Adaptive Density Ship Trajectory Clustering Based on AIS data

Changsheng Wu*, Shuchang He

Dalian Maritime University, Dalian, Liaoning 116026, China

*Corresponding author's e-mail: wcs@dlmu.edu.cn

Abstract. Ship trajectory clustering is one of the main methods for mining ship feature trajectories based on AIS data. However, there are two main problems in trajectory clustering: First, the clustering algorithm itself has the problems of difficult to determine the parameters and poor noise recognition ability; second, the trajectory similarity metric, most of the measurements are only similarity metrics for the ship's position, and do not take into account the other dimensional information of the ship's trajectory. In order to solve these problems, this paper proposes a fast adaptive density clustering method for ship trajectories, which integrally considers multiple attributes of ship position, heading and speed to construct similarity metrics between ship trajectories; introduces Silhouette Coefficient (SC) and Davies-Bouldin Index (DBI) The DBTCAN algorithm is constructed to evaluate the comprehensive CMI index, which in turn realizes the adaptive selection of clustering parameters. An example study was conducted using AIS data of real waters, and the results show that the method can adaptively cluster ship trajectories to match the traffic situation of real waters.

Keywords: AIS data, trajectory clustering, similarity measurement, adaptive

1 Introduction

With the development of navigation digitalization, the Automatic Identification System (AIS) has been widely used, and the AIS data contains a large number of ship motion characteristics [1], which is often used for traffic flow feature extraction [2], route planning [3] and other studies, which cannot be separated from the analysis of ship behavior patterns and traffic flow distribution [4]. These researches can not be separated from the understanding of ship behavior pattern analysis and traffic flow distribution characteristics, and trajectory clustering is an effective means to realize this kind of research. Scholars at home and abroad have conducted a series of studies on ship trajectory clustering, which can be divided into three categories:

The first category is to cluster the data of individual trajectory points. The position of AIS data is mainly used to analyze the ship trajectory movement pattern. Yan et al [5] divided the ship navigation state into two states, route and docking, and extracted the ship's route by clustering analysis of the trajectory points of the two states, respec-

G. Zhao et al. (eds.), *Proceedings of the 2024 7th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA 2024)*, Advances in Engineering Research 241,
https://doi.org/10.2991/978-94-6463-514-0_19

tively. Xu et al [6] divided the complex waters by using the ship's position, speed, and heading, and combined the three attributes to the traditional DBSCAN algorithm for the extension, and finally proposed the extraction method of typical trajectory using the idea of vector point representation. Clustering based on trajectory points is the most commonly used method, but it ignores the spatio-temporal correlation between neighboring points, which is not conducive to the description of overall motion characteristics.

The second category is clustering with the whole ship trajectory. Mou Junmin et al [7], in order to realize the fast clustering of ship trajectories, designed a similarity measure function based on Hausdorff distance with automatically selected scale parameters, and used a spectral clustering algorithm to cluster the ship trajectories; Zhao et al [8], in order to improve the clustering performance of the ship trajectory data characterized by the large amount of data and the distributional complexity, proposed the DP compression-based and improved adaptive DBSCAN algorithm to extract the ship trajectories. Adaptive DBSCAN algorithm to extract maritime traffic features. Yang et al [9] proposed a density-based trajectory clustering of applications with noise (DBTCAN) algorithm that can directly cluster complete ship trajectories. algorithm, which will be generalized from traditional point clustering to line clustering and carried out the research of route identification.

The third category is clustering with ship trajectory segments. Xiao Xiao [10] proposed to obtain a typical motion model of a ship by clustering ship trajectory segments based on AIS data and extracting typical trajectories. Jiang Yuling et al [11] take the speed and heading as the information metrics to divide the ship trajectory, and use the DBSCAN algorithm to cluster the trajectory segments. Gao et al [12] cluster the ship sub-trajectory segments through the spectral clustering algorithm, and identify the representative trajectories of the ship's maneuvering behavior. Liu Yu et al [13] combined the heading and speed change rate to obtain feature points for trajectory segmentation, and obtained the typical motion trajectory of the ship by DBSCAN algorithm. Most segmentation methods of trajectory segment clustering only consider a single index, and different segmentation criteria lead to very different clustering results with great uncertainty.

