

# Data Mining Algorithm for Moving Object Trajectory

Yudong Guo\*

Information Cernet, Jinzhong University,  
Jinzhong, 030600  
China

Jinping Zuo

Department of Information technology and Engineering, Jinzhong University,  
Jinzhong, 030600  
China

Received: June 21, 2021. Revised: December 19, 2021. Accepted: January 11, 2022. Published: January 12, 2022.

**Abstract**—Aiming at the poor effect and long recognition time of data mining algorithm for moving target trajectory recognition, a data mining algorithm based on improved Hausdorff distance is proposed. The position and angle of abnormal trajectory data are detected by calculating the distance between trajectory classification and sub trajectory line segments, and the trajectory unit is established by using the improved Hausdorff distance algorithm to optimize the similarity matching structure. Experimental results show that the algorithm has low error pruning rate in identifying moving target trajectory, improves the detection efficiency of moving target trajectory recognition data, and ensures the quality of moving target trajectory recognition data mining.

**Keywords**—Data Mining, Improved Hausdorff Distance, Moving Objects, Running Track.

## I. INTRODUCTION

IN recent years, with the rapid development of various wireless communication technologies, more and more mobile devices have been applied in various fields [1]. This enables that a large number of mobile data including location information, also known as trajectory data, are stored in various location service application systems, such as mobile location services, GPS vehicle navigation, wildlife tracking system and others [2]. The continuous accumulation of location data has made people realize that effective mining of operation rules from mobile data set plays a decisive role in the further development of positioning application system [3].

The complexity of trajectory data makes traditional data mining techniques and algorithms unable to use directly in this field [4]. Therefore, the researchers have done a thorough

research on the algorithm for the data mining of moving objects' running trajectory. Literature [5] proposed an algorithm for mining moving object's trajectory based on multi-scale space division and road network modeling. Through the moving track and moving process sequence, the moving global pattern and process pattern are identified, the road network topological relationship model of moving track is constructed, and the information of data conversion and pattern is mined, but the error of the algorithm is large, which affects the quality of data mining. An evolutionary computation based algorithm for moving object's trajectory mining was proposed [6]. From the point of view of trajectory direction and density, the anomaly is detected by evolutionary computation. However, this algorithm has a slow detection speed, which affects the efficiency of data mining. A position prediction algorithm for moving object's trajectory mining was proposed [7]. The improved pattern mining model is used to extract the track frequent pattern, and the prediction algorithm is proposed to calculate the best matching degree to obtain the predicted position of the moving object track. However, in the detection process of this algorithm, the result of data mining is affected by the poor accuracy of abnormal trajectory recognition. Literature [8] proposed a trajectory mining algorithm for moving objects based on fuzzy constraints. The average dynamic characteristics of specific data are extracted, and the fuzzy neural network is trained to adjust the membership function parameters to mine the specific data in the database. But the effect of this algorithm is poor, which affects the results of data mining.

In order to solve the problems of poor recognition effect and long recognition time in trajectory recognition data mining of current methods, we propose a moving object trajectory mining data mining algorithm based on improved Hausdorff distance. The overall structure is as follows:

1) The trajectory data detection is realized through the trajectory classification.

2) The improved Hausdoff distance algorithm is used to match the similarity of the trajectory, and the point distance algorithm is introduced to realize the data mining of the moving object's trajectory.

3) The experimental data and analysis are carried out to demonstrate the feasibility and effectiveness of the improved Hausdoff distance and realize the data mining of the moving object's running trajectory.

4) The prospect on the basis of the existing research is put forward.

## II. ABNORMAL DETECTION OF MOVING OBJECT'S TRAJECTORY BASED ON TRAJECTORY CLASSIFICATION

Trajectory's abnormal detection is an important research field of moving object's trajectory data mining, mainly used to find those data are significantly different from their adjacent trajectories. Trajectory classification can effectively identify the trajectory. By classifying trajectories into several sub segments, abnormal detection algorithm based on trajectory classification is used to detect abnormal trajectories of sub trajectories set, and achieve abnormal detection of moving objects.

The distance of the trajectory's subsegment after the trajectory classification is calculated. The distance between the two subsegments is an important measure of trajectory's abnormal detection. The distance equation consists of three parts: 1) vertical distance ( $d_{\perp}$ ); 2) horizontal distance ( $d_{\parallel}$ ); 3) angular distance ( $d_{\theta}$ ).

There are two line segments  $L_i = s_i e_i$  and  $L_j = s_j e_j$ , of which  $s_i$ ,  $e_i$ ,  $s_j$ , and  $e_j$  represent two endpoints of the line segment  $L_i$  and  $L_j$ , respectively. For no loss of generality, it is assumed that  $L_i$  is longer than  $L_j$ ; The projections of the end points  $s_j$  and  $e_j$  of the line segment  $L_j$  on the line segment  $L_i$  is  $p_s$  and  $p_e$  respectively.

