Mining Ship Spatial Trajectory Patterns from AIS Database for Maritime Surveillance

ZHU Feixiang

Navigation College

Dalian Maritime Unviersity

Dalian, 116026, China

dlmuzhufeixiang@163.com

Abstract—With the development of AIS (Automatical Identification System), more and more ships are equipped with AIS. The messages transmitted by AIS have thus become an abundant and inexpensive source of information for maritime surveillance. In view of this, this paper applied Electronic Chart System (ECS), database management, Data Warehouse (DW) and Data Mining (DM) technologies to facilitate the discovery of hidden and valuable information in a huge amount of AIS data. With enlighten of business intelligence, Association Rules Algorithm is applied to deal with discovering the ship trajectory pattern. The results show that the knowledge mined by this paper is interesting to maritime safety administration. The application domain can range from current Vessel Traffic Service (VTS), maritime surveillance, marine traffic engineering, ship's behavior and traffic flow study to homeland security.

Keywords – Marine Traffic; AIS; Data Warehouse; Data Mining; Association Rule

I   INTRODUCTION

Maritime surveillance of ports and coastlines is of paramount importance for China. Considering that more than 90% of imports and exports to China are carried by sea, potential maritime traffic accident in ports or waterways could cause serious disruptions with significant economic implications. Currently used methods to maritime surveillance include visual observation, audio exchange, radar, infrared and video cameras installed on fixed ground location. A number of self-reporting maritime systems have been introduced, mainly for the purposed of safety in navigation and collision avoidance. The most important self-report maritime system is the AIS, which has been made compulsory by the International Maritime Organization's (IMO) International Convention for the Safety of Life at Sea (SOLAS) for the most commercial ships(cargo, passenger, tankers, tugs, pilot vessel, etc.). Recently there has been an increased interest in the analysis of AIS data for maritime surveillance [1,2]. The trajectory data available from the AIS transmissions are spatial series, describing the movement of ships entering a very large sea port. Mining the trajectory pattern is useful to improve the existing traffic, as well as to plan future traffic developments, and to minimize the risk of collision. This paper is devoted to analysis of AIS data for the purpose of mining ship spatial trajectory patterns in ship motion.

A.   AIS

The IMO's SOLAS requires AIS to be fitted aboard international voyaging ships with the gross tonnage (GT) of 300 or more tons, and all passenger ships regardless of size [3,4]. The AIS is an automated tracking system used on ships and by VTS for identifying and locating vessels by electronically exchanging data with other nearby ships and AIS base stations in VTS [5]. AIS messages are automatically broadcasted with a reporting frequency directly proportional to the speed of the vessel. There are various types of AIS messages which are broadly classified as static information (name, type, size, etc. of a vessel) and dynamic information (position in geodetic coordinates, speed, course, heading, destination, estimated time of arrival, etc.) [4]. AIS enables the VTS not only to offer commonly known functions such as identification, tracking and monitoring of vessels, but also to provide much real-time information that is useful for marine traffic investigation, statistical analysis and theoretical research. With the development of AIS, more and more ships are equipped with AIS. The messages transmitted by AIS have thus become an abundant and inexpensive source of information for maritime surveillance. The ship historical trajectory data available from the AIS dynamic information is very huge. This paper is devoted to using data mining and data warehouse technologies for the purpose of mining ship spatial trajectory patterns.

B.   ECS

ECS [6] is the charting software system for presentation of electronic chart data on a computer screen, which intended to improve the safety of navigation and help mariner in his day-to-day work. ECDIS (Electronic Chart Display and Information System) which should comply with IMO Resolution, International Hydrographic Organization Specification and International Electrotechnical Commission International Standard is the internationally adopted system for presentation of electronic chart data on a computer screen, and intended to replace paper charts as the legal base for safe navigation. There is no strict standard for ECS in contrary to ECDIS. ECS can show marine spatial information (including land area, sea area, anchorage area, fairway, traffic lane,

lighthouse, etc.) which is very useful to the purpose of ship navigation safety. In this paper, geographic grids and mining results are shown based on ECS.

C. Data Warehouse and Data Mining

A DW is a central repository for all or significant parts of the data that an enterprise's various business systems collect [7]. Typically, a DW is housed on an enterprise mainframe server. Data warehouse emphasizes the capture of data from diverse sources for useful analysis and access, but does not generally start from the point-of-view of the end user or knowledge worker who may need access to specialized, sometimes local databases.

