

## Research Article

# Clustering Algorithm for Urban Taxi Carpooling Vehicle Based on Data Field Energy

**Xiao Qiang** <sup>1,2</sup> and **Yao Shuang-Shuang**<sup>1</sup>

<sup>1</sup>*School of Economics and Management, Lanzhou Jiao Tong University, Lanzhou 730070, China*

<sup>2</sup>*School of Traffic and Transportation, Lanzhou Jiao Tong University, Lanzhou 730070, China*

Correspondence should be addressed to Xiao Qiang; [lzjt\\_xq@126.com](mailto:lzjt_xq@126.com)

Received 4 March 2018; Revised 19 May 2018; Accepted 30 July 2018; Published 9 August 2018

Academic Editor: Giuseppe Musolino

Copyright © 2018 Xiao Qiang and Yao Shuang-Shuang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A clustering algorithm for urban taxi carpooling based on data field energy and point spacing is proposed to solve the clustering problem of taxi carpooling on urban roads. The data field energy function is used to calculate the field energy of each data point in the passenger taxi offpoint dataset. To realize the clustering of taxis, the central point, outlier, and data points of each cluster subset are discriminated according to the threshold value determined by the product of each data point field values and point spacing. The classical algorithm and proposed algorithm are compared and analyzed by using the compactness, separation, and Dunn validity index. The clustering results of the proposed algorithm are better than those of the classical clustering algorithm. In the case of cluster numbers 25, 249, 409, and 599, the algorithm has good clustering results for the taxi trajectory dataset with certain regularity in space distribution and irregular distribution in time distribution. This algorithm is suitable for the clustering of vehicles in urban traffic roads, which can provide new ideas and methods for the cluster study of urban traffic vehicles.

## 1. Introduction

The acceleration of urbanization and the rapid increase in the number of travel vehicles have caused traffic congestion on urban roads to become a major problem that must be solved in urban development. Accordingly, several scholars have suggested the idea of “carpool” which has rapidly stimulated academic resonance. Empirical data show that this concept demonstrates policy effectiveness in alleviating urban traffic congestion and solving the plight of vehicle operation. Researchers from the Massachusetts Institute of Technology (MIT) analyzed a week-long run of New York City taxis in Manhattan in March 2013. Out of the 13,600 taxis in New York, approximately 10,000 were used during rush hours. Manhattan requires only 3,000 shared taxis to satisfy its 98% ride requirements [1]. This study showed that an effective ride-sharing system cannot only alleviate the traffic congestion in cities but also increase the passenger-carrying rate of operating vehicles and the operating income of drivers. In addition, energy consumption and environmental pollution must be reduced [2]. Therefore, implementing carpool is

an effective means of improving the quality of urban traffic [3, 4].

The study of the Vehicle Routing Problem (VRP) has given rise to major developments in the fields of exact algorithms and heuristics. In particular, highly sophisticated exact mathematical programming decomposition algorithms and powerful metaheuristics for the VRP have been put forward in recent years [5]. In 2014, Lin [6] provided a classification of Green Vehicle Routing Problem (GVRP) that categorizes GVRP into Green-VRP, Pollution Routing Problem, and VRP in Reverse Logistics and suggests research gaps between its state and richer models describing the complexity in real-world cases. In 2018, Musolino [7] presented a procedure for the solution of the Vehicle Routing Problem (VRP) based on reliable link travel times, the equation of reliable link travel times is composed of a congestion term, expressing the traditional congested link travel times (or generalized costs), and a reliability term, which depends on the fundamental diagram of the link and the Network Fundamental Diagram (NFD) of the homogeneous cluster of adjacent links; NFDs data are used in the proposed link travel

time function to calculate reliable travel times. The reliable link travel times are used for the solution of VRP to obtain optimal routes of freight vehicles.

Taxi carpooling as a way to alleviate taxi traffic congestion has been adopted and implemented in some cities. Therefore, as a hot topic of urban transportation research, carpooling has attracted the attention of many scholars. In the research of carpool service problem (CSP), in 2011, Agatz [8] studied the matching problem of drivers and passengers in a dynamic environment, and optimization method was proposed to minimise vehicle running mileage and individual travel cost. Shinde [9] proposed a genetic algorithm for multiobjective optimization based on the carpooling path matching. The algorithm effectively reduces the computational complexity and processing time and improves the carpooling effect. In 2015, Pelzer [10] proposed a dynamic decision algorithm based on network partition; the algorithm divides and numbers the road network and uses the spatial routing search algorithm to realize the matching of passengers and vehicles. In 2015, Jiao [11] implemented the carpooling path matching in a short time through genetic algorithm and realized the carpooling path matching scheme of low complexity and low memory. In 2015, Huang [12] propose a fuzzy-controlled genetic-based carpool algorithm by using the combined approach of the genetic algorithm and the fuzzy control system, with which to optimize the route and match assignments of the providers and the requesters in the intelligent carpool system. In 2016, Chou [13] developed a particle swarm carpool algorithm based on stochastic set-based particle swarm optimization (PSO); the set-based PSO (S-PSO) can be realized by local exploration. Method yielded the best result in a statistical test and successfully obtained numerical results for meeting the optimization objectives of the CSP.