The vertical distance between the line segment  $L_i$  and the line segment  $L_j$  is a two order Lehmer mean, and the vertical distance between the line segment  $L_i$  and  $L_j$  can be expressed as:

$$d_{\perp}(L_i, L_j) = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}} \quad (1)$$

Where,  $l_{\perp 1}$  is the Euclidean distance between point  $s_j$  and  $p_s$ ;  $l_{\perp 2}$  represents the Euclidean distance between point  $e_j$  and  $p_e$ . The horizontal distance between the line segment  $L_i$  and the line segment  $L_j$  can be expressed as:

$$d_{\parallel}(L_i, L_j) = \min(l_{\parallel 1}, l_{\parallel 2}) \quad (2)$$

Among them,  $l_{\parallel 1}$  is the smaller value of the distance between the projection point  $p_s$  to the two ends  $s_i$  and  $e_i$  of the line segment  $L_i$ , and it is known as the length of the line segment  $s_i p_s$  through the analysis. In the same way,  $l_{\parallel 2}$  is the small value of the distance between the projection point  $p_e$  to the two ends  $s_i$  and  $e_i$  of the line segment  $L_i$ . By analysis, it is known

as the length of the line segment  $p_e e_i$ . The angular distance between line segment  $L_i$  and  $L_j$  is defined as:

$$d_{\theta}(L_i, L_j) = \begin{cases} \|L_j\| \times \sin(\theta), & 0^\circ \leq \theta \leq 90^\circ \\ \|L_j\|, & 90^\circ \leq \theta \leq 180^\circ \end{cases} \quad (3)$$

Where,  $\|L_j\|$  is the length of the line segment  $L_j$ , and  $\theta(0^\circ \leq \theta \leq 180^\circ)$  is the smaller angle between the line segment  $L_i$  and the line segment  $L_j$ . The distance between the line segment  $L_i$  and the line segment  $L_j$  can be defined as:

$$\text{dist}(L_i, L_j) = w_{\perp} \cdot d_{\perp}(L_i, L_j) + w_{\parallel} \cdot d_{\parallel}(L_i, L_j) + w_{\theta} \cdot d_{\theta}(L_i, L_j) \quad (4)$$

Where,  $w_{\perp}$ ,  $w_{\parallel}$ , and  $w_{\theta}$  represent the vertical, horizontal and angular distance weights of the line segment  $L_i$  and the line segment  $L_j$ , respectively.

For a trajectory set  $TR = \{TR_1, TR_2, \dots, TR_n\}$ ,  $n$  of which is the total number of trajectories. After detecting by abnormal trajectory detection algorithm, it can get an abnormal trajectory set  $OTR = \{OTR_1, OTR_2, \dots, OTR_m\}$ , of which  $m$  is the total number of abnormal trajectories.

The setting of moving object's trajectory is a sequence of multidimensional data points, defined as  $TR_i = p_1, p_2, \dots, p_j, \dots, p_{len_i} (1 \leq i \leq n)$ , where  $p_j (1 \leq j \leq len_i)$  is a multi-dimensional spatial location point, and  $len_i$  is the length of trajectory  $TR_i$ , that is, the number of points in a sequence. The sequence  $p_j$  is the subtrajectory of the trajectory  $TR_i$ .

The subtrajectory line segment is a line segment  $p_i p_j (i < j)$ , of which  $p_i$  and  $p_j$  are points selected from the same trajectory. The abnormal trajectory is a trajectory that contains an abnormal trajectory's subsegment. If there are not enough adjacent trajectories for a single trajectory's subsegment, it is defined as an abnormal.

The running trajectory of an object is the most basic unit of the trajectory. A basic unit is the smallest and meaningful trajectory unit in a given domain, which needs to consider the interval in the location of the moving object. If the recording interval is longer than the minimum trajectory unit, the basic unit can be each single recording point [9].

Set the trajectory  $TR_i = p_1, p_2, \dots, p_j, \dots, p_{len_i} (1 \leq i \leq n)$ , and the obtained trajectory's subsegment based on fine-grained classification is  $p_i p_{i+b}$ ,  $b$  of which is a basic unit, expressed in lowercase letters, the obtained trajectory's subsegment based on coarse-grained classification is  $p_i p_{i+j*b} (j \geq 1)$ , and is written in capitals. A combination algorithm of coarse-grained and fine-grained is applied to classify each trajectory into a set of coarse grain line segments and select possible abnormal trajectory's subsegment, and classify them according to fine-grained method. The algorithm satisfies Minimum Description distance (MDL) principle [10]: By giving a hypothesis set, a data sequence  $d$ , certain hypotheses or combinations of certain hypotheses are determined to maximize

the compression of data sequence  $d$ .

The MDL cost consists of two parts:  $L(H)$  and  $L(D|H)$ , of which  $H$  is the hypothesis, and  $D$  is data.  $L(H)$  is the description length of the hypothesis  $H$ , and  $L(D|H)$  is the description length of the data sequence  $D$  encoded in the premise of  $H$ . The optimal scheme is determined by calculating the assumptions and data obtained when the minimum cost function  $L(H) + L(D|H)$  takes the minimum value.

In the coarse-grained scheme, the set of the trajectory's subsegment after classification corresponds to hypothesis  $H$ , and that before the classification corresponds to the data  $D$ . Measure precision and accuracy with  $L(H)$  and  $L(D|H)$ , respectively. Then the optimal granularity classification strategy is transformed into the optimal assumption problem by using the MDL principle. Compared with the linear simplification algorithm, the advantage of the algorithm is that it does not require any additional incoming parameters.

