International Journal of Engineering & Technology, 7 (2.12) (2018) 167-170

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Mining of missing ship trajectory pattern in automatic identification system

Kwang Il Kim [1], Keon Myung Lee [1*], Jang Young Ahn [2]

[1] *Dept. of Computer Science, Chungbuk National University, Cheongju, 28644, Republic of Korea*
[2] *College of Ocean Science, JejuNational University, Jeju, 63243, Republic of Korea*
**Corresponding author E-mail: kmlee@cbnu.ac.kr*

Abstract

Background/Objectives: Ship trajectories in Vessel Traffic Service (VTS) system are generated by integrating the Automatic Identification System (AIS) or Radar system. However, the AIS system has missing data section caused by AIS device problems, radio jamming, and so on. These data have been confusing ship navigators and VTS operators.
Methods/Statistical analysis: In order to extract missing AIS data, time intervals of sequent points from each ship trajectory are calculated. The section with missing AIS data is above a threshold time limit defined by characteristics. Using k-means algorithm, missing AIS data were clustered into several clusters stored by ship's ID and sailing direction. Using association rule mining analysis, meaningful association pattern were calculated by missing AIS dataset.
Findings: As a result of the association rule mining, we found several missing AIS situation patterns. In case of the west route, the probability of missing AIS situation is high when they enter the east and passenger routes. Also, the probability of missing AIS situation of passing the passenger route is high when that ship enter the LNG, east and west routes.
Improvements/Applications: These results can be used to predict the probability of missing AIS data in VTS system.

Keywords: *Automatic Identification System; Missing AIS Data; Association Mining; K-Means; Data Mining*

1. Introduction

In Vessel Traffic Service (VTS) systems, a ship's position and vector, such as course and speed are obtained by using the Automatic Identification System (AIS) or Radar system. The AIS system is installed in almost all navigating ships; it automatically broadcasts the ship-movement information through GPS and receives information regarding other ships, updating the information at a rate of 3–10 s[1]. The radar, which is another device installed on shore-based station, uses an electromagnetic sensor for object detection via reflected radio waves to determine the range, altitude, direction, or speed of objects[2]. Both these systems are important to determine ship's movement using the VTS operating system.
In meantime, many AIS studies on maritime area are performed. Abbas et al.[3] separated AIS characteristics into VTS-based, data-mining, proactive-AIS and analyzed AIS data reliability and human error. The static information are incorrect with real ship status especially in case of ship type and navigation status. In order to recover AIS trajectory by missing AIS situation, Kim at al.[4] suggest trajectory interpolation methods using neural network module. Until now, however, the study of missing AIS pattern does not attempt. Missing AIS data—which might confuse a mariner—occurs when data is not updated at the update rate required for AIS. This can be caused by AIS device problems, radio jamming, and so on. Therefore, to determine the data pattern for the missing AIS data, the system needs to know the distribution of missing AIS data patterns. In this study, we extract missing AIS data using k-means clustering and find missing pattern rules using the association rule mining method.

The remainder of this paper is organized as follows. Section 2 describes target integration with AIS in VTS. Section 3 explains extraction of missing AIS dataset. Extracted missing AIS dataset are analyzed with sequential pattern mining. Finally, the conclusions are presented in section 5.

2. Target integration with ais in vts

Ships with more than 300 gross tonnages are equipped with an AIS device to broadcast the ship movement information and receive the movement information of other ships with AIS as well[5]. AIS information are consist of dynamic and static information. The static information include ship's name, type, width, length and so on. The dynamic information consist of receiving time, latitude, longitude, speed, course and heading direction and so on. These information are updated by ship movement status as follows:

- Less than 14 knots: 10 s
- 0−14 knots and changing course: 3.3 s
- 14−23 knots: 6 s
- 14−23 knots and changing course: 2 s
- More than 23 knots: 2 s

These AIS data are also collected by a VTS center, which is a shore-based station. The AIS data provides a source of information for tracking ship movement in the VTS's area. Further, it can identify a ship's information using dynamic as well as static information instead of inquiring information from each individual ship. Moreover, AIS contributes to VTS's allied services and port operations[6,7].
VTS target management systems integrate both radar and AIS targets. These systems present target vectors and positions with high accuracy by implementing target fusion algorithms[8]. However,

though AIS data has update rate 3–10 s based on the with ship movements, sometimes the AIS update might be delayed for a considerable time—even longer than 5 min. This problem is primarily caused by time slot capacity, onboard device problem, and radio interference by object such as island or ship structure. Figure 1 shows the comparison of AIS and radar trajectories, which radio interfered by island.