In the research of taxi carpool problem (TCP), In 2013, Cheng [14] took the benefit of travellers and drivers as the optimization objective and established a multidynamic taxi carpooling model. The genetic algorithm was used to solve the problem of the carpooling algorithm. In 2013, Ma [15] proposed a large-scale taxi ride-sharing service; it efficiently serves real-time requests sent by taxi users and generates ride-sharing schedules that reduce the total travel distance significantly. In 2014, Xiao et al. [16] created membership function by using three factors of driving route, driving time, and the number of passengers to realize carpooling fuzzy clustering and recognition of passengers and taxis. In 2015, Ma [17] devised a mobile-cloud architecture based taxi-sharing system. Taxi riders and taxi drivers use the taxi-sharing service provided by the system via a smart phone App. In 2017, Zhang [18] presented the first systematic work to design a unified recommendation system for both the regular and the carpooling services, called CallCab, based on a data-driven approach; this recommendation system has been done to assist passengers to find a successful taxicab ride with carpooling.

It can be seen from the research of domestic and foreign carpool service problem and taxi carpool problem that the solution of the carpool problem is mainly realized by using multiobjective programming algorithm or intelligent

algorithms. In practical application, because of the large amount of carpool data, the calculation time is very long when the algorithm is used to determine the carpool scheme. Therefore, a two-stage algorithm is proposed to solve the carpool problem in big data environment to reduce the computational time of the algorithm. Häme [19] proposed an adaptive insertion algorithm in 2011, using the idea of clustering to sort the order of passengers on and off, which greatly simplifies the complexity of the problem and the difficulty of the solution and is helpful in addressing the problem of taxi carpooling. In 2012, Manzini [20] proposed phased clustering model. Cluster factors such as route, distance, user information, and carpool were clustered separately. Then, the decision support system was used to judge the factors and provide decision support for the carpooling. In 2013, Shao [21–23] proposed a two-stage clustering heuristic matching strategy to solve the problem of multivehicle carpooling. In 2016, Yang [24] proposed a kind of carpooling model in distributed parallel environment and used the two-stage distributed estimation algorithm to solve the carpooling scheme.

From the above research, we can see that, in the two-stage algorithm, the first-stage carpool factors (such as route, distance, and time) are clustered. In the second stage, according to the results of the clustering, the algorithm of matching, multiobjective programming, and intelligent algorithm are used to solve the combinatorial problem. The main content of this paper is to study the first-stage clustering problem.

In the present application of the clustering algorithm, clustering centre and range are keys in the accurate application of the clustering algorithm, but the classic clustering algorithm on clustering effect is poor because of the city road network distribution. To realize taxi clustering in a city road, a clustering algorithm of urban taxi carpooling based on data field energy and point spacing is proposed for the clustering problem of taxi carpooling on urban roads. The data field energy function is used to calculate the field energy of each data point in the passenger taxi offpoint dataset. The central point, outlier, and data points of each cluster subset are discriminated according to the threshold value determined by the product of each data point field values and point spacing, and the taxi clustering can be realized to provide a basis for theoretical research on city taxis.

The rest of the paper is organized as follows. Section 2 introduces the data field theory. Section 3 proposes carpool taxi clustering algorithm based on **data field energy** and point spacing. Section 4 presents a case of Nanjing taxi data that illustrates the application of the proposed approach. Section 5 provides the conclusions.

## 2. Data Field Energy

*2.1. Definition of Data Field Energy.* From field theory, the field is used to describe the interaction between substances, such as gravitational field, electric field, and magnetic field. Each field is known to decay or increase with distance, and the distribution of field energy can be described by either a scalar or a vector function [25]. The method treats each object in space as a particle with a certain mass. There is a spherical symmetric virtual gravitational field around it,

and any object in the field will be subjected to the joint action of other objects, so a data field is determined in the whole space. Similar to the vector strength function and scalar potential function description of physical field, this method introduces the definition of potential function and field strength function of data field and realizes the self-organizing hierarchical aggregation of data object by simulating the interaction and motion of object in data field. The proposed method is not dependent on the careful selection of user parameters and can identify nonspherical clustering of arbitrary size and density. It is insensitive to noise data and has an approximate linear time complexity. According to this principle, the data point in the large data set is considered a kind of data particle in space, where a virtual field is observed around the data particle. The data particles in the field will be subjected to field forces. Therefore, the data points interact to form a data field. In the data field, the distance between data points is near, and the interaction forces between the data are strong. Then, the data field can be large and vice versa.

$D = \{x_1, x_2, \dots, x_n\}$ , which is a set of data, where  $x_i$  refers to the data point from 1 to  $n$ .  $I_s = \{1, 2, \dots, N\}$ , which is a new index, where  $N$  refers to the number of data points. The data point of potential energy is defined as follows [26]:

$$E(x_i) = M \cdot \sum_{j \in I_s} e^{(-\|x_j - x_i\|/d_c)^R} \quad (1)$$

where  $\|x_j - x_i\|$  is the distance between data points  $x_j$  and  $x_i$ ,  $d_c$  is the interaction factor of data points,  $M$  is the quality of data objects (in this study, the value of  $M$  is 1), and  $R$  is the distance index. The data of spatial distribution mainly depend on the object interaction process or the radius of influence. Thus, spatial distribution has nothing to do with the specific form of the potential function or the selection of distance index, so distance index has minimal influence on the description of structural characteristics [25, 26]. Let  $R = 2$ ,  $E(x_i)$  be the Gauss nuclear field, and the data point of potential energy  $E(x_i)$  is defined as follows:

$$E(x_i) = \sum_{j \in I_s} e^{(-\|x_j - x_i\|/d_c)^2} \quad (2)$$

Gauss field potential function is shown in Figure 1.