Set  $L(H)$  be the sum of the length of the line segment after the coarse-grained classification, then it can be calculated by the next formula.

$$L(H) = \sum_{i=1}^n \log_2 p_{len_i} \quad (5)$$

It can be seen from the upper formula that the ISO trajectory line segment is detected according to the relative distance between the trajectories, which is independent of the end point coordinates of the line segment. Therefore, the length of the line segment is used to define  $L(H)$  without using the end point coordinates of the line segment.

The sum of the distance between the trajectory and the line segment is set as  $L(D|H)$ , which can be expressed as:

$$L(D|H) = \sum_{j=1}^{len_i} \sum_{i=1}^n \{\log_2(d_{\perp}(L_i, L_j)) + \log_2(d_{\theta}(L_i, L_j))\} \quad (6)$$

In order to get the upper and lower bounds of the distance, the fine-grained classification of the trajectory is also needed. In the process of calculation, it is assumed that the line segment obtained by fine-grained classification is approximately parallel to its coarse grain line segment, the angle between them is usually very small. Therefore, the angular distance can be ignored, which contains only vertical distance  $d_{\perp}$  and horizontal distance  $d_{\parallel}$ . The formula for calculating the upper and lower bounds of  $dist(L_i, L_j)$  can be expressed as:

$$lb(L_i, L_j, dist) = lb(L_i, L_j, d_{\perp}) + lb(L_i, L_j, d_{\parallel}) \quad (7)$$

$$ub(L_i, L_j, dist) = ub(L_i, L_j, d_{\perp}) + ub(L_i, L_j, d_{\parallel}) \quad (8)$$

Where,  $lb(L_i, L_j, dist)$  and  $ub(L_i, L_j, dist)$  are the upper and lower bounds of  $dist(L_i, L_j)$ ,  $lb(L_i, L_j, d_{\perp})$  and  $ub(L_i, L_j, d_{\perp})$  are the upper and lower bounds of  $d_{\perp}(L_i, L_j)$ , and  $lb(L_i, L_j, d_{\parallel})$  and  $ub(L_i, L_j, d_{\parallel})$  are the upper and lower bounds of  $d_{\parallel}(L_i, L_j)$ , respectively.

According to the results of the upper and lower bounds of the calculated coarse-grained segment distance, whether

fine-grained classification is needed is classified. For a determined distance threshold  $D$ , when condition  $lb(L_i, L_j, dist) > D$  is met, neither fine-grained segment  $L_i$  nor fine-grained segment  $L_j$  belong to approximate trajectories. When condition  $lb(L_i, L_j, dist) < D$  is met, both fine-grained segment  $L_i$  and fine-grained segment  $L_j$  belong to approximate trajectories. When  $ub(L_i, L_j, dist) \leq D$ , all fine-grained line segments in  $L_i$  and any fine-grained line segments in  $L_j$  are approximate trajectories, and vice versa; When  $lb(L_i, L_j, dist) \leq D < ub(L_i, L_j, dist)$ , part of fine-grained line segments in  $L_i$  and part of fine-grained line segments in  $L_j$  are approximate trajectories, and vice versa.

In the process of trajectory abnormal detection, it needs to detect adjacent trajectories, and set a sub trajectory

$L_j \in P(TR_j)$ , if  $\left(\sum_{L_i \in TR_i} len_i\right) \geq len_i$ ,  $TR_i \neq TR_j$ , then trajectory

$TR_i$  is a adjacent trajectory of  $L_j$ . It is shown that only if there are enough subtrajectories in one path to be close to  $L_j$ , can they be considered as the adjacent trajectories of  $L_j$ . For one trajectory's subsegment  $L_i \in TR_i$ , if it has a small enough adjacent trajectory, it is considered an abnormal trajectory's subsegment, which can be expressed as:

$$|TR(L_i, D)| \leq [(1-p) \times |TR|] \quad (9)$$

Where,  $TR$  represents the whole set of trajectories, and the  $p$  is a percentage threshold.

For the trajectory  $TR_i$ , its abnormal factor is the ratio of the sum of the length of the abnormal trajectory's subsegment to the sum of the length of all the subsegments, and can be expressed as by the (10):

$$ofrac(TR_i) = \frac{L}{M} \geq F \quad (10)$$

$F$  of them is a given threshold for abnormal factor, 2 is the sum of the length of all abnormal subsegments in the trajectory  $TR_i$ .  $M$  is the sum of the length of all the subsegments in the trajectory  $TR_i$ . When the value of  $ofrac(TR_i)$  is large enough, the trajectory is abnormal, and the abnormal trajectory detection of the moving object is realized.

Through the above discussion, a trajectory abnormal detection algorithm based on trajectory classification is proposed. A coarse-grained and fine-grained trajectory classification strategy is adopted to classify each trajectory into several subsegments, which ensures the effectiveness of segmented results and the efficiency of segmentation process.

### III. DATA MINING ALGORITHM FOR MOVING OBJECT'S TRAJECTORY BASED ON IMPROVED HAUSDORFF DISTANCE

In order to realize the mining of the running trajectory data of the moving object, the improved hausdorff distance algorithm is use to analyze the similarity between trajectories by measuring the distance between segments. The point distance algorithm is

used to optimize the data mining algorithm of the moving object's trajectory, and the specific process is shown as follows.