In this paper, we first construct a multi-dimensional comprehensive similarity metric, the trajectory similarity metric not only considers the position of the trajectory, but also integrates the ship's speed and heading information, the DBTCAN algorithm is affected by the selection of parameters to influence the final clustering effect, and we introduce SC and DBI to construct a comprehensive clustering evaluation index CMI to evaluate the clustering algorithm, and determine the optimal number of clusters in an adaptive way.

2 Trajectory Clustering Model

DBTCAN trajectory clustering algorithm is based on the DBSCAN algorithm from the traditional point clustering promoted to line clustering. The algorithm calculates the similarity between trajectories through the sequence similarity metric algorithm to construct the ship trajectory segment clustering model can be better adapted to the clustering of ship trajectories [14,15]. The comprehensive similarity measure that integrates the speed and heading features on the basis of Hausdorff spatial distance is used to measure the similarity between trajectories, and the DBTCAN algorithm is utilized to analyze the clustering of ship trajectories.

2.1 Trajectory Similarity Metric

The similarity measure of ship trajectories is the basis for realizing the clustering of ship trajectories. In this paper, in addition to the spatial distance between trajectories as the similarity metric, the speed and heading difference characteristics between trajectories are also taken into account, constituting a comprehensive similarity metric between trajectories to quantify the similarity between trajectory segments.

Spatial. Distance.
Hausdorff distance calculates the distance between two sets of point sets and describes the degree of similarity. The larger the Hausdorff distance between two sets of point sets, the lower the similarity. Given 2 trajectories consisting of a number of ordered trajectory points, the Hausdorff distance between set A and set B is:

$$D_d\left(A,B\right)=max(h(A,B),h(A,B)) \tag{1}$$

$$h\left(A,B\right)=\underset{a_i\in A}{max}(\underset{b_{i\in B}}{min}\|a_i-b_i\|) \tag{2}$$

$$h\left(B,A\right)=\underset{b_{i\in B}}{max}(\underset{a_i\in A}{min}\|b_i-a_i\|) \tag{3}$$

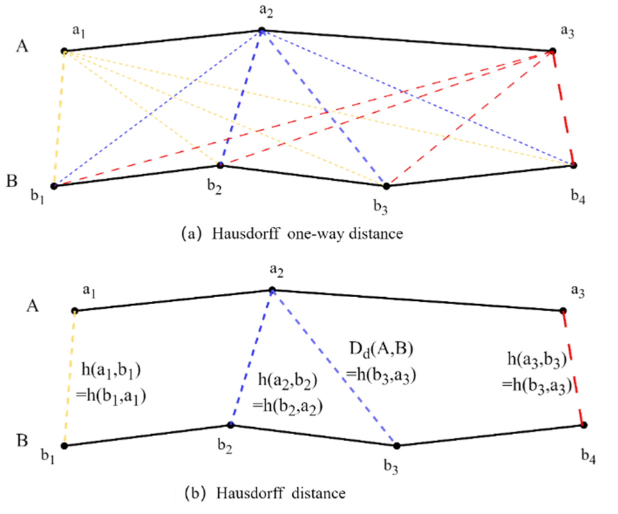

Fig. 1. Diagram of Hausdorff distance calculation

h(A, B) Firstly, the distance between each point ai in the point set A to the point in the set B which is closest to this point ai is ranked, and then the maximum value of this distance is taken as the value of h(A, B), and similarly, h(A, B) is obtained. In this paper, the Hausdorff distance is used to calculate the spatial distance between ship trajectories of different lengths as a similarity measure in spatial location. The Hausdorff distances between trajectories of different lengths are shown in Fig. 1.

Speed and Course distance.

