A Trajectory Clustering Method Based on Moving Index Analysis and Modeling

YUQING YANG1, JIANGHUI CAI1,2, (Member, IEEE), HAIFENG YANG1, (Member, IEEE), XUJUN ZHAO1, (Member, IEEE), AND JING LIU1

1School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China
2School of Computer Science and Technology, North University of China, Taiyuan 030051, China

Corresponding author: Jing Liu (1984016@tyust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1931209, in part by the Key Research and Development Projects of Shanxi Province under Grant 201903D121116, in part by the Central Government Guides Local Science and Technology Development Funds under Grant 20201070, and in part by the Fundamental Research Program of Shanxi Province under Grant 20210302123223 and Grant 202103021224275.

ABSTRACT Aiming at the problem of low trajectory clustering accuracy caused by only focusing on the characteristics of Stop Points, this paper analyses the features of both the Stop and the Move Points and proposes a trajectory clustering method based on the moving index analysis and modelling. Firstly, the different characteristics of the trajectory points are explored, and each feature is analysed and evaluated by experiments. On this basis, the PD (Point Density) and MC (Movement characteristic) are selected to define a new moving index (MPD) to evaluate the movement performance of different types of points. Secondly, a trajectory clustering algorithm called PMS (Points Moving Index Analysis and Modelling) is proposed. This algorithm finds the Stop Points by the following steps. (1) Obtaining the candidate move points with the help of PD. (2) Calculating the MPD of all the points to approximate the trajectory points. (3) Establishing a MIGM (Moving Index Gaussian model) model based on the MPD representation. (4) Fitting all the trajectories and extracting the points that are not fitted by MIGM. (5) Judging whether the extracted points satisfy the convergence condition. If the convergence condition is satisfied, the extracted points are Stop Points. Otherwise, adjust the radius R and repeat the above four steps. Experimental results show that this method can reduce error merging of adjacent clusters and find the trajectory clusters of Stop Points with different shapes.

INDEX TERMS Trajectory clustering, moving index, stop points extraction, moving index Gaussian model.

I. INTRODUCTION

Trajectory data [1] describes the change of the position of moving objects with time, and contains valuable knowledge. A common trajectory of a person’s daily life is shown in Fig.1. This person left his home to work at the office, after some time went shopping at the supermarket, and finally returned home. Normally, two types of points are included in trajectory data, namely Stop Points and Move Points [2]. The stop parts in a trajectory are the trajectory segments produced by a person who stays in certain locations for a period of time, and the points included in these stop parts are the Stop Points. The moving parts correspond to the trajectory segments resulting from the movement of a person between stop positions, and the points within these trajectory segments are noted as Move Points.

Stop Points correspond to some specific geographical places, or the locations where certain important events occur, such as the Office and the Supermarket in Fig.1, or schools, shopping malls and the locations where parades, assemblies and traffic accidents occur, etc. These special locations can be used to analyse the behaviour patterns [3], [4] of moving objects and predict the next occurrence time of the event. Therefore, the analysis and extraction of the Stop Points is of great importance. Stop Points clustering (or extraction) is an important study in trajectory clustering [1], and which is capable of clustering a person’s stop location by assigning different constraints to different features.

A common problem suffered by most Stop Points clustering methods is that they ignore the active role played by
Move Points, and only pay attention to the Stop Points. Under the above problem, it is difficult to avoid the following three dilemmas for a large number of algorithms.

1) Difficult to truly grasp the characteristics of Stop Points. Most clustering based Stop Points extraction methods focus on characteristic analysis of Stop Points, such as the speed, density, movement characteristic, direction angle, etc. However, in actual data distribution, the speed, density, movement characteristic and direction angle of Stop Points are quite different and irregular, which leads to considerable fluctuation in the accuracy of Stop Points extraction on different trajectory data. Therefore, it is difficult to truly grasp the characteristics of Stop Points only from the perspective of Stop Points feature analysis.

2) Trouble in dealing with different scales of Stop Points clusters. The scale of the Stop Points in the same or different trajectories differs greatly from each other, and the extraction accuracy of these points is often affected by clustering the Stop Points with different aggregation scales using a globally unique radius threshold. In contrast, the Move Points in a trajectory usually correspond to a certain path, which connects two adjacent clusters. The difference between the density of the Move Points is relatively small, so the clustering effect of different scales of Stop Points may be improved by analyzing the Move Points.

3) Adjacent clusters are easily mis-merged. The Move Points play the role of connecting two sets of adjacent Stop Points, and if we focus only on Stop Points, some small adjacent clusters may be incorrectly combined into a big one, making the clustering accuracy decrease. Therefore, considering the characteristics of Move Points can attenuate the false merging phenomenon of adjacent clusters.

If we can extract the Move Points through their feature analysis, then the set of Stop Points can be obtained through further clustering analysis of the remaining points after extraction. The above method can not only improve the accuracy of Stop Points extraction, but also be used as a preprocessing method for Stop Points extraction, which can reduce the amount of data involved in Stop Points clustering and improve clustering accuracy.

A. MOTIVATIONS

The motivations of this study can be summarized as follows:

- Existing Stop Points clustering methods only focus on characteristic analysis of Stop Points, which makes it difficult to grasp the characteristics of trajectory data, resulting in large fluctuations in the extraction accuracy of the Stop Point.
- Some Stop Points extraction methods fail to deal with different scales of Stop Points clusters due to the diversity of Stop Points cluster sizes in the same or different trajectories.
- Some methods ignore the characteristics of Move Points and often encounter the problem of false merging of adjacent clusters. Finding a simple and effective method is important to solve the above problem.

B. CONTRIBUTIONS

Inspired by the above motivations, the different characteristics of the trajectory points are explored, and each feature is analyzed and evaluated by experiments, then a trajectory Stop Points extraction method called PMS is proposed based on points moving index analysis and modelling. In this paper, the PD (Point Density) and MC (Movement characteristic) are selected to define a new moving index (MPD) to evaluate the movement performance of different types of points. Then, all of the following four steps are performed until convergence is reached. (1) Obtaining the candidate Move Points with the help of PD. (2) Calculating the MPD all the points to approximate the trajectory points. (3) Establishing a MIGM model based on the MPD representation. (4) Fitting all the trajectories and extracting the points that are not fitted by MIGM. (5) Observing the structure of the extracted points, and judging whether the extracted points satisfy the convergence condition. If the convergence condition is satisfied, the extracted points are the Stop Points. Otherwise, adjust R and repeat the above process. Instead of focusing on the characteristics of Stop Points, the features of the Move Points are captured through the analysis of various characteristics of trajectories. The difficulty of Stop Points extraction is reduced by the above method, and the accuracy of the Stop Points extraction is improved by adjusting the radius R.
The main contributions of this study are as follows:
- A moving index - MPD - and a MIGM model are established to abstract the characteristics of Move and Points, so that the characteristics of data points can be accurately grasped.
- A PMS algorithm based on points moving index analysis and modelling is proposed. Through the above algorithm, the Move and Stop Points are distinguished more accurately and the error merging of adjacent clusters is also alleviated.
- The clusters with different aggregation scales and shapes can be found by adjusting the values of parameter R in MIGM.
- This method also provides an effective preprocessing technology for Stop Points extraction. Firstly, the Move Points are roughly culled, and then the characteristics of the remaining points can be further analyzed.