Generally, DM (sometimes called knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining as a process is depicted in figure 1 and consists of the following steps [8,9,10]:

Step One: Data cleaning (to remove noise and inconsistent data)

Step Two: Data integration (where mulltiple data sources may be combined)

Step Three: Data Selection (where data relvent to the analysis task retrieved from the database)

Step Four: Data transformation (where data are transformed or consolidataed into forms appropriate for mining by performing summary or aggregation operation, for instance)

Step Five: Data mining (an essential process where intelligent methods are applied in order to extract data patterns)

Step Six: Pattern evaluation (to identify the truly interesting patterns representing knowledge basde on some interestingness measures)

Step Seven: Knowledge presentation (where visualization and knowledge representation techniques are used to presents the mined knowledge to the user)
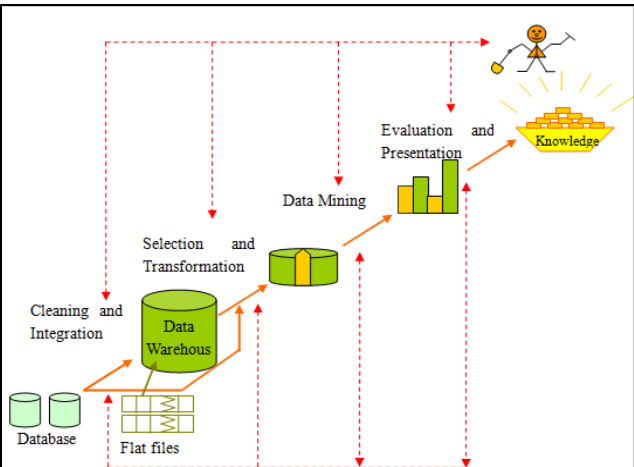

Fig 1.   Steps in the process of data mining

The paper is organised as follows. Section I introduces the pre-knowledge of AIS, ECS, DW and DM. Section II describes the pre-processing of AIS data before mining. Section III describes the analysis of AIS data and uses the association algorithm to obtain the marine traffic knowledge. Section IV describes the experimental result.

II   AIS DATA PRE-PROCESSING

A. AIS Message Decoding

The AIS allows ships to be tracked in realtime based on information transmitted by each ship. They are equipped with either a Class A or Class B AIS transponder which includes a GPS for accurate position and speed reporting. AIS receivers decode the transmitted information and output the data as AIVDM messages. The structure of the AIVDM message is described in International Electrical Commission (IEC) 61993-2 and it is a variation of the NMEA 0183 sentence format that includes the raw data encoded in a 6-bit format. A Message example looks like this: !AIVDM,1, 1,,B,19NS7Sp02wo?HETKA2K6mUM20<L=,0*27.

The meaning of each data element in the AIS messages is covered by the International Telecommunication Union (ITU)M.1371 and IEC 62287 documents. Using Visual Studio C++ 6.0, this paper realized the decoder for AIS messages in NMEA (National Marine Electronics Association) format. For more detailed information about decoder, please see the literature [4].

B. Assigning Geographic Grids

In order to suppress the computational complexity, this paper proposed that the surveillance area should be discretized into a uniformly sized grid to facilitate the integration of AIS data and spatial analysis. Results are overlaid on a chart in the form of grids. This paper uses the dimensions $0.02° \times 0.02°$ as the size of a grid to divide up the area (as shown in figure 2) and give each grid a unique number based on ECS.
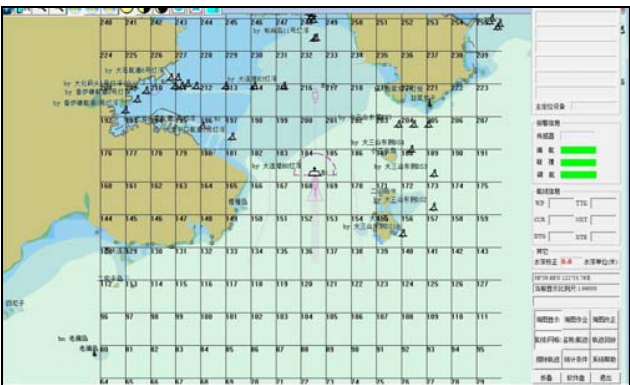

Fig 2.   Geographic grid based on ECS

So, every voyage will cross the grids and every voyage is transformed into a series of grids. The ship trajectory such as $\{ (lat_1; lon_1) , (lat_2; lon_2) , (lat_3; lon_3) \ldots (lat_i; lon_i) \ldots \}$ is

transformed into a series of grids such as $\{N_1, N_2, N_3 \ldots N_i, \ldots \}$. It is useful to find the trajectory pattern. The example of corresponding relationship is shown as in table 1.

