A Trajectory Abnormal Detection Method Based on Segmentation and Clustering

Wanting Qin¹, Jun Tang¹*, Cong Lu¹ and Songyang Lao¹

¹College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, 410073, China
*Corresponding author’s e-mail: tangjun06@nudt.edu.cn

Abstract. The early warning of abnormal behaviour is helpful to explore the potential risk and avoid the occurrence of accidents. Therefore, it is of great significance to excavate abnormal trajectory patterns from a large number of trajectories. From the perspective of unsupervised learning, a two-level trajectory abnormal detection method based on segmentation and clustering is proposed in this paper. Firstly, the method considers the influence of a time factor in the stages of segmentation, and segments by the constraint of trajectory state. Then, it redefines the distance between the trajectory segments according to the temporal and spatial distance to cluster after segmentation. Finally, abnormal detection is carried out according to clustering results. In the abnormal detection phase, a two-level detection method is adopted. Firstly, the abnormal trajectory segment is found by first-level coarse-grained abnormal detection. Then, the abnormal subtrajectory segment is detected by second-level fine-grained abnormal detection. The real Atlantic Ocean hurricane data is used for the simulation experiments. Compared with the classical TRAOD detection method, proposed method is more consistent with the actual situation.

1. Introduction
The pattern mining based on trajectory data aims to extract the common features of many moving objects from the massive trajectory set, and the corresponding type is the anomaly discovery pattern oriented to trajectory big data[1]. An anomaly is also called an outlier. Abnormal trajectories can be viewed as events that do not conform to certain expected patterns, or as behaviors that differ from those of other objects according to similarity criteria (such as travel time or data distribution) [2]. On October 12, 2019, the typhoon ‘Hagibis’ landed on the Izu Peninsula in Shizuoka-ken, Japan. This typhoon killed 88 people, and 7 people went missing. More than 3900 people were affected during this typhoon. The immediate needs cannot be met if emergency measures are taken only after a disaster occurs. Therefore, in order to reduce the losses caused by such disasters as hurricanes, it is urgent to shift from post-disposal to pre-anomalous warning of anomalous behavior. The purpose of trajectory abnormal detection is to discover abnormal patterns with common characteristics which are different from the common ones from a large number of trajectory data. By mining abnormal patterns in the trajectory, potential risk conflicts can be effectively discovered, and measures can be prepared in advance to avoid some accidents [3,4].

Common abnormal detection methods include distance-based detection method [5-10], density-based detection method [11,12], mesh-based detection method [13-16] and classification-based detection method [17,18]. Distance-based detection method is the most commonly used one. This method is mainly based on some distance measurement standard to find the number of adjacent trajectory and the trajectory to be detected in the historical trajectory. When the adjacent trajectory is
less than a certain threshold value, it is judged as abnormal. Knorr [5,6] et al. first proposed the traditional distance-based trajectory abnormal method. This method utilizes the global attributes of trajectories and ignores the local differences between trajectories; hence, it is suitable for simpler trajectories. The threshold selection of distance-based detection method is usually a global threshold, which cannot take into account local anomalies. Therefore, someone proposes a density-based detection method. Liu et al. [11] proposed an abnormal detection method combining density and distance (DBTOD). Mesh-based detection method is to divide the moving region of the moving object into equally large meshing region. The abnormal grid sequences are identified from these grid regions. Chen [13,14] defines the number of grid cells required for anomaly detection as the window size, and proposes an iBOAT method based on fixed window size and adaptive window size. Classification-based detection methods can be roughly divided into two stages: training stage and testing stage. In the training stage, the classifier is trained by using labeled data. In the test stage, the test data is divided into normal and abnormal according to the trained classifier, but this method usually requires the data to carry abnormal label. Reference [19-20] proposed a classification method based on SVM. The theoretical premise of this method is that the abnormal data only account for a small number in the whole data set, and if the target trajectory is judged as abnormal by the classifier, it is an abnormal trajectory. Classification-based detection is relatively more accurate, but this method requires the abnormal label. Most of the trajectory data does not have abnormal label, which brings difficulty for trajectory abnormal detection. The problem that there is no abnormal label on the trajectory can be solved by manual annotation, but manual annotation by experts is needed when marking data. If you want to get accurate data label will lead to a lot of manpower and time cost. Unsupervised learning can also solve the problem of trajectory without abnormal label. Clustering, as a kind of unsupervised learning, is often used to solve the problem of data without labels[21]. The purpose of clustering is to classify similar data in the data set into a category, which is often used for pattern recognition and information retrieval [22,23]. Clustering can also be used for abnormal detection, such as credit card fraud, etc. Therefore, clustering has a high application value in trajectory abnormal detection. At present, some scholars have carried out researches such as Reference [24]. At present, the common clustering algorithm is clustering after trajectory segmentation. Compared with the whole trajectory clustering, it is more conducive to the comparison of local features, such as TRACLUS method[8]. In this paper, segmentation and clustering methods are used for trajectory abnormal detection, and the main contributions are mainly divided into the following parts: 1. In the segmenting stage, the characteristics of trajectory time change are considered, state constraints are added, and the distance between trajectory segments is redefined. 2. Clustering is applied to abnormal detection to realize unsupervised learning, and the historical trajectory is divided into normal and abnormal trajectory through clustering. 3. In the abnormal detection stage, it is divided into two-level abnormal detection. The first level of coarse-grained abnormal detection is used to find the abnormal trajectory segments, and the second level of fine-grained abnormal detection is used to find the abnormal subtrajectory segments.