When a ship navigates, its speed and heading will change due to geographical constraints and the relevant provisions of ship traffic management systems such as the ship routeing system. There is also a difference between the speed and heading of upstream and downstream ships traveling on the same route[1]. Therefore, the speed and heading characteristics are one of the important considerations, so in addition to the spatial distance, it is necessary to compare the speed difference and heading difference between trajectory segments. In this paper, the metric of speed distance is considered in terms of maximum speed, minimum speed and average speed combined.

$$D_V = \frac{1}{3}\left(v_{max}\left(a,b\right) + v_{mean}\left(a,b\right) + v_{min}\left(a,b\right)\right) \tag{4}$$

The course distance measure is a combination of maximum course, minimum course, and average course, i.e:

$$D_C = \frac{1}{3}\left(c_{max}\left(a,b\right) + c_{mean}\left(a,b\right) + c_{min}\left(a,b\right)\right) \tag{5}$$

where: $v_{\max}\left(a,b\right) = \left|v_{\max}(a) - v_{\max}(b)\right|$ is the difference value of the maximum speed between trajectory a and trajectory b; similarly, $v_{mean}\left(a,b\right)$, $v_{min}\left(a,b\right)$, and $c_{\max}\left(a,b\right)$, $c_{mean}\left(a,b\right)$, $c_{min}\left(a,b\right)$ are the difference values of the average speed, minimum speed, and the maximum heading, average heading, and minimum heading between trajectory a and trajectory b, respectively. Combining the three distances calculated for space, heading, and speed requires assigning different weights to these three distances, where $w = \{w_d, w_c, w_v\}$. The normalized distances are summed by assigning the corresponding weights to obtain the composite distance with the following formula:

$$D_{dis}\left(a,b\right) = w_d * d_d^{'} + w_c * d_c^{'} + w_v * d_v^{'} \tag{6}$$

2.2 DBTCAN trajectory clustering

DBTCAN can cluster ship trajectories of different lengths. The idea is to select any one trajectory segment in the trajectory set, and find all trajectories with similarity to this trajectory less than or equal to the neighborhood distance threshold ε through computation: if the number of trajectories is less than the neighborhood density threshold, the trajectory is labeled as a noise trajectory; if the number of trajectories is greater than or equal to it, the trajectory is a core trajectory. Then judge whether the trajectories within the neighborhood of this trajectory are core trajectories, if so, these trajectories belong to a cluster C. Then judge the other unselected trajectories within the set of trajectories until all the trajectories have been judged to complete the clustering.

The DBTCAN algorithm uses a parameter sum to determine the threshold value for dividing the set of high-density trajectories, and different combinations of parameters have a large impact on the final clustering effect. To realize the adaptive determination of parameters, the key lies in determining a suitable set of candidate threshold parameters. In this paper, using the spatial distribution characteristics of the trajectory set itself, based on the K-average nearest neighbor (KANN) algorithm [11] and the mathematical expectation method, respectively, to generate the input candidate parameters and sets of the DBTCAN algorithm, which are input into the DTSCAN algorithm to perform the clustering analysis on the trajectory set, and the number of cluster numbers can be obtained under different values of K. Then, the number of clusters under different values of K is calculated separately, and the number of clusters under different values of K is calculated. Then the clustering evaluation indexes SC, DBI and CM of each clustering cluster under different K values are calculated respectively, and the average score is output as the final clustering evaluation score of the clustering results. Finally, the class with the highest CM clustering evaluation index score is selected as the optimal clustering cluster.

2.3 Clustering Effectiveness Measures

Since the selection of the parameters of the DTSCAN algorithm affects the final clustering effect, in order to adaptively determine the optimal clustering parameters, and considering the practicality and objectivity of the clustering evaluation indexes, two standard evaluation of the Silhouette Coefficient (SC) and Davies-Bouldin Index (DBI) are combined to evaluate the performance and clustering effect of the clustering method. indexes are integrated to evaluate the performance and clustering effect of clustering methods. Comprehensive clustering index (CMI) takes into account the similarity of the samples within the class and the difference of the samples between the classes, and the larger the value, the better the clustering effect is, and the CMI is defined as follows.

$$\mathrm{CMI} = SC + \frac{1}{DBI} \tag{7}$$

3 Results and Analysis of Clustering Experiments

A set of simulated trajectory datasets is first used to experimentally analyze and validate the effectiveness of the fused multi-feature trajectory similarity metric and the adaptive DBTCAN algorithm. The simulated cross-trajectory data containing different moving directions are used in the experiments, and the adaptive parameter algorithm is used to cluster the simulated trajectory data sets respectively, and there are 10 clusters obtained with the corresponding neighborhood trajectory distance threshold e=0.33 and the neighborhood density threshold Lmin=8 for K=25, and the more ideal clusters are shown in Figure 2.