C. ORGANIZATION
The rest of this paper is organized as follows. Section II briefly reviews some related works. Section III gives a detailed theoretical analysis of the characteristics in trajectory data. The processing steps of the proposed method are also presented in Section III. Section V shows the experimental results. Finally, a conclusion is given in Section VI.

II. RELATED WORK
The main approaches used in trajectory Stop Points extraction can be divided into two categories: static methods and dynamic methods [2]. The first methods take the existing geography information as background knowledge of the system, only the Stop Points contained in existing information are found. In dynamic technologies, Stop Points can be identified as long as the spatial-temporal constraints set in various dynamic algorithms are satisfied. Through continuous development, dynamic methods extend many kinds of technologies. This paper mainly focuses on some typical dynamic methods and some newly published algorithms.

For example: A sequence oriented clustering approach was presented in [5]. In this paper, a SOC (Sequence Oriented Clustering) algorithm was developed to automatically extract stops from a single trajectory. Based on traditional DBSCAN [6] (Density-Based Spatial Clustering of Application with Noise) algorithm, reference [7] presented a DJ-Cluster (Density-and-Join-based) algorithm to extract the position in trajectories. In [2], the DBSCAN algorithm was also improved, and the density measurement was replaced by a new density calculation method by means of a new concept of Move Ability and data domain theory. In [8] the CB-SMoT (Clustering based Stops and Moves of Trajectories) algorithm was proposed, in this algorithm the definition of neighbour points and density in the traditional clustering methods were expanded to time dimension according to the space-time characteristics of trajectory data. Reference [9] proposed a two-step clustering approach to extract individuals’ locations, this method found different types of Stop Points using the spatio-temporal clustering. From the perspective of density clustering, reference [10] proposed a TAD algorithm to extract trajectory Stays based on spatial-temporal density analysis of data. In this algorithm, a noise tolerance factor and a NMAST function was defined to calculate the space-time density of trajectory.

In addition to some typical methods mentioned above, some newly published methods have emerged to find valuable information. For example, a framework for AIS (automatic identification system) data-driven vessel trajectory prediction based on a long short term memory (LSTM) network was proposed in literature [11]. Reference [13] introduced a data field-based cluster analysis technique to detect hot spots. A local maximum density (LMD) approach to identify the hot spots was proposed in [14]. LMD can not only identify multiple local hotspots in highly popular regions, but also detect potential hotspots in less popular regions. A novel algorithm for semantic trajectory clustering based on community detection (STCCD) in networks was given in [15]. This method can better measure the semantic similarity of trajectories and capture global relationships among trajectories from the perspective of the network. One of the most important challenges in achieving reliable mining results for massive vessel trajectories is how to efficiently compute the similarity between different ship trajectories. Reference [16] proposed an unsupervised learning method to automatically extract low-dimensional features by convolutional auto-encoder (CAE), and then obtained the final trajectory similarity by computing the similarity between low-dimensional features. A trajectory regression clustering method [17] was proposed to reduce local information loss of trajectory data and avoid getting stuck in the local optimum. In this method, the Lagrange-based method and Hausdorff-based Kmeans++ were integrated in fuzzy C – means (FCM) clustering to maintain the stability and the robustness of the clustering process. Trajectories were clustered by learning deep representations in [18]. In this method, a sliding window and a sequence-to-sequence auto-encoder were used to learn fixed-length deep representations. Density-based and swarm intelligence based clustering [19] is also a hotspot in the field of clustering research. In order to solve the problem that most of the existing trajectory clustering methods focus on some spatial properties of trajectories and some vital information of nodes is lost, reference [12] proposed a joint spatial-temporal trajectory clustering method (JSTTCM) to exploit some spatial-temporal properties of the trajectory points.

In the above methods, the Move Points are often overlooked in trajectory data analysis. However, since the characteristics of Stop Points are usually complicated and variable, it is difficult to directly analyse and extract them. Importantly, the size of the cluster where the Stop Points are located is quite different, especially in different trajectories. The accuracy of the methods using global unique radius R to find the Stop Points is not satisfactory. In order to achieve more desirable clustering results, the value of R must be
TABLE 1. Notations and their description.

| Notation | Description |
|----------|-------------|
| D        | A dataset consists of multiple trajectories |
| n        | The number of points in Tra |
| k        | The number of clusters in Tra |
| R        | The neighbourhood radius |
| C        | The set of stop points clusters |
| P_i      | A point in Tra |
| X_i      | The latitude of P_i |
| radX_i   | The radian value of X_i |
| T_i      | The acquisition time of P_i |
| A_{1r}   | The mean latitude of the points in Cluster i |
| Dist(P_i, P_j) | The distance between P_i and P_j |
| DDist(P_i, P_j) | The displacement between P_i and P_j |
| m        | The number of points in CMP |
| SP       | Stop Points |
| MC       | Move Characteristics |
| PD       | Point Density |
| NA       | Average Neighbourhood Radius |
| AS       | Average Speed |
| NMA      | Neighbourhood move ability |
| MIGM     | Moving index Gaussian model |
| σ        | The variance of MIGM |
| PMS      | Points moving index analysis and modelling |
| TSP      | Total stop points |
| FSP      | Found stop points |

| Notation | Description |
|----------|-------------|
| Tra      | A moving trajectory |
| n'       | Average number of neighbours of the points in Tra |
| N        | The number of points in a cluster |
| r        | The radius of the earth in meters |
| E        | The model convergence condition |
| N(P_i)   | The set of points within the R neighbourhood of P_i |
| Y_i      | The longitude of P_i |
| radY_i   | The radian value of Y_i |
| ρ(P_i)   | The point density of P_i |
| A_{1y}   | The mean longitude of the points in Cluster i |
| CMP      | The set of candidate move points in Tra |
| cmn P_i  | An element in CMP |
| MP       | Move Points |
| SI       | Spatial Influence |
| MPD      | An enhance PD under moving index |
| DA       | Direction Angle |
| IS       | Instantaneous Speed |
| NMAST    | Neighbourhood Move Ability and Stay Time |
| μ        | The mean of MIGM |
| t        | Number of model adjustments |
| DT       | The maximum point density threshold |
| RSP      | Realised stop points |
| USP      | Undetected stop points |

From the above point of view: the characteristics of Move Points are simpler and easier to grasp in terms of characteristic complexity. In order to further demonstrate the above conclusions and provide the basis for the later work of this paper, this section then gives a detailed theoretical analysis of various characteristics of trajectories.

1) THEORETICAL ANALYSIS OF CHARACTERISTICS OF THE STOP AND MOVE POINTS

Trajectory data includes two types of characteristics: space and time. The space characteristics mainly include Movement Characteristics (MC), Spatial Influence (SI), Point Density (PD), Average Neighbourhood Radius (NA), Direction Angle (DA), etc., and the time characteristics mainly include Average Speed (AS), Instantaneous Speed (IS), Space-time Density, etc., some brief descriptions of the above characteristics are given below.

1) Movement Characteristic (MC, [2]) estimates the moving features of points by analysing the ratio of the displacement and distance between neighbours. The greater the ratio is, the greater the possibility of them being Move Points; conversely, the mobility of the neighbours will be smaller, and the neighbours are less likely to be Move Points. Fig. 3 shows MC of two trajectory segments. Since the ratio of displacement to distance from point A to B in Fig. 3(b) is significantly smaller than that in Fig. 3(a), the MC of the stopping segment is significantly smaller than the MC of the moving segment. Through MC, the Move Points and the Stop Points can be divided roughly, but the regular movement of a moving object in a stopping area which makes the mobility artificially low can not be identified. In order to improve the ability of MC in adjusted appropriately according to the shape and size of clusters.