The organizational structure of this paper is as follows: In Sect.1, the basic idea and the aim of trajectory abnormal detection are introduced, the research status of trajectory abnormal detection is also described. In Sect.2, definitions related to trajectories and some related methods of abnormal detection is discussed. The trajectory abnormal detection method is elaborate in Sect.3. Finally, the Atlantic hurricane data are used to simulate and verify the rationality of proposed method in Sect.4

2. Related work

2.1. Trajectory Definition
Trajectory data store large amounts of position information of moving objects at different times. A collection of data in chronological order is called a trajectory $Trj$.

Definition 1 Trajectory. Moving objects move in geospatial space, and their positions change with time. The ordered sequence of $m$ discrete position points and related auxiliary information is defined as trajectories.
where \( m \) represents the spatial position information of the \( m \)-th trajectory point \( p_m \). \( s_m \) represents the relevant information of the \( m \)-th trajectory point \( p_m \). Taking hurricane data as an example, \( s_m \) can represent the wind speed and pressure.

Definition 2 Subtrajectory. A subtrajectory refers to the ordered set of trajectory points within a certain trajectory, as shown in equation (2).

\[
\text{Seq} = \{(P_{m},S_m),(P_{m+1},S_{m+1}),\ldots,(P_{m+k-1},S_{m+k-1})\}
\]  

Where \( k \) is the length of subtrajectory.

Definition 3 Trajectory segment. The trajectory segment refers to the line segment formed by the connection of any two trajectory points within a certain trajectory, as shown in equation (3).

\[
\text{segTrj} = \{p_i=(P_{i},S_i), p_j=(P_{j},S_j)\}
\]

Where \( 1 \leq i < j \leq m \).

Definition 4 Subtrajectory segment. Subtrajectory segment refers to the line segment formed by the connection of adjacent coordinate points on the track space position, as shown in equation (4).

\[
\text{segSeq} = \{p_i=(P_{i},S_i), p_{i+1}=(P_{i+1},S_{i+1})\}
\]

The Trajectory and Subtrajectory are composed of numerous Subtrajectory segment, for example \( \text{Trj} = \{\text{segSeq}_1,\text{segSeq}_2,\ldots,\text{segSeq}_n\} \).

2.2. Distance of trajectory segment

LEE [8] et al. defined the distance between two trajectory segments from the view of spatial location, as indicated in figure 1. Vertical distance \( d_{p_i} \), parallel distance \( d_{l_{ij}} \), and angular distance \( d_{\theta} \) are proposed.

![Figure 1. Distance definition diagram[8].](image)

The vertical distance between \( \text{segTrj}_i \) and \( \text{segTrj}_j \) is defined in equation (5). The parallel distance between \( \text{segTrj}_i \) and \( \text{segTrj}_j \) is defined in equation (6). The angular distance between \( \text{segTrj}_i \) and \( \text{segTrj}_j \) is defined in equation (7).

\[
d_{p_i}(\text{segTrj}_i,\text{segTrj}_j) = \frac{l_{i}^{2} + l_{j}^{2}}{l_{i} + l_{j}}
\]

\[
d_{l_{ij}}(\text{segTrj}_i,\text{segTrj}_j) = \text{MIN}(l_{i},l_{j})
\]

\[
d_{\theta}(\text{segTrj}_i,\text{segTrj}_j) = \left\{\begin{array}{ll}
\|\text{segTrj}_i\| \times \sin \theta, & \text{if } 0^\circ < \theta < 90^\circ \\
\|\text{segTrj}_j\|, & \text{if } 90^\circ < \theta < 180^\circ
\end{array}\right.
\]

2.3. Minimum Description Length (MDL)

To speed up computation, a trajectory is usually segmented. The goal of segmentation is to identify the feature points whose trajectory behaviour changes significantly. Using these feature points, the trajectory is divided into several trajectory segments, and a new trajectory is obtained by connecting the
feature points in turn. By dividing the trajectory by these feature points, the local features of the trajectory can be effectively maintained without losing the global features of the trajectory. The main idea of trajectory segmentation based on MDL criterion is to regard the local optimum as a global optimum. The principle of MDL was first proposed in the study of general encoding. To save storage space, a model is used to compress the coding. MDL consists of two parts: \( L(H) \) and \( L(D/H) \), where \( H \) is the hypothesis and \( D \) is the data. \( L(D/H) \) is the length, in bits, of the description of the data when encoded using the hypothesis.

