

## *Retraction*

# **Retracted: Automatic Mining Method of Group Behavior Patterns Based on Incremental Spatiotemporal Trajectory Big Data**

## **Security and Communication Networks**

Received 8 December 2022; Accepted 8 December 2022; Published 26 January 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Security and Communication Networks* has retracted the article titled “Automatic Mining Method of Group Behavior Patterns Based on Incremental Spatiotemporal Trajectory Big Data” [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Editorial Board.

## **References**

- [1] X. Chen and Y. Liu, “Automatic Mining Method of Group Behavior Patterns Based on Incremental Spatiotemporal Trajectory Big Data,” *Security and Communication Networks*, vol. 2022, Article ID 1222250, 7 pages, 2022.
- [2] L. Ferguson, “Advancing Research Integrity Collaboratively and with Vigour,” 2022, <https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/>.

## Research Article

# Automatic Mining Method of Group Behavior Patterns Based on Incremental Spatiotemporal Trajectory Big Data

Xinfang Chen  and Yiqing Liu 

*College of Information Engineering, Institute of Disaster Prevention, Sanhe 065201, Hebei, China*

Correspondence should be addressed to Yiqing Liu; [yiqingliu2022@outlook.com](mailto:yiqingliu2022@outlook.com)

Received 25 January 2022; Revised 12 February 2022; Accepted 25 February 2022; Published 16 April 2022

Academic Editor: Muhammad Arif

Copyright © 2022 Xinfang Chen and Yiqing Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a problem of unclear group clustering in group behavior pattern mining, which leads to a long mining time. An automatic group behavior pattern mining method based on incremental spatiotemporal trajectory big data is proposed. The grid sequence of each road segment and the road segment information included in each grid are obtained using the group behavior pattern trajectory network. Using incremental trajectory data, the properties of incremental spatiotemporal trajectory big data are retrieved, and the group behavior pattern is grouped. In the obtained class, all data element records are categorized according to their data elements. Multiple attribute dimensions, such as data definition, limitations, and feature words, are used to standardize the spatiotemporal trajectory data pieces. To complete the autonomous mining of group behavior patterns, all subsequences are visited, computed, and compared. The test results show that when the group size threshold is 20, the running time of the group behavior pattern automatic mining method based on incremental spatiotemporal trajectory big data is 311.66, which is 141.29 s and 148.66 s shorter than that based on DBSCAN and K-means, respectively. Therefore, this method has higher execution efficiency.

## 1. Introduction

In recent years, with the development of global positioning technology and wireless communication technology, mobile terminals (such as GPS devices, smart phones, and so on) have produced a large amount of space-time trajectory data. Each spatiotemporal trajectory line is composed of trajectory sampling points in chronological order. Each trajectory point stores the moving object's spatiotemporal information as well as the object's interaction with the geographical environment and depicts the moving object's movement characteristics, behavior preferences, and activity law in the spatiotemporal environment. Because of the widespread use of intelligent mobile terminals, the storage, analysis, and research of spatiotemporal trajectory data have gotten a lot of attention in academia and industry, and they have been used in a variety of fields like traffic coordination and management, tourism route recommendation, natural disaster early warning, and environmental protection. Data mining is the process of

extracting hidden, unknown, but possibly important information and knowledge from a significant amount of incomplete, noisy, ambiguous, and random data from a practical application. However, unlike conventional data mining (such as text data mining, data retrieval, and so on), spatiotemporal trajectory data mining includes information about time and space dimensions and specifies the geographical, temporal, spatial, and other aspects of moving objects [1]. At the microindividual level, the high-precision individual behavior trajectory provides a new observation perspective for behavior pattern mining and the perception of the internal state characteristics and changes of objects. At the level of large-scale group aggregation, massive spatiotemporal trajectory data not only study the group movement patterns and laws but also analyze the dynamic evolution of geographical processes, socioeconomic characteristics, interaction mechanism, and constraints between behavioral activities and geographical environment. We can discover the motion law and behavior pattern of moving objects by analyzing these

