A Summary of the Research on AIS Track Clustering Methods

Tong Jiang, Fuliang Guo*,Gang Zhou
School of Electronic Engineering, University of Naval Engineering
Wuhan, China
yjzjh0601@qq.com, fuliang_guo@163.com

***Abstract*—There are a lot of water traffic characteristics in the track data of ship Automatic Identification System (AIS) [1]. Lots of effective and potential information can be obtained by using clustering methods to analyze and study the historical track data of AIS. Based on the relevant literature at home and abroad, firstly, this paper briefly introduces the AIS track data and sums up the general steps of AIS track clustering, and then classifies and summarizes the common algorithms of AIS track clustering and their advantages and disadvantages. Finally, the future development trend and main problems of the study are prospected from the perspectives of data preprocessing, the influence of environmental factors and the application in the military field. This paper has certain reference significance to the related research.**

Keywords—AIS track; Clustering methods; Abnormal target detection; Maritime intelligent supervision

I. INTRODUCTION

In recent years, the Automatic Identification System (AIS) has been continuously improved and widely used, and the data of the AIS track with lots of useful information has been generated. At present, the domestic and foreign experts and scholars mainly use the method of clustering to extract the ship motion pattern and analyze the characteristics of the ship's behavior from these data and carry out the relevant research in different fields [2-3]. In respect of abnormal behavior detection, Pan et al. proposed an abnormal behavior detection method based on multi-dimensional track characteristics. The method was effect- tively verified on the simulation military scene and the real civil scene, as shown in Fig. 1 [4]. Li et al. generated the knowledge base of target motion characteristics in a specific area based on the cumulative data clustering. They calculated the abnormal target set by the real-time data test analysis [5]. In respect of intelligent maritime supervision, Yan et al. classified the track points into two behavior states: in-flight and anchor and discussed the choose of main route path and anchor site location [6]. CHEN et al. established the statistical inference model by using the trajectory grid method. They inferred the main track bands of the two kinds of traffic flows in the Taiwan Strait and identified the significant conflict areas, so as to adjust the existing warning area setting scheme [7].

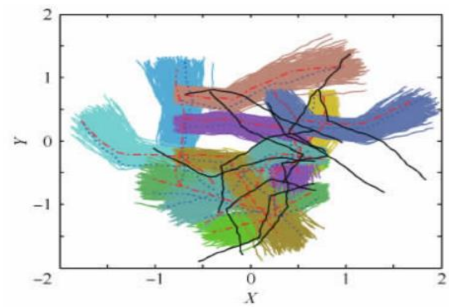

Fig. 1. Schematic diagram of 2000 tracks in the experimental data set.

In this paper, the research progress of ship AIS track clustering method and its application are systematically reviewed [8~10]. Firstly, briefly introduce the track data of the AIS and induce general steps of AIS track clustering, then sum up the common algorithms of AIS track clustering and their advantages and disadvantages, and finally look forward to the development trend and challenges of this research in the future.

II. AN OVERVIEW OF AIS TRACK CLUSTERING

A. AIS track data introduction

AIS is a new type of navigation system applied to the maritime safety and communication between the ship and the shore, the ship and the ship. It is usually composed of a VHF communication machine, a GPS locator and a communication control unit. The International Maritime Organization (IMO) stipulates that all international navigation ships of 300 gross tons and over, non-international navigation ships of 500 gross tons and over and all passenger ships should be equipped with one AIS as required. The number of AIS report types are 13, including ship position report, base station report, channel management and so on, with length bits ranging from 168 to 1192. It records static information such as Maritime Mobile Service Identification (MMSI), IMO number, length and width of ship and ship type, as well as dynamic information such as ship position, ground speed, ground heading and navigation status.

B. General steps of track clustering

Track clustering is the process of dividing all tracks into different clusters, which makes the objects in the cluster similar to each other, but not similar to those in other clusters. The main contents to be studied include research object, feature selection, similarity measurement and clustering algorithm, as shown in Fig. 2.