TABLE 1   POINTS OF VOYAGE AND CORRESPONDING GRID ID

Position Points of Voyage (latitude;longitude)	Series of Grid Grid ID
(38.77422; 121.97697)	261
(38.767603; 121.981407)	230
(38.757475; 121.987997)	198
…	…

III   MINING SHIP SPATIAL TRAJECTORY PATTERNS

A. Association Rule

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases, otherwise termed Marker Basket Analysis. The following tab 2 is an example of transaction data.

TABLE 2   TRANSACTION DATA RECORDS

TID Transaction	Items				
T1	X1,	X2,	X3,	X4,	X5
T2		X2,	X3,	X4	
T3			X3,	X4,	X5
T4	X1,	X2			
T5	X1,		X3,		X5
T6	X1,		X3,	X4,	X5

Association rules, first introduced in 1993 [11], are used to identify relationships among a set of items in a database. A formal statement of the association rule problem is [12]:

Definition 1: Let $I = \{I_1, I_2, \ldots, I_m\}$ be a set of m distinct attributes, also called literals. Let D be a transaction database, where each record (tuple) T has a unique identifier, and contains a set of items such as $T \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \subset I$, are sets of items called itemsets, and $X \cap Y = \varnothing$. Here, X is called antecedent, and Y consequent.

Two important measures for association rules, support (s) and confidence (α), can be defined.

Definition 2: The support (s) of an association rule is the ratio (in percent) of the records that contain to the total number of records in the database.

Definition 3: For a given number of records, confidence (α) is the ratio (in percent) of the number of records that contain to the number of records that contain X.

Apriori uses a "bottom up" approach, where frequent itemsets are extended one item at a time (a step known as candidate itemsets generation), and groups of candidate itemsets are tested against the data. We will introduce the apriori algorithm.

Key Concepts:

Frequent Itemsets: The sets of item which has minimum support (denoted by L_i for i^{th} -itemset).

Join Step: To find L_K, a set of candidate k-itemsets (denoted by C_K) is generated by joining L_{K-1} with itself.

Purne Step: Any (K-1)-itemset that is not frequent cannot be subset of a frequent K-itemset.

Pseudo code:

$L_1 = \{$frequent items$\}$;

for (k=1; $L_K \neq \varnothing$; k++) **do begin**

C_{K+1} =candidates generated from L_K;

for each transaction t in database **do**

Increment the count of all candidates in C_{K+1} that are contained in t

L_{K+1} =candidates in C_{K+1} with min_support

end

return $\cup_K L_K$;

B. Discovering Ship Spatial Trajectory Patterns based on Association Rule Algorithm

If we consider the present research as analogous to the transaction data, the corresponding associations would be shown as in tab 3.

TABLE 3   TRANSACTION DATA AND CORRESPONDING AIS DATA

Transaction Data	AIS Data
Product	Grid
Product ID	Grid ID
Customer	vessel
Transaction	voyage
Transaction data	Position points of a voyage (or grids passed by a voyage)

All the grids in the area of study can be viewed as a collection of all retail products. The grid ID corresponds to the product ID. Every transaction (i.e trajectory of a voyage) is different. Through data mining, we can discover voyage (trajectory) behaviour and discover the popular (high traffic density) frequent grids. By applying this method to marine traffic, this paper can discover ship navigation knowledge (ie frequent spatial trajectory patterns) and traffic distribution, to be used as reference for marine traffic management.

IV EXPERIMENTAL RESULT

We adopt Microsoft SQL Server 2005 as DW and Visual Studio C++6.0 as the development tool for algorithm program library. In this paper, the prototype system adopts the C/S mode and web mode because the web mode has excellent performance in platform interchanges. Experimental AIS data are collected from 2010-9-1 to 2010-9-30 by AIS base station on land which covers the Dalian harbour sea area.

The popular (high traffic density) frequent grids are shown as in tab 4. For example, the support of frequent item (gird ID={178}) is equal to 227.