spatiotemporal trajectory data. In spatiotemporal trajectory pattern mining, group behavior patterns have long been a hot issue. It can detect hidden patterns and possible rules in the movement of moving objects, such as frequent patterns, adjoint patterns, and aggregation patterns. Urban planning, traffic management, public safety, and animal migration studies may all benefit from these models and rules. Moving items' positions will shift over time. As a result, while modelling and mining spatiotemporal trajectory data, it is often important to consider the temporal and geographical properties of objects in order to analyze the reasons of their development and forecast the future. When we mine trajectory data in a streaming context, for example, the data arrive in a never-ending stream. We must consider the data size, data quality, computation results timeliness, and other factors in the algorithm. The diversity and differences in structure, granularity, mode, breadth, expression, characteristics, types, and quality of trajectory big data, as compared to traditional spatiotemporal data, will inevitably lead to changes in research paradigm, research objectives, research contents, and research methods, and then new propositions for spatiotemporal trajectory big data mining are proposed [2]. In terms of research paradigm, while establishing the big data-driven research paradigm, it is still necessary to use observation, experiment, and simulation to verify the correctness of the mined laws and patterns. Therefore, this paper proposes an automatic mining method of group behavior patterns based on incremental spatiotemporal trajectory big data to find reliable feature groups and their behavior patterns.

## 2. Automatic Mining Method of Group Behavior Patterns Based on Incremental Spatiotemporal Trajectory Big Data

### 2.1. Constructing Group Behavior Pattern Trajectory Network.

For different group behavior patterns, we first need to select a mining area, so we can determine the grid size according to the size, geography, and other characteristics of the area. The mining area is represented by a grid. Firstly, it reduces the amount of data processed in the partition stage, the amount of transmission when distributing data to each partition and node, and the number of clustering tasks in the clustering stage, which can effectively improve the efficiency of the algorithm [3]. Road network spatial grid is a local geospatial grid with road network characteristics, that is, only the geographical location of the road section in the road network area can be mapped with the corresponding spatial grid, and each grid has a unique number. The construction framework of group behavior pattern trajectory network is shown in Figure 1.

The grid sequence of each road section and the road section information included in each grid may be acquired by partitioning the road network map into a series of two-dimensional grids. Each grid may, of course, be provided with geographic semantics, vehicle stay duration, vehicle passing amount, and other information as required. The moving item is mapped to the matching grid in the streaming environment for the time frame received at each time. The following is the formula for calculating it:

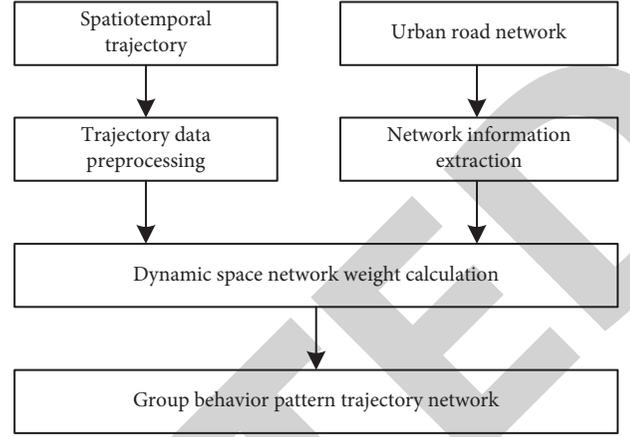


FIGURE 1: Group behavior pattern trajectory network.

$$\begin{cases} a = \lceil \frac{p - p_{\min}}{\varphi} \rceil \\ b = \lceil \frac{q - q_{\min}}{\varphi} \rceil \end{cases}, \quad (1)$$