II. MATERIALS AND METHODS

Motivated by the above mentioned difficulties in Stop Points extraction, this paper analyses different characteristics of trajectory points and proposes a points-moving index-analysis-and-modelling based method called PMS to extract Stop Points. In this section, we first analyse several characteristics of trajectory data theoretically, then, give some related definitions. Finally, a MIGM model and a PMS algorithm are proposed. The notations used in this paper and their descriptions have been summarised in Table 1.

A. CHARACTERISTIC ANALYSIS OF MOVE AND STOP POINTS

As shown in the trajectory in Fig. 1 in Section I, two types of points, Stop Points(SP) and Move Points (MP), are included in a trajectory, and the distribution of Move Points is relatively regular. A real trajectory of user ‘000’ in GeoLife trajectory data [20]– [22] is given in Fig. 2(a). This user moves from position A to position B along the arrow direction. Fig. 2(b) is the schematic diagram of the user’s position (indicated by a black circle) at each time, the blue and orange boxes show the enlarged images in the corresponding color areas of this trajectory. The aggregation degree of the orange points in orange box is relatively high and the distribution of these points is irregular, these points are Stop Points. In addition, the points in the blue box in Fig. 2(b) are evenly spaced, and these points are the Move Points. Thus we can obtain the same conclusion as in Fig. 1, that the distribution of Move Points is relatively regular.
identifying different categories of trajectory points, an extended MC named NMA (Neighbourhood Move Ability) is proposed in [10]. The experiments in [10] also demonstrate that NMA is an effective way to measure the spatial characteristics of trajectory data points.

2) Spatial Influence (SI) [2], [23] evaluates the concentration of trajectory points in a certain area from the perspective of neighbors’ spatial distance and time. The main idea of SI comes from the data domain theory [24] in geography. This theory believes that all points in space will be affected by their neighbours, and the influence is inversely proportional to distance. The Move and Stop Points in a trajectory will also be affected by their neighbours within a certain range. Theoretically, the SI of the Move Points is less than that of the Stop Points (the density of Stop Points in their neighbourhood range is much higher than that of Move Points). Fig. 4 shows two examples of SI. In Fig. 4(a), points $P_2$ and $P_4$ can be affected by point $P_3$ within the radius of $R_1$, and points $P_1$, $P_2$, $P_4$ and $P_5$ can be affected by point $P_3$ within the radius of $R_2$. While in Fig. 4(b), points $P_1$ and $P_2$ can be affected by point $P_6$ within the radius of $R_1$, and points $P_1$, $P_2$ and $P_3$ can be affected by point $P_6$ within the radius of $R_2$. Through the above analysis, we can be found that SI is more vivid and easy to understand, but its calculation process is sensitive to the neighbourhood, as the SI of a point obtained under different neighbourhood radius may show great fluctuation.

3) Point Density (PD) [25] is also a feature indicator based on spatial distance. The PD indicator regrades the points within a point’s radius as neighbours, and counts the number of neighbours to approximate the density of the point. Due to the different aggregation degrees of the Stop Points and Move Points, their PD values also show obvious differences. The concept of PD is easy to understand and the calculation process is also very simple, while its disadvantage is that it is also sensitive to neighbourhood.

4) Average Neighbourhood Radius (NA) [26] obtains the neighbours of a point, and then approximately replaces the point by calculating the average distance between these neighbours and the point. The choice of neighbours greatly affects the effectiveness of NA. If the neighbour selection accuracy is high, this indicator may be able to distinguish the Stop and the Move Points. On the contrary, NA may not reflect any characteristic of trajectories. The average distance calculation may also overwhelm some features, causing positive and negative offsets. In addition, NA is sensitive to noise.

5) Direction Angle (DA) [27], [28] shows the features of trajectories by detecting the change of the rotation angle of the moving objects. Fig. 5 gives the DA of a moving and a stopping segment in a trajectory, where the blue dashed arrow is the movement direction of the trajectory. Ideally, a moving object may perform some special activities (shopping, rallying, etc.) in the stopping area, and the direction angle may change greatly and irregularly. Comparing Fig. 5(a) and
6) Velocity feature. The velocity feature [29], [30] includes Average Speed (AS) and Instantaneous Speed (IS). AS reflects the overall speed of a moving object in a period of time, while IS reflects the movement of a moving object at a certain time or position. AS can be measured by the ratio of the distance travelled by the moving object to the travel time, and IS of a point can be measured by the ratio of the distance between the predecessor and the successor to the travel time. Theoretically, the AS and IS can distinguish the Stop and Move Points, as the speed of Stop Points is relatively low and that of the Move Points is relatively high. The speed indicator is intuitive and easy to understand, its calculation process is relatively simple, but the low-speed points in the moving parts (caused by traffic jams, signal lights, etc.) may be incorrectly identified as Stop Points.

7) Space-time density [31] is a more effective measurement to calculate the temporal and spatial density of trajectory data. A more effective space-time density measurement appears currently is the NMAST (Neighbourhood Move Ability and Stay Time) density function given in [10]. Under this function, a Stop Point corresponds to a higher value, while others correspond to a lower value. Therefore, this function can also be applied to distinguish the Move and Stop Points. However, for complex trajectory, it may expense a lot to calculate the NMAST function value.

A comparison of the above different characteristics is given in Table 2. n is the number of data points in a trajectory and \(n'\) is the average neighbour of each data point under the neighbourhood radius \(R\). The complexity in Table 2 is the time overhead for computing the value of each indicator for an arbitrary point.

In Table 2, the time complexity of PD and IS is low, but IS is unreliable in distinguishing between Move and Stop Points. The NMAST, SI, PD and MC are all theoretically good at distinguishing between Move and Stop Points, but the time complexity of NMAST, SI is relatively high, so PD and MC are theoretically better choices. In addition MC and MC can play a certain complementary role, as the special clusters that cannot be found by MC satisfy the constraints of PD. To further verify the above theoretical analysis, experimental validation is performed, and the results of the experiments will be given in Section V-D.

