving with knowledge from the physical world[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:316-324.