where  $(p, q)$  represents the longitude and latitude coordinates of the moving object;  $(a, b)$  represents the network coordinates of group behavior patterns;  $p_{\min}$  and  $q_{\min}$  are the minimum longitude and latitude of the excavation area; and  $\varphi$  is the grid side length. After the complete road network space is divided into grids, any track point in the track can be mapped to one grid and can only be mapped to one grid. Therefore, a track can be represented by a series of passing spatial grids. Then, through the coordinates of all moving objects in the grid, the average value of longitude and latitude is calculated as a new grid coordinate, so that the grid coordinate changes with the distribution of objects in the grid, which can reflect the distribution of moving objects in the current grid and avoid some extreme situations, for example, all moving objects are distributed in the corners of the grid [4]. After map matching, the input is an original track, and the output is a new track. The spatial grid is utilised to represent the driving trajectory data of each vehicle in order to decrease the dimension of large trajectory data. However, owing to the geographical layout of the urban road network, there is a blank region between roads, resulting in a significant number of needless grids in the typical GIS spatial grid index in the form of continuous numbering, considerably increasing the spatial complexity. Find all road sections in its grid and neighbouring grids for each location point in the source track and then use the projection point of the location point on the best matching road section as the map matching point [5, 6]. Therefore, this paper will use the road network spatial grid for indexing according to the characteristics of the road network. The constructed grid index can be expressed as

$$W = \{g_1, t, g_2, \eta\}, \quad (2)$$

where  $W$  is the grid index;  $g_1$  represents the grid coordinates after calculating the average longitude and latitude of all

objects;  $t$  is the time window number;  $g_2$  represents the original grid coordinates obtained by grid mapping; and  $\eta$  is the weight of the grid. In the distributed environment, this paper uses the grid as the clustering object for clustering, but there are no moving objects in the formed cluster. To acquire the matching moving items, the grid must match the grid index established during grid formation. The problems of trajectory data may be overcome via trajectory grid vector modelling. The trajectory sparsity can be lowered, the tolerance of position drift may be increased, and the data dimension can be considerably reduced by modifying the grid accuracy [3].

**2.2. Extracting Big Data Features of Incremental Spatiotemporal Trajectory.** Spatiotemporal trajectory is the collection of position information in the process of object moving in space, and it is the function of spatial characteristics about time. The basic characteristics of spatiotemporal data are as follows. (1) Time attribute: time information represents when to obtain the moving track object. This time is often specific and accurate to seconds, which is converted according to the research granularity. (2) Spatial attributes and the position information of spatiotemporal trajectory data reflect the spatial information of moving objects. They are expressed in longitude and latitude. Different coordinate systems obtain different position information, and coordinate conversion is often required in the process of visualization. (3) Other attributes, including object speed, direction, and height, reflect the state of the moving object in the moving process. The temporal and geographical distribution properties of trajectory data must be studied and analyzed further. Data quality is a need for successful data mining. Various objective causes such as vehicle terminal equipment failure and tunnel obstruction often cause data irregularities and data confusion during the collection of traffic trajectory data. When raw, unprocessed data are utilised for mining and analysis, it is frequently difficult to accurately depict real-world urban road operations. Data missing, redundancy, and irregularity are some of the issues with traffic trajectory big data [7]. The track noise is removed using generic methods, and the track segments are then filtered using track motion parameters (track point rotation angle, speed, and so on), and the filtered track segments are adaptively interpolated. Semantic trajectory is a series of dwell points and movement points, and trajectory is a spatiotemporal function that records the location changes of moving objects in space through time. The track's dwell point is a crucial semantic component. A dwell point indicates that the user has spent time in a certain geographic location and may have engaged in some kind of activity. It is reasonable to assume that he or she is particularly engaged in this topic, implying that the user's actions in this area are more significant [8]. Individual vehicles move in the road network constraint space, and the recorded trajectory data represent the geometric characteristics and topology of the road network [9]. The subtrack between the dwell points is the moving point sequence. Dwell point refers to the set composed of at least three continuous low-speed points,

which meets the requirements of time threshold, distance threshold, and average direction difference threshold [10]. The calculation formula of dwell point is as follows:

$$s = \frac{1}{|\alpha|} \sum_{n=1}^{|\alpha|-1} \vartheta_n \beta, \quad (3)$$

where  $s$  represents the dwell point;  $\alpha$  indicates the low-speed point;  $n$  indicates the low-speed point number;  $\beta$  represents the center coordinate of the track end; and  $\vartheta_n$  represents the weight of coordinates. The weight is calculated as follows:

$$\vartheta_n = \frac{1/(m + \gamma)}{\sum_{n=1}^{|\alpha|-1} [1/(m + \gamma)]}, \quad (4)$$

where  $m$  represents the average velocity of the track segment and  $\gamma$  is the smoothing weight adjustment parameter, which is set as the standard deviation of the average speed in this paper. The road geometric data contained in the vehicle trajectory data can be divided into road plane (lane level polygon, road network polygon, ancillary facilities polygon, and so on), road line (road center line, lane level road route, crosswalk, road boundary line, and so on), and road point (road intersection, traffic light location point, split/merge point, traffic ancillary facilities, and so on). The behavioral features of mobile users may be summarised via stay analysis of spatiotemporal trajectory data. It may be established if a user is in an active stay state or has a certain stay behavior by examining information such as the amount of time the user stays, the location of the stay, and the features of the stay area [11]. Because road data have several scales, road information retrieved from trajectory data has multiple granularities as well. The topological connectivity of the road network is represented by the traffic relationship in the trajectory. When there are vehicle tracks connecting any two road nodes in a road network, it means that the nodes are topologically linked.

**2.3. Group Behavior Pattern of Big Data Clustering Based on Incremental Spatiotemporal Trajectory.** Incremental spatiotemporal trajectory data are essentially high-dimensional dynamic data integrating time series data model and graph model under the constraints of road network structure. At the same time, with the rapid development of emerging information technologies such as mobile Internet and communication technology, traffic space-time trajectory data accumulate rapidly and the data scale increases sharply. The parts in the clustering process that need a lot of computing are performed on the data source node, and parallel clustering is done on separate data source nodes. The data are physically split into data blocks at each data source node, and the parallel clustering operation is performed in the unit of data blocks, considerably improving clustering performance. Each piece of data comprises spatiotemporal markers that may be linked to specific objects, as well as moving item spatiotemporal behavior characteristics. Calculate the cosine distance between the cluster center and the incremental spatiotemporal trajectory data element record to be grouped in the data block. The formula for the computation is as follows:

$$H = \frac{\sum_{i=1}^u f_i f_k}{\sum_{i=1}^u f_i^2 \sum_{i=1}^u f_k^2}, \quad (5)$$

where  $H$  represents the distance;  $u$  indicates the number of attribute dimensions;  $i$  represents the sequence number of tuples;  $k$  represents the cluster center; and  $f$  represents a data element record. The data are divided into blocks, and the clustering results of each iteration are merged locally according to a certain strategy to form an intermediate result with a small amount of data and transmitted to the central node, so as to greatly reduce the amount of data to be moved and reduce the cost of data movement [12]. The discovered and retrieved predefined behavior trajectory data are mapped to geographical units based on the static viewpoint in order to investigate static attribute semantics such as the coupling interaction features of particular behavior events and location, location range, and so on. The spatiotemporal trajectory data pieces that fulfil an adequate similarity criteria are dynamically mapped to produce a semantic fusion dynamic reference table [13]. The semantic unification of diverse spatiotemporal trajectory data pieces and associated data records is done via the dynamic reference table. Traffic trajectory data have high-dimensional dynamics, big data size, road network accessibility, spatiotemporal asynchrony, and nonstationary spatiotemporal distribution, among other properties. When individual objects are aggregated to the group level, a significant number of group samples represent the aggregation level's group behavior activity mode. This rule of group activity behavior is inextricably linked to the urban geographical setting in which humans dwell. It may assist individuals in seeing geographical location features, comprehending behavior semantics, and revealing and characterising dynamic changes in semantics and socio-economic factors. The data source node may dynamically join or quit the global public semantic reference model without affecting the dynamic semantic fusion mechanism [14]. This research examines the change state and evolution process of semantic theme and behavior events in place from the standpoint of spatiotemporal dynamic change. For example, through the semantic understanding of long-time series of spatiotemporal events in the activity place, the microchange characteristics, driving mechanism, and influence range can be analyzed at a fine granularity. When the clustering operation is completed, all data element records in the obtained class are grouped according to their data elements, and the ratio of the number of records of a data element in the class to the total number of records of the data element is defined as the clustering accuracy [15]. The calculation formula of clustering accuracy is as follows:

$$\begin{cases} \theta = \frac{R}{R'} \\ |\theta - \theta_{T-1}| \leq \theta_0 \end{cases}, \quad (6)$$

where  $\theta$  is the global clustering result;  $\theta_0$  represents the given precision parameter;  $T$  represents the number of iterations; and  $R$  and  $R'$  are the input blocking results and the output results of secondary merging. When the clustering accuracy

is greater than a given threshold, the data element is considered to belong to this category. Different data elements in the class have the same semantics, so as to realize the clustering and fusion of group behavior patterns of heterogeneous spatiotemporal trajectory data elements with the same semantics.

*2.4. Designing an Automatic Mining Algorithm for Group Behavior Patterns.* On the basis of clustering group behavior patterns based on incremental spatiotemporal trajectory big data, an automatic mining algorithm of group behavior patterns is designed. The mining algorithm framework is shown in Figure 2.

Multiple attribute dimensions such as data definition, limitations, and feature words are used to standardize heterogeneous spatiotemporal trajectory data components of diverse data source nodes, and particular data records are standardized depending on the attribute dimensions of data elements [16, 17]. The automated mining technique for group behavior patterns' fundamental function is to link cluster sets of two neighbouring times in a comparable way. All cluster pairings with cluster similarity relationships are obtained from the cluster set at two distinct periods, according to the requirements of member similarity and spatial similarity. Cluster similarity connection is the term for this procedure [18]. The connection condition that members have similar needs to be met can be expressed as:

$$\lambda(A, B) = \frac{|A \cap B|}{|A \cup B|} \geq \varepsilon, \quad (7)$$

where  $\lambda$  represents similarity;  $A, B$  represent a set of two connected clusters; and  $\varepsilon$  represents the overlap rate of a given member. In this process, the most time-consuming part is cluster similarity search. The data in the sliding window can be divided into two parts: the new data just collected and some old data already processed. The size of the population in the cluster (i.e., the length of the cluster) is used for pruning, and the clusters whose length in the candidate set does not meet the conditions are directly filtered out [19]. The pruning conditions are as follows:

$$\lceil |A| \varepsilon \rceil \leq |B| \leq \lfloor \frac{|A|}{\varepsilon} \rfloor. \quad (8)$$

The cluster in the candidate set is trimmed further based on the sequence of moving items in the cluster and the location of the same object in the prefix. Traverse the index table for each object in the prefix of the cluster to be searched, acquire the matching index item, and use it as the cluster candidate [20, 21]. The secondary merging of intermediate findings is finished after merging separate data blocks in the central node, and the global clustering results are created. Although the old data were processed in the previous updated sliding window, the new data reconstituted a new whole in the current sliding window and must still be included in the computation. After that, all candidates check the cluster similarity link and deliver the final result, which includes length and position trimming. To produce candidate patterns, all cluster pairings with cluster similarity are detected in the

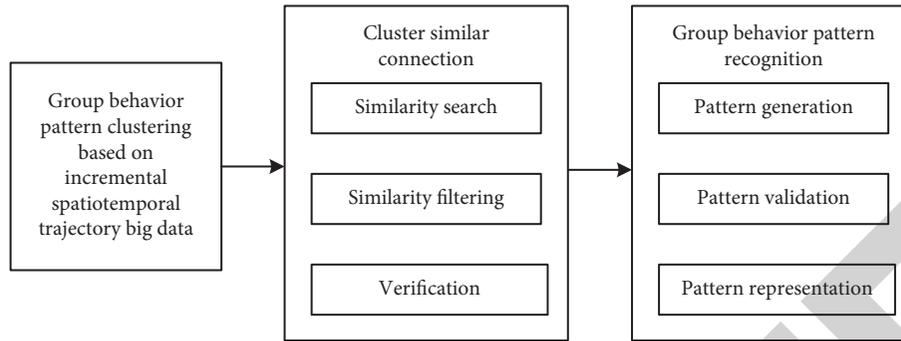


FIGURE 2: Framework for automatic mining of group behavior patterns.