Hausdorff distance is used to calculate the distance between two-point sets. In the field of pattern recognition, it is often used to compare the shapes between binary images [11]. Although the trajectory and image have the same representation, the image is generally disordered. The point set of the trajectory usually contains a certain operation mode, that is, orderly. Therefore, it is necessary to improve the Hausdorff distance to make it suitable for the trajectory's characteristics of the moving object.

For the trajectory's segment set  $TR = \{TR_i | 1 \leq i \leq a\}$ , where  $a$  represents a number of trajectory segments, the trajectory segment  $TR_i = \{q_{i1}, \dots, q_{ib}\}$ , the length is record as  $|TR_i| = b_i$ ,  $TR_{ij} = \{q_{in}, \dots, q_{i(n+k-1)} | 1 \leq n \leq b_i - k + 1, k \leq b_i\}$  is a basic comparison trajectory unit belongs to target trajectory's subsegment  $TR_i$ ,  $TR_{jx} = \{q_{jm}, \dots, q_{j(m+k+1)} | 1 \leq m \leq b_j - k + 1, k \leq b_j, j \neq i\}$  is a basic comparison trajectory unit belongs to target trajectory's subsegment  $TR_j$ , and the basic comparison trajectory unit is formed by  $(TR_{ij}, TR_{jx})$ .

In order to make distance measurement conform to the requirement of trajectory feature, the distance of Hausdorff is improved.

The idea that Hausdorff distance is introduced into translation, that is,  $TR_{ij}$  is fixed, only  $TR_{jx}$  is allowed to move, then the Hausdorff distance based on the translation can be written as:

$$Ha(TR_{ij}, TR_{jx} \oplus g) = \max lb(L_i, L_j, dist) \quad (11)$$

Where,  $Ha$  represents the Hausdorff distance,  $\oplus$  is a standard Minkowski symbol, and  $g$  is a translation amount. For the translation amount  $g$ , the average value  $\bar{g}$  of the vector distance between each point of the center is replaced by the basic trajectory comparison unit, and then there are:

$$\bar{g} = \frac{1}{k} \sum_{i=0}^{k-1} (q_i - q_j) \quad (12)$$

Then, the distance between the two running sub trajectories can be defined as:

$$dist(TR_{ij}, TR_{jx} \oplus \bar{g}) = \sum_{i=0}^n (dist(q_{i(n+r)} - q_{j(m+r)}) + \bar{g}) \quad (13)$$

According to the characteristics of mobile objects, due to the different running rules of mobile objects in different regions, it needs to match the trajectories that do not exceed a certain range when matching trajectories. Then the distance of  $dist(q_{i(n+r)} - q_{j(m+r)})$  can be written as:

$$dist(q_{i(n+r)} - q_{j(m+r)}) = \begin{cases} dist(q_{i(n+r)} - q_{j(m+r)}) & dist(q_{i(n+r)} - q_{j(m+r)}) \leq w \\ \infty & dist(q_{i(n+r)} - q_{j(m+r)}) > w \end{cases} \quad (14)$$

It can be seen from the above that in comparison with the shape of the trajectory's fragment, the operating modes contained in  $dist(TR_{ij}, TR_{jx} \oplus \bar{g})$  are compared. The improved Hausdorff distance is more suitable for the matching between the trajectories.

The trajectory's subsegment is generally composed of more

than  $k$  locus points, so a number of basic comparison units can be formed between the two trajectories. When the trajectory's subsegment is matched, it is first to find all the basic comparison units between the two trajectories, and the basic comparison unit is as the smallest comparison unit. By determining whether the basic comparison unit is matching or not, the matching of the trajectory's subsegments is determined. According to the matching result, the running trajectory of the moving object is mined [12].

For a local match threshold  $w$ , if the distance between the two basic locus units  $TR_{ij}$  and  $TR_{jx}$  is  $dist(TR_{ij}, TR_{jx} \oplus \bar{g}) \geq w$ , it is indicated that  $TR_{ij}$  and  $TR_{jx}$  are matched to each other and recorded as  $TR_{jx} \in LM(TR_{ij})$ .

A trajectory subsegment  $TR_i$  and a trajectory subsegment  $TR_j$  are given,  $G(TR_{ij} | TR_{jx}) = \{q_{ie} | q_{ie} \in TR_{ij}, TR_{ij} \in TR_i\}$ ,  $TR_i = \{TR_{ij} | \forall TR_{jx} \in TR_j, TR_{jx} \in LM(TR_{ij})\}$ . When the (15) is established, the matching trajectory subsegment of  $TR_i$  is  $TR_j$ .

$$G(TR_{ij} | TR_{jx}) \geq S + b_i \quad (15)$$

Where,  $b_i$  represents the number of trajectories of the plan, and  $S$  is a given global matching threshold.

The basic comparison unit is the minimum unit of distance measure. Searching the basic comparison unit correctly and quickly is the key to realize the matching of the subsegment of the trajectory. The basic comparison unit pair is made up of  $k$  consecutive points pairs. By introducing the point distance feature matrix, the fast search of the basic comparison unit pair is realized. The point distance feature matrix is to store the subsegment of the trajectory to be matched in the matrix in the form of point pair. Each slash on a matrix is made up of a continuous point pair. If the number of continuous point pairs is greater than  $k$ , the continuous point pair is a basic comparison unit [13].