### B. MOVING INDEX GAUSSIAN MODEL

1) POINT DENSITY

For each trajectory point \(P_i(X_i, Y_i, T_i)\), \(X_i\) and \(Y_i\) are the altitude and longitude of \(P_i\), respectively. \(T_i\) is the acquisition time of \(P_i\), \(0 \leq i \leq n\), \(n\) is the number of points in trajectory \(Tra\) (\(Tra = \{P_0, P_1, \ldots, P_i, \ldots, P_n\}\)). The Point Density of \(P_i\) (denoted as \(\rho(P_i)\)) is calculated as follows.

\[
\rho(P_i) = \begin{cases} \rho(P_i) + 1, & \text{Dis}(P_i, P_j) \leq R \\ \rho(P_i) + 0, & \text{Dis}(P_i, P_j) > R \\ \end{cases}, \quad 0 \leq i \neq j \leq n
\]

In Eq.(1), \(\text{Dis}(P_i, P_j)\) is the distance between \(P_i\) and \(P_j\) in \(Tra\). Drawing on the calculation method in [10], \(\text{Dis}(P_i, P_j)\) can be obtained according to Eq.(2); \(R\) is the neighbourhood radius; if the distance between \(P_i\) and other points in \(Tra\) is less than \(R\), the Point Density of \(P_i\) plus 1. The initial value
of \( \rho(P_i) \) is set to 0. Eq.(1) counts the number of trajectory points contained within the R domain of each point, and uses the number of points as the density of each point, which helps to reflect the difference in the degree of aggregation and scale between the Stop Points and Move Points in a trajectory.

\[
\text{Dis}(P_i, P_j) = 2\arcsin\sqrt{\sin^2(a/2) + A\sin^2(b/2)\gamma * r}
\]  
(2)

Eq.(2) [10] presents the method of calculating the distance between any two points in latitude and longitude coordinates. In Eq.(2) \( r \) is the radius of the earth in meters. \( A \) and \( B \) are the cosine values of \( \text{rad}X_i \) and \( \text{rad}X_j \), respectively. \( \text{rad}X_i \) and \( \text{rad}X_j \) are the radian values of the latitude of \( P_i \) and \( P_j \); \( a \) and \( b \) are respectively the difference of the radian values corresponding to the latitude of \( P_i \) and \( P_j \), their formula are presented in Eq.(3) and Eq.(4).

\[
A = \cos(\text{rad}X_i)
B = \cos(\text{rad}X_j)
\text{rad}X_i = X_i \times \pi / 180
\text{rad}X_j = X_j \times \pi / 180
a = \text{rad}X_i - \text{rad}X_j
b = \text{rad}Y_i - \text{rad}Y_j
\]  
(3)

2) MOVING INDEX
A candidate move point of \( \text{Tra} \), denoted as \( \text{CMP}(	ext{Tra}) \), is a subset of \( \text{Tra} \). \( \text{CMP}(	ext{Tra}) = \{ \text{cm}P_0, \ldots, \text{cm}P_j, \ldots, \text{cm}P_m \} \) (\( m \) is number of elements in \( \text{CMP} \), \( 0 \leq j \leq m \leq n \)). If the Point Density of a point is less than \( DT \), then the point can be added to \( \text{CMP} \). \( \text{CMP}(	ext{Tra}) = \{ \text{cm}P_0 | \rho(P_i) \leq DT, 0 \leq j \leq n \} \).

Since the actual trajectory points in real life are affected by many factors, such as outliers or noise, the quality of trajectory is often unpredictable. A role of \( DT \) is to attenuate the effect of noise. As the noise points are scattered and their density is much smaller than \( DT \), the noise points are also regarded as candidate Move Points. The above approach is able to separate the noise points from the Stop Points, thus avoiding the interference of noise points on Stop Point extraction. In addition, the points with density exceeding the \( DT \) limit are not necessarily the Stop Points. For example, a large number of points in a traffic jam are concentrated in a small area, but these points are false stops which do not correspond to a specific location. Similarly, points with density less than the \( DT \) limit are not necessarily moving points. For example, the blockage of buildings makes the density of points in some areas less than \( DT \), but the area may be a meaningful geographical location. Thus, it is necessary to further cull the candidate mobile points.

We already know that \( PD \) and \( MC \) are more effective in distinguishing Move and Stop points. The calculation method of \( PD \) has been given in Eq.(1); \( MC \) is the theory proposed in [2] describing the moving properties of the points. In order to further reflect the moving ability of the trajectory segment on the basis of examining the distribution characteristics of trajectory points, we correct \( PD \) with the help of \( MC \) and propose an enhanced \( PD \) under moving index, named \( MPD \). The \( MPD \) of point \( P_i \) in trajectory can be calculated as follows.

\[
\begin{align*}
\text{MPD}(P_i) &= e^{-\frac{MC(P_i)}{\pi}} \times \rho(P_i) \\
\text{MPD}(P_i) &= \frac{DDis(P_{j1}, P_{jn})}{CDis(P_{j1}, P_{jn})}, P_{j1}, P_{jn} \in \text{NP}(P_i) \\
\end{align*}
\]  
(5)

In Eq.(5), \( \text{NP}(P_i) \) is the set of neighbours within the R-neighbourhood of \( P_i \); \( P_{j1} \) and \( P_{jn} \) are the first and last neighbours of point \( P_i \), respectively. \( DDis(P_{j1}, P_{jn}) \) is the displacement between \( P_{j1} \) and \( P_{jn} \); \( CDis(P_{j1}, P_{jn}) \) is the total travelling distance from \( P_{j1} \) to \( P_{jn} \); \( \text{MC}(P_i) \) is the mobility value of point \( P_i \), its calculation method is the same as in [2]. According to this equation, the \( MPD \) of all the points in a trajectory can be seen as a function with respect to \( MC \) and \( \rho \) when \( R \) is constant. At this time, \( MPD \) is inversely proportional to \( MC \), and positively proportional \( \rho \). This reflects the distribution property of points in a trajectory, i.e., the fewer the number of neighbours in the \( R \) range of a point and the greater the mobility of neighbours, the greater the probability that these points are mobile points. The above function allows us to further refine the CMP, i.e., the greater the movement capacity the smaller the density is the more likely to be a Move Point.

3) MOVING INDEX GAUSSIAN MODEL
For all candidate Move Points in \( \text{CMP}(	ext{Tra}) \), calculate the mean and variance of the \( MPD \) of these points, denoted as \( \mu \) and \( \sigma \). The \( \text{MIGM} \) model is as follows.

\[
\text{MIGM}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\text{MPD}(P_i) - \mu)^2}{2\sigma^2}}
\]  
(6)

In which \( \text{MPD}(P_i) \) is the enhanced \( PD \) of \( P_i \) in \( \text{Tra} \), it can be calculated according to Eq.(5), \( n \) is the number of points in \( \text{Tra} \). Since the \( \text{MIGM} \) model is constructed using the \( MPD \)
Algorithm 1: A Trajectory Clustering Method Based on Points Moving Index Analysis and Modelling (PMS)

Require: \( D \): trajectory dataset; \( R \): neighbourhood radius; \( DT \): the maximum point density threshold.

Ensure: \( C \): Stop Points cluster.

1. Initialize \( R=10 \) and \( DT=35 \);
2. repeat
   3. Calculate each point’s \( PD \) value according to Eq.(1);
   4. Mark the points with \( PD \) values less than \( DT \);
   5. Obtain \( MPD \) of all trajectory points according to Eq.(5);
   6. Calculate the \( \mu \) and \( \sigma \) of the \( MPD \) of all marked points;
   7. Establish \( MIGM \) model of \( \mu \) and \( \sigma \) according to Eq.(6);
   8. Fit the trajectories with \( MIGM \) and extract the points that are not fitted by \( MIGM \);
9. until The convergence condition in Eq.(7) is satisfied;

IV. THE PMS ALGORITHM

This section presents the working steps of \( PMS \). The pseudo-code of \( PMS \) and some brief explanations are given.

A. THE WORKFLOW OF PMS

The pseudo-code for the above procedure in \( PMS \) algorithm is given in Algorithm 1.

B. ANALYSIS OF PMS

1) MIGM MODEL ESTABLISHMENT

In Algorithm 1 step3 and step4, the Point Density of each point is calculated according to Eq.(1) with the help of \( R \), and then the points with \( PD \) of each point is added \( CMP \), its time complexity is \( o(n^2) \).