temporal domain. Then, to locate all closed group behavior patterns, check whether these candidate patterns fit the concept of group behavior patterns [22, 23]. When the global clustering results meet the given conditions, the clustering operation ends. Otherwise, the central node outputs the comparison parameters and distributes them to each data source node for a new round of clustering analysis. Incremental group detection is carried out in combination with sliding time window. Each sliding time window slides on the time window, and group detection is carried out in the sliding time window. At each moment, density-based clustering is carried out and an effective snapshot cluster set is obtained. The latest snapshot cluster in each candidate's cluster sequence is then found using the cluster similarity relationship, and the connection action is performed. Then, decide whether or not the applicant should be given another chance. If this is the case, extend the existing candidate and add it as a new candidate mode to the candidate set. The computed data results from the previous sliding time window will be utilised to update the groups in the current sliding time window with the new time window. Each lower bound function is connected in series to form a cascade lower bound that ranges from low to high in time complexity, and all subsequences in the data set are traversed, calculated, and compared to eliminate dissimilar subsequences as much as possible, reduce calculation time, and improve retrieval efficiency [24]. Condition judgement and mode verification are carried out in the absence of this. If the interruption requirements are satisfied, the candidate mode's interruption time is updated. Otherwise, it signifies that the candidate cannot be enlarged any more (closed), and the group behavior pattern verification is performed. Finally, the currently unassigned snapshot clusters may form an aggregate movement mode at a later point. Therefore, all unassigned snapshot clusters at the current time should be added to the candidate set as new candidates. Thus, the design of group behavior pattern automatic mining method based on incremental spatiotemporal trajectory big data is completed.

### 3. Experiment

**3.1. Experimental Environment and Setting.** In this section, Windows Server 2008 system, R language development environment, and PC equipped with Intel Core, 2.6 GHz, and 128 GB RAM are selected as the experimental platform.

Dataset refers to the data provided by the "traffic line access time prediction" algorithm competition held on the big data competition platform of data castle. The dataset contains information of 14000+ taxis, more than 1 billion GPS records, and the daily data flow is about 2.8 GB. The official cleans out the repeated and abnormal records, with a sampling interval of 30 seconds. The track data include vehicle number, geographic location (longitude and latitude), speed, direction, time stamp, and other information. In this paper, the passenger status information irrelevant to the experiment is ignored. In this experiment, a user trajectory is randomly selected from the real trajectory dataset, and the three-month driving trajectory data are divided according to the time interval of continuous trajectory points (30 min), and 725 subtrajectories can be obtained. When the query track query is randomly selected from 725 subtracks, the remaining 724 subtracks will be used as the target track. The time domain of the moving object dataset used includes 30 time slices, the cluster member threshold is 10, the group lifetime threshold is 4, the sliding time window width is 5, and the sliding distance is 1. The maximum number of members in the partition is set to 1000, the distance threshold is 200 m, the grid size is generally  $100 \times 100$ , and the group clustering distance threshold is 400. The experiment is repeated for 5 times, and the average value is taken. After each group of experiments is completed, the optimal parameters will be selected as the default value for the next group of experiments.

**3.2. Experimental Results and Analysis.** The mining aggregation pattern entails mining a collection of aggregated groupings. When there is a vast quantity of data, the algorithm's mining performance will be drastically lowered, and the whole system will eventually be halted. This experiment will assess the mining efficiency of an autonomous mining approach for group behavior patterns based on incremental spatiotemporal trajectory big data. The running time of the approach proposed in this study is compared to DBSCAN and K-means-based automated mining algorithms for group behavior patterns. Different population size thresholds were set, which were 10 and 20, respectively. The population size threshold affects the size of the cluster set at each time, thus affecting the execution time of the connection operation. The running time results of each group behavior pattern mining method are shown in Tables 1 and 2.

TABLE 1: Running time under population size threshold 10 (s).