Suppose a trajectory contains only coordinate information \( T_{\text{trj}} = \{p_1, p_2, p_3, \ldots, p_n\} \). A series of characteristic points of this trajectory are \( \{p_{i_1}, p_{i_2}, p_{i_3}, \ldots, p_{i_m}\} \). The calculation results of \( L(H) \) and \( L(D/H) \) are displayed in figure 2. The calculation of \( L(H) \) is displayed in equation (8) and the calculation of \( L(D/H) \) is displayed in equation (9). \( \text{len}(p_{i_1}, p_{i_2}) \) represents the Euclidean distance of point \( p_{i_1} \) and point \( p_{i_2} \). If \( p_i \) and \( p_j (i < j) \) are the characteristic points of the trajectory, then the MDL cost of \( p_i \) and \( p_j \) is \( \text{MDL}_{\text{par}}(p_i, p_j) = L(H) + L(D/H) \).

\[
L(H) = \sum_{j=1}^{m} \log_2(\text{len}(p_{i_j}, p_{i_{j+1}})) \tag{8}
\]
\[
L(D/H) = \sum_{j=1}^{m} \sum_{k=j}^{m} \{\log_2(d_1(p_{i_j}, p_{i_{j+1}}, p_{i_{k+1}})) + \log_2(d_0(p_{i_j}, p_{i_{j+1}}, p_{i_{k+1}}))\} \tag{9}
\]

Figure 2. MDL calculation method.

3. Proposed method

3.1. Trajectory segmentation based on state constraints

MDL only considers the spatial distance of trajectory; it ignores the temporal variation of trajectory. The temporal variation can be reflected by the state parameters of the sampling points. Because the trajectory states change at certain locations, such as wind speed and pressure when a hurricane lands, these locations have important research value. Therefore, not only the change of trajectory in space but also the change of trajectory in time are considered.

\( T_{\text{trj}} = \{(P_1, S_1), (P_2, S_2), \ldots, (P_m, S_m)\} \) contains some auxiliary information, which can be used to describe the state change of trajectory with time change. \( S = (S_{a_1}, S_{a_2}, \ldots, S_{a_s}) \) represents the auxiliary information in the trajectory point, which is called the state parameter. \( S_{m,k} \) represents the information of the \( k \)-th state parameter of the \( m \)-th trajectory point. The experimental data in this paper is Atlantic Ocean hurricane data. The state parameter contains the Maximum Sustained Wind Speed and Minimum Central Pressure, which can used to represent the change of sampling points in time, that is \( S = (W, P) \), \( W \) represents Maximum Sustained Wind Speed and \( P \) represents Minimum Central Pressure.
Define 5: The state parameters of subtrajectory segment \( S_{\text{segSeq},i,k} \), represents the information of the \( k \)-th state parameter of \( \text{segSeq} \). It is represented by the average value of the \( k \)-th state parameter of the starting and ending points of \( \text{segSeq} \), as shown in the equation (10).

\[
S_{\text{segSeq}} = (S_{i,k} + S_{i+1,k}) / 2
\]  

(10)

Where, \( S_{x,k} \) and \( S_{y,k} \) represent the value of the \( k \)-th state parameter of the \( i \)-th and \( (i+1) \)-th trajectory points respectively.

Definition 6. State discrete index \( \sigma^2 \). This index is used to describe the stability degree of subtrajectory \( \text{Seq} \) that composed of \( n \) subtrajectory segments. The greater the value of \( \sigma^2 \), the more unstable the trajectory state, as equation (11).

\[
\sigma^2 = \frac{\sum_{j}^{n} (S_{\text{segSeq},j} - \bar{S})^2}{n} 
\]  

(11)

where \( S_{\text{segSeq},j} \) represents the value of the \( j \)-th state parameter of \( \text{segSeq} \) and \( \bar{S} \) represents the average value of the \( j \)-th state parameter of all subtrajectory segments in \( \text{Seq} \). In this paper, only Maximum Sustained Wind Speed and Minimum Central Pressure are considered; then, the state discrete index is displayed in equation (12)

\[
\sigma^2 = \left( \frac{\sum_{i}^{n} (W_{\text{segSeq}} - \bar{W})^2}{n} + \frac{\sum_{i}^{n} (P_{\text{segSeq}} - \bar{P})^2}{n} \right) / 2
\]  

(12)

Where \( \bar{W} \) represents the average value of the wind speed of all subtrajectory segments in subtrajectory \( \text{Seq} \) and \( \bar{P} \) represents the average value of central pressure of all subtrajectory segments in subtrajectory \( \text{Seq} \).

