

A DBSCAN based Algorithm for Ship Spot Area Detection in AIS Trajectory Data

Bao Lei^{1,a}

¹ *Computer Science Department, Wuhan Donghu University, Wuhan, China*

Abstract. The big data acquired by AIS system contains abundant maritime traffic information. With the wide application of data mining in various fields in recent years, the mining on AIS data has draw attention of related researchers. Based on the ship AIS location data, this paper studies the relevant spot area detection algorithm. Firstly, the sample data are pre-processed from the original data, and the residence point of each ship is identified according to the ship speed and course change. Then a DBSCAN based clustering algorithm is used to cluster several latitude and longitude lattice, that is spot areas. The experiments on real AIS data sets shows that the algorithm is efficient and correct.

1 Introduction

Maritime transportation is the most important transportation nowadays. The analysis and research of maritime location data sets can find out the information of location characteristics, movement rules and behavior patterns hidden behind the large location data, so as to guide the efficient development of marine activities [1]. The Automatic Identification System (AIS) data is one of these data set, which is a tracking and self-reporting system used by maritime vessels to exchange information with other ships, AIS base stations, and satellites. The International Maritime Organization (IMO) adopted performance standards for AIS and made AIS installation compulsory on all large maritime platforms around the world at 2000, which enable AIS to provide a wealth of valuable surveillance data for vast decision support applications.

How to use the AIS data reliably, accurately and efficiently, and how to effectively discover knowledge from spatiotemporal large data are hot issues in the related field of research. One of the topic is the spot area detection, that is, mining spatial areas frequently visited by ships. The purpose of this paper is to propose a spot area detection approach from AIS data, which can provide information support for relevant maritime traffic control, infrastructure planning, and maritime security surveillance.

2 Related works

At present, the research of spot area mining method is mainly based on GPS data. The early extraction approach is to detect an indoor spot area by using the feature that the GPS signal disappears when the user enters a building

[2]. This method is effective for smaller indoor areas, but not for larger outdoor hot spots. Ashbrook and Starne use K-means clustering algorithm [3] to detect user's residence area in the mobile process. Although their method solves the size limitation of spot areas, it is necessary to know the number of final clusters and be sensitive to noise data before clustering. Kang uses incremental clustering algorithm to mine the residence area [4]. This method can identify any shape area from WI-FI data. The disadvantage is that its time clustering needs frequent GPS data sample data. Density-based clustering is an improvement of K-means clustering algorithm, which is an improvement of K-means clustering algorithm [5]. It strengthens the processing of noise and anomalous data. Compared with K-means clustering method, density-based clustering method is more suitable for clusters of arbitrary shape. [6] introduces a density-based clustering algorithm to mine all spot areas of different granularity on the basis of residence point set, but it also needs a large amount of computation and time-consuming algorithm. In this paper, we propose a method to identify the spot area by using time information. If a target stays at a certain place for a period of time, the point will be regarded as a residence point, and then all residence points will be clustered into a data set to get the residence area.

3 Ship trajectory stop points extraction

The stop points in ship trajectory are the points where a ship's detention time in a certain sea area exceeds a certain range. Compared with other AIS points, these data contain more important semantic information. The AIS data are usually vast and contains a lot of noise data. In order to solve these problems, we do not use the original

^a Corresponding author: blnj2000@163.com

AIS data in data mining, but use the data set which contains all the stop points extracted from each ship's AIS trajectory. These stop points contain more abundant semantic information and are more representative.

The common method to extract stop points is to determine whether the ship's stay time in the region exceeds the set threshold. This method needs to calculate the distance between any two points, involving huge computation, which complicates the problem and makes the size of parameters difficult to determine. The average speed of AIS data message is defined as the ratio of the actual displacement of ship to the time interval of message in the process of generating two copies of AIS equipment. Considering the influence of ocean currents, even the berthing ship will move. Therefore, it is indirectly considered that when the average speed of ship is lower than the set threshold, the range of motion of ship is small and it is regarded as staying state. Sometime is some heavy traffic area, not only the ship's velocity changes, but also their course change. These turning points are identified by setting the threshold of the rate of change in navigation.