Number of tests	Automatic mining method of group behavior patterns based on incremental spatiotemporal trajectory big data	Group behavior pattern mining method based on DBSCAN	Group behavior pattern mining method based on K-means
1	164.45	290.05	300.46
2	182.51	285.44	286.58
3	175.82	280.88	311.80
4	178.24	284.69	305.65
5	169.69	291.26	294.28

TABLE 2: Running time under population size threshold 20 (s).

Number of tests	Automatic mining method of group behavior patterns based on incremental spatiotemporal trajectory big data	Group behavior pattern mining method based on DBSCAN	Group behavior pattern mining method based on K-means
1	310.43	445.49	460.41
2	300.67	450.85	462.87
3	306.82	458.52	458.68
4	315.75	457.64	454.52
5	324.63	452.23	465.13

When the group size threshold is 10, the running time of the group behavior pattern automatic mining method based on incremental spatiotemporal trajectory big data is 174.14 s, which is 112.32 s and 125.61 s shorter than that of the mining methods based on DBSCAN and K-means, respectively.

When the group size threshold is 20, the running time of the group behavior pattern automatic mining method based on incremental spatiotemporal trajectory big data is 311.66, which is 141.29 s and 148.66 s shorter than that of the mining methods based on DBSCAN and K-means, respectively. The time it takes for an algorithm to mine group behavior patterns increases as the group size barrier rises. Moving object points are used as clustering objects in the automated mining technique of group behavior patterns. A big quantity of data must be partitioned and clustered as the amount of data grows. Distributing data to relevant nodes will take a long time during the data splitting procedure. From the above experimental results, it can be seen that the running time of this design method is less than that of the mining methods based on DBSCAN and K-means, so it has higher execution efficiency.

#### 4. Conclusion

Various sensors, such as remote sensing equipment, vehicle GPS, smart phones, and so on, record the temporal and spatial trajectories of various moving objects. Through the study of these trajectories, we can mine valuable group behavior patterns and understand group movement behavior. This paper proposes an automatic mining method of group behavior patterns based on incremental spatiotemporal trajectory big data. This method can effectively shorten the running time of mining and improve the execution efficiency. The group mobility model proposed in this paper only considers the temporal and spatial information of mobile objects and does not integrate the information of environment, social networking, hot events, and so on.

Integrating spatiotemporal trajectory data with meteorological, microblog, and other data to study the group movement pattern under multivariate data is a problem worthy of research in pattern mining in the future. Follow-up research can be carried out on this issue to meet the needs of different levels.

#### Data Availability

The data used to support the findings of this study are included within the article.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This study was supported by Langfang Science Technology Research Self Financing Project (2019011054), Central University Basic Scientific Research Business Fee Project, and China Scholarship Fund Project.

#### References

- [1] S. Zhu, G. Sun, Q. Jiang, W. Xia, and R. Liang, "Microscopic visual analysis of high-density group trajectory data," *Journal of Computer-Aided Design & Computer Graphics*, vol. 32, no. 12, pp. 1871–1880, 2020.
- [2] L. Yu, L. Xiang, S. Sun, X. Guan, and H. Wu, "kNN query processing for trajectory big data based on distributed column-oriented storage[J]," *Geomatics and Information Science of Wuhan University*, vol. 46, no. 5, pp. 736–745, 2021.
- [3] W. Zhang, C. Ji, and S. Xie, "Pattern mining of spatiotemporal behavior trajectories by complex network analysis[J]," *Scientia Geographica Sinica*, vol. 41, no. 9, pp. 1505–1514, 2021.
- [4] D. Govender and J.-R. Tapamo, "Spatio-temporal scale coded bag-of-words," *Sensors*, vol. 20, no. 21, p. 6380, 2020.

