Clustering and Analyzing Vessel Sailing Routes

Subjects: Engineering, Marine
Contributor: Fangli Mou , Zide Fan , Xiaohe Li , Lei Wang , Xinming Li

A vessel automatic identification system (AIS) provides a large amount of dynamic vessel information over a large coverage area and data volume. The AIS data are a typical type of big geo-data with high dimensionality, large noise, heterogeneous densities, and complex distributions. This poses a challenge for the clustering and analysis of vessel sailing routes.

automatic identification system (AIS) | vessel sailing routes clustering and analyzing | ocean engineering | clustering

1. Introduction

With the continuous development of marine information technology, automatic identification system (AIS) equipment has been applied to ship navigation and has become one of the main navigational aids of ships. AIS is a system that can track and report the sailing status of ships, which is applied to maritime safety and communication between ships and shore and between ships and plays a significant role in maritime traffic management and ship navigation safety [1]. The AIS data contain the objective law of maritime traffic flow and have high application value. However, unlike vehicle trajectories on land, ships have high degrees of freedom when sailing, and AIS data are characterized by a large amount of data, large noise, heterogeneous densities, and complex distributions. This increases the difficulty of AIS data mining technology [2].

Vessel sailing route clustering was performed by analyzing the AIS records of the regions of interest. By clustering sailing routes, traffic flow information and traffic management references can be provided to the marine traffic supervision department. In addition, vessel sailing route clustering and analysis technologies are the foundation of sailing path planning [3][4], vessel trajectory prediction [5][6][7], and detection of abnormal trajectories [8][9].

The practical situation mainly lies in the following aspects: First, the sea routes and the traffic flow characteristics of ships can be revealed, which have attracted wide attention in the shipping industry. Second, the clustered sea routes and their densities that can guide the ships' navigation safety, which has always been the focus of continuous research in marine engineering. Furthermore, ship traffic monitoring can be performed, since main sea routes and sailing patterns are identified, and this is one of the biggest challenges in maritime law enforcement and emergency management. Hence, acquiring a ship's route can help understand ship behaviors and reveal ship regular movement patterns, and it is of great significance for ship anomaly detection, route planning, navigation safety, and maritime situation awareness.

Currently, there are two main problems in vessel sailing route clustering. The first problem is the low computational efficiency and the inability to effectively conduct clustering with a large amount of uniform density data. The other problem is the similarity measurement of the trajectories; most measurement methods consider the global features of trajectories and ignore the overall motion trend or local features of ship trajectories.

The main difference between the local and global features of the trajectories is whether the entire trajectory or trajectory segments were used to calculate the similarity measurement. The global features of the trajectories mainly include the overall direction and motion trend, and the similarity measurement is calculated using the entire trajectory. Correspondingly, the local features of trajectories primarily consist of the characteristics (such as density and trend) of local trajectory segments, and the similarity measurement uses similar local features between trajectory segments.

2. Clustering and Analyzing Vessel Sailing Routes

Related work on vessel trajectory clustering mainly uses similarity measurements of the trajectories or trajectory segments to perform clustering and analysis. Related methods are introduced as follows:

Currently, research on AIS trajectory clustering methods can be divided into two main categories: point-based clustering and trajectory-based clustering. The difference between the categories lies primarily in the different measurement methods for trajectory similarity or trajectory distance [10]. The point-based clustering method calculates the similarity of each trajectory point, which is easy to conduct but usually ignores the spatiotemporal correlation of vessel track points and is not conducive to the characterization of the overall motion characteristics of vessels [11][12]. The trajectory-based clustering method chooses the entire trajectory, or a trajectory segment composed of continuous trajectory points as the clustering object, and the similarity measure of the overall trajectory or the similarity measure of sub-trajectory segments is calculated [13][14]. The dynamic time warping (DTW) method, Hausdorff distance, and Frechet distance are commonly used to measure the similarities between trajectories. However, the DTW algorithm does not consider the characteristics of local trajectory segments, which may lead to over-stretching and over-compression [15][16]. The Hausdorff distance can identify the trajectory shape but ignores the timing features and direction of the trajectory, which is affected by trajectories near the transportation hub [17][18][19]. The Frechet distance can reflect the temporal characteristics of the trajectory and obtain better clustering results; however, it is not suitable for cluster analysis of long trajectory segments because of its high computational complexity [20].

After calculating the similarity measurement, clustering algorithms are used to determine the vessel sailing routes. Common clustering algorithms can be divided into the following categories: spatial clustering, hierarchical clustering, clustering based on density clustering, grid clustering, and model-based independent clustering algorithms [21]. The k-means method has the advantages of a simple process and fast calculation speed and is suitable for largescale datasets. However, it has poor processing ability for non-spherical clusters and is easily affected by the selection of initial cluster centers and the need to specify the number of clusters in advance [22][23][24]. Some widely used methods, such as the spectral clustering method, have disadvantages similar to those of the k-means method [25].

