A Spatiotemporal ROI Mining Algorithm on AIS Trajectory Data

Bao Lei

1Computer Science Department Wuhan Donghu University Wuhan, China
Email: blnj2000@163.com

Abstract. As one of the most important transportation in international logistics, marine transportation generate vast amount of data. At present, we have been able to collect ship navigation data from many approaches such as AIS data. And extracting accurate and valuable potential information from these data for further application is a new and important topic in related areas. In this paper, AIS data is selected as the experimental data and a Region of Interest mining algorithm based on historical AIS trajectory data is proposed. Unlike the clustering based approaches, this algorithm locates the ROI from location-time data in AIS history trajectories, which can detect those ROI where ships didn't stay long, but visited frequently. Experiments and analysis on real AIS data shows that the proposed approach is effective and correct.

1. Introduction
Maritime information analysis is an important means for supervision and traffic control. With the vigorous development of maritime industry, the Automatic Identification System’s emergence and widespread use in ship identification system make it possible to obtain real-time information of ship navigation. The analysis based on AIS data can mine accurate and valuable potential information from a large number of disorderly data. Introducing methods of maritime information analysis based on AIS data with higher reliability and more universality has become a hot topic in related area of research[1].

In this paper, the algorithms of mining Region of Interested are studied. Region of Interest, that is, the sea area that a ship is more interested in, refers to the sea area that a ship frequently arrives in a longer period of time. For example, fishing vessels are interested in the waters rich in fishery resources, and cargo ships are interested in the waters around major ports. By mining the ROI of ships, the characteristics of ships can be well described, which provides support for waterway management and maritime traffic planning and supervision.

2. Related works
In related spatiotemporal clustering, Eric Hsueh-Chan Lu and Vincent S. Tseng proposed a transaction clustering method, CO-Smart-CAST [2], to form a user clustering model for mobile transactions. Riveiro [3] attempted to extract abnormal points of ship trajectories in clustering algorithm (SOM) and Gaussian hybrid model (GMM) using self-organizing mapping neural network. The time stamp, ship position (longitude and latitude), direction, speed and ship type information in AIS information are taken as input parameters of the model, and three different ship types (grocery ship, fishing vessel and cargo ship) are selected for experiments. This method has been successfully applied by Maria Riveiro and others in Sweden. On the research of movements pattern mining, Giannotti proposed the trajectory pattern of moving objects [4][5]. Extracts ship motion patterns from historical AIS data of ports and important waterways, and constructs an algorithm for abnormal detection and prediction of ship motion.
patterns. The basic idea is to extract the four-dimensional vector space (longitude, latitude, direction and size of velocity) of ship motion pattern from a large number of historical AIS data, and construct the boundary of ship trajectory. Xue Ran [6] detects user's region of interest, uses frequent sequence pattern technology to mine user's frequent trajectory patterns, and constructs user's movements attributes. Movements behavior prediction is also one of the important applications of behavior pattern research. Jeung proposed a prediction method called HPM (Hybrid Prediction Model) to estimate where an object will be at some point in the future. [7] proposed two different methods for mobile user location prediction: real-time prediction of mobile user location using local linear prediction method in the case of less location information of mobile users; non-real-time prediction of mobile user location based on data mining in the case of mobile users having a large number of location information.

This paper introduces an algorithm for mining the ROI in ship AIS trajectories. The ROI is obtained by merging several AIS locations, by means of an algorithm can adapt to AIS data and other similar historical location-time data.

3. The ROI in ship AIS trajectories
The ROI is the area where ships stay for a long time or visit frequently for a long period of time. Different ships have different degrees of interest in different sea areas, just as fishing boats often haunt areas rich in fishery resources, and cargo ships travel between major cargo ports. Ships should not be included if they pass only occasionally or stay briefly. This requires that the algorithm be able to exclude the sea areas that the ship normally sails through and stays for various reasons, and only excavate the sea areas that the ship is really interested in. Our algorithm is based on historical data and mine the ROI from trajectories data for a period of time. Therefore, the algorithm does not require ships to stay in a certain area continuously, but pays more attention to the number of ships appearing in the whole time range.

The distance threshold and sequence length threshold are set in this algorithm, where the distance threshold is used to specify the size of the ROI, and the sequence length threshold is used to ensure that each ROI has sufficient data support.

3.1. Algorithm
Step 1. Starting from the starting point S, the distance between the starting point S and the next point is calculated. If the distance threshold D is limited, the point is stored in the sequence and deleted from the original data matrix to obtain a series of position points.

As shown in the figure 1, S is the starting point, D is the distance threshold, and black represents the point stored in the sequence. Any point that meets the distance threshold limit can be accessed into a sequence without requiring a continuous point.

![Figure 1. ROI mining](image)

Step 2. The sequence length is calculated. If the sequence length threshold is reached, the sequence is merged into an ROI.

Step 3. The starting point is changed to the next one in sequence and the calculation continue until the number of remaining points can no longer meet the requirements of sequence length.