- [5] B. Cao, H. Feng, J. Liang, and X. Li, "Hilbert curve and cassandra based indexing and storing approach for large-scale spatiotemporal data[J]," *Geomatics and Information Science of Wuhan University*, vol. 46, no. 5, pp. 620–629, 2021.
- [6] A. Bne, O. Colliot, and S. Durrleman, "Learning the spatio-temporal variability in longitudinal shape data sets[J]," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 2873–2896, 2020.
- [7] H. O. N. G. Yue-hua, "Group behavior patterns mining algorithm for distributed big data[J]," *Journal of Yulin Normal University*, vol. 40, no. 2, pp. 145–151, 2019.
- [8] Y. Sun, X. Liu, and W. Wang, "Predicting the event-driven evolution behavior of online social groups[J]," *Journal of Intelligence*, vol. 38, no. 6, pp. 110–117, 2019.
- [9] S. Wang, Y. Zhu, H. Wang et al., "Multi-modal group activity recognition method combining motion trajectory features[J]," *Microelectronics & Computer*, vol. 38, no. 11, pp. 7–13, 2021.
- [10] Z. Zhu, W. Huang, and X. Zhang, "Research on trajectory similarity of college students' network information exchange behavior[J]," *Information Studies: Theory & Application*, vol. 44, no. 10, pp. 158–164, 2021.
- [11] H. Huang, Y. Zeng, Yi Zhang et al., "Prediction method of abnormal crowd aggregation based on group behavior analysis[J]," *Computer Engineering*, vol. 46, no. 3, pp. 292–298, 2020.
- [12] C. H. E. N. Chuan-yi and W.-jun Dai, "High dimensional data hiding pattern mining based on bayesian network[J]," *Computer Simulation*, vol. 38, no. 1, pp. 287–290, 2021.
- [13] S. Sumalatha and R. Subramanyam, "A MapReduce solution for incremental mining of sequential patterns from big data[J]," *Expert Systems with Applications*, vol. 133, pp. 109–125, 2019.
- [14] C. Sweetlin Hemalatha, R. Pathak, and V. Vaidehi, "Hybrid decision trees for data streams based on Incremental Flexible Naive Bayes prediction at leaf nodes," *Evolutionary Intelligence*, vol. 12, no. 4, pp. 515–526, 2019.
- [15] M. R. Mohamed, A. A. Nasr, I. F. Tarrad et al., "Exploiting incremental classifiers for the training of an adaptive intrusion detection model[J]," *International Journal on Network Security*, vol. 21, no. 2, pp. 275–289, 2019.
- [16] R. Davashi, "ILUNA: single-pass incremental method for uncertain frequent pattern mining without false positives," *Information Sciences*, vol. 564, no. 1, pp. 1–26, 2021.
- [17] Y. A. N. G. Jing-rong and J. Liu, "Application of spatiotemporal association rules data mining based on ordered tree[J]," *Computer Technology and Development*, vol. 31, no. 6, pp. 19–23, 2021.
- [18] X. Li, Y. Wang, and L. I. Dong, "Application of quantum computing in incremental parallel mining of large data[J]," *Journal of Jilin University(Information Science Edition)*, vol. 38, no. 4, pp. 457–466, 2020.
- [19] W. Andrzejewski and P. Boinski, "Parallel approach to incremental co-location pattern mining," *Information Sciences*, vol. 496, pp. 485–505, 2019.
- [20] S. Bardak, T. Bardak, H. Peker, E. Sözen, and Y. Çabuk, "Predicting effects of selected impregnation processes on the observed bending strength of wood, with use of data mining models," *Bioresources*, vol. 16, no. 3, pp. 4891–4904, 2021.
- [21] M. Y. N. Attari, B. Ejlaly, H. Heidarpour, and A. Ala, "Application of data mining techniques for the investigation of factors affecting transportation enterprises," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–16, 2021.
- [22] Y. Mei, "Research on dynamic incremental distributed mining method of network information big data[J]," *Journal of Anyang Institute of Technology*, vol. 19, no. 2, pp. 41–44, 2020.
- [23] G. Taranto-Vera, P. Galindo-Villardón, J. Merchán-Sánchez-Jara, J. Salazar-Pozo, A. Moreno-Salazar, and V. Salazar-Villalva, "Algorithms and software for data mining and machine learning: a critical comparative view from a systematic review of the literature," *The Journal of Supercomputing*, vol. 77, no. 10, pp. 11481–11513, 2021.
- [24] N. Nan, "Efficient distributed incremental sequence mining in big data environment[J]," *Journal of South China Normal University (Social Science Edition)*, vol. 45, no. 11, pp. 80–85, 2020.