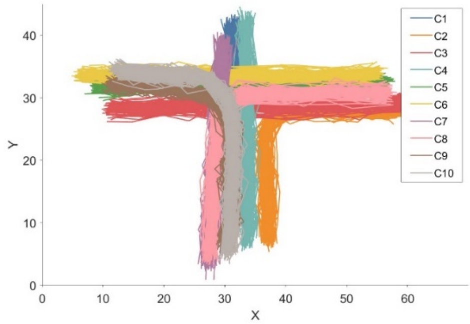

Fig. 2. Clustering effect of simulated trajectories

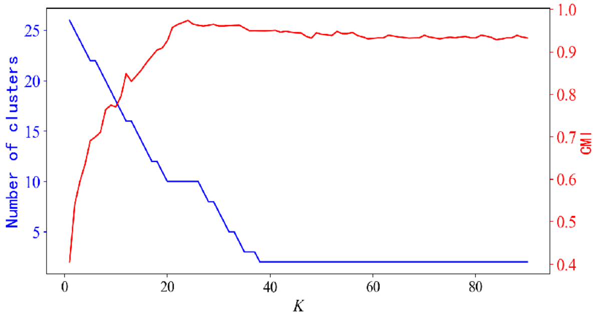

Fig. 3. Composite clustering indicator vs. K

Figures 3 and 5 show the correspondence between the K value and the clustering evaluation index value CMI, respectively. A comparison between Fig. 2 and Fig. 4 reveals that Fig. 2 shows the integrated trajectory similarity metric proposed in this paper that takes into account position, direction, and velocity improvements, which can cluster and distinguish trajectories in different directions, and 10 clusters of trajectory clusters in different directions are obtained. However, Fig. 4 shows the effect of clustering using the traditional Hausdorff distance metric of trajectory similarity, with only 7 trajectory clusters, which can be seen that only trajectories with closer positional distances are clustered into one class, and the direction of motion of the trajectories is not considered. Therefore, the integrated trajectory similarity measure constructed based on trajectory position, direction and velocity has some advantages in distinguishing trajectories with similar positions but different directions.