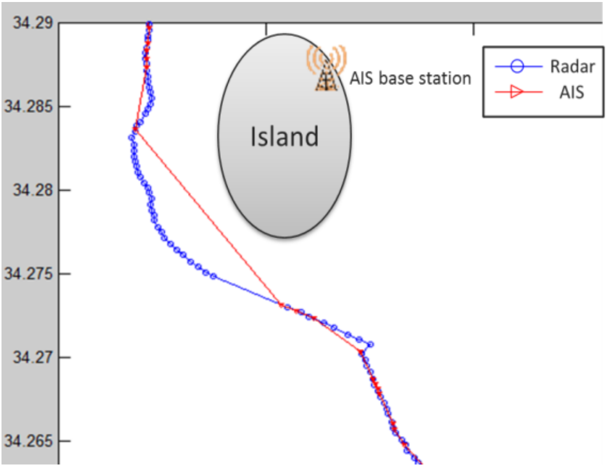

Fig.1:.Route Gate Lines and Integrated Ship Trajectory.

3. Extraction of missing ais dataset

In order to extract missing or abnormal data from the AIS dataset, target data needs to be sorted according to ship ID and sailing directions. Then, time intervals of sequent points from each ship trajectory are calculated. The section with missing AIS data is determined using interval times above a threshold time limit that is defined by characteristics of the water in the area.

The saved data were clustered into several clusters using the k-means clustering algorithm[9]. K-means algorithm is a partitioning clustering method that divides a dataset into similar data by calculating cluster centroids. To use k-means algorithms, the number of clusters k needs to be set. In this study, we experimented the cluster number by dividing the number of total datasets into unit cluster number 50. The target data used in this study is the AIS data of the Incheon water area for duration of 1 month. The saved data were clustered into 50 clusters using the k-means clustering algorithm. These clusters include the location label for sequence pattern analysis. Figure 2 and 3 shows non-clustered and clustered missing AIS data points.

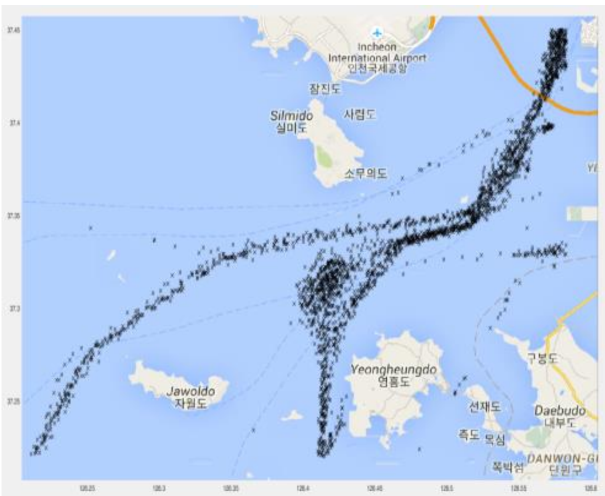

Fig. 2: Non-Clustered Missing AIS Data Points.

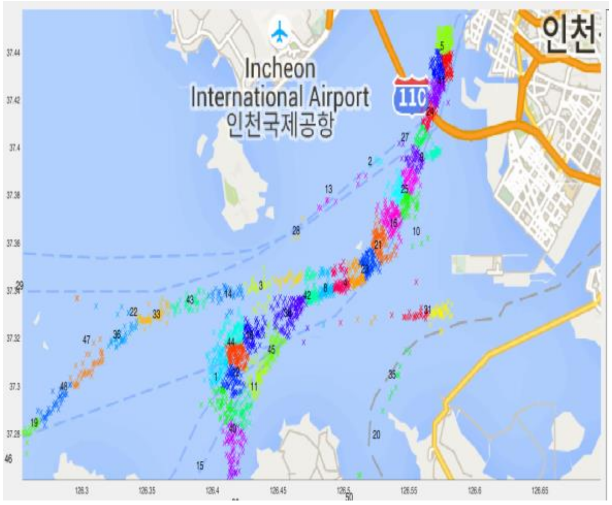

Fig. 3: Clustered Missing AIS Data Points.

After extracting missing AIS data cluster, these clusters are rearranged by each combination of ship's ID and sailing direction of arrival or departure[10]. Table 1 represents missing AIS pattern transaction. A data transaction means a case of either the arrival or departure of ship with ship ID. Data size indicates the number of missing data clusters and the missing AIS data cluster numbers are included in data transactions sequentially.