As shown in Figure 1, if  $\|x_j - x_i\|$  gradually grows larger (which represents that the distance between data points  $x_j$  and  $x_i$  is becoming increasingly farther), then  $E(x_i)$  is becoming smaller. By contrast, if  $\|x_j - x_i\|$  gradually grows smaller (which represents that the distance between data points  $x_j$  and  $x_i$  is becoming increasingly closer), then  $E(x_i)$  is becoming larger. From the data interaction factors, we can find the change; if  $d_c$  is small, then the distance of data points  $x_j$  and  $x_i$  interaction is smaller; if  $d_c$  is large, then the distance of data points  $x_j$  and  $x_i$  interaction is larger. Thus, parameter  $d_c$  can control the range and size of data clustering.

**2.2. Distance between Data Points.** The distance between data points reflects the degree of density between the data points and also the size of the data field between the data points. If

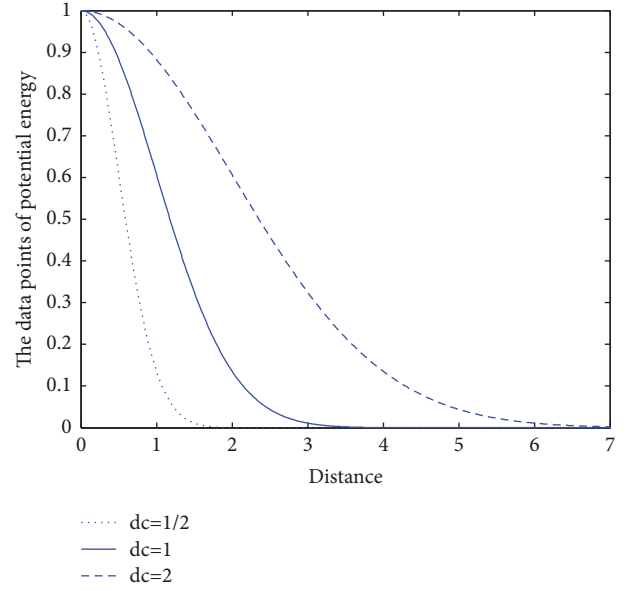


FIGURE 1: Gauss field potential function.

the distance between data points is smaller, then the density of the data set is higher and the data field can be greater. By contrast, if the distance between data points is larger, then the density of the data set is low and the data field can be smaller. Thus, the data field energy of data points can be determined by the distance between data points. The distance  $d_{ij}$  between data points  $x_j$  and  $x_i$  is defined as follows:

$$d_{ij} = \|x_j - x_i\| = \sqrt{(x_j - x_i)^2} \quad (3)$$

Data point spacing matrix  $H$  is denoted as follows:

$$H = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ & & \vdots & & \\ d_{m1} & d_{m2} & d_{m3} & \cdots & d_{mn} \end{bmatrix} \quad (4)$$

In (4),  $d_{mn} = d_{nm}$ ,  $m \neq n$ .

### 3. Carpool Taxi Clustering Algorithm

This section combines the aforementioned data field energy and the distance between data points to identify the clustering centres and cluster subsets of passenger taxi data by using GPS (Global Positioning System) trajectory data. In addition, the computing procedure is presented in detail.

We established passenger taxi GPS data set  $D1$  at  $T_i$  time.  $D1 = \{t_1, t_2, \dots, t_n\}$ , where  $t_i$  is the  $i$  passenger taxi that can provide carpool.  $t_i = \{long_{t_i}, lat_{t_i}, time_{t_i}\}$ , where  $long_{t_i}$  is the taxi longitude,  $lat_{t_i}$  is the taxi latitude, and  $time_{t_i}$  is the taxi location time.  $E = \{E(t_1), E(t_2), \dots, E(t_n)\}$ , where set  $E$  represents the data points of potential energy and  $E(t_1)$  represents the data points  $t_i$  of potential energy. These data

points are sorted according to the size of potential energy [27–29].  $I_p = \{I_{p1}, I_{p2}, \dots, I_{pn}\}$ , where set  $I_p$  is the ranking index of potential energy.  $E_p = \{E_{I_{p1}}, E_{I_{p2}}, \dots, E_{I_{pn}}\}$ , where set  $E_p$  is the ranking of potential energy. The distance  $u_i$  between different data points is defined as

$$u_i = \begin{cases} \min_{k \in I_p} \{d_{ik}\}, & E_{I_{pi}} < E_{I_{pk}} \\ \max_{k \in I_p} \{d_{ik}\}, & E_{I_{pi}} = \max_{I_{pk} \in I_p} \{E_{I_{pk}}\} \end{cases} \quad (5)$$

If  $E_{I_{pi}} = \max_{I_{pk} \in I_p} \{E_{I_{pk}}\}$ , then the potential energy of this data point  $t_{I_{pi}}$  will reach the maximum in the local data field, which means that  $u_i$  is the maximum distance between the cluster centre and the  $t_{I_{pi}}$  in local data set D1. If  $E_{I_{pi}} < E_{I_{pk}}$ , then in the local data field potential,  $t_{I_{pi}}$  is less than the maximum value of local data field, which means that  $u_i$  is the minimum distance between data points and  $t_{I_{pi}}$  in local data set D1.