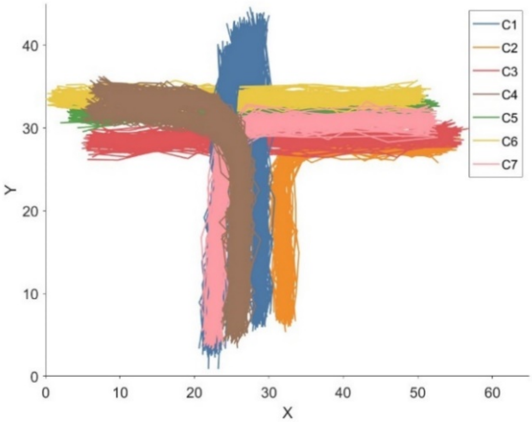

Fig. 4. Clustering effect of simulated trajectories

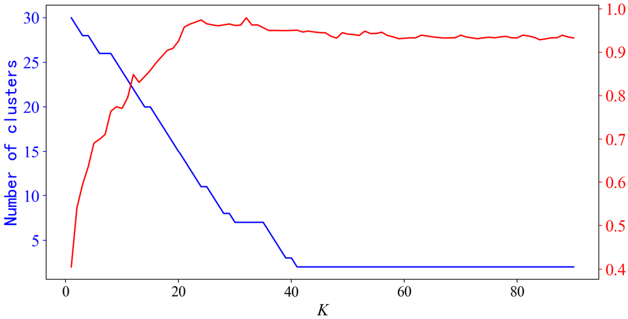

Fig. 5. Composite clustering indicator vs. K

Taking part of the ship AIS data in Zhoudai Bridge waters as the research object, using the parameter adaptive determination method to analyze the ship trajectory data, we get the relationship between the number of clustering clusters and the value of K. See Fig.6, starting from K=73, the number of clustering clusters began to have a period of time not changing and tend to be in a stable state, until the K is greater than 86 and then change and gradually become smaller until the last tends to be stable and no longer change. Then calculate the CMI clustering evaluation index of the clustering results under the parameter corresponding to each K value, as shown in Figure 6, then the K corresponding to the highest CMI score is the best K value parameter. The overall trend of the CM clustering evaluation index is first rising and constantly fluctuating, and finally tends to stabilize, and the CM clustering evaluation index scores the highest score when K = 86, indicating that the clustering results at this time is better, and at this time the corresponding K average nearest-neighbor distance that is, the optimal neighborhood trajectory distance threshold, and then calculate the neighborhood density threshold corresponding to this distance threshold, that is, the optimal neighborhood density threshold.

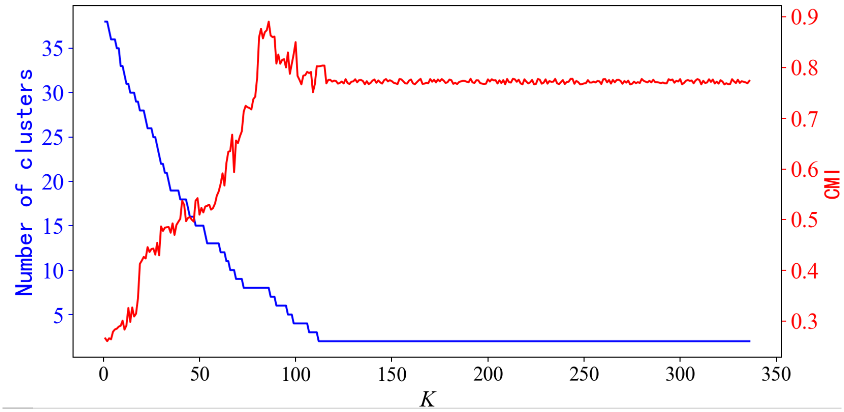

Fig. 6. Composite clustering indicator vs. K

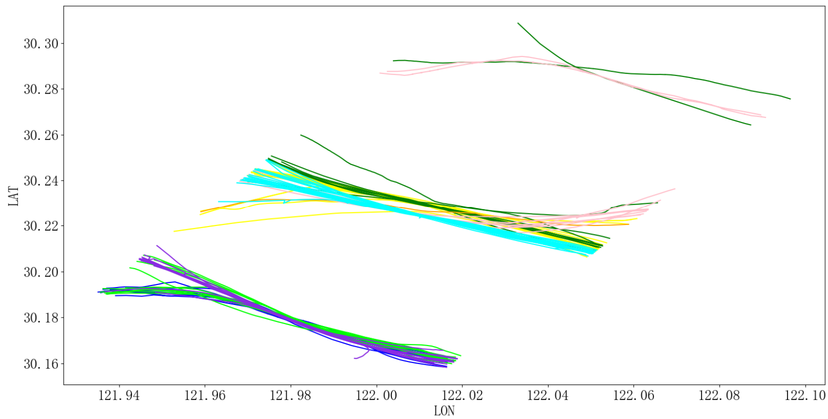

Fig. 7. Clustering effect diagram of ship trajectory

It is finally found that when the distance weight of trajectory similarity metric is set to 0.6, the distance weight of speed is 0.1, and the distance weight of heading is 0.3, the neighborhood trajectory distance threshold of 0.23 and the neighborhood density threshold of 6 corresponding to K=86 can obtain better experimental results, eight trajectory clusters in different directions can be obtained, and a more ideal clustering effect is achieved.

Zhoudai Bridge is divided into the south navigable hole, the main navigable hole and the north navigable hole, in which the main navigable is bi-directional navigable, as can be seen in Fig. 7 can be clustered and divided into different directions of traffic flow. The different traffic flows analyzed by clustering can provide reference suggestions for ship navigation, and can also be used as the basis for abnormal detection of ship trajectory, if a ship trajectory is inconsistent with the direction of the existing traffic flows analyzed by clustering, with a certain degree of deviation, it can be regarded as an abnormal trajectory.