The density-based spatial clustering of applications with noise (DBSCAN) is one of the most widely used density-based clustering algorithms for analyzing AIS data [26]. The core idea of the DBSCAN algorithm is that a given data point belongs to a cluster if its density reaches a certain threshold; otherwise, it is considered a noise point. The DBSCAN method efficiently discovers clusters of arbitrary shapes and processes noisy data. In addition, it can automatically identify the number of clusters without specifying it in advance. However, DBSCAN faces challenges in terms of parameter sensitivity and time complexity. It has the problems of long clustering convergence time and low accuracy in the case of a large amount of data. Additionally, the clustering effect may be poor for datasets with large density differences. In addition, the hyperparameters of DBSCAN must be adjusted and optimized. To improve clustering performance, several studies have adopted many methods to determine the best parameters. Mohammad et al. proposed an adaptive DBSCAN method to identify clusters with varying densities; however, the number of data clusters must be predetermined, and the algorithm is less generalized [27]. Yang et al. proposed an adaptive semi-supervised method that uses both labeled and unlabeled data [28]. Liu et al. proposed a novel adaptive density trajectory cluster algorithm that computes the cluster radius using the density of the data distribution; however, the problem of time complexity remains [29]. Yaohui et al. introduced the concept of k-nearest neighbors and proposed an adaptive clustering algorithm to overcome the shortcomings of density peak-based clustering algorithms [30]. C Marques et al. proposed a clustered method to automatically perform a robust determination of cluster numbers and distributions; however, noisy data can be recognized [31]. Yang et al. proposed a density-based trajectory clustering of applications with noise (DBTCAN) algorithm to consider the spatiotemporal correlation between neighboring points on the same ship trajectory, which can be applied to pattern recognition of ship behavior in different ranges of waters [32]. Because the Hausdorff distance between every two trajectories is needed, the efficiency of DBTCAN is greatly reduced in large numbers of AIS data. Tang et al. proposed a FOLFST method to solve the problem of ignoring the overall motion trend or local features of ship trajectories [33], which also solved the problems of difficulty in parameter setting and sensitivity to a dataset with uneven density. However, the FOLFST still needs to simultaneously measure the similarity between the overall trajectories and calculate the single sub-trajectory distance, which requires more time. Yan et al. proposed a semantic extraction method for a large-area maritime route based on graph theory [34] that can be effectively used to analyze network characteristics, such as key nodes, edges, and network evolution. Waypoint analysis is important for this method.

References

1. Yan, R.; Wang, S.; Cao, J.; Sun, D. Shipping Domain Knowledge Informed Prediction and Optimization in Port State Control. Transp. Res. Part B Methodol. 2021, 149, 52–78.

2. Zhang, C.; Bin, J.; Wang, W.; Peng, X.; Wang, R.; Halldearn, R.; Liu, Z. AIS data driven general vessel destination prediction: A random forest based approach. Transp. Res. Part C-Emerg. Technol. 2020, 118, 102729.
3. Shen, C.; Shi, Y.; Buckham, B. Path-Following Control of an AUV: A Multi objective Model Predictive Control Approach. IEEE Trans. Control Syst. Technol. 2019, 27, 1334–1342.
4. Zhang, T.; Fu, M.; Song, W.; Yang, Y.; Wang, M. Trajectory Planning Based on Spatio-Temporal Map With Collision Avoidance Guaranteed by Safety Strip. IEEE Trans. Intell. Transp. Syst. 2022, 23, 1030–1043.
5. Chen, J.; Chen, H.; Zhao, Y.; Li, X. FB-BiGRU: A Deep Learning model for AIS-based vessel trajectory curve fitting and analysis. Ocean Eng. 2022, 266, 112898.
6. Xiao, Z.; Fu, X.; Zhang, L.; Goh, R.S.M. Traffic Pattern Mining and Forecasting Technologies in Maritime Traffic Service Networks: A Comprehensive Survey. IEEE Trans. Intell. Transp. Syst. 2020, 21, 1796–1825.
7. Chen, J.; Zhang, J.; Chen, H.; Zhao, Y.; Wang, H. A TDV attention-based BiGRU network for AIS-based vessel trajectory prediction. iScience 2023, 26, 106383.
8. Anneken, M.; Fischer, Y.; Beyerer, J. Evaluation and comparison of anomaly detection algorithms in annotated datasets from the maritime domain. In Proceedings of the 2015 SAI Intelligent Systems Conference (IntelliSys), London, UK, 10–11 November 2015; pp. 169–178.
9. Rong, H.; Teixeira, A.; Soares, C.G. Data mining approach to shipping route characterization and anomaly detection based on AIS data. Ocean Eng. 2020, 198, 106936.
10. Yang, D.; Wu, L.; Wang, S.; Jia, H.; Li, K.X. How big data enriches maritime research—A critical review of Automatic Identification System (AIS) data applications. Transp. Rev. 2019, 39, 755–773.
11. Bo, L.; Souza, E.; Matwin, S.; Sydow, M. Knowledge-based Clustering of Ship Trajectories Using Density-based Approach. In Proceedings of the IEEE International Conference on Big Data, Washington, DC, USA, 27–30 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 603–608.
12. Yan, W.; Rong, W.; Zhang, A.N.; Yang, D. Vessel Movement Analysis and Pattern Discovery Using Density-based Clustering Approach. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; IEEE: Piscataway, NJ, USA, 2017; pp. 3798–3806.
13. Zhang, G.; Zhang, J. Trajectory Clustering Based on Trajectory Structure and Longest Common Subsequence. DEStech Trans. Comput. Sci. Eng. 2018.
14. Zhao, L.; Shi, G. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. Ocean Eng. 2019, 172, 456–467.