The algorithm of stop point extraction is shown as below:

ALGORITHM 1 THE EXTRACTION ALGORITHM OF SHIP STOP POINTS

Algorithm1:

Stop_point_extraction($v_threshold, turnrate_threshold, trajectories$)

- 1: Initialize *stop_pts* set
 - 2: for each ship trajectory *s* in Trajecotries do
 - 3: If $s.velocity < v_threshold$ or $s.rateofcourse < turnrate_threshold$
 - 4: CandidateSet $\leftarrow s$;
 - 5: for each points *p* in CandidateSet do
 - 6: If the 4 neighbor points *pi* have $pi.velocity < v_threshold$
 or $pi.rateofcourse < turnrate_threshold$
 - 7: **stop_pt** $\leftarrow p$;
 - 8: return stop_pts;
- End

4 The clustering on ship stop points

Based on DBSCAN algorithm, we can do clustering on all the stop points of each ship into different clusters, which can be treated as spot areas. In traditional DBSCAN, the method of clustering usually uses circle to represent the residence area. Although this method is more accurate, it needs coordinate system transformation to calculate the distance. As AIS data provides latitude and longitude information, the distance in the clustering process is defined as the difference between latitude and longitude, which avoids the inconvenience caused by the circular clustering method.

Clustering algorithm:

Step1: a new cluster is created by selecting a core object point of an unlabeled category.

Step2: all objects starting from the core object are searched and labeled as clusters based on two parameters (a, MinPts) density reachable.

Step3: Repeat the process until all the objects are processed, that is, there are no core objects in the unlabeled cluster.

Examples:

Given the search range Eps_j, Eps_w of longitude and latitude and threshold $Minpts=4$, there are eight points(A,,B,C,D,E,F,G,H) to scan them in turn. Scanning A, there are five points in its neighborhood, greater than the threshold 4, so it is the core point, five points belong to the same category.(B,C,D,E) are the objects to search next, Take each point as core, check the number of points in its neighborhood, and when calculating E, new point F was added for further searching.