TABLE 4   FREQUENT GRIDS

Support	Frequent item
227	Grids ID={178}
212	Grids ID={179}
…	…
139	Grids ID={72, 56}
…	…
123	Grids ID={40, 72, 56}
…	…

From above tab 4, we know that grids (ID=178,179) and grids (ID=40, 72, 56) are high traffic density as shown in figure 3. The higher traffic density area is the danger area.The administrative departments should pay more attention to this area.
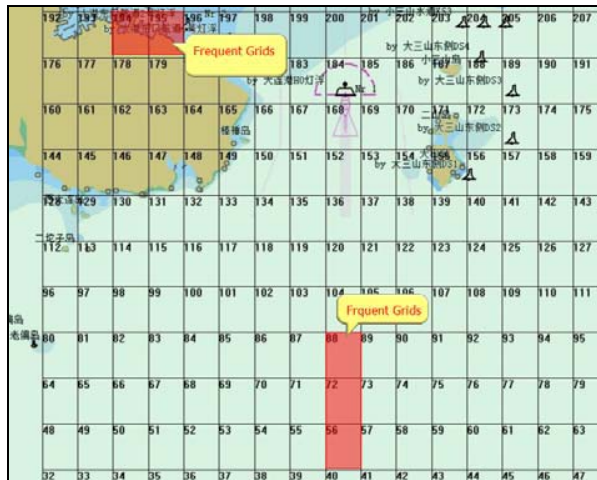

Fig 3.   Frequent grid based on ECS

The rules and confidence values obtained from association rule mining are shown as in tab 5.

TABLE 5   ASSOCIATION RULES AND CORRESPONDING CONFIDENCE

Confidence	Rule
0.884	135->119
0.868	56, 72->40
0.820	213, 214->229
…	…

Take the rule 56, 72->40 as an example: the rule means that after passing through grid (ID=56) and grid (ID=72), grid (ID=40) will be passed through.

V CONCLUSION

With the development of AIS, more and more ships are equipped with AIS. The messages transmitted by AIS have thus become an abundant and inexpensive source of information for maritime surveillance. Therefore, there is a need to perform effective and systematic analysis to uncover the hidden valuable information. In this paper, by combining techniques such as ECS, database management, data warehouse and spatial data mining, processing and analysis operation on AIS data are performed with greater efficiency. Interesting insights are obtained as well as an understanding of Dalian sea area such as high traffic density area and spatial pattern of traffic flow. The application domain can range from current VTS, marine traffic engineering, ship behavior and traffic flow study to homeland security. In the future, we will extend the observation periodic of AIS data and add other data mining models to facilitate the maritime surveillance.

REFERENCES

[1] C. Carthel, S. Coraluppi, and P. Grignan, "Multisensor tracking and fusion for maritime surveillance," in Proc. Int. Conf. Information Fusion, Qu´ebec, Canada, July 2007.

[2] B. J. Rhodes, N. A. Bomberger, and M. Zandipour, "Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness," in Proc. Int. Conf. Information Fusion, Qu´ebec, Canada, July 2007.

[3] B.Ristic, B.La Scala, M.Morelande, N.Gordon, "Statistical analysis of motion patterns in AIS Data: Anomaly detection and motion prediction," Proceeding of 2008 11th International Conference on information fusion, pp. 1-7, June 2008.

[4] ITU-R Recommendation M.1371-1, "Technical characteristics of a universal shipborne automatic identification system using time-division multiple access in the maritime mobile band", 2001.

[5] Shwu-Jing Chang, "Development and analysis of AIS applications as an efficient tool for Vessel Traffic Service", Proceedings of IEEE Oceans'04/ TECHNO-Ocean'04, pp. 2249-2253, 2004

[6] www.dkart.ru/index3.htm.

[7] William H. Inmon. Building the data warehouse [M]. Beijing: China Machine Press, 1998.

[8] ZHU Ming. Data Mining [M] . HeFei: University of science and technology of china press, 2002.

[9] HAN J, KAMBER M. Data mining: concepts and techniques [M]. San Francisco: Morgan Kaufmann Publishers, 2001.

[10] ZHU Feixiang, ZHANG Yingjun. "Research on Marine Traffic Data Mining System Based on AIS", Proceeding of 2010 International Conference on Future information Technology, pp. 350-353,December 2010.

[11] R.Agrawal, T.Imielinski, A.Swami, "Mining Association Rules between Sets of Items in Large Databases", SIGMOD'93, pp. 207-216, 1993.

[12] R.Agrawal,R.Srikant,Fast algorithms for mining association[C].In:Proc of the 20th Int'l Conf on Very Large Database, pp. 487-499,1994.