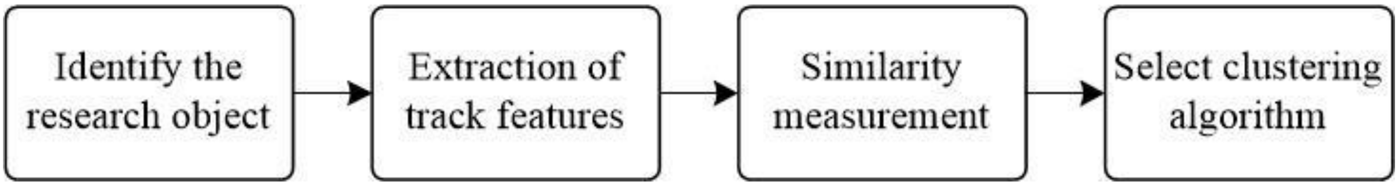

Fig. 2. General steps of track clustering.

(1) Identify the research object. In the aspect of the research object of track clustering, It is originally a point trace, that is, a position point in a track. LIU et al. proposed a method based on the IMO rules to cluster the AIS trace of ships as an auxiliary means to extract the features of the "class" [11]. However, point-based clustering can not reflect the spatial continuity, spatial-temporal correlation and local pattern similarity between adjacent points on the same track [12]. Since then, the focus of the study turned to the sub-track section. Through the analysis and comparison of the sub-track segments, the similarity measurement between them is established, and the clustering results can well describe the trajectory motion patterns of different forms and categories. Lee et al. proposed a framework of track clustering "TRACLUS" [13]. The framework divides the tracks into sub-track segments, and the similarity between the balance segments is processed by linearizing the shape and angle of the track segments. This method has been widely used for reference by subsequent related research. Xiao et al. also put forward that the track should be divided into sub-track segments, and take the point where the course changes greatly as the breakpoint of the track [14]. Wei et al. and Peng et al. realized the linear segmentation and clustering of the ship's AIS track by a similar method [15~16]. How to segment becomes a key factor that affects the clustering effect.

(2) Extraction of track features. On the feature selection of track clustering, Li et al. took the whole track as the cluster object to establish the spatial position distance matrix between the tracks, and then realized the class division in different spatial distribution and direction [17]. However, they ignored the dynamic ship information such as heading and speed of each point on the track. Zhang et al. proposed a similarity model of ship behavior, which merges dynamic information such as position, speed, course and so on, designed a ship track clustering method based on behavior feature similarity. Then they used the AIS track data of the south channel section of the Changjiang Estuary to verify and analyze the algorithm. The results showed that it can effectively distinguish the difference of trajectory in spatial position distribution [18].

(3) Similarity measurement. In the research of track similarity measurement, there are several commonly used similarity calculation methods.

a. Based on Hausdorff distance. The Hausdorff distance (HD) was originally designed to calculate the distance between two sets in the metric space. This similarity measurement method does not consider the timing of trajectory points, and the similarity of trajectory sequences can not be well expressed [19]. Cao et al. improved the Hausdorff distance to effectively solve the problem that any single isolated point leads to excessive distance between sets and disordered points in the set [20]. Chen et al. introduced the idea of translation into Hausdorff distance to eliminate the common deviation between trajectories, and compared trajectories by point-to-point to make the set of points orderly [21].

b. Based on the longest common subsequence distance. Wei et al. proposed a longest common subsequence (LCS) distance algorithm. In processing unequal length data, it is possible to calculate the length of discontinuous longest common subsequence between trajectories to measure the inter-trajectory of similarity, and does not need to match all trajectory points and noise robust, etc. [22].

c. Other measurement methods: The structured distance gives different weights to the different attributes of the trajectory according to the actual situation, so as to judge the similarity between the two trajectories synthetically [23]. In addition, there are similarity measures based on editing distance (ED), single term distance and discrete Frechet distance [24~26]. However, if only distance is considered, it is easy to lose local feature information of trajectory, so it needs to be combined with clustering algorithm.