After obtaining \( CMP \) of trajectory data, the Point Density model of Move Points, called \( MIGM \), can be obtained according to Algorithm 1 step5 and step7. The establishment of \( MIGM \) includes the following steps:

- Traversing all the \( MPD \) of all points, its time complexity is also \( o(n) \).
- Calculating the mean value \( \mu \) and variance \( \sigma \) of all the marked elements in \( MPD \), and establishing the \( MIGM \) model of \( \mu \) and \( \sigma \), its time complexity is \( o(n) \).

2) MIGM MODEL FITTING

Fitting all trajectories with \( MIGM \) according to Algorithm 1 step8 (its time complexity is \( o(n) \)). In this process, each point in trajectories corresponds to a specific \( MPD \) value. This value is used to indicate the possibility that the point can be regarded as a Move Point. When this value is between a certain interval (fitting interval), the probability of the point being a Move Point is 1, otherwise it is 0.

According to the \( 3\sigma \) criterion of Gauss function, a set of test data is assumed to contain only random errors and the standard deviation can be obtained by calculating and processing them. It is considered that the errors beyond the fitting interval are gross errors, and the data containing gross errors are not considered when building the \( MIGM \) model (only the mean and variance of \( MPD \) values of candidate Move Points are calculated). The corresponding data of gross errors in this method are the Stop Points we are looking for.

Using the \( 3\sigma \) distribution characteristics, the value of the elements in \( MPD \) approximate representation is introduced into the \( MIGM \) (\( \mu, \sigma^2 \)). The probability that the fitting value of each element in \( MPD \) approximate representation is located in the interval \((\mu - \sigma, \mu + \sigma)\) is 0.6826; the probability that the fitting value is located in the interval \((\mu - 2\sigma, \mu + 2\sigma)\) is 0.9544; and the probability that the fitting value is located in the interval \((\mu - 3\sigma, \mu + 3\sigma)\) is 0.9974. Therefore, the distribution of fitting values is almost all concentrated in intervals \((\mu - 3\sigma, \mu + 3\sigma)\), and the possibility of exceeding this interval is less than 0.3%. So, the lower bound of fitting is taken as \( \mu - 3\sigma \) or \( \mu + 3\sigma \). Meanwhile, the corresponding value of \( \mu - 3\sigma \) or \( \mu + 3\sigma \) is equal according to the symmetrical distribution of Gauss function. Therefore, we take \((\mu, \mu + 3\sigma)\) as the upper and lower bounds of the fitting interval of \( MIGM \) model. The points that do not satisfy the upper and lower bounds are regarded as Stop Points.

3) MIGM MODEL ADJUSTMENT

In order to achieve more desirable extraction results, the value of \( R \) needs to be changed to find a relatively optimal \( MIGM \) model. In this step, the following processes need to be performed:

- Extracting the points which do not satisfy the upper and lower bounds of the fitting interval;
- Judging whether the points extracted in the above step satisfy the convergence condition. If these points satisfy the convergence condition, then these points are the Stop Points. Otherwise, the value of \( R \) will be adjusted and steps3 to step8 in Algorithm 1 will be executed. In this paper the convergence condition \( E(((A_{1x}, A_{1y}), (A_{2x}, A_{2y}), \ldots, (A_{kx}, A_{ky}))_k) \) (denoted as \( E \)) is given below.

\[
E = \frac{1}{2j=1} \sum_{j=1}^{N} ((X_{ji} - A_{ji})^2 + (Y_{ji} - A_{ji})^2)
\]

In which \( k \) is the number of clusters that the Stop Points formed, \( N \) is the number of points in cluster i, \( A_{1x} \) is the mean latitude of the points in the first cluster, and \( A_{1y} \) is the mean longitude of the points in the first cluster, so, \((A_{1x}, A_{1y})_1\) stands for the central point of the first cluster.
TABLE 3. Preprocessed target trajectory data.

| Points | Latitude | Longitude | Time          |
|--------|----------|-----------|---------------|
| 1      | 39.97770000000000 | 116.304011666667 | 2012-04-10,14:38:54 |
| 2      | 39.97756333333333 | 116.303376666667 | 2012-04-10,14:38:59 |
| 3      | 39.97748333333333 | 116.303250000000 | 2012-04-10,14:39:04 |

\( x_i \) and \( y_i \) are the latitude and longitude of the \( i \)-th point in cluster \( j \).

The above formula determines whether the model converges by examining the change of centroid position \((A_{ji}, A_{jy})\) in each Stop Points cluster. When the cluster centres are relatively fixed, the found cluster structure has been relatively stable. If the convergent value of the extracted Stop Points in Eq. (7) does not fluctuate greatly with the change of \( R \), these extracted points are the Stop Points in trajectories.

The time complexity of convergence judgement approximate \( o(Nk) \), \( Nk < n \). As the optimal \( R \) of each trajectory is different, the number of model adjustments may be different. Assuming that the number of model adjustments is \( t \), the overall time complexity of PMS approximate \( o(t(n^2 + 4n)) \).

V. EXPERIMENTS AND RESULTS

In this section, the effectiveness of several trajectory characteristics in Section III-A and the performance of PMS are evaluated by using GeoLife dataset [20]–[22]. The experiments are conducted on a PC with core Intel(R)Xeon(R) E-2186M processor with 2.9 GHz and 32 GB RAM running Windows 10 Ultimate service pack-1.

A. DATASET DESCRIPTION

The GeoLife dataset was collected from April 2007 to August 2012 by Microsoft Research Asia. This data set contains 1,7621 trajectories of 182 users and is represented by a series of time-stamped points recorded every 5 − 10m or 1 − 5s. Through data preprocessing, 200 trajectories from different users are selected to carry out our experiments, and a preprocessed trajectory data is given in Table 3.

B. COMPARISON ALGORITHMS

In this study, some classical algorithms in trajectory clustering are used to compare with the proposal of this paper, these selected algorithms are briefly described below.

1) DBSCAN (Density-Based Spatial Clustering of Applications with Noise, [6]). It is a more representative density-based clustering algorithm, which defines a cluster as the largest set of densely connected points and discovers clusters of arbitrary shapes in a spatial database of noise. DBSCAN requires two parameters: the radius-\( \text{Eps} \) and the minimum number of neighbours-\( \text{minPts} \). Considering that the experimental dataset is the activity trajectory of urban residents, the values of the above two parameters are set to 50 and 35 respectively.

2) DJ-Cluster (Density-and-Join-based algorithm, [7]). It is an extension of the DBSCAN algorithm. For each point in trajectory, it calculates its neighbourhood with the help of parameters \( \text{Eps} \) and \( \text{minPts} \). For any point, a new cluster is created if it has no neighbours belonging to an existing cluster, or it is joined with an existing cluster if it has neighbours belonging to an existing cluster. According to the parameter discussion in [7], \( \text{Eps} \) and \( \text{minPts} \) take the values of 20 and 18, respectively.

3) CB-SMoT (Clustering based Stops and Moves of Trajectories, [8]). It is a two-step algorithm to extract stops and moves. First, some potential stops are identified by an improved DBSCAN algorithm. Second, clusters are found considering the geography behind the trajectories. The \( \text{Eps} \) parameter is involved in CB-SMoT. In order to solve the difficulty of selecting \( \text{Eps} \), a parameter called \( \text{area} \) is introduced to compute \( \text{Eps} \). The value of \( \text{area} \) is determined as 0.4 with reference to the usage in [8].