Both the MDL criterion and the discrete state index are considered to extract the feature points. If \( p_i \) and \( p_j \) \((i < j)\) are characteristic points of a trajectory, then \( \sigma^2 \) must be less than a fixed threshold \( \sigma^2_{\text{max}} \). By adding the state discrete index to extract the feature points, the segmented trajectory maintains the local characteristics of the trajectory in temporal and spatial.

3.2. Trajectory clustering based on spatial-temporal distance

The similarity of trajectories can be determined by the degree of phase of trajectories in time and space. By comparing the references [8, 25-26], density-based clustering can adapt to trajectory clustering effectively. In this paper, the trajectory abnormal detection must be performed by clustering the results. First, the historical trajectory is segmented as the training set \( \Omega = \{ \text{segTrj}_1, \ldots, \text{segTrj}_{\text{num}} \} \). Then, the normal trajectory and abnormal trajectory are obtained by density-based clustering. Classes whose trajectories exceed the threshold are regarded as normal classes, representing the normal trajectory patterns of local motion. And the classes less than the threshold are regarded as abnormal, representing abnormal trajectory patterns in the training set.

The spatial distance \( (\text{spatial} \_\text{dist}(\text{segTrj}_i, \text{segTrj}_j)) \) between trajectory segment \( \text{segTrj}_i \) and \( \text{segTrj}_j \) is shown in equation (13). The trajectory data has not only spatial characteristics but also temporal characteristics. Therefore, the temporal distance \( (\text{temporal} \_\text{dist}(\text{segTrj}_i, \text{segTrj}_j)) \) between trajectory segment \( \text{segTrj}_i \) and \( \text{segTrj}_j \) defined in equation (14).

\[
\text{spatial} \_\text{dist}(\text{segTrj}_i, \text{segTrj}_j) = d_x + d_y + d_\theta
\]  

(13)
temporal \_dist(\text{seg\text{Trj}}, \text{seg\text{Trj}}) = \frac{1}{1 - (\sum_{k=1}^{n} S_k \_dist) / m}, S_k \_dist = \frac{\overline{S_{\text{seq\text{Trj}}}}_{k} - \overline{S_{\text{seg\text{Trj}}}}_{k}}{\overline{S_{\text{seq\text{Trj}}}}_{k} + \overline{\overline{S_{\text{seg\text{Trj}}}}}_{k}} \quad (14)

Where $S_k \_dist$ represents the distance of the $k$-th state parameter between trajectory segment $\text{seg\text{Trj}}_k$ and $\text{seg\text{Trj}}$. $\overline{S_{\text{seq\text{Trj}}}}_{k}$ represents the average value of the $k$-th state parameter of the subtrajectory corresponding to the $\text{seg\text{Trj}}$. The distance between trajectory segments $\text{seg\text{Trj}}_i$ and $\text{seg\text{Trj}}_j$ is calculated as indicated in equation (15).

$$
\text{dist}(\text{seg\text{Trj}}_i, \text{seg\text{Trj}}_j) = W_s \cdot \text{spatial \_dist}(\text{seg\text{Trj}}_i, \text{seg\text{Trj}}_j) + W_t \cdot \text{temporal \_dist}(\text{seg\text{Trj}}_i, \text{seg\text{Trj}}_j) \quad (15)
$$

Where $W_s$ and $W_t$ represent the weights of spatial and temporal distances respectively.

3.3. Two-level trajectory abnormal detection method

First, the detection trajectories $\Gamma = \{D\text{Trj}, \cdots D\text{Trj}_{\text{data}}\}$ are segmented into trajectory segments $F = \{D\text{seg\text{Trj}}, \cdots D\text{seg\text{Trj}}_{\text{num\_total}}\}$. By comparing the neighbourhood of $D\text{seg\text{Trj}}_{i}$ in training set, we can determine if it belongs to abnormal trajectories. If the majority of trajectories in the neighbourhood of $D\text{seg\text{Trj}}_{i}$ are abnormal, the trajectory segment $D\text{seg\text{Trj}}_{i}$ is regarded as abnormal; otherwise, the trajectory segment is regarded as normal. In this paper, a two-level trajectory abnormal detection method is proposed. The first level detection is coarse-grained detection, which detects the abnormality of trajectory segment $D\text{seg\text{Trj}}_{i}$. If $D\text{seg\text{Trj}}_{i}$ is an abnormal, second level detection is performed. The second level is fine-grained detection, which detects the abnormality of subtrajectory segments that corresponds to trajectory segment $D\text{seg\text{Trj}}_{i}$.

3.3.1 First-level coarse grained abnormal detection

The first level detection is coarse-grained detection. It can be determined according to the clustering results. For a trajectory segment $D\text{seg\text{Trj}}_{j} \in F$, the neighbourhood ($D\text{N}_{i}(D\text{seg\text{Trj}}_{j})$) of $D\text{seg\text{Trj}}_{j}$ in the training set is defined as in equation (16).

$$
D\text{N}_{i}(D\text{seg\text{Trj}}_{j}) = \{\text{seg\text{Trj}} | \text{dist}(D\text{seg\text{Trj}}_{j}, \text{seg\text{Trj}}_{j}) < e, D\text{seg\text{Trj}}_{j} \in F, \text{seg\text{Trj}}_{j} \in \Omega\} \quad (16)
$$

where the calculation of $\text{dist}(D\text{seg\text{Trj}}_{j}, \text{seg\text{Trj}}_{j})$ is given in equation (15).