Table 1: Missing AIS Pattern Transaction

Ship's ID(MMSI)	Time stamp	Data size	Data transaction
440320***	2014-12-31 10:24	2	[11 42]
440452***	2014-12-31 13:32	4	[11, 45, 38, 8]
440369***	2014-12-31 14:38	3	[13, 24, 1]
440721***	2014-12-31 12:24	1	[11]
440452***	2014-12-31 15:11	2	[11, 31]
440369***	2014-12-31 05:32	2	[31, 22]
440321***	2014-12-31 08:24	1	[5]
440352***	2014-12-31 23:05	2	[12, 4]
440369***	2014-12-31 11:34	3	[36]
440620***	2014-12-31 05:24	2	[6,24]
440452***	2014-12-31 01:20	2	[33, 40]
:	:	:	:

4. Association mining

In data mining, association mining is a method used to discover useful or unexpected relations in large databases. In order to find interesting rules, the concept of support and confidence is used as the method of measurement[11]. Support means the ratio of target data to the total number of dataset. Confidence means the ratio of the number of frequent X∪Y to the number of records X. In general, confidence represents the strength of the association rules. For instance, If the support of an specific event X→Y is 0.2 and its confidence is 60%, it has 20 percent of frequency and 60 percent of event Y occur after X is occurred.

Assuming that event Y occurs after event X has occurred, the support and confidence calculation formula are as follows:

$$\mathrm{supp}(X \rightarrow Y) = \frac{\mathrm{frq}(X \cup Y)}{N},\ \mathrm{conf}(X \rightarrow Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X)}.$$

This method of frequent pattern analysis is used to find inter-session patterns assuming that a specific ship tends to omit transmitting the AIS data in several locations. In order to analyze large scale of database, many algorithms for finding association rules have been developed. In this study, we used Apriori algorithm which is well-known algorithm using breadth-first search strategy[12-13].

Missing AIS dataset arranged by for one ship's trajectory is applied to the Apriori algorithm with more than 0.01 support. Figure 4 represents the support value of each cluster.

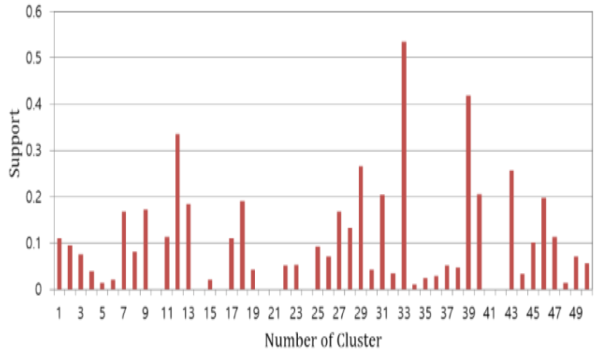

Fig. 4: Support Value of Each Cluster.

Using association rule mining analysis[14], meaningful sequential pattern were found; these are shown in Figure 5. The confidence value indicates the ratio that when missing AIS data occurs in cluster A, then it is also missing in cluster B. For example, if missing AIS data occurred in cluster 47, then this ship will data missing in cluster 17 with a probability of approximately 25.5 percent.

Passenger Route		
Start Cluster	Sequent Cluster	Confidence
2	30	0.116279
2	11	0.139535
2	31	0.139535
2	19	0.116279
2	26	0.116279
27	13	0.144737
27	18	0.131579
27	9	0.105263
27	39	0.105263
28	47	0.183333
28	17	0.15
28	18	0.116667

West Route		
Start Cluster	Sequent Cluster	Confidence
3	12	0.205882
33	9	0.082988
33	18	0.06639
33	46	0.058091
33	39	0.049793
33	43	0.033195
33	7	0.029046
33	13	0.029046
33	50	0.029046
33	12	0.020747
47	17	0.254902
47	18	0.098039
43	7	0.172414
43	46	0.094828
43	9	0.068966
43	18	0.068966

Middle Route		
Start Cluster	Sequent Cluster	Confidence
1	47	0.26
1	28	0.2
8	17	0.162162
8	18	0.162162
9	18	0.269231
23	6	0.291667

LNG Route		
Start Cluster	Sequent Cluster	Confidence
31	50	0.108695668
31	45	0.086956535
31	18	0.076086943
31	19	0.076086943
31	38	0.076086943
31	9	0.065217401
31	33	0.065217401
31	39	0.05434781
31	46	0.05434781

East Route		
Start Cluster	Sequent Cluster	Confidence
11	30	0.156863
11	31	0.117647
11	39	0.117647
11	40	0.098039
11	50	0.098039
25	43	0.166667
25	7	0.119048
25	32	0.119048
40	7	0.204301
40	43	0.139785
40	18	0.053763
40	46	0.053763

Fig. 5: Sequential Pattern Analysis of Incheon Port.