4) TAD (Trajectory clustering algorithm based on density analysis, [10]). This algorithm gives two new metrics - NMAST (Neighbourhood Move Ability and Stay Time) density function and NT (Noise Tolerance) factor - to find stop points clusters in trajectories. Neighbourhood radius \( R \), neighbourhood moving ability feature \( \sigma \) and minimum density threshold \( mD \), minimum noise threshold \( mNT \) are the parameters contained in TAD. According to the results of parameter analysis in [10], we set the values of the above parameters as 50, 0.5, 0.08, 0.35.

5) LMD (Local Maximum Density, [14]). LMD identifies local hotspots with a sufficient number of stops, and three steps: local maximum determination, neighbourhood reshaping, and popular local hotspot determination are involved in its main procedure. A minimum number of stops (\( \text{minPts} \)) is needed to determine whether a local hotspot is popular. The \( \text{minPts} \) is set to 35 in our comparison study.

6) The algorithm in [2]. This algorithm combines the move ability and the data fields theory to construct a density measure, and then improves the DBSCAN algorithm with the newly defined density measure to extract interesting clusters in trajectories. Three parameters of \( \sigma \text{MA} \) (moving ability feature), \( \sigma_1 \) (interactive impact factor between trajectory points) and \( \text{Nap} \) (the number of adjacent points considered in density calculation) are used to guide the operation of the algorithm. In our comparison study, these parameters are assigned as 0.5, 0.3, 51, respectively.

C. PARAMETER ESTIMATING

Two parameters, \( DT \) and \( R \), are contained in this paper. Parameter \( DT \) is a truncation threshold used to control whether a point can be considered as a candidate Move Point. If a point is regarded as a candidate Move Point, it means...
that the two regions connected by this point are the clusters of the Stop Points and these two clusters can not be merged.

In addition, $DT$ can also play a role in reducing the effect of noise. The noise points are also considered as candidate move points due to their small density, which can distinguish the noise from the Stop Points and avoid the interference of noise points on the Stop Points clustering. The value of $DT$ is set to 35 according to [2] in this paper.

$R$ is used to determine the size of the neighbourhood range to be searched for when searching for candidate Move Points. If $R$ is too large, the search space of candidate Move Points will be enlarged. If $R$ is too small, the detection of candidate Move Points will be incomplete. Therefore, an appropriate $R$ is very important to improve time efficiency and extraction accuracy. Considering that the test data are the travel data of the residents in the urban area, and the data acquisition accuracy of GPS positioning equipment in urban areas is usually about 50m [2], so, the value of $R$ should not exceed 50 m in theory. In order to find clusters of different sizes, we start from a small initial $R$ value so as to prevent missing neighbour information of the small-scale cluster, the range of $R$ is 10-50m. Fig. 6 shows the specific parameter adjustment process of a certain trajectory.

Fig. 6(a) and Fig. 6(d) are the fitting image of the $MIGM$ model when $R$ is 10m. In these figures, the points marked by the black circle are the points fitted by the $MIGM$ model, which is the Move Points; the points marked by the asterisk are the Stop Points. Fig. 6(d) shows a magnified image of the red circle in Fig. 6(a). It can be observed that most of the Stop Points are fitted by $MIGM$, so $R$ needs to be further adjusted. Fig. 6(b) shows the fitting image when $R$ is 30m. Fig. 6(b) changes to some extent on the basis of Fig. 6(a). Further observation of the enlargement in Fig. 6(e) finds that more Stop Points are recognized when $R$ is 30m. However, the points extracted at this time do not satisfy the convergence condition, we increase the value of $R$ to 50m, and give the fitting image in Fig. 6(c) and Fig. 6(f). Combining Fig. 6(c) and Fig. 6(f), the Stop Points in the red region have not been fitted and the structure of the cluster located in the red circled area has been basically found. When $R$ is further adjusted, the value of the convergence condition does not fluctuate. Therefore, the final value of $R$ is 50m.

D. EXPERIMENTAL EVALUATION OF THE TRAJECTORY CHARACTERISTICS

This section randomly selects multiple trajectories of user No.000 in GeoLife trajectory datasets to further compare the characteristics described in Section III-A, and draws feature curves of each trajectory under different characteristics. The parameters contained in the indicators are set according to the discussion in [2] and [10]. The specific feature curves are given in Fig. 7.

As shown in Fig. 7 (Trajectory1, 2, ... represent the first, second trajectory of user '000'), each subgraph shows eight feature curves of $NMAST$ (left first curve of each subgraph), $SI$(left second curve of each subgraph), $PD$(left third curve of each subgraph), $MC$(left fourth curve of each subgraph), $DA$(right first curve of each subgraph), $NA$(right second curve of each subgraph), $IS$(right third curve of each subgraph), $AS$(right fourth curve of each subgraph). The $NMAST$, $SI$ and $PD$ curves show significant positive correlation, and the $MC$
curves show obvious negative correlation with the above three types of curves. The main reason is that the NMAST, SI and PD indicators are all density related measures. The peaks in the NMAST, SI and PD curves are the potential clusters that may exist in trajectories. In these potential clusters, the points gather in a short time. Therefore, the points in the peak areas in the first three curves are the Stop Points, while others are the Move Points. As the moving ability of Stop Points
is smaller than that of the Move Points, the peak positions in the NMAST, SI, and PD curves are depressed in the MC curves. Based on the above analysis, it can be drawn that the NMAST, SI, PD and MC have relatively stable performance, some characteristics still contained in these indicators with the change of trajectory.

Carefully observe the characteristic curves with relatively stable performance mentioned above, we notice that the SI and PD curves are very similar, as they both reflect the concentration of trajectory points. Their differences are that SI absorbs the data field theory in geography to show the concentration degree of data, while in PD the number of neighbours is counted to achieve the above goal. Compared with the above stable indicators, the laws of DA, NA, IS and AS curves are not obvious. For the DA curves in Fig. 7, only the peak areas in Fig. 7(b), 7(d), 7(h), 7(j), 7(k) are negatively correlated with that in NMAST, PD, and SI curves. Thus, the assumption that the direction angle change of the Stop Points may be relatively irregular does not hold. The performance of DA indicator in distinguishing the Stop and Move Points is not stable, the goal of distinguishing the Stop and Move Points cannot be achieved by using DA alone.

In all NA curves from Fig. 7(a) to Fig. 7(o), Only Fig. 7(e), Fig. 7(k) and Fig. 7(o) have no obvious protrusion regions, which show weak corresponding relationship with NMAST, PD and SI. Few characteristics can be seen in other NA curves except that in Fig. 7(e), Fig. 7(k) and Fig. 7(o). So, the method of using average distance of neighbour points to reflect a certain point in NA can not work very well in describing trajectory features. Even for the trajectory with obvious characteristics, the laws in the NA curves will be weakened. Thus, the performance of the NA indicator is extremely unstable, and it can not be used to distinguish the Stop and Move Points.