Detect whether each trajectory segment $D\text{seg\text{Trj}}_{j}$ in the neighbourhood of $\text{seg\text{Trj}}_{j} \in D\text{N}_{i}(D\text{seg\text{Trj}}_{j})$ is abnormal, and get the abnormal score of $D\text{seg\text{Trj}}_{j}$ as shown in equation (17).

$$
\rho(D\text{seg\text{Trj}}_{j}) = \frac{\eta_{\text{all\_ir}}(\text{seg\text{Trj}}_{j})}{\eta_{\text{total\_ir}}(\text{seg\text{Trj}}_{j})} \quad (17)
$$

$\eta_{\text{total\_ir}}(\text{seg\text{Trj}}_{j})$ represents the total number of neighbor ($|D\text{N}_{i}(D\text{seg\text{Trj}}_{j})|$) in the training set. $\eta_{\text{all\_ir}}(\text{seg\text{Trj}}_{j})$ is the number of abnormal trajectory trajectories in $D\text{N}_{i}(D\text{seg\text{Trj}}_{j})$. If $\rho(D\text{seg\text{Trj}}_{j}) \geq \rho_{\text{threshold}}$ or $\eta_{\text{total\_ir}}(\text{seg\text{Trj}}_{j}) < \text{Min\_Lns}$, the trajectory segment $D\text{seg\text{Trj}}_{j}$ is regarded as an abnormal. Otherwise it is regarded as normal.

3.3.2 Second-level fine grained abnormal detection

The first-level coarse grained abnormal detection is to detect if trajectory segment $D\text{seg\text{Trj}}_{i}$ is an abnormal. The second-level fine grained abnormal detection is to detect every subtrajectory segment $\text{Dseg\text{Seq}}_{i}$ of subtrajectory corresponding to $D\text{seg\text{Trj}}_{i}$. When $D\text{seg\text{Trj}}_{i}$ is an abnormal, the following method is used for the abnormal detection of each subtrajectory segment $\text{Dseg\text{Seq}}_{i}$ of the subtrajectory.
\[ \text{Seq}_i = \{p_1, p_2, \ldots, p_n\} \] corresponding to \( D_{\text{segTrj}} \). First, the starting point of \( D_{\text{segSeq}} \) and \( D_{\text{segTrj}} \) are connected to form trajectory segment \( P_{\text{prior}} \), and then the ending point of \( D_{\text{segSeq}} \) and the starting point of \( D_{\text{segTrj}} \) are connected to form trajectory segment \( P_{\text{latter}} \). Finally, the first-level coarse-grained detection of \( P_{\text{prior}} \) and \( P_{\text{latter}} \) is performed. If both \( P_{\text{prior}} \) and \( P_{\text{latter}} \) are detected to be normal, then \( D_{\text{segSeq}} \) is normal; otherwise, \( D_{\text{segSeq}} \) is an abnormal. The second-level fine-grained abnormal detection is displayed in figure 4.

As indicated in figure 3, through the first-level coarse-grained abnormal detection of \( P_1 \) and \( P_4 \), we can determine whether subtrajectory segment \( D_{\text{segSeq}} \) is an abnormal. If both \( P_1 \) and \( P_4 \) are abnormal, then \( D_{\text{segSeq}} \) is an abnormal trajectory.

4. Experiments and Analysis

Atlantic Ocean hurricane data (https://www.nhc.noaa.gov/data/) released by the National Hurricane Center of the United States is used. The data includes the longitude and latitude of the hurricane, maximum sustained wind speed, minimum central pressure, sampling time, and other information. In this paper, we extract data from 1990 to 2009 as the training data for 20 years. The data contains 322 hurricane trajectories and 9537 sampling points. First, the data is preprocessed. When data is missing, the approximate value of the adjacent data is used to complete the missing data. Duplicate data is also cleaned and noise data is minimized. For example, there can be an unnamed hurricane in the data, i.e., the data Name field displays “UNNAMED”; this typically offers minimal sampled points and the missing data is severe. Therefore, such hurricane data is deleted. Longitude, latitude, maximum sustained wind speed, and minimum central pressure are selected for this experiment.