4 Conclusions

This study takes the ship trajectory in the actual waters of Zhodai Bridge as the research object, and the improved trajectory similarity metric based on the trajectory position, direction and speed can distinguish the trajectory direction compared with the traditional Hausdorff distance, and applies the comprehensive clustering performance index CMI to the process of trajectory clustering of DBTCAN algorithm to determine the optimal input parameters, and the ship trajectories of the navigational holes of the region near Zhodai Bridge are is divided into several different direction clusters, and the results of the distinguished upstream and downstream channels all match with the electronic chart channels, indicating that the algorithm is able to find similar trajectories in a large number of complex ship trajectories and cluster them, and the clustering results are consistent with the actual traffic flow situation. Ship trajectory clustering is an important method to study ship behavior pattern, which can provide certain reference basis for relevant departments in ship behavior monitoring, route planning and ship routing system in the research waters.

References

1. Liu Wen, Wang Wen Bo, A vessel trajectory reconstruction method based on low-rank minimization matrix denoising. Journal of Transportation Systems Engineering and Information Technology, 2022, 22, pp: 106-114.
2. Rong H, Teixeira A P and Soares C G, Maritime traffic probabilistic prediction based on ship motion pattern extraction. Reliability Engineering & System Safety, 2022, 217, pp:20-32.
3. Shen Yu, Wang Fuping and Zhao Peimin, Ship route planning based on particle swarm optimization.2019 34RD Youth Academic Annual Conference of Chinese Association of Automation (YAC). 2019, pp:216-20.10.1109/yac.2019.8787628.
4. Alizadeh, D, Alesheikh, AA and Sharif, M, Vessel Trajectory Prediction Using Historical Automatic Identification System Data. Journal of Navigation, 2021, 74(1), pp: 156-174.
5. Rong H, Teixeira A P and Soares C G, Data mining approach to shipping route characterization and anomaly detection based on MS data. Ocean Engineering, 2020, pp: 198-216.
6. Zhang Chunwei, Ma Jie, NIU Yuanmiao and Llu Qi, Clustering method of ship trajectory based on similarity of behavior pattern. Journal of Wuhan University of Technology (Transportation Science &. Engineering), 2019, 43, pp: 517-621.
7. Yan Z J, Xiao Y J and Cheng L, Exploring AIS data for intelligent maritime routes extraction. Applied Ocean Research, 2020,36, pp: 101-112.
8. Xu X F, Cui D A and Li Y, Research on ship trajectory extraction based on multi-attribute DBSCAN optimisation algorithm. Polish Maritime Research, 2021, 28(1), pp: 136-148.
9. Yang J X, Liu Y and Ma L Q, Maritime traffic flow clustering analysis by density based trajectory clustering with noise. Ocean Engineering, 2022, 249, pp: 767-792.
10. Mou Jun Min, Chen Pengfei and He Yixiong, Fast self-tuning spectral clustering algorithm for AIS ship trajectory. Journal of Harbin Engineering University, 2018, 39, pp: 428-432.
11. Zhao L B, Shi G Y. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. Ocean Engineering, 2019, 172,pp: 456-467.

12. Zhang C, Liu S T and Guo M Z, A novel ship trajectory clustering analysis and anomaly detection method based on AIS data. Ocean Engineering, 2023, pp: 288-296.
13. Xiao Xiao, Shao Zheping and Pan Jiacai, Ship trajectory clustering model based on AIS data and its application. Navigation of China, 2015, 38, pp: 82-86.
14. Jiang Yuling, Xiong Zhennan and Tang Jihong, Ship trajectory clustering algorithm based on DBSCAN. Navigation of China, 2019, 42, pp: 1-5.
15. Liu Yu and Peng Pengfei, Research on clustering method of vessel trajectory based on improved DBSCAN. Ship Electronic Engineering, 2022, 42, pp: 57-63.