(4) Select clustering algorithm. In order to get better clustering effect, it is particularly important to select the appropriate clustering algorithm. At present, the commonly used track clustering algorithms mainly include density-based, statistics-based and grid-based trajectory clustering algorithms. The representative algorithm and its advantages and disadvantages will be introduced in detail in the next section.

III. COMMON ALGORITHMS FOR TRACK CLUSTERING

A. Density-based trajectory clustering algorithm

DBSCAN (Density Based Spatial Clustering of Applications with Noise) is a classical density-based clustering algorithm based on high density connected region [27]. It can divide the region with high enough density into classes and find any shape clustering in space database. How to set the neighborhood radius of clustering object and the number of samples in the neighborhood is the key to reduce the error. Zhao et al. proposed an adaptive hierarchical clustering method of ship trajectory based on DBSCAN by analyzing the characteristics of algorithm. They determined the optimal parameters pass points according to the inherent distribution law of data set and the variation law of quasi-clustering effect and took the AIS trajectory data of Qiongzhou Strait as an example to verify, as shown in Fig. 3 [28]. PALLOTTA et al. clustered of the steering point in the ship's AIS track data by using the increment DBSCAN algorithm to analyze the traffic flow patterns of ships [29]. Dong simplified the clustering of trajectory segments to clustering about vector points based on DBSCAN algorithm, by drawing lessons from the thinking of segmentation and grouping. They greatly reduced the complexity of the algorithm as shown in Fig. 4 [30].

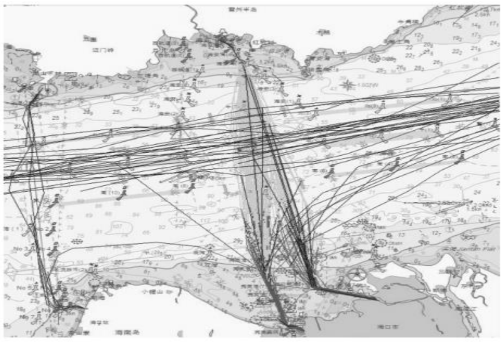

Fig. 3. The cluster results under the optimal parameters

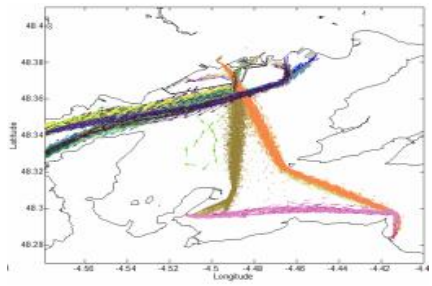

Fig. 4. Colour-coded oriented route objects.

In order to overcome the disadvantage of using a group of global parameters in cluster analysis, the OPTICS (object sort recognition clustering structure) algorithm which identifies the cluster structure by the point-ordering emerges as the times require [31]. Song took into account the course of the ship based on the OPTICS algorithm and realized the excavation of the cross and overlapping traffic flow of the ship [32]. Hu et al. combined an improved OPTICS algorithm and the technology of relational analysis [33].

B. Trajectory clustering algorithm based on statistics

Based on the mixed gaussian model (GMM) is a trajectory modeling approach that is often used in trajectory clustering [34]. Based on the measured data from Yantai Port to Dalian Bay, jiang et al. adopted an analysis method for constructing GMM and principal component analysis model for cluster analysis of target trajectory behavior [35]. The GMM is based on the multivariate normal distribution, without prior knowledge, and selects the parameter with the largest probability density according to the maximum likelihood criterion. But parameter values require a large amount of observational data to determine [36~37].

Kernel density estimation (KDE) is another effective trajectory clustering method. KDE is a nonparametric density estimation method derived from statistics, which is used to estimate unknown density functions. Prior knowledge is not needed to analyze the distribution characteristics of data, so it is widely used in the application of statistics. Ristic, LAXHAMMAR, LAMPE et al. used the observed values to characterize the total ship motion mode by selecting the appropriate kernel function and window width [38~40].