As shown in Fig. 7(a) - Fig. 7(o), the IS curves in eight subgraphs (Fig. 7(a), Fig. 7(b), Fig. 7(c), Fig. 7(e), Fig. 7(f), Fig. 7(j), Fig. 7(k)) hardly contain any valuable information. The remaining eight IS curves contain some particularly obscure bulges, some of which have higher speed and roughly correspond to the moving parts in trajectories. Comparing the IS and AS curves in Fig. 7, we find that most of these curves are similar. But the laws in the AS curves are more obvious than those in the IS curves, and some of the AS curves (as shown in Fig. 7(h) and Fig. 7(o)) are directly proportional to the MC curves. The performance of the AS and IS are both unstable, but the overall performance of AS is better. Thus, using speed to define the trajectory points may not rigorous, and we cannot completely rely on the speed indicator to distinguish different types of trajectory points.

A summary of the above analysis is given in Table 4. The NMAST, SI, PD, and MC are relatively stable, and the changes in trajectory complexity have little effect on them. However, when more clusters with different sizes or aggregation degrees are contained in trajectories, the differences between the peaks in the NMAST curves become large, and the computational complexity of NMAST and SI increases a lot. Another important finding is: the difference in stopping area is obviously greater than that in the moving areas, as there are multiple peaks in the stopping areas, while the moving area is relatively smooth. That is to say, the characteristics of Move Points are relatively simple. Therefore, PD and MC are used to analyse the characteristics of trajectory points, and finally complete the task of extracting Stop Points.

### E. EXPERIMENTAL EVALUATION OF THE PMS ALGORITHMS

In this section, the performance of PMS algorithm is evaluated comparing with the algorithms introduced in Section V-B. Table 5 shows the Precision, Recall and Error Rate value of five different algorithms. In this table, Total - SP is the total number of clusters of Stop Points in 200 trajectories; Found - SP, Realised - SP and Undetected - SP stand for the number of clusters of Stop Points found, realised and undetected by these five algorithms, respectively. In this table, we can see that the Precision of PMS is the highest, and the Precision of DBSCAN is the lowest. This phenomenon shows that all the algorithms except DBSCAN have certain advantages in processing trajectory data. The reason why PMS algorithm has high Precision is that we have adjusted the value of the parameter R in the process of model fitting, which enables this method to find clusters of different scales with higher Precision. On the contrary, the R values in other methods are relatively fixed so that some clusters are not completely found or even missed.

The TAD algorithm has the highest Recall value, but the Precision value of TAD is significantly lower than that of PMS. The main reason is that the TAD detects more stop clusters than the real existing clusters, while the number of clusters detected by the PMS algorithm is close to the real
cluster number. Evidently, the PMS and TAD algorithms in this paper have their own advantages. TAD is very good at discovering clusters, but the cluster detection accuracy of PMS is higher than that of TAD.

In addition to the above algorithms, the Precision of the LMD algorithm in Table 5 is relatively high, and its Recall value of it also ranks third in all methods. This is mainly due to the advantages of both time-based clustering and distance-based clustering are leveraged to effectively extract significant locations. Combining Precision and Recall, we find that the performance of PMS is satisfactory. This is mainly due to the fact that the Stop Points are extracted from the characteristics of points’ feature analysis, which can grasp the characteristics of the Stop Points more accurately.

To further analyze the results of PMS, this paper gives the fitting images of PMS on different trajectories.

Fig. 8 shows the PMS fitting images on different trajectories, and the dotted lines with different colors indicate the regions (or clusters) where the Stop Points are located. Compared with these sub-pictures, we find that there is only one area marked with the red line in Fig. 8(a), 8(b), 8(c) and 8(i). The points in these areas are the Stop Points as they are highly significant.
aggregated. In addition, the distributions of the remaining unselected points in Fig. 8(a), 8(b) and 8(c) are relatively regular and orderly, while the distributions of unselected points in Fig. 8(i) are scattered and disorderly. All the clusters in the above figures are found. It can be seen that PMS has good robustness.

Further observation of Fig. 8(d), 8(e) and 8(f) show that the red and blue regions in these pictures are relatively close. From the perspective of Stop Points analysis and extraction, these clusters that are closer to each other may be wrongly merged. By analysing the data characteristics of Move Points, this paper uses PD to build a model to fully reflect the characteristics of Move Points. Through the above strategies, the PMS algorithm in this paper can accurately segment the clusters which are close to each other, avoid the wrong merging of clusters, and improve clustering accuracy.

Comparing the clusters marked with different colors in Fig. 8, it is found that not only the size of the clusters in different trajectories are different, but also the clusters in the same trajectory. Using the global unique $R$ to process the above trajectories may lead to the situation that the cluster discovery is incomplete or misidentified or even unrecognized. In this paper, by adjusting $R$, clusters of different scales are found and the cluster members are relatively complete. For the case where the size of clusters in the same trajectory is very different (as shown in Fig. 8(h)), the characteristics of large-scale clusters are shown as extremely large characteristic values, which masks the characteristics of small-scale clusters, thus directly neglecting small-scale clusters. The PMS algorithm accurately identified the two clusters with very different scales presented in Fig. 8(h). The main reason for the above phenomenon is that this paper analyses the Move Points’ characteristics, and realizes the discovery of high-precision Stop Points, so as to solve the above problems from the other side. In addition, the shapes of the clusters in Fig. 8 are irregular and diverse, which fully demonstrates the ability of PMS to find clusters of different shapes.

From all the above analysis, it can be concluded that PMS is a simple and effective Stop Points extraction method. By accurately grasping the characteristics of trajectory points, PMS can not only find clusters of different shapes, but also reduce the error merging of adjacent clusters, so as to improve the clustering accuracy.

VI. CONCLUSION

We have presented a trajectory clustering method based on Stop and Move Points feature analysis and modelling. Starting from the analysis of various trajectory characteristics, this paper defines a moving index based on the PD and MC indicator to reflect the features of trajectory points, and proposes a clustering algorithm called PMS. PMS abstracts the multidimensional trajectory data into a MPD approximate representation, then establishes a Point Density Gauss Model (MIGM) based on the moving index, and eliminates the Move Points in trajectories through MIGM to extract the Stop Points. Compared with the existing technologies, the main advantages of this method include: (1) From the point of view of feature analysis of trajectory Points, a moving index and a MIGM model are established to cluster the Stop Points. By this method, the Stop Points and Move Points can be distinguished more accurately, reducing the error combination of adjacent clusters, and obtaining more accurate Stop Points extraction results. (2) By adjusting the values of $R$, the accuracy of Stop Points extraction is improved. It can find the Stop Points with different aggregation scales in the same or different trajectories, and is not limited by the shape of clusters. (3) This method can also be used as a preprocessing technology for Stop Points extraction in many existing methods. Firstly, the Move Points in a trajectory are roughly culled, and then the characteristics of the remaining points can be further analysed.

We plan to extend our work in several directions. First, although the adjustment process of the neighbourhood radius $R$ is discussed in the paper, its adjustment process relies on priori knowledge. Therefore, we will further investigate a method that automatically adjusts the parameter $R$. Second, the clustering efficiency usually decreases when there is too much trajectory data. Studying the parallel strategy of the PMS algorithm will help to alleviate the above phenomenon. Finally, the algorithm proposed in this paper is only used for clustering human activity trajectory datasets, and it will be the subsequent research direction of this paper to explore the application of this method on other types of trajectory datasets, so as to solve the trajectory clustering problems in different fields.