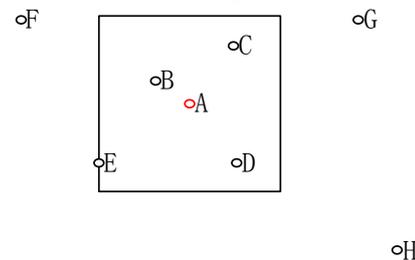


Figure 1a. Clustering Example Step1

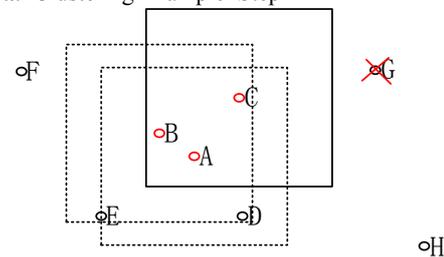


Figure 1b. Clustering Example Step2

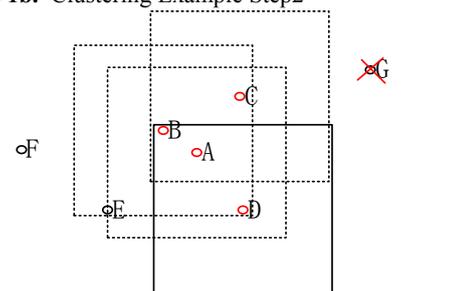


Figure 1c. Clustering Example Step3

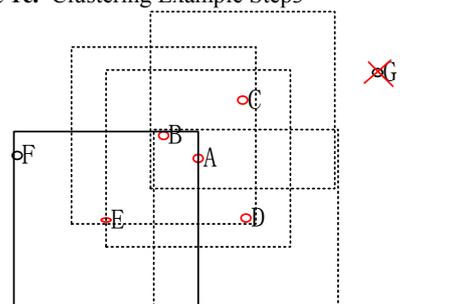


Figure 1d. Clustering Example Step4

ALGORITHM 2 THE DBSCAN ALGORITHM ON STOP POINTS

```

Algorithm2: DBSCAN(stop_pts,j_c,w_c,minpts)

1: stop_pts = Stop_point_extraction;
2: ClusterId=1;
3: for i=1 to n do
4: Point:= stop_pts.Get ( i );
5: If Point.ClusterId=0{
6: CandidateSeeds:=stop_pts.regionquety(Point);
7: If CandidateSeeds.size<minpts
8: Point.ClusterId= 0;
9: else
10: Point.ClusterId=ClusterId;
11: ALLCluster(stop_pts,j_c,w_c,
minpts,CandidateSeeds,ClusterId );
12: print Clusters;
End
    
```

The subfunction AllCluster() obtains all density reachable data objects of the core object point through region queries.

ALGORITHM 3 ALLCLUSTER

```

Algorithm 3: AllCluster(stop_pts,j_c,w_c,minpts)

1: While CandidateSeeds <> Empty do
2: CurrentP:=CandidateSeeds.first;
3: Result:=stop_pts.regionquety(CurrentP,a);
4: If Result.Size>=minpts
5: For each point p in Result do
6: If p.Clid= UNCLASSIFIED
7: CandidateSeeds.append ( p );
8: If p.Clid= UNCLASSIFIED or NOISE
9: p.ClusterId=ClusterId;
10: CandidateSeeds.delete(currentP);
End
    
```

5 Experiments

5.1 Environment and data preparation

The experimental data were selected from real AIS historical data. The local AIS data near NanSha on April 2012 are selected, with a data volume of about 80 GB, with the latitude and longitude within 120E-125E,25N-29N.

AIS data generally includes static information, dynamic information, navigation-related information. The data content is relatively rich, shown in Table 1. Most of the analysis is based on part of AIS data content. For example, only longitude and latitude information is needed for the display of ship spatial distribution, and only longitude and latitude and time stamp information is needed for the fitting of ship trajectory. Therefore, before AIS data analysis, it is necessary to extract and preprocess source data for different data content analysis needs.

Table 1 The content of AIS data

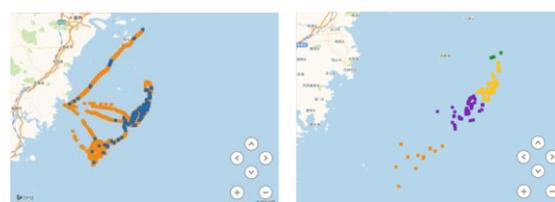
Type	Contents
Statis data	IMO id,MMSI,class, name
Dynamic data	System_timestamp,Nav_status,Longitude, Latitude,COG,SOG,True_head,ROT
Auxiliary data	Drainage, port of destination, type of cargo and route plan
Security data	Navigation aid reports, safety-related radio messages

Considering that the stop points extraction needs a lot of original data, less than 700 ships with AIS information are filtered out from the sample data, and the remaining ships are selected as the research object.

5.2 Results

5.2.1 Clustering on single trajectory

Using the algorithm proposed, the spot area of each ship was calculated. Taking a ship coded 998511033 as an example, the threshold of recognition speed is 0.5, the threshold of course change rate is 0.5, the threshold of change point is 0.5, the yellow is the ship track, and the blue is the stop point/course change point of ship activity. The total number of nine ship points provided by four AIS messages before and after the stop point is calculated, which is regarded as the stop point if it is larger than the support count 6. After filtering, it is shown in Figure 2. The clustering algorithm divides the residence points into different clusters. Different colors in Figure 2 represent different clusters. Some residence points are abandoned by the algorithm because of noise. The clustering results are several clusters, i.e. spot area. Each cluster will be regarded as a whole in spatio-temporal mining, while irregular regions will inevitably increase the difficulty of mining. If the maximum and minimum longitude and latitude of each residence area are taken as grid boundary, it is more feasible to design the algorithms for further analysis. From the algorithm result, the ship 998511033 have 4 spot area shown as rectangles.