C. Grid-based trajectory clustering algorithm

The grid-based clustering adopts the space-driven method to divide the embedded space into units which are independent of the input object distribution. CLIQUE (Clustering In QUEst) divides each dimension into non-overlapping intervals. Thus, the whole embedded space of the data object is divided into the units [41]. STING (Statistical Information Grid), which is a multi-resolution clustering technique based on the network, divides the spatial region of the input object into rectangular elements [42].

D. Other trajectory clustering algorithms

Du et al. used the hierarchical clustering method to cluster the ship trajectory by extracting the ship navigation characteristic points from the ship AIS trajectory data. They pruned the cluster tree with the appropriate pruning method, so as to obtain the ship track classification in a target sea area [43]. ETIENNE et al. proposed a ship AIS trajectory clustering method based on graph theory. In this method, the starting port and destination port of the ship are regarded as the nodes in the map, and the trajectory of the same kind of ship is unified into a specific directed path [44]. KROODSMA et al. use con-volution neural network (CNN) to cluster the global fishing vessel trajectory and extract the location of hot fishing vessels with different fishing methods [45].

Several clustering algorithms and their advantages and disadvantages are shown as Table 1.

TABLE I. SEVERAL CLUSTERING ALGORITHMS AND THEIR ADVANTAGES AND DISADVANTAGES

Clustering algorithm	**Representation**	**Advantages**	**Disadvantages**
Based on density	DBSCAN OPTICS	Find a cluster of trajectories of any shape; Strong robustness to abnormal trajectory	The clustering effect is poor when the trajectory density is not uniform and the class spacing is very different
Based on statistics	GMM KDA	The inference and prediction are made based on the quantitative analysis of the data.	The cost of calculation and computation complexity are high.
Grid-based	CLIQUE STING	Fast processing speed	The clustering effect depends heavily on the division of network and network units.

IV. The Prospect

The development trend and challenges of AIS track clustering method in the future will be embodied in the following aspects.

A. Date processing

In the research of data mining, data preprocessing accounts for more than 70% of the workload, and many research methods can not get good research results because of the problem of data quality. How to combine related algorithms with big data processing platform such as hadoop is very important to improve the cluster effects.

B. *Environmental factors*

Environmental factors have an important influence on clustering effect. The natural environmental factors represented by visibility, wind flow and ocean current, and the social environmental factors represented by shipping market and inter-country situation need our comprehensive consideration in the future research.

C. *Military field*

Due to many factors, such as small data samples, the research of AIS track in military field is still in its infancy. How to improve the credibility of simulation data is a serious difficulty in this field.

V. THE CONCLUSION

With the improvement of onshore base station and the development of satellite communication technology, the data generated by ship Automatic Identification System (AIS) is increasing day by day, and the research on data mining of track data has been paid more and more attention in the past ten years. On the basis of briefly introducing AIS track data, this paper sums up the general steps of AIS track clustering, classifies and summarizes the common algorithms and advantages and disadvantages of AIS track clustering, and looks forward to the development trend and challenges of this research in the future. It is helpful to master AIS track data clustering process, common algorithms and related research.

REFERENCES

[1] S.H. Ma. Route generation based on AIS data [D]. Dalian: Dalian Maritime University, 2017.

[2] Y Chen. The application and analysis of cluster analysis in ship collision[J]. Ship Science and Technology, 2016(14): 163-165.

[3] Y Chen. A clustering mining analysis method for target track data [J]. Radio Engineering, 2015, 45(3): 22-24.

[4] X.L. Pan, H.P. Wang, Y. He, et al. The abnormal behavior detection method based on the multi-dimensional track feature [J]. Journal of Aeronautics, 2017, 38(4): 249-258.

[5] X.M. Li,J Bai.Target behavior anomaly detection based on clustering [J]. Computer and Network, 2018, 44(14): 62-64.