Thus, we can obtain data pairs  $(E(t_i), u_i), i \in I_t$ .  $w_i = E(t_i)u_i, i \in I_s$ , where  $w_i$  is the discriminant value of the number of clusters in the dataset. According to formula (5), if  $t_i$  is the central point in a local data field, then  $E(t_i)$  and  $u_i$  are the maximum value and the value of  $w_i$  is relatively large. If  $t_i$  is not the central point in a local data field, then the values of  $E(t_i)$  and  $u_i$  are relatively small and the value of  $w_i$  is relatively small. According to the  $w_i$  value changes, the cluster centre number can be determined.  $\{q_i\}_{i=1}^N$  represents a descending subscript order of  $\{w_i\}_{i=1}^N$ , and the number of clustering centre is defined as

$$g = \begin{cases} g + 1 & w_{q_i} > w_{q_j}, q_j < q_i \\ 0 & w_{q_i} \approx w_{q_j}, q_j < q_i \end{cases} \quad (6)$$

where  $g$  is the number of clustering centres. If  $w_{q_i} > w_{q_j}$ , then a distinct jump is observed between  $w_{q_i}$  and  $w_{q_j}$ , which means that  $q_i$  is a data point clustering centre. If  $w_{q_i} \approx w_{q_j}$ , then no distinct jump is observed between  $w_{q_i}$  and  $w_{q_j}$ , which means that  $q_i$  is not a data point clustering centre. The threshold points between jump points and nonjumping points can be determined, and we can find the data potential values  $E(t_g)$  and distance  $u_g$  corresponding to the threshold points.

The centre of clustering is determined by the number of cluster centres, and the subsets of each point are determined according to the distance from each point to the centre. A subset of  $C$  is defined as

$$C = \begin{cases} t_{ck} & u_i \leq u_g, i \in I_t \\ t_{cli} & u_i > u_g, i \in I_t \end{cases} \quad (7)$$

where  $t_{ck}$  is a subset of the data  $t_{ci}$ , which indicates that the distance from the data point to the maximum point of the region data potential energy is less than the threshold  $u_g$ .  $t_{cli}$  is an outlier, which indicates that the distance from the data point to the maximum point of the region data potential energy is greater than the threshold  $u_g$ .

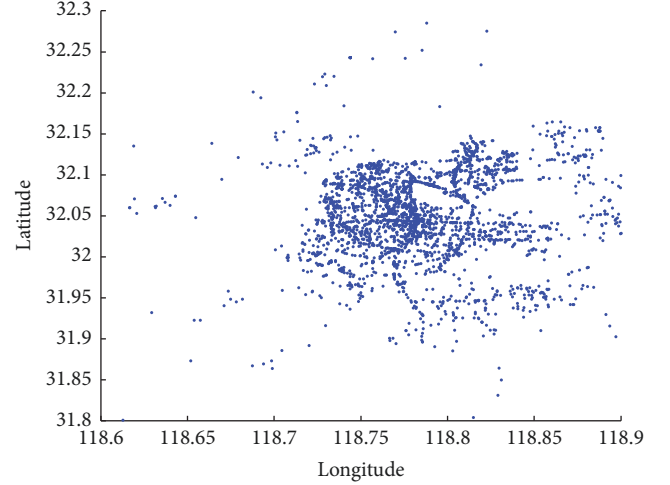


FIGURE 2: Passenger taxi to get-off point scatter diagram in Nanjing City between 9:00 and 9:10.

## 4. Case Study

**4.1. System Description.** This study uses a GPS history of Nanjing urban taxi data on September 1, 2014. We took the data on 3,391 passenger taxi drop-off points as experimental data from 9:00 to 9:10 and the taxi GPS trajectory data preprocessing to obtain the scatter distribution, as shown in Figure 2.

As shown in Figure 2, the main points of the passenger taxis in Nanjing are concentrated in the central area of the city at 9:00. If we move to the urban area, then fewer passengers get off the bus, thereby implying a lower aggregation rate. Moreover, as shown in Figure 2, although the get-off point distribution of the passenger taxi is out of order, the get-off point is the distribution of road networks in the city. Therefore, the dataset of the taxi get-off point has a certain spatial distribution, but it is disordered in the time distribution.

**4.2. Experimental Process and Results.** The horizontal axis is the data field energy and the vertical axis is the point spacing in Figure 3. As shown in the upper part of Figure 3, the longitudinal axis of the right data points in the data field to  $E(t_i)$  and interval of  $u_i$  are large which shows the clustering centre data point of the local data subset. The potential energy  $E(t_i)$  of the data point near the vertical axis is smaller, but the point spacing  $u_i$  is larger, which indicates that the point is the outlier in the data. As shown in the lower part of Figure 3, energy  $E(t_i)$  and the point spacing  $u_i$  are less close to the horizontal axis, which shows the subset of data in the local clustering centres.

Figure 4 presents a descending  $w_i$  value chart of the Nanjing passenger taxi get-off point at 9:00–9:10. In the cluster centre above the threshold,  $w_i$  has a descending order from large to small. In the cluster centre below the threshold, the  $w_i$  value is almost equal. The threshold can be determined by the number of cluster centres and distance  $u_g$ . By calculation,  $u_g$  equals 1.116272.