4.1. Trajectory Segmentation Experiment

In this experiment, the trajectories from 1990 to 2009 are segmented by trajectory segmentation based on state constraints. The threshold of state discretization index is set to \( \sigma_{\text{max}} = 25 \). Finally, 2346 trajectory segments are obtained by this method. Figure 4 displays the original trajectories of hurricanes from 1990 to 2009. Figure 5 displays the trajectories after segment by proposed. It can be observed from figure 5 that the trajectories after segment are virtually identical to original trajectories. However, the order of magnitude of the subtrajectory segments after segmentation is not consistent with the original trajectory. The number of original subtrajectory segments is 9181. And 3097 subtrajectory segments are obtained after segmentation. The subtrajectory segments are compressed to 33.7%, which improves the speed of trajectory clustering and preserves the local characteristics of the trajectories in time and space.
4.2. Trajectory clustering

The abnormal trajectory detection method proposed in this paper is based on clustering. Hence, the hurricane trajectories from 1990 to 2009 are clustered as a training set. The parameters were set to $\varepsilon = 2.6, MinLns = 5$, $\varepsilon = 3.0, MinLns = 6$, and $\varepsilon = 4.0, MinLns = 5$. The clustering results are displayed in figure 6, 7, and 8 (the graph indicates only the normal trajectory clusters). The result of the trajectory clustering in figure 6 has 23 clusters. From figure 7, we can observe that there are numerous trajectory clusters, although they are scattered, and there are abundant abnormal trajectories. The results of the trajectory clustering in figure 8 is nine clusters, and the trajectory clusters are relatively small, yet relatively concentrated. The trajectory clustering results in figure 8 indicate two clusters. It can be observed from figure 8, that the majority of the trajectories are grouped into one class, with the other abnormal trajectories also grouped into one class. Therefore, the clustering results are different for different parameter settings.
The results of the clustering of hurricane trajectories from 1999 to 2009 under different parameters are analysed. As indicated in figure 9 parameter $\varepsilon = 5$, the number of clusters varies with $\varepsilon$. From figure 9, the number of trajectory clusters is 2 with the increase of the value of $\varepsilon$. At this point, there are only two kinds of trajectory clustering: normal and abnormal pattern. As indicated in figure 10, when the parameter $\varepsilon = 4$, the number of clusters varies with $\text{MinLns}$. From figure 10, the number of trajectory clusters increases with the increase of $\text{MinLns}$. When $\text{MinLns}$ is less than 6, the number of trajectory clusters is also stable in two categories. Therefore, as the value of $\varepsilon$ increases, the number of
trajectory clusters decreases; as the value of $MinLns$ increases, the number of trajectory clusters increases gradually.

4.3. Trajectory Abnormal Detection Experiment

4.3.1. Parameter selection of trajectory abnormal detection
The data for hurricane trajectories from 1990 to 2009 is clustered as a training set, and the abnormal detection of hurricane trajectories from 2010 to 2012 is performed. The two-level abnormal detection method is applied. There are 1808 subtrajectory segments from 2010 to 2012. For parameter $\epsilon = 2.6, MinLns = 5, p_{threshold} = 0.7$, the detection result of abnormal trajectories is displayed in figure 11. The lines marked in red in figure 11 are abnormal trajectories. There are 1448 abnormal subtrajectory segments, and the proportion of abnormal subtrajectories segments accounts for more than 80%, which is clearly unreasonable. This is because with this parameter there are 23 clusters, which is a large number, and they are relatively dispersed. Under this parameter, the results of the trajectory abnormal detection are relatively inaccurate.

To select suitable parameters, the results of the hurricane abnormal detection in 2010–2012 under different parameters are analyzed. For $MinLns = 5, p_{threshold} = 0.7$, the variation of the proportion of abnormal hurricane subtrajectory segments in 2010–2012 with $\epsilon$ is indicated in figure 12. It can be
observed from figure 12 that with the increase of $\varepsilon$, the proportion of abnormal subtrajectory segments decreases gradually, and the decreasing range tends to be stable. From figure 9, when $\varepsilon$ is 5.0, and 5.5, the trajectory clusters converge into two classes. In figure 12, when the value of $\varepsilon$ is 5.0, and 5.5, the proportion of abnormal subtrajectory segments is less than 0.2. When $\varepsilon = 4, p_{\text{threshold}} = 0.7$, the variation of the proportion of abnormal subtrajectory segments in 2010–2012 with $\text{MinLns}$ is indicated in figure 13. The proportion of abnormal subtrajectory segments increases with the increase of $\text{MinLns}$. As can be observed from figure 10, when $\text{MinLns}$ is 3, 4, 5, and 6, the trajectory clusters converge into two classes. In figure 13, when the value of $\text{MinLns}$ is 3, 4, 5, and 6, the proportion of abnormal subtrajectory segments detected is less than 0.25. Therefore, when choosing parameters, the least number of clusters must be considered as possible. We typically choose parameters that can force the trajectory aggregate into two clusters. Under this parameter, the trajectory can only be divided into two categories that are normal cluster and abnormal cluster.