REFERENCES

[1] Z. Yu, “Trajectory data mining: An overview,” ACM Trans. Intell. Syst. Technol., vol. 6, no. 3, pp. 1–41, May 2015.
[2] T. Luo, X. Zheng, G. Xu, K. Fu, and W. Ren, “An improved DBSCAN algorithm to detect stops in individual trajectories,” ISPRS Int. J. Geo-Inf., vol. 6, no. 3, pp. 1–16, Feb. 2017.
[3] C. Comito, D. Falcone, and D. Talia, “Mining human mobility patterns from social geo-tagged data,” Pervasive Mobile Comput., vol. 33, pp. 91–107, Dec. 2016.
[4] Z. Zheng, D. Xia, X. Zhao, L. Tan, H. Li, L. Chen, and W. Liu, “Spatial–temporal travel pattern mining using massive taxi trajectory data,” Phys. A, Stat. Mech. Appl., vol. 501, pp. 21–41, Jul. 2018.
[5] L. Xiang, M. Gao, and T. Wu, “Extracting stops from noisy trajectories: A sequence oriented clustering approach,” ISPRS Int. J. Geo-Inf., vol. 7, pp. 1–18, Mar. 2016.
[6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Proc. Int. Conf. Knowl. Discovery Data Mining, vol. 96, Jan. 1996, pp. 226–231.
[7] C. Zhou, F. Dan, P. Ludfrog, S. Shekhar, and L. Terveen, “Discovering personally meaningful places: An interactive clustering approach,” ACM Trans. Inf. Syst., vol. 25, pp. 12–42, Jul. 2007.
[8] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, “A clustering-based approach for discovering interesting places in trajectories,” in Proc. ACM Symp. Appl. Comput., Mar. 2008, pp. 863–868.
[9] Z. Fu, Z. Tian, Y. Xu, and C. Zhao, “A two-step clustering approach to extract locations from individual GPS trajectory data,” ISPRS Int. J. Geo-Inf., vol. 5, no. 10, pp. 1–17, Sep. 2016.
[10] Y. Yang, J. Cai, H. Yang, J. Zhang, and X. Zhao, “TAD: A trajectory clustering algorithm based on spatial-temporal density analysis,” Expert Syst. Appl., vol. 139, pp. 112846.1–112846.16, Jan. 2020.
X. Zhou, F. Miao, H. Ma, H. Zhang, and H. Gong, "Deep learning-powered vessel trajectory prediction for improving smart traffic services in maritime Internet of Things," IEEE Trans. Netw. Sci. Eng., early access, Jan. 7, 2022, doi: 10.1109/TNSE.2022.3140529.

J. Tang, L. Liu, J. Wu, J. Zhou, and Y. Xiang, "Trajectory clustering method based on spatial-temporal properties for mobile social networks," J. Intell. Inf. Syst., vol. 56, no. 1, pp. 73–95, Feb. 2021.

K. Qin, Q. Zhou, T. Wu, and Y. Xu, "Hotspots detection from trajectory data based on spatial-temporal data field clustering," Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., vol. XLII-2/W7, pp. 1319–1325, Sep. 2017.

X.-J. Chen, Y. Wang, J. Xie, X. Zhu, and J. Shan, "Urban hotspots detection of taxi stops with local maximum density," Comput., Environ. Urban Syst., vol. 89, pp. 101661.1–101661.15, Sep. 2021.

C. Liu and C. Guo, "STCDD: Semantic trajectory clustering based on community detection in networks," Expert Syst. Appl., vol. 162, pp. 113689.1–113689.20, Dec. 2020.

M. Liang, R. W. Liu, S. Li, Z. Xiao, X. Liu, and F. Lu, "An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation," Ocean Eng., vol. 225, pp. 108803.1–108803.16, Apr. 2021.

X. Zhou, F. Miao, H. Ma, H. Zhang, and H. Gong, “A trajectory regression clustering technique combining a novel fuzzy C-means clustering algorithm with the least squares method,” ISPRS Int. J. Geo Inf., vol. 7, pp. 1–23, Apr. 2018.

D. Yao, C. Zhang, Z. Zhu, Q. Hu, Z. Wang, J. Huang, and J. Bi, “Learning deep representation for trajectory clustering,” Expert Syst., vol. 35, no. 2, pp. e12252.1–e12252.16, Feb. 2018.

J. Cai, H. Wei, H. Yang, and X. Zhao, “A novel clustering algorithm based on DPC and PSO,” IEEE Access, vol. 8, pp. 88200–88214, 2020.

Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, “Mining interesting locations and travel sequences from GPS trajectories,” in Proc. 18th Int. Conf. World Wide Web (WWW), Madrid, Spain, Apr. 2009, pp. 791–800.

Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, “Understanding mobility based on GPS data,” in Proc. ACM Int. Conf. Ubiquitous Comput., Seoul, South Korea, 2008, pp. 312–321.

Y. Zheng, X. Xie, and W.-Y. Ma, “GeoLife: A collaborative social networking service among user, location and trajectory,” IEEE Data Eng. Bull., vol. 33, nos. 3–4, pp. 32–40, Jun. 2010.

L. Xiand and X. Shao, “Visualization and extraction of trajectory stops based on kernel-density,” Acta Geodaeticae Cartographica Sinica, vol. 45, pp. 1121–1131, Apr. 2016.

X. W. Li, C. X. Cao, and C. Chang, “The first law of geography and spatial temporal proximity,” Chin. J. Nature, vol. 29, pp. 69–71, Apr. 2007.

S. Q. Li, C. Y. Sheng, and J. C. Wang, “An algorithm of cartographical generalization for point cluster features based on point density analysis and adaptive difference detection,” Geog. Geo-Inf. Sci., vol. 35, pp. 1–5 and 151, Mar. 2019.

Y. Takei, R. Grasso, and A. Berthoz, “Quantitative analysis of human walking trajectory on a circular path in darkness,” Brain Res. Bull., vol. 40, nos. 5–6, pp. 491–495, Jan. 1996.

H. Zhang, X. Hu, W. Bai, H. Zhang, Y. Zhou, and Y. Lin, “An efficient IoT trajectory compression method in vehicle terminals using width-direction-angle,” IEEE Access, vol. 7, pp. 71447–71458, 2019.

H. Ailin, L. Zhong, and Z. Dechao, “Cluster-based trajectory overall trend extraction,” in Proc. 36th Chin. Control Conf. (CCC), Dalian, China, Jul. 2017, pp. 103–107.

R. Zhang, P. Xie, C. Wang, G. Y. Liu, and S. H. Wan, “Classifying transportation mode and speed from trajectory data via deep multi-scale learning,” Comput. Netw., vol. 162, pp. 1–13, Oct. 2019.

Y. Minegishi, “Experimental study of the walking behavior of crowds mixed with slow-speed pedestrians as an introductory study of elderly-mixed evacuation crowds,” Fire Saf. J., vol. 120, pp. 103041.1–103041.10, Mar. 2021.

J. Cai, H. Wei, H. Yang, and X. Zhao, “A novel clustering algorithm based on DPC and PSO,” IEEE Access, vol. 8, pp. 88200–88214, 2020.

J. Cai, H. Wei, H. Yang, and X. Zhao, “A novel clustering algorithm based on DPC and PSO,” IEEE Access, vol. 8, pp. 88200–88214, 2020.

J. Cai, H. Wei, H. Yang, and X. Zhao, “A novel clustering algorithm based on DPC and PSO,” IEEE Access, vol. 8, pp. 88200–88214, 2020.