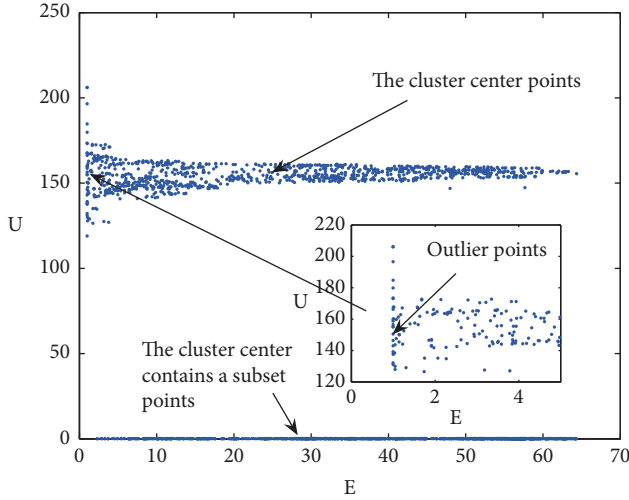


FIGURE 3: Data field energy and point spacing distribution diagram of passenger taxi to get-off point.

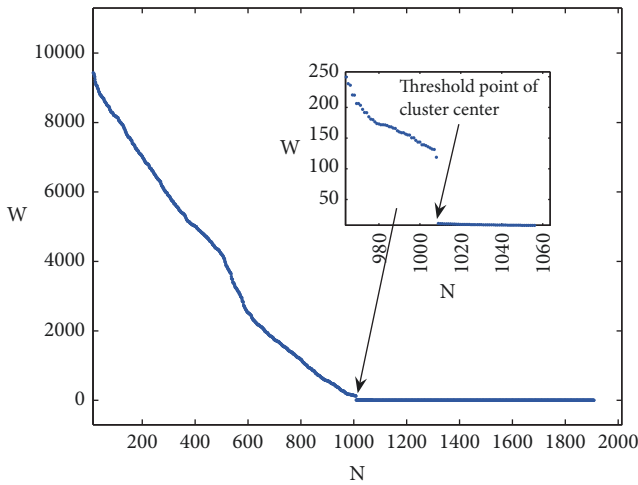


FIGURE 4: Schematic diagram of the  $w_i$  descending order.

As shown in Figure 5 and partial enlargement, the clustering algorithm proposed in this study can realize taxi clustering. The advantages of the algorithm are expressed in the following aspects:

(1) Based on the effectiveness of the algorithm, it can identify that the clustering centre is very good and, to determine the number of clusters, the clustering of arbitrary shape does not need to experience the selection of the number of cluster centres, which can be better applied to the trajectory data clustering.

(2) From the point of view of special treatment, the points are divided into a new cluster according to the number of edge points for concentrated or outliers, based on data field and distance discrimination.

(3) From the clustering result, the data clustering is basically distributed along the road network of the city to meet the requirements of cluster city taxi vehicles.

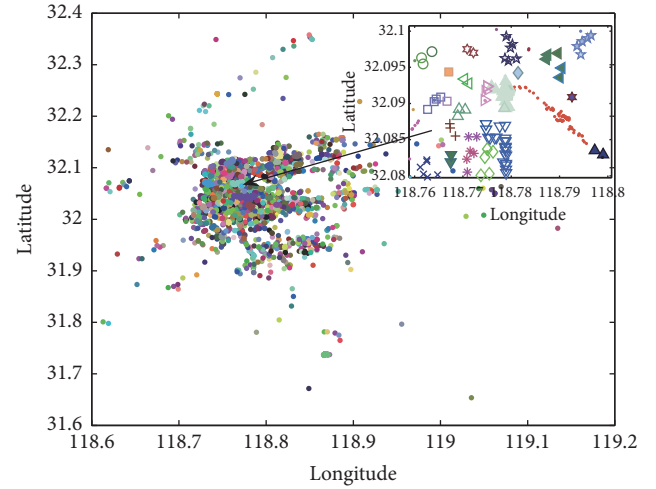


FIGURE 5: Clustering graph based on data field energy algorithm.

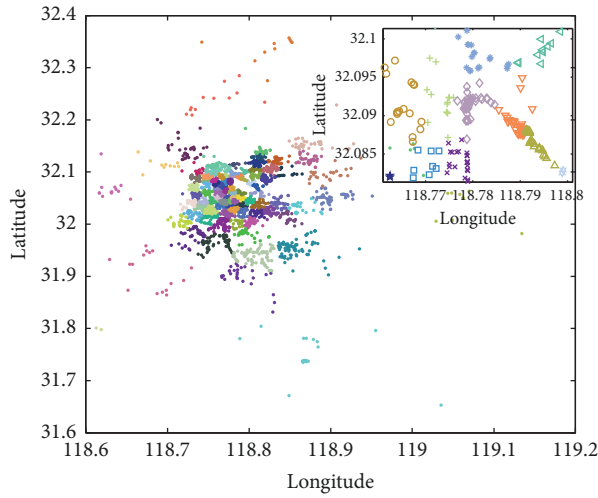
**4.3. Algorithm Evaluation.** We use classical clustering algorithm, which includes K-means algorithm, fuzzy C-means (FCM) clustering algorithm, and hierarchical clustering algorithm, and the proposed algorithm in this is used for comparative analysis. These algorithms are evaluated by compactness (CP), separation (SP), and Dunn validity index (DVI).