15. Zhao, L.; Shi, G. A Novel Similarity Measure for Clustering Vessel Trajectories Based on Dynamic Time Warping. J. Navig. 2018, 72, 290–306.
16. Li, H.; Liu, J.; Yang, Z.; Liu, R.W.; Wu, K.; Wan, Y. Adaptively constrained dynamic time warping for time series classification and clustering. Inf. Sci. 2020, 534, 97–116.
17. Zhen, R.; Jin, Y.; Hu, Q.; Shao, Z.; Nikitakos, N. Maritime Anomaly Detection within Coastal Waters Based on Vessel Trajectory Clustering and Naïve Bayes Classifier. J. Navig. 2017, 70, 648–670.
18. Wang, L.; Chen, P.; Chen, L.; Mou, J. Ship AIS Trajectory Clustering: An HDBSCAN-Based Approach. J. Mar. Sci. Eng. 2021, 9, 566.
19. Wei, Z.; Meng, X.; Li, X.; Zhang, X.; Gao, Y. Vessel manoeuvring hot zone recognition and traffic analysis with AIS data. Ocean Eng. 2022, 266, 112858.
20. Efrat; Guibas; Har-Peled, S.; Mitchell; Murali. New Similarity Measures between Polylines with Applications to Morphing and Polygon Sweeping. Discret. Comput. Geom. 2002, 28, 535–569.
21. Tang, C.; Wang, H.; Wang, Z.; Zeng, X.; Yan, H.; Xiao, Y. An improved OPTICS clustering algorithm for discovering clusters with uneven densities. Intell. Data Anal. 2021, 25, 1453–1471.
22. Zhou, Y.; Daamen, W.; Vellinga, T.; Hoogendoorn, S.P. Ship classification based on ship behavior clustering from AIS data. Ocean Eng. 2019, 175, 176–187.
23. Liu, C.; Liu, J.; Zhou, X.; Zhao, Z.; Wan, C.; Liu, Z. AIS data-driven approach to estimate navigable capacity of busy waterways focusing on ships entering and leaving port. Ocean Eng. 2020, 218, 108215.
24. Zhang, W.; Goerlandt, F.; Montewka, J.; Kujala, P. A method for detecting possible near miss ship collisions from AIS data. Ocean Eng. 2015, 107, 60–69.
25. Gao, M.; Shi, G.Y. Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms. Ocean Eng. 2020, 205, 106919.
26. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise; AAAI Press: Washington, DC, USA, 1996.
27. Khan, M.M.R.; Siddique, M.A.B.; Arif, R.B.; Oishe, M.R. ADBSCAN: Adaptive Density-Based Spatial Clustering of Applications with Noise for Identifying Clusters with Varying Densities. In Proceedings of the 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, 13–15 September 2018; pp. 107–111.
28. Yang, Y.; Li, Z.; Wang, W.; Tao, D. An adaptive semi-supervised clustering approach via multiple density-based information. Neurocomputing 2017, 257, 193–205.

29. Liu, F.; Zhang, Z. Adaptive density trajectory cluster based on time and space distance. Phys. A Stat. Mech. Its Appl. 2017, 484, 41–56.

30. Liu, Y.; Ma, Z.; Yu, F. Adaptive density peak clustering based on K-nearest neighbors with aggregating strategy. Knowl. Based Syst. 2017, 133, 208–220.

31. Marques, J.C.; Orger, M.B. Clusterdv, a simple density-based clustering method that is robust, general and automatic. Bioinformatics 2018, 35, 2125–2132.

32. Yang, J.; Liu, Y.; Ma, L.; Ji, C. Maritime traffic flow clustering analysis by density based trajectory clustering with noise. Ocean Eng. 2022, 249, 111001.

33. Tang, C.; Chen, M.; Zhao, J.; Liu, T.; Liu, K.; Yan, H.; Xiao, Y. A novel ship trajectory clustering method for Finding Overall and Local Features of Ship Trajectories. Ocean Eng. 2021, 241, 110108.

34. Yan, Z.; Xiao, Y.; Cheng, L.; He, R.; Ruan, X.; Zhou, X.; Li, M.; Bin, R. Exploring AIS data for intelligent maritime routes extraction. Appl. Ocean Res. 2020, 101, 102271.