In case of ships navigating through the LNG route, missing AIS situations most often occurred on cluster 31 area in the LNG passage. Since then, the sequent missing AIS situation occurred on cluster 18, 45 and 38 in the east route with the probability of 24%; cluster 33, 43 and 47 in the west route with the probability of 16%. In middle route, missing AIS situation often occurred on cluster 11, 25 and 40. Since then, the sequent missing AIS situation occurred on cluster 47 with the probability of 26% in the west route and cluster 28 with the probability of 20% in the Passenger route and cluster 18 with the probability of 43% in the east route. In east route, missing AIS situation often occurred on cluster 11, 25 and 40. Since then, the sequent missing AIS situation occurred on cluster 31 with the probability of 12% in the LNG route and cluster 7 with the probability of 25% in the east route. In middle route, missing AIS situation often occurred on cluster 1, 8, 9 and 23. Since then, the sequent missing AIS situation occurred on cluster 9 with the probability of 8.3% in the middle route and cluster 12, 18 and 43 with the probability of 49% in the east route. In passenger route, missing AIS situation often occurred on cluster 2, 27 and 28. Since then, the sequent missing AIS situation occurred on cluster 31 with the probability of 14% in the LNG route and cluster 18 with the probability of 36% in the East route and cluster 19 and 47 with the probability of 30% in the East route.

5. Conclusions

With the development of modern ship electronics, mariners increasingly depend on onboard electronic devices. Therefore, mariners need to know the characteristics of the water in the area and the onboard devices. In this paper, we found missing AIS data pattern rules using association rule mining. The proposed approach will predict the probability of missing AIS data.

Based on the results of data mining, we found several missing AIS situation patterns. In case that VTS center does not receive ship's data in the west route, the probability of missing AIS situation is high when that ship enter the east and passenger routes. Also, the probability of missing AIS situation of passing the passenger route is high when that ship enter the LNG, east and west routes.

Acknowledgment

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation (NRF) of Korea (Grant no.: NRF-2017M3C4A7069432) and by Basic Science Research Programs through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2016R1A6A3A11935806).

References

[1] International Maritime Organization. Guidelines for the onboard operational use of shipborne automatic identification systems (AIS). IMO Resolution A.917, 2002.

[2] F. Heymann, P. Banys and C. Saez, Radar Image Processing and AIS Target Fusion, The International Journalon Marine Navigationand Safety of Sea Transportation, 2015, 9 (3), pp.443-448.

[3] Abbas H., Alan W., Philip N. and Jon W., Automatic Identification System (AIS): Data Reliability and Human, the Journal of Navigation, 2007 September 60 (3), pp. 373-389.

[4] Kim K., Jeong J. and Park G., Development of grid projection algorithm of vessel trajectories for e-Navigation, 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems, SCIS, 2014 February, 1 (1), pp. 210-213.

[5] Kim K. and Lee K., Ship Encounter Risk Evaluation for Coastal Areas with Holistic Maritime Traffic Data Analysis, Advanced Science Letters, 2017 October, 23 (10), pp.9565-9569.

[6] Jeong J., Park G., Kim K., Risk assessment model of maritime traffic in time-variant CPA environments in waterway, Journal of Advanced Computational Intelligence and Intelligent Informatics, 2012 November, 16 (7), pp.866-873.

[7] Kim K., Lee K., Context-Aware Information Provisioning for Vessel Traffic Service Using Rule-Based and Deep Learning Techniques, International Journal of Fuzzy Logic and Intelligent System, 2018 February, 18 (1), pp. 13-19.

[8] Ingo H., AIS Adding New Quality to VTS Systems, Journal of Navigation,2000 September, 53 (3), pp. 527-539.

[9] D. M. Kim et al, A music recommendation system with a dynamic k-means clustering algorithm, Sixth International Conference on Machine Learning and Applications, 2008 February, 1(1).

[10] Kim, J., Jeong. J., Extraction of Reference Seaway through Machine Learning of Ship Navigational Data and Trajectory, International Journal of Fuzzy Logic and Intelligent System, 2017 July, Vol.17, No.2, pp.82-90.

[11] Jyoti A. , Nidhi B.and Sanjeev R.,A Review on Association Rule Mining Algorithms, International Journal of Innovative Research in

Computer and Communication Engineering,2013 July, 1 (5), pp.1246-1251.

[12] Bhardwaj, S.; Ozcelebi, T.; Lukkien, J.J.; Lee, K., Semantic Interoperability Architecture for Smart Spaces. International Journal of Fuzzy Logic and Intelligent System, 2018 February, 18 (1), pp. 13-19.

[13] Lee S., Lee K. and