Figure 6 shows the clustering results of K-means algorithm, FCM algorithm, and hierarchical clustering algorithm. As shown in Figure 6(a), in the K-means algorithm, the number of initial centre points determines the number of clusters. For a class of clusters with a complex shape, it divides some edge points into near clusters. From the subgraph (a) of the partial enlargement, the K-means algorithm cannot accurately be clustered on the edge point. For subgraph (b), the clustering effect of the FCM algorithm is better than that of the K-means algorithm, but it is limited by the number of initial centres and the influence of isolated points. The FCM algorithm is similar to the K-means algorithm, but it cannot exactly cluster the edge points. From the subgraph, the hierarchical clustering algorithm is limited by the algorithm itself, and the clustering of the outlier data points is not accurate.

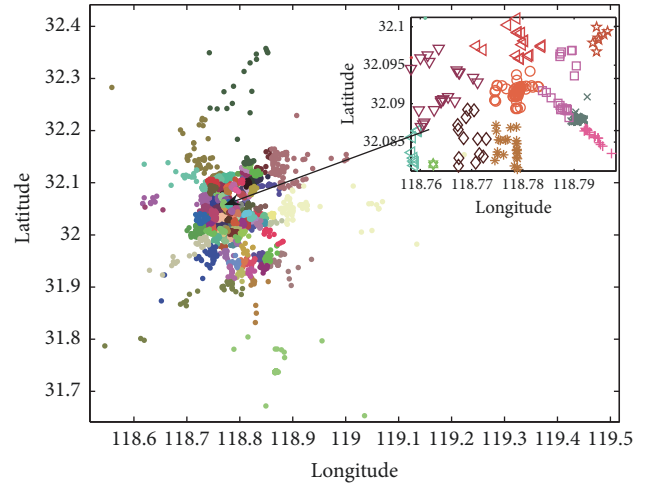
As shown in Table 1 and Figure 7 and on the basis of the K-means algorithm, FCM algorithm, hierarchical clustering algorithm, and data field energy clustering algorithm, the CP index will be reduced with the increase in the number of clusters. However, the CP index of the data field clustering algorithm is less than the CP index of other clustering algorithms. The proposed algorithm in this study is better than that of other clustering algorithms in the CP index. The SP index decreases with the increase of cluster numbers, which shows that the class spacing decreases as the number of clusters increases. However, the SP index is higher than that of other clustering algorithms, and the proposed algorithm in this study is better than other clustering algorithms in the SP index. Based on the overall clustering evaluation of the DIV index, the clustering algorithm in this study is better than other clustering algorithms in the clustering of GPS data in urban taxis.

TABLE 1: Comparison table of clustering effect evaluation index of each clustering algorithm.

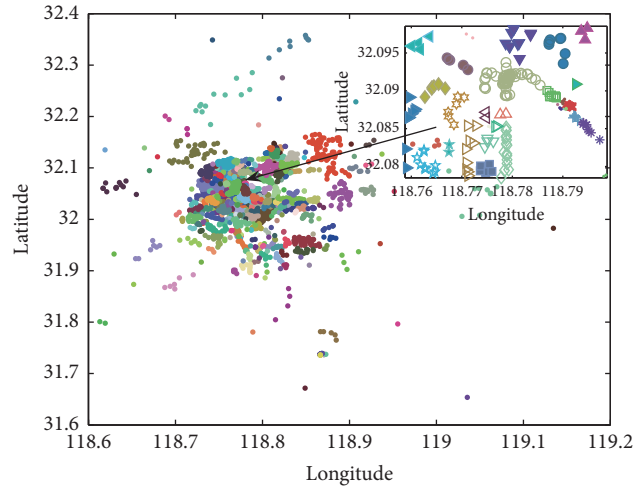
Clustering number	Clustering algorithms	CP	SP	DVI
25	K-means	0.0392	0.3644	$2.543 \times 10^{-4}$
	FCM	0.0201	0.1233	$1.001 \times 10^{-4}$
	Hierarchical	0.0123	0.6069	$2.717 \times 10^{-4}$
	Data field energy	0.0088	0.6666	$1.32 \times 10^{-2}$
249	K-means	0.0047	0.1414	$7.008 \times 10^{-5}$
	FCM	0.005	0.0597	$2.749 \times 10^{-5}$
	Hierarchical	0.0042	0.181	$2.724 \times 10^{-5}$
	Data field energy	0.0016	0.2319	$5.152 \times 10^{-4}$
409	K-means	0.0034	0.1213	$3.845 \times 10^{-5}$
	FCM	0.0033	0.0558	$2.041 \times 10^{-5}$
	Hierarchical	0.0028	0.1398	$2.1742 \times 10^{-5}$
	Data field energy	0.0013	0.1847	$3.846 \times 10^{-4}$
599	K-means	0.0021	0.119	$3.6076 \times 10^{-5}$
	FCM	0.0023	0.0569	$2.3121 \times 10^{-5}$
	Hierarchical	0.0019	0.1242	$1.086 \times 10^{-5}$
	Data field energy	0.001	0.1685	$1.823 \times 10^{-4}$



(a) K-means algorithm clustering graph



(b) FCM algorithm clustering graph



(c) Hierarchical clustering graph

FIGURE 6: Classical algorithm clustering graph.