[6] W. Yan, R. Wen, A.N. Zhang, et al. Vessel movement analysis and pattern discovery using density-based clustering approach [C]. International Conference on Big Data. IEEE, 2016: 3798-3806.

[7] J.H. Chen, F. Lu, G. Peng. A quantitative approach for delineating principal fairways of ship passages through a strait [J]. Ocean Engineering, 2015, 103: 188-197.

[8] G.D. Chou, M. He, J. Yang, et al. A summary of the key technology of the early warning and prediction of the abnormal track data [J]. Journal of System Simulation, 2017, 29 (11): 2608-2617.

[9] L.K. Xu, L.Z. Ren, S.B. Zhou. A review of the research progress of ship AIS trajectory clustering methods [J]. Journal of Guangzhou Maritime College, 2019, 27(2): 7-12.

[10] N.X. Mu, Y.J. Xu, H.C. Zhang, et al. A Review of the Mobile Trajectory Clustering Methods [J]. Bulletin of Surveying and Mapping, 2018(1): 1-7

[11] B. Liu, de Souza E N, Matwin S, et al. Knowledge-based clustering of ship trajectories using density-based approach [C]. International Conference on Big Data. IEEE, 2014: 603-608.

[12] Burzigotti P, Ginesi A. Automatic identification system receiver and satellite payload comprising the same: U.S. Patent 9,008,233 [P]. 2015-4-14.

[13] Lee J G, Han J, Whang K Y. Trajectory clustering: a partition-and-group framework [C]. Proceedings of the 2007 ACM SIGMOD international conference on Management of data. ACM, 2007: 593-604.

[14] X. Xiao, Z.P. Shao, J.C. Pan, et al. Ship trajectory clustering model based on AIS information and its application [J]. Chinese navigation, 2015, 38(2): 82-86.

[15] Z.K. Wei, K. Zhou, M. Wei, et al. Ship motion pattern recognition and application based on AIS data [J]. Journal of Shanghai Maritime University, 2016 (2): 17-22.

[16] X.W. Peng, S. Gao, X.M. Chu, et al. Spark - based clustering method for ship trajectory [J]. Navigation of China, 2017, 40(3): 49-53.

[17] H Li, J Liu, R Liu, et al. A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis [J]. Sensors, 2017 (17): No.8.

[18] C.W. Zhang, J. Ma, Y.M. Niu, et al. Ship trajectory clustering method based on behavior feature similarity [J]. Journal of Wuhan University of Technology (Traffic Science and Engineering Edition), 2019, 43(3): 517-521.

[19] F.X. Zhu, Y.J. Zhang, Z.J. Gao. Research on Ship Behavior Based on Data Mining [J]. China Navigation, 2012(2): 50-54.

[20] Y.Y. Cao, Z.M. Cui, J Wu. A vehicle trajectory Model Learning method based on improved Hausdorff distance and Spectral clustering [J]. Computer application and software, 2012, 29(5): 38-40.

[21] J.Y. Chen, J.T. Song, L.X. Liu, et al. Trajectory clustering method based on improved Hausdorff distance [J]. Computer Engineering, 2012(17): 157-161.

[22] L.X. Wei, X.H. He, Q.Z. Teng, et al. The track classification of the Hausdorff distance and the longest common sub-sequence is combined [J]. Electronic and Informatics, 2013, 35(4): 784－790.

[23] G. Yuan, S.X. Xia, L. Zhang, et al. Trajectory clustering algorithm based on structural similarity [J]. Journal of Communication, 2011, 32(9): 103-110.

[24] VRIES G D, SOMEREN M V.. Unsupervised Ship Trajectory Modeling and Prediction Using Compression and Clustering [C]. Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning, 2009: 7-12.

[25] W.Y. Ma, Z.L. Wu, J.X. Yang, et al. Motion pattern recognition of spectral clustering ship based on one-way distance [J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 2015, 34(5): 130－134.