a) Ship routes and stop points

b) The stop points clusters

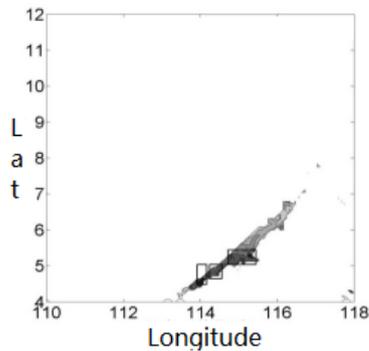


C) The spot areas

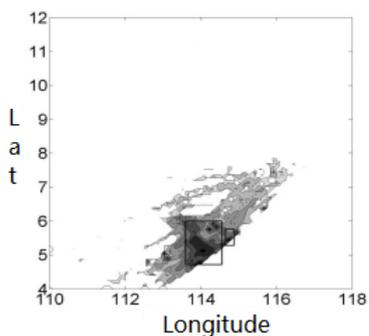
Figure 2. The Clustering Result of Ship 998511033

5.2.2 Clustering on multiple trajectories

The figure 3 shows the clustering results on different trajectory data sets. The figure 3a is the result on the data near coastal area, and the figure 3b is the data on high sea, shown in gray-scaled figures.



a)Trajectories near coastal area



b)Trajectories near high sea area

Figure 3. Clustering results on multiple trajectories

6 Conclusion

Spot area detection is an important topic in AIS data mining. The purpose is to accurately mine the geographical areas frequently visited by many ships, and to provide strong information support for relevant service objects. In this paper, we propose a stop points extraction model to select typical locations from large AIS data, and

use DBSCAN to cluster the residence areas of each ship during its voyage. Our algorithm can work on real AIS data, and the experiment prove its effectiveness.

This work is just a first step, many challenges lies ahead. The algorithm proposed in this paper only takes the spatial attributes into computation, cant use the information on ships' customs, habits and even the characteristic of a certain region. Future research can be made on integrating the trajectory semantics of vessels with stop point detection. Moreover, future work should be done on adaptation on MapReduce or Spark architecture on the algorithms to incorporate huge data sets. Due to the limitation of computing power of the platform, the DBSCAN algorithm with less time overhead is used in this paper. The clustering should be used for multi-granularity clustering to achieve better results.

References

1. Xiao F, Ligteringen H, Van Gulijk C, et al. AIS data analysis for realistic ship traffic simulations model[J]. Shanghai Jiao Tong University, 2012, Proceedings of IWNTM'2012:44-49.
2. Y. Liu, C. W. Sun. An analysis method of multi-granularity hotspots based on the clustering of staypoint. Microcomputer information, 2012, 09:295-297.
3. Cao H, Mamoulis N, Cheung D W. Mining frequent soatio-temporal sequential patterns [C]//Proc of the 15th IEEE Int conf on Data Mining . Piscataway, NJ:2015:82-89.
4. Kang. Ju-Young, Yong. Hwan-Seung. Mining Spatio-Temporal Patterns in Trajectory Data[J]Journal of information processing systems, 2010, 6(4):521.
5. Lee A J T, Chen Y-A, Ip W-C. Mining Irequent trajectory patterns in spatial-temporal databases [J]. Information Sciences, 2015, 179(13): 2218-2231.
6. Giannotti F, Nanni M, Pinelli F, et al. Trajectory pattern mining[C]//Proc of the 13th AC'M SIUKDD Int C'onf on Knowledge Discovery and Data Mining. New York: ACM, 2017:330-339.