Supposing there are a target trajectory subsegment  $TR_i = \{q_{i1}, \dots, q_{in}\}$  and a comparative trajectory subsegment  $TR_j = \{q_{j1}, \dots, q_{jm}\}$ , the point distance matrix of the trajectory subsegments  $TR_i$  and  $TR_j$  can be expressed as:

$$Z = \begin{Bmatrix} (q_{i1}, q_{j1}) & \cdots & (q_{in}, q_{j1}) \\ \vdots & (q_{ie}, q_{jf}) & \vdots \\ (q_{i1}, q_{jm}) & \cdots & (q_{in}, q_{jm}) \end{Bmatrix} \quad (16)$$

Where, the difference between the matrix line number and the column number of the matrix element  $(q_{ie}, q_{jf})$  is called the positive diagonal sequence number, and is recorded as  $pssn(q_{ie}, q_{jf}) = e - f$ .

If the  $k$  elements  $(q_{ie_1}, q_{jf_1}), \dots, (q_{ie_k}, q_{jf_k})$  of the matrix are adjacent to the  $k$ -positive diagonal, then the matrix is satisfied:  $pssn(q_{ie_1}, q_{jf_1}) = \dots = pssn(q_{ie_k}, q_{jf_k})$  (17)

It can be seen that the  $k$ -diagonal adjacent is composed of two basic comparative fragments. If the element with a distance greater than  $w$ , it is set into an empty value in the matrix element, a non empty value element adjacent to the  $k$ -positive

diagonal corresponds to a basic comparison fragment pair. As long as we find out the combination of all the non empty elements adjacent to the  $k$ -diagonal in the matrix, we can find out all the basic comparison fragment pairs of possible local matching.

In order to reduce the complexity of the algorithm, the elements in the matrix with a distance greater than  $w$  are quickly determined, and the R-tree index is quoted. A R-tree index is set up for every sub segment of the target trajectory to find the  $w$ -domain of the trajectory point, that is, the trajectory point set with the distance to the target trajectory less than  $w$ . The trajectory points in the  $w$ -domain and target trajectory are set up as the matrix elements to save and the other elements are set into the empty value. Because the sub segment of target trajectory is similar to that of  $w$ -domain, the effect of mining is improved by comparing the local similarity of moving objects' running trajectory.

Supposing  $A = \{a_1, a_2, a_3, a_4, a_5\}$  is the target trajectory and  $T_l = \{T_{l01}, T_{l02}, T_{l03}, T_{l04}, T_{l05}\}$  is the trajectory to compare. The point distance matrix between  $A$  and  $T_l$  can be expressed as:

$$Z_z = \begin{pmatrix} (a_1, t_{l01}) & (a_1, t_{l02}) & (a_1, t_{l03}) & (a_1, t_{l04}) & (a_1, t_{l05}) \\ (a_2, t_{l01}) & (a_2, t_{l02}) & (a_2, t_{l03}) & (a_2, t_{l04}) & (a_2, t_{l05}) \\ (a_3, t_{l01}) & (a_3, t_{l02}) & (a_3, t_{l03}) & (a_3, t_{l04}) & (a_3, t_{l05}) \\ (a_4, t_{l01}) & (a_4, t_{l02}) & (a_4, t_{l03}) & (a_4, t_{l04}) & (a_4, t_{l05}) \\ (a_5, t_{l01}) & (a_5, t_{l02}) & (a_5, t_{l03}) & (a_5, t_{l04}) & (a_5, t_{l05}) \end{pmatrix} \quad (18)$$

The point pair in the matrix element with a distance greater than  $w$  is set as an empty value. The R-TREE index is used to quickly identify these elements. The R-TREE index is established for each trajectory point of the target trajectory, and the  $w$ -domain of the target trajectory point is found. The  $w$ -domain of all trajectory points of the target trajectory subsegment  $A$  can be represented as:

$$N_w(A) = \{t_{l01}, t_{l02}, t_{l03}, t_{l04}, t_{l05}\} \quad (19)$$

If the element is not located in the  $w$ -domain of the target trajectory point, it is directly set into an empty value, then the point distance matrix between  $A$  and  $T_l$  becomes:

$$Z_z = \begin{pmatrix} (a_1, t_{l01}) & \phi & \phi & \phi & \phi \\ \phi & (a_2, t_{l02}) & \phi & \phi & \phi \\ \phi & \phi & (a_3, t_{l03}) & \phi & \phi \\ \phi & \phi & \phi & (a_4, t_{l04}) & \phi \\ \phi & \phi & \phi & \phi & (a_5, t_{l05}) \end{pmatrix} \quad (20)$$

The basic comparison segments of the two trajectories of the point distance moment are  $\{(a_1, t_{l01}), (a_2, t_{l02}), (a_3, t_{l03}), (a_4, t_{l04})\}$  and  $\{(a_2, t_{l02}), (a_3, t_{l03}), (a_4, t_{l04}), (a_5, t_{l05})\}$ . The similarity between the two trajectories is analyzed by (20). According to the similarity matrix, the running trajectory data of the moving object is mined [14].