Figure 12. $\text{MinLns} = 5, p_{\text{threshold}} = 0.7$, variation of proportion of abnormal hurricane subtrajectory segments in 2010–2012 with $\varepsilon$

Then, the parameter $\varepsilon = 5.0, \text{MinLns} = 5, p_{\text{threshold}} = 0.7$ is selected for abnormal detection. Under this parameter, the hurricane trajectories are grouped into two categories. The number of abnormal subtrajectory segments are 285. The detection results are displayed in figure 14. Under this parameter, the detection results of abnormal trajectories are relatively reasonable, and the proportion of abnormal subtrajectory segments is approximately 0.16. Further, hurricanes deviating from the normal trajectory can be effectively detected. In figure 14, where indicated by circles, the trajectories either turn or the area where the hurricane appears deviates from the normal area, which are clearly abnormal.

Figure 13. $\varepsilon = 4, p_{\text{threshold}} = 0.7$, variation of proportion of outlying hurricane subtrajectory segments in 2010–2012 with $\text{MinLns}$

Figure 14. Results of hurricane abnormal trajectory detection in 2010–2012 with parameter $\varepsilon = 5.0, \text{MinLns} = 5, p_{\text{threshold}} = 0.7$. 
4.3.2. Comparing with TRAOD method

In this paper, proposed method and TRAOD[9] method are compared, and the results are shown in figure 15. The presented method sets the parameter to $\varepsilon = 5.5$, $MinLas = 5$, $p_{threshold} = 0.7$. The hurricane trajectories from 1990 to 1999 are clustered as training sets, and then two-level outlier detection is performed for the hurricane trajectories from 2000 to 2006. The results are shown in figure 15 (a). The results of outlier trajectory detection in 2000-2006 based on TRAOD are shown in figure 15 (b).

![Figure 15. Proposed method and TRAOD algorithm for outlier trajectory detection.](image)

(a) Proposed method (b) TRAOD method

There are 356 abnormal subtrajectory segments detected by two level abnormal detection method. As can be observed from Figure 16, the proposed method and TRAOD have the same detection results. However, there are also differences, as indicated in the black box in figure 15(a). The trajectories detected by the proposed method in the area indicated by the black box are outliers, whereas the TRAOD method indicates these trajectories as only partially outlying. Compared with the original trajectories of from 1990 to 2009 in figure 4, the number of hurricane trajectories appearing in the region during the 20 years is extremely small. Therefore, the trajectories of the hurricane appearing in this region are detected as abnormal outliers. The trajectory indicated by the arrow in figure 15(a) is clearly different from that of its adjacent areas. The two-level trajectory abnormal detection method detects it as an outlying trajectory, whereas TRAOD does not. In this paper, outlying trajectories are detected based on the clustering of historical trajectories; hence, historically, regions with fewer hurricane trajectories are considered outliers. The TRAOD algorithm only uses spatial distance to determine if there is abnormal. Therefore, the proposed method is more reasonable than TRAOD. Further, numerous discontinuous trajectory segments are marked as abnormal in figure 15(a) because the trajectory state has changed significantly with time, and hence, the trajectory segments are detected as abnormal. At the same time, TRAOD cannot independently detect new trajectories. Therefore, we can conclude that the trajectory abnormal detection method proposed in this paper is more accurate, and the detected abnormal trajectories are closer to the actual situation.

5. Conclusion

Abnormal detection of trajectory has broad application prospects and research value in numerous fields, such as climate detection, traffic management, safety monitoring, and urban planning. To mine outlying patterns of trajectories from big data effectively, a two-level trajectory abnormal detection method based on segmentation and clustering was proposed in this paper. In summary, the main contributions of this paper are as follows: 1. The time-varying characteristics of a trajectory are considered in the segmentation phase and the segmentation of the trajectory is based on state constraints. 2. In the clustering phase, the distance between the trajectory segments, which is a combination of temporal and spatial distance, is redefined. 3. In the abnormal detection phase, a two-level trajectory-abnormal detection is proposed. Firstly, abnormal trajectory segments are identified by first-level coarse grained abnormal detection, and then abnormal subtrajectory segments are determined by second-level fine-grained abnormal detection for the original trajectory corresponding to the outlying segments. 4. Finally, the simulation results confirm that the two-level trajectory abnormal detection results are more accurate and reasonable. The state parameters selected in this paper were restricted to the maximum sustained wind speed and minimum central pressure. In a future study, we will consider a variety of state...
parameters to improve the accuracy of the test results. There is no abnormal label in the original data; hence, the results of the abnormal detection cannot be quantified. In future research, we will consider quantifying the results.

Acknowledgements
This work was supported by the National Natural Science Foundation of China (62073330), Natural Science Foundation of Hunan Province (2019JJ20021), NUDT Scientific Research Project (ZK18-02-12), Huxiang Young Talents (2018RS3079) and Innovation project for graduate students in Hunan Province (CX20190046).