[26] Y.L. Jiang, Z.N. Xiong, J.H. Tang. The algorithm of ship trajectory clustering based on the track segment DBSCAN [J]. China's navigation, 2019, 42(3): 1-5.

[27] D. Wu, Y.Y. Du, J.W. Yi, et al. Spatio-temporal clustering analysis of trajectory based on density [J]. Journal of Earth Information Science, 2015, 17(10): 1162-1171.

[28] L.B. Zhao, G.Y. Shi, J.X. Yang. Adaptive hierarchical clustering of ship trajectory based on DBSCAN algorithm [J]. Chinese Navigation, 2018, 41(3): 53-58.

[29] Pallotta G, Vespe M, Bryan K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction [J]. Entropy, 2013, 15(6): 2218-2245.

[30] Y.Q. Dong. Spatial-temporal trajectory clustering analysis and research based on DBSCAN improved algorithm [D]. Tianjin: Tianjin University, 2018.

[31] Y.L. Zeng, H.B. Xu, S. Bai. An Improved OPTIS Algorithm and Its Application in Text Clustering [J]. Journal of Chinese Information Processing, 2008(1): 51-55.

[32] P. Song. Ship trajectory clustering and application based on C-OPTICS algorithm [D]. Dalian: Dalian Maritime University, 2017.

[33] W.B. Hu, W. Huang, G.C. Hu. Trajectory adjoint pattern analysis based on OPTICS clustering and association analysis [J]. Computer and Modernization, 2017 (12): 82-87.

[34] Riveiro M J. Visual analytics for maritime anomaly detection [D]. Sweden: University of Skovde, 2011.

[35] B.C. Jiang, J. Guan, W. Zhou, et al. Mining and recognition Analysis of ship abnormal behavior in Maritime Traffic [J]. Computer Simulation, 2017, 34(6): 329-334.

[36] J.H. Chen, F. Lu, G.J. Peng. The Progress of Research in Mari- time Vessel Trajectory Analysis [J]. Chinese Navigation, 2012, 35 (3): 53-57.

[37] J.H. Chen, F. Lu, M.X. Li. Optimizing the Precautionary Areas of Coastal Ships Routing in Western Taiwan Strait with Statistical Inference Method for Delineating the Boundary of Maritime Principal Fairways [J]. Journal of Geo-Information Science, 2015, 17(10): 1196-1206.

[38] Ristic B, La Scala B, Morelande M, et al. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction [C]. 11th International Conference on Information Fusion. IEEE, 2008: 1-7.

[39] Laxhammar R, Falkman G, Sviestins E. Anomaly detection in sea traffic -a comparison of the gaussian mixture model and the kernel density estimator [C]. 12th International Conference on Information Fusion. IEEE, 2009: 756-763.

[40] Lampe O D, Kehrer J, Hauser H. Visual analysis of multivariate movement data using interactive difference views[C]. 15th International Workshop on Vision, Modeling and Visualization, 2010: 315-322.

[41] X.Q. Xiang, H. Li, and S.B. Chen. Analysis and Research on CLIQUE Algorithm [J]. Journal of Hefei College (Natural Science Edition), 2011, 21(1): 54-58.

[42] Wang W, Yang J, Muntz R. STING: A statistical information grid approach to spatial data mining [C]. 23rd International Conference on Very Large Databases, 1997: 186-195.

[43] L. Du, Y.Q. Wen, C.S. Xiao, et al. Collision probability calculation for ship sailing in free navigational sea area [J]. China Safety Science Journal, 2015, 25(1): 53-59.

[44] Etienne L,Devogele T, Bouju A. Spatio-temporal trajectory analysis of mobile objects following the same itinerary [J]. Advances in Geo-Spatial Information Science, 2012,(10): 47-57.

[45] Kroodsma D A, Mayorga J, Hochberg T, et al. Tracking the global footprint of fisheries [J]. Science, 2018, 359: 904-908.