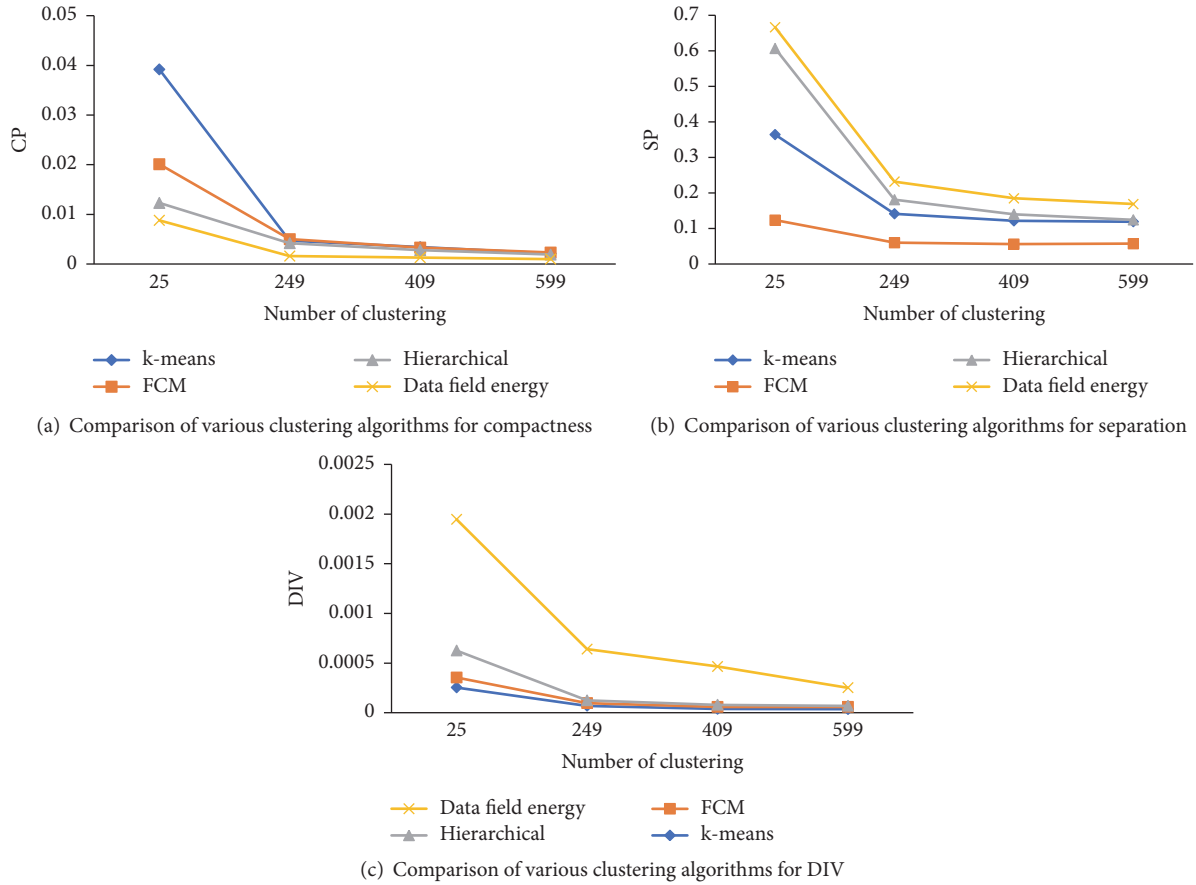


FIGURE 7: Comparison of various clustering algorithm evaluation index.

## 5. Conclusion

The main conclusions of this study are presented as follows:

(1) The use of field theory and city taxi clustering algorithm is established based on data field and space. The algorithm uses the dataset of taxi get-off points in Nanjing to achieve the clustering of taxis carpooling on urban roads.

(2) By comparing the classical clustering analysis, the proposed algorithm is found more suitable for the taxi data clustering, which has a certain spatial distribution shape. However, the time distribution of uncertainty provides new clustering methods and ideas for urban traffic data clustering.

(3) The clustering algorithm of urban taxi carpooling vehicles based on data field energy realized the taxi carpooling clustering. However, the taxi clustering data in this study are conducted in a static environment. In the future, we will discuss the clustering model and algorithm of taxi carpooling in a dynamic environment in combination with the dynamic change of the urban taxi get-off point.

## Data Availability

The datasets analyzed during the current study are available from DATATANG (<http://www.datatang.com/>) but restrictions apply to the availability of these data, which were used under license from DATATANG and so are not publicly

available. Data are however available from the authors upon reasonable request and permission of DATATANG. The downloading data used in the study are available in the DATATANG: <http://www.datatang.com/product/solution/ff808081593ef56001593fab75080046>.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work has been supported by the National Natural Science Foundation of China (nos. 61364026 and 51408288)

## References

- [1] A. Afian, A. Odoni, and D. Rus, "Inferring Unmet Demand from Taxi Probe Data," in *Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems, ITSC 2015*, pp. 861–868, September 2015.
- [2] W. He, K. Hwang, and D. Li, "Intelligent carpool routing for urban ridesharing by mining GPS trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2286–2296, 2014.