References
[1] Zheng Y. (2015) Trajectory Data Mining: An Overview. ACM Transactions on Intelligent Systems and Technology. 6(3) 1-41.
[2] Meng F, Yuan G, Lv S, Wang Z., Xia S. (2018) An overview on trajectory abnormal detection. Artificial Intelligence Review, 2 1-20.
[3] Tang J., Piera M. A., Guasch T. (2016) Coloured Petri net-based traffic collision avoidance system encounter model for the analysis of potential induced collisions. Transportation research part C: emerging technologies, 67: 357-377.
[4] Tang J., Zhu. F, Piera M. A. (2016) A causal encounter model of traffic collision avoidance system operations for safety assessment and advisory optimization in high-density airspace. Transportation research part C: emerging technologies, 96: 347-365.
[5] Knorr E. M., Ng R. T. (1998) Algorithms for mining distance-based abnormalities in large datasets. In: Proceedings of the international conference on very large data bases. New York City. pp.392-403.
[6] Knorr E. M., Raymond T., Tucakov V. (2000) Distance-based outliers: algorithms and applications. The VLDB Journal, 8(3):237-253.
[7] Lee J. G., Han J., Li X. (2008) Trajectory outlier detection: a partition-and-detect framework. In: IEEE International Conference on Data Engineering. Cancun.
[8] Lee J. G., Han J. W., Whang K. Y. (2007) Trajectory clustering: A partition-and-group framework. In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data. Beijing. pp. 593-604.
[9] Mao J., Wang T., Jin C., et al. (2017) Feature Grouping-Based Abnormal Detection Upon Streaming Trajectories. IEEE Transactions on Knowledge & Data Engineering, 29(12) :2696-2709.
[10] Adam N. R., Janeja V. P., Atluri V. (2004) Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets. In: Proceedings of the 2004 ACM symposium on Applied computing. New York. pp. 576-583.
[11] Liu Z., Pi D., Jiang J. (2013) Density-based trajectory outlier detection algorithm. Journal of Systems Engineering & Electronics, 24(2):335-340.
[12] Yu Q., Luo Y., Chen C., Wang X. (2018) Trajectory outlier detection approach based on common slices sub-sequence. Applied Intelligence, 48(1):2661-2680.
[13] Chen C., Zhang D., Castro P. S., et al. (2011) Real-time detection of anomalous taxi trajectories from GPS traces. In: Mobile and Ubiquitous Systems: Computing, Networking, and Services. Berlin Heidelberg.
[14] Chen C., Zhang D. Q., Castro P. S., Li N., Sun L., Li S. J., Wang Z. H. (2013) iBOAT: Isolation-Based online anomalous trajectory detection. IEEE Transactions on Intelligent Transportation Systems, 14(2):806-818.
[15] Birant D., Kut A. (2006) Spatio-Temporal Outlier Detection in Large Databases. In: IEEE 28th International Conference on Information Technology Interfaces. Croatia. pp.291-297.
[16] Sun L., Zhang D., Chen C., et al. (2013) Real Time Anomalous Trajectory Detection and Analysis. Mobile Networks and Applications, 18(3):341-356.
[17] Chawla S., Zheng Y., Hu J. (2012) Inferring the Root Cause in Road Traffic Anomalies. In: IEEE International Conference on Data Mining. Belgium. pp.141-150.

[18] Pang L. X., Chawla S., Liu W., et al. (2013) On detection of emerging anomalous traffic patterns using GPS data. Data & Knowledge Engineering, 87(9):357–373.

[19] Piciarelli C., Foresti G. L. (2007) Anomalous trajectory detection using support vector machines. In: 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, London. pp. 153-158.

[20] Piciarelli C., Micheloni C., Foresti G. L. (2008) Anomalous trajectory patterns detection. In: International Conference on Pattern Recognition. Tampa. pp. 1-4.

[21] Kumar D., Wu H., Rajasegarar S., Leckie C., Krishnaswamy S., Palaniswami M. (2018) Fast and scalable big data trajectory clustering for understanding urban mobility. IEEE Transactions on Intelligent Transportation Systems, 19(11): 3709-3722.

[22] Zhao L., Shi G. (2019) A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. Ocean Engineering, 172: 456-467.

[23] Li L, Leckie C. (2016) Trajectory pattern identification and anomaly detection of pedestrian flows based on visual clustering. In: International Conference on Intelligent Information Processing. Springer, Cham, pp.121-131.

[24] Zhou P., Ding Q., Luo H., Hou Q. (2017) Trajectory outlier detection based on DBSCAN clustering algorithm. Infrared and Laser Engineering, 46(05): 238-245.

[25] Nanopoulos A., Theodoridis Y., Manolopoulos Y. (2001) C2P: Clustering based on Closest Pairs. In: International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc. San Francisco. pp.331–340.

[26] Chang C., Zhou B. (2009) Multi-granularity Visualization of Trajectory Clusters Using Subtrajectory Clustering. In: IEEE International Conference on Data Mining Workshops. Miami. 2009.