$$F = \text{dist}(TR_{ij}, TR_{jk} \oplus \bar{g}) \times Z \quad (21)$$

Based on the above discussion, aiming at the directivity and speed of moving objects, the distance of Hausdorff is improved, and the basic comparison fragments are applied when matching trajectories, to calculate them in the form of corresponding point to point. The R-tree index is used to search for the basic

comparison fragments, to determine the similarity between the trajectories, and to realize the data mining of the moving objects.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to prove the validity and practicability of the algorithm for moving trajectory data mining based on the improved Hausdorff distance, an experiment is carried out. The experiment data of the taxi running trajectory in a city is taken as the experimental data, and the experiment process is simulated by MATLAB software. Using Windows 10 system, equipped with i7 processor, the running memory is 16 G. Through the experiment, the results are described as follows. In order to ensure the accuracy of the analysis, data mining algorithms based on the minimum outsourced rectangular distance and Euclidean distance are introduced. The comparison results of different algorithms are shown in Figs. 1 to 4.

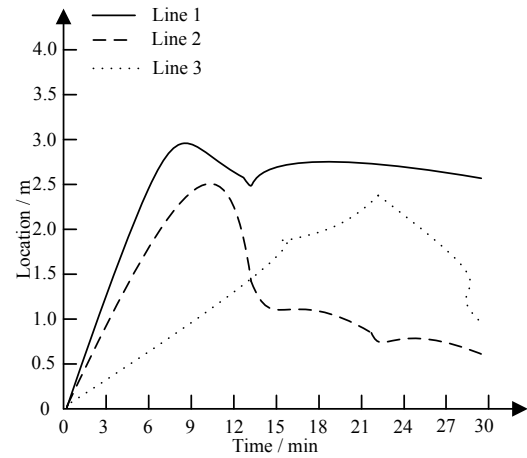


Fig. 1 Original trajectory

Fig. 1 shows the original trajectory. Based on this, the trajectory recognition effects of the improved Hausdorff distance based trajectory recognition algorithm, the minimum outer rectangle distance based trajectory recognition method and the Euclidean distance based trajectory recognition method are compared.

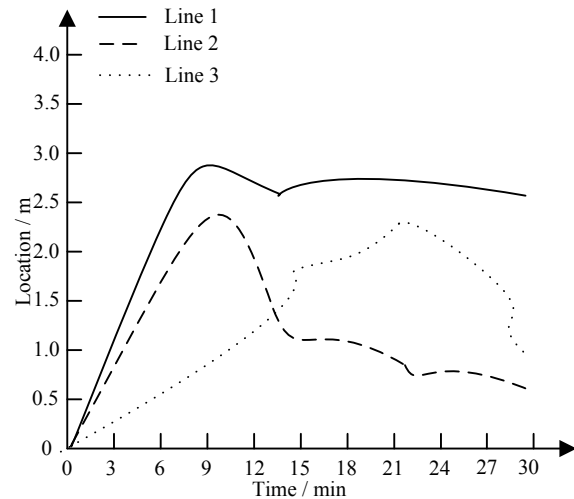


Fig. 2 Trajectory recognition based on improved Hausdorff distance

As can be seen from Fig. 2, compared with the original trajectory, the three trajectory lines deviated slightly from the original trajectory 15 minutes ago, but began to coincide with the original trajectory 15 minutes later, indicating that the algorithm has high recognition accuracy, small error and fast recognition time, and can be applied to practical applications.

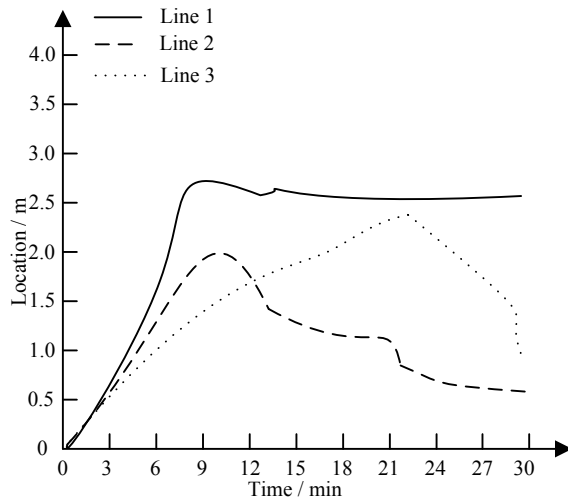


Fig. 3 Trajectory recognition based on minimum outsourced rectangular distance

As can be seen from Fig. 3, compared with the original line, only individual points of track 1 and track 2 of the track recognition algorithm based on the minimum outsourcing rectangular distance coincide with the original track, and track 3 is consistent with the original track from 22min to 29min, with poor accuracy, large error and poor recognition effect.