- [3] S.-C. Huang, M.-K. Jiau, and C.-H. Lin, "A genetic-algorithm-based approach to solve carpool service problems in cloud computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 352–364, 2015.
- [4] S. Huang, M. Jiau, and Y. Liu, "An ant path-oriented carpooling allocation approach to optimize the carpool service problem with time windows," *IEEE Systems Journal*, vol. 99, pp. 1–12, 2018.
- [5] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, no. 4, pp. 408–416, 2009.
- [6] C. Lin, K. L. Choy, G. T. S. Ho, S. H. Chung, and H. Y. Lam, "Survey of green vehicle routing problem: past and future trends," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1118–1138, 2014.
- [7] G. Musolino, A. Polimeni, and A. Vitetta, "Freight vehicle routing with reliable link travel times: a method based on network fundamental diagram," *Transportation Letters-The International Journal of Transportation Research*, vol. 10, no. 3, pp. 159–171, 2018.
- [8] N. A. H. Agatz, A. L. Erera, M. W. P. Savelsbergh, and X. Wang, "Dynamic ride-sharing: A simulation study in metro Atlanta," *Transportation Research Part B: Methodological*, vol. 45, no. 9, pp. 1450–1464, 2011.
- [9] T. Shinde and B. Thombre, "An effective approach for solving carpool service problems using genetic algorithm approach in cloud computing," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 3, no. 12, pp. 29–33, 2015.
- [10] D. Pelzer, J. Xiao, D. Zehe, M. H. Lees, A. C. Knoll, and H. Aydt, "A partition-based match making algorithm for dynamic ridesharing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 1–12, 2015.
- [11] M.-K. Jiau and S.-C. Huang, "Services-oriented computing using the compact genetic algorithm for solving the carpool services problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 1–12, 2015.
- [12] S.-C. Huang, M.-K. Jiau, and C.-H. Lin, "Optimization of the carpool service problem via a fuzzy-controlled genetic algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1698–1712, 2015.
- [13] S.-K. Chou, M.-K. Jiau, and S.-C. Huang, "Stochastic set-based particle swarm optimization based on local exploration for solving the carpool service problem," *IEEE Transactions on Cybernetics*, vol. 46, no. 8, pp. 1771–1783, 2016.
- [14] J. Cheng, Z. Tang, J. Liu, and L. Zhong, "Research on dynamic taxipooling model based on genetic algorithm," *Journal of Wuhan University of Technology (Transportation Science and Engineering)*, vol. 37, no. 1, pp. 187–191, 2013.
- [15] S. Ma, Y. Zheng, and O. Wolfson, "T-share: A large-scale dynamic taxi ridesharing service," in *Proceedings of the 29th International Conference on Data Engineering, ICDE 2013*, pp. 410–421, April 2013.
- [16] Q. Xiao, R.-C. He, W. Zhang, and C.-X. Ma, "Algorithm research of taxi carpooling based on fuzzy clustering and fuzzy recognition," *Journal of Transportation Systems Engineering and Information Technology*, vol. 14, no. 5, pp. 119–125, 2014.
- [17] S. Ma, Y. Zheng, and O. Wolfson, "Real-time city-scale taxi ridesharing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 7, pp. 1782–1795, 2015.
- [18] D. Zhang, T. He, Y. Liu, S. Lin, and J. A. Stankovic, "A carpooling recommendation system for taxicab services," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 254–266, 2017.
- [19] L. Häme, "An adaptive insertion algorithm for the single-vehicle dial-a-ride problem with narrow time windows," *European Journal of Operational Research*, vol. 209, no. 1, pp. 11–22, 2011.
- [20] R. Manzini and A. Pareschi, "A decision-support system for the car pooling problem," *Journal of Transportation Technologies*, vol. 2, no. 2, pp. 85–101, 2012.
- [21] Z. Z. Shao, H. G. Wang, H. Liu, and et al, "Research on service requirement distribution algorithms in carpooling problems," *Journal of Tsinghua University*, vol. 53, no. 2, pp. 252–258, 2013.
- [22] Z. Z. Shao, H. G. Wang, H. Liu et al., "Single carpooling problem based on matching degree clustering algorithm," *Journal of Software*, vol. 23, supplement 2, pp. 204–212, 2012.
- [23] Z. Z. Shao, H. G. Wang, H. Liu, C. Song, C. Meng, and H. Yu, "Heuristic optimization algorithms of multi-carpooling problem based on two-stage clustering," *Journal of Computer Research and Development*, vol. 50, no. 11, pp. 2325–2335, 2013.
- [24] Z. J. Yang, Z. Wang, Y. Wang et al., "Two-stage estimation of distribution algorithm to solve multi-vehicle carpooling problem," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 2, pp. 164–169, 2016.
- [25] W. Y. Gan, D. Y. Li, and J. M. Wang, "An hierarchical clustering method based on data fields," *Acta Electronica Sinica*, vol. 34, no. 2, pp. 258–262, 2006.
- [26] N. Zhang, T. T. Chen, and X. X. He, "A single partition clustering algorithm based on data field," *Journal of Zhe Jiang University of Technology*, vol. 44, no. 1, pp. 52–57, 2016.
- [27] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [28] S. Wang, D. Wang, C. Li, Y. Li, and G. Ding, "Clustering by fast search and find of density peaks with data field," *Chinese Journal of Electronics*, vol. 25, no. 3, pp. 397–402, 2016.
- [29] R. Mehmood, G. Zhang, R. Bie, H. Dawood, and H. Ahmad, "Clustering by fast search and find of density peaks via heat diffusion," *Neurocomputing*, vol. 208, no. 6191, pp. 210–217, 2016.