3.2. Threshold control
There are two thresholds involved in the algorithm, namely distance threshold and sequence length.
threshold. The distance threshold can be set to 8000 meters according to experience. In order not to miss out every area of interest as much as possible without false reporting of the area of interest, it is necessary to reasonably determine the sequence length threshold through appropriate estimation. Table 1 shows the time needed for transverse 8000m in different velocity.

Table 1 Time needed for transverse the distance threshold

| Ship velocity (knot) | time (second) |
|----------------------|---------------|
| 5                    | 3112.0        |
| 10                   | 1556.8        |
| 15                   | 1038.4        |
| 20                   | 777.6         |

It can be seen from the table that even if the ship passes through the sea area at a slower average speed of 5 knots, the time can be controlled within about 3000 seconds. According to the ship report frequency table, the highest frequency of AIS data report is 2 seconds per time. The maximum number of AIS records reported for the normal passage of a ship at a distance of 8000 meters shall not exceed 1500. If more than 1500 location data have been detained in the sea area, or have visited the sea area many times, they can be regarded as ROI, namely, the sequence length threshold is set to 1500.

4. Experiments

4.1. Environment and data preparation

The experimental data were selected from real AIS historical data. AIS data from three ship trajectories was selected, they distributed near NanTong on Yangtz River and Shandong Yingkou, on April 2012.

4.2. Results

From the algorithm above, the ROI of three ships are listed in Table 2.

Table 2 ROI Results

| MMSI       | Center of ROI                  | Location points in ROI |
|------------|--------------------------------|------------------------|
| 999000011  | 39.840103° N, 124.11853° E     | 2759                   |
|            | 40.309265° N, 121.97065° E     | 4918                   |
|            | 40.290352° N, 122.061455° E    | 90428                  |
|            | 40.211079° N, 121.437706° E    | 1586                   |
|            | 40.247894° N, 121.678986° E    | 1637                   |
|            | 40.181862° N, 121.552856° E    | 13513                  |
|            | 40.144447° N, 121.633492° E    | 4236                   |
|            | 40.212868° N, 121.942932° E    | 2634                   |
| 977777620  | 32.018936° N, 120.819901° E    | 117669                 |
| 977777903  | 32.020432° N, 120.818748° E    | 103607                 |

4.3. Analysis and Optimization

From the results, the results of some low speed ships trajectories have a relatively small range of activities and concentrated location points, so the algorithm attributes almost all points to the same ROI. Different distance thresholds should be determined according to the different range of movement range. If the range of ship activity is large, the distance threshold can be properly raised to avoid the occurrence of too many areas of interest excavated and confusing the primary and secondary situation; if
the range of ship activity is small, the distance threshold can be reduced appropriately to excavate the areas of real interest of ships in a limited area.

In view of the small range of activities of ship 9777903, it is necessary to appropriately reduce the distance threshold for mining the areas of interest of these two ships. Here they are all changed to 1000 meters. The results of mining new areas of interest are obtained as shown in the table 3.

| Longitude     | Latitude     | Location points |
|---------------|--------------|-----------------|
| 120.81874800000 | 32.0204320000000 | 37376           |
| 120.80080400000 | 32.0170210000000 | 38182           |
| 120.79079400000 | 32.0204700000000 | 23896           |

As shown in table 3, the new results are more detailed and can better reflect the ship's behavior pattern.

5. Conclusion
With the rapid development of information technology and the continuous improvement of data processing ability, it is the general trend to make assistant decision by analyzing the information obtained from a large amount of maritime location data. The algorithm proposed in this paper are mainly based on historical location-time data, which can be well adapted to AIS data in the traffic control of ships, waterways and important ports. The ROI of different ships can be used to achieve more reasonable and orderly management.

Future work should be done on adaptation on big data set, including introduce MapReduce or Spark architecture on the algorithms to incorporate huge data sets.

Reference
[1] Eric Hsueh-Chan Lu, Vincent S. Tseng. Mining Cluster-based Mobile Sequential Patterns in Location-Based Service Environments. Tenth International Conference on Mobile Data Management: System, Services and Middleware, 2012.
[2] Riveiro, M., Johansson, F., Falkman, G. Supporting Maritime Situation Awareness Using Self Organizing Maps and Gaussian Mixture Models[C]//The Proceedings of SCAI 2016 of the Tenth Scandinavian Conference on Artificial Intelligence. Frontiers in Artificial Intelligence and Applications. USA: IOS Press, June 2016.
[3] V. S. Tseng and C. F. Tsui. Mining Multi-Level and Location-Aware Associated Service Patterns in Mobile Environments. IEEE Trans. On Systems, Man and Cybernetics: Part B, 2014, 34(6).
[4] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Mining Mobility Behavior from Trajectory Data. International Conference on Computational Science and Engineering. IEEE Press, 2015. 948-951.
[5] Xue Ran. Construction and comparison of user mobility attributes based on geographic location services[D]. Shandong University, 2013.
[6] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou. A Hybrid Prediction Model for Moving Objects. IEEE 24th International Conference on Data Engineering, 2016.
[7] Xue Dan. Research on Mining and Location Prediction of User Mobile Patterns[D]. Donghua University, 2015.