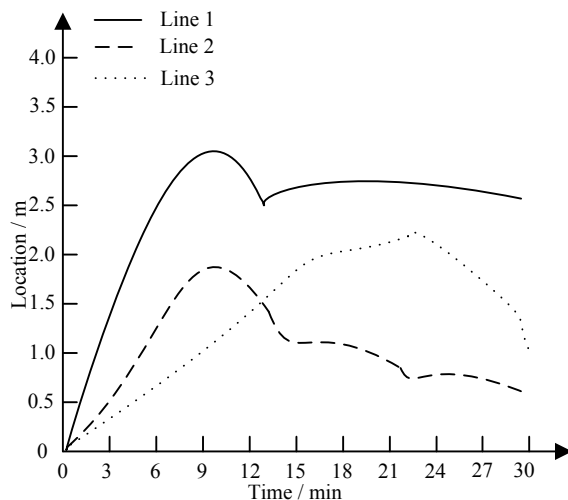


Fig. 4 Trajectory recognition based on Euclidean distance

As can be seen from Fig. 4, compared with the original track, after using the track recognition method based on Euclidean distance, track 1 has a large error from 8min to 13min, track 2 has a large error from the beginning to 13min, and track 3 has an

error as a whole. In general, the recognition effect is better than the track recognition algorithm based on the minimum outer rectangular distance, but the recognition effect is general.

As can be seen from Fig. 1 to Fig. 4, the error of the trajectory recognition algorithm based on the minimum outer rectangular distance is the largest, the error of the trajectory recognition algorithm based on Euclidean distance is moderate, the error of the trajectory recognition algorithm based on the improved Hausdorff distance is the smallest, and the recognition trajectory is the closest to the original trajectory. It can be seen that the trajectory recognition algorithm based on the improved Hausdorff distance has low error rate and high recognition accuracy. Because the algorithm can track the trajectory from three aspects: vertical distance, horizontal distance and angular distance, the accuracy of trajectory recognition can be improved and the recognition effect is better.

By identifying the different trajectory data of different taxis, there is only one abnormal in each group of taxi trajectory data. The time (s) for identifying the abnormal is analyzed by different algorithms, and the results obtained by experiments are shown in Table I.

Table I. Time consuming for abnormal trajectory recognition of different algorithms

Data / group	Improved Hausdorff distance / s	Minimum outsourced rectangular distance / s	Euclidean distance / s
1	2.5	3.2	3.5
2	2.3	2.9	3.2
3	2.7	3.4	3.7
4	2.4	3.0	3.5
5	2.8	3.5	3.7
6	2.6	3.3	3.6
7	2.4	3.1	3.4
8	2.5	3.3	3.5

Table I shows that the time of using the improved Hausdorff distance based data mining algorithm for trajectory abnormal detection is the shortest, indicating that speed of the improved algorithm for moving object's trajectory abnormal detection is the fastest, it can accurately identify the trajectories of moving objects, and reduce the complexity of trajectory abnormal detection.

The performance test results of the granularity partition strategy are shown in Fig. 5 during the abnormal detection of the trajectory of a moving object. Among them, Total is the proportion of segments that are pruned by partitioning strategy in all coarse-grained sub track segments, False is the proportion of false pruning, and Optimal is the proportion of pruning to actual needs.

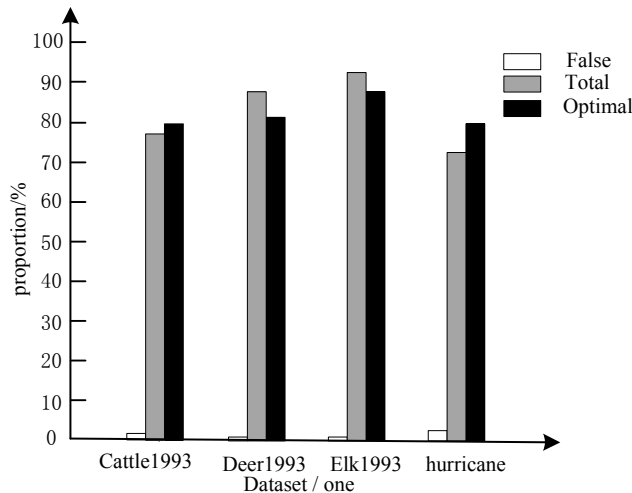


Fig. 5 The pruning effect of the partitioning strategy

As can be seen from Fig. 5, the error rate of the granularity division strategy combining coarse granularity and fine granularity is low, and the recognition result is closer to the actual requirements. The results show that the strategy has good recognition effect and high accuracy, ensures the accuracy of moving object anomaly detection, and can improve the quality of recognition data mining. The time-consuming results combined with the abnormal trajectory test show that the moving target trajectory anomaly detection algorithm based on trajectory Division has better detection effect and faster speed, which provides a technical guarantee for improving the data mining effect of moving targets.

From the above experiments, it can be seen that the moving target data mining algorithm based on the improved Hausdorff distance can track the trajectory more accurately, the segmentation strategy has better effect, and ensures the quality of moving object trajectory data mining. At the same time, the abnormal trajectory detection speed is faster, which improves the data mining speed of moving objects and ensures the data mining effect of moving objects.

## V. CONCLUSION

Because of the popularity of GPS equipment, mobile phone and other automatic positioning equipment, the trajectory data is not only longer and longer, but also the amount of data is also increasing rapidly. The traditional algorithm has been unable to meet the mining requirements of the existing trajectory data. To solve this problem, a new algorithm for data mining of moving objects based on improved Hausdorff distance is proposed. According to the development trend of mobile data processing and trajectory data mining, and some research achievements obtained in this paper, several further research works are proposed.

1) in order to track the motion pattern as much as possible, we need to further find a more suitable distance measure for trajectories.

2) based on the improved Hausdorff distance, the trajectory data mining algorithm for mobile objects is two-dimensional

trajectory data, but in many areas the trajectory is three-dimensional or even high dimensional. Therefore, it is necessary to analyze the trajectory data outside the two dimensions.

3) the amount of the moving object's trajectory data increases as time goes on. The purpose of mass data mining is to find the trajectory of a moving object with the same motion characteristics from the mass data according to the set conditions. The dynamic online data mining technology is a new method of online data analysis, which is to reduce the time of mining. On the basis of offline data mining of mobile objects, we adjust the mining results according to the new trajectory characteristics, so as to avoid remining and achieve a fast processing purpose.

## REFERENCES

- [1] C. C. Hung, W. C. Peng, and W. C. Lee, "Clustering and aggregating clues of trajectories for mining trajectory patterns and routes," *The VLDB Journal*, vol. 24, no. 2, pp. 169-192, 2015.
- [2] M. Li, X. Li, and J. Yin, "TORD problem and its solution based on big trajectories data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1666-1677, 2016.
- [3] C. Ruhhammer, M. Baumann, and V. Protschky, H. Kloeden, F. Klanner, and C. Stiller, "Automated intersection mapping from crowd trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 1-12, 2017.
- [4] G. Yuan, P. Sun, and J. Zhao, "A review of moving object trajectory clustering algorithms," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 123-144, 2017.
- [5] L. Wang, K. Y. Hu, and T. Ku, "Mining urban moving trajectory patterns based on multi-scale space partition and road network modeling," *Acta Automatica Sinica*, vol. 41, no. 1, pp. 47-58, 2015.
- [6] M. M. Tang, G. L. Ji, and B. Zhao, "Parallel algorithm for detecting trajectory outliers based on evolutionary computation," *Journal of Data Acquisition and Processing*, vol. 32, no. 2, pp. 382-389, 2017.
- [7] J. Deng, Y. L. Wang, and Z. J. Dong, "Dynamic trajectory pattern mining facing location prediction," *Application Research of Computers*, vol. 34, no. 10, pp. 2984-2988, 2017.
- [8] H. L. Wu, and X. Y. Ren, "Specific data mining algorithm based on fuzzy constraint database," *Computer Simulation*, vol. 33, no. 10, pp. 240-243, 2016.
- [9] J. Sawamoto, K. Watanabe, and H. Yajima, "Proposal of hybrid data mining method for early detection of depression," *WSEAS Transactions on Biology and Biomedicine*, vol. 16, pp. 114-120, 2019.
- [10] H. Su, K. Zheng, J. Huang, H. Wang, and X. Zhou, "Calibrating trajectory data for spatio-temporal similarity analysis," *The VLDB Journal*, vol. 24, no. 1, pp. 93-116, 2015.
- [11] C. Loglisci, "Using interactions and dynamics for mining groups of moving objects from trajectory data," *International Journal of Geographical Information Science*, vol. 32, no. 7, pp. 1436-1468, 2018.

- [12] Q. Y. Wu, Z. H. Hu, H. Zhang, "Fine division and identification of urban traffic status based on multi-source trajectory data," *Journal of Transportation Systems Engineering and Information Technology*, 2020, vol. 20, no. 1, pp. 83-90, 2020.
- [13] P. Zhang, "Analysis of travel pattern recognition based on trajectory information," *Computer fan*, no. 01, pp. 157, 2019.
- [14] A. Alshreef, J. L. Yuana and L. Lin, "Based SVM distinct stages framework data mining technique approach for text extraction," *WSEAS Transactions on Information Science and Applications*, vol. 16, no. 12, pp. 100-110, 2019.

Yudong Guo and Jinping Zuo proposed a data mining algorithm based on improved Hausdorff distance. Yudong Guo calculated the distance between trajectory classification and sub trajectory line segments, Jinping Zuo detected the position and angle of abnormal trajectory data, and cooperated to establish trajectory units by using the improved Hausdorff distance algorithm to optimize the similarity matching structure. Yudong Guo tested the recognition effect of trajectory under different algorithms by comparing with other methods, and proved the feasibility of the proposed method. Yudong Guo and Jinping Zuo recorded the experimental data and wrote a manuscript. All authors read and approve the final manuscript.

### **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0  
[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)



Yudong Guo, male, born in September 1974, and he is a professor. He had got the B.S. in 1999, from Shanxi University, majoring in computer application. Then he was awarded the M.S. in 2007, from the computer department of Shanxi University, majoring in computer application. Now He is working in the Department of information technology and engineering of Jinzhong University. And His research areas include: database technology, big data and cloud computing. He has published 15 academic papers. At the same time, he has participated in five cooperative research projects.



Jinping Zuo, female, born in November 1975, He is an associate professor. She had got the B.S. in 1999, from Shanxi University, and majoring in computer application. Then she was awarded the M.S. in 2006, from the computer department of Shanxi University, majoring in computer application. Now she is working in the computer department of Jinzhong University. Her research areas include: Theory of computation, software engineering. She has published 10 academic papers. At the same time, she has participated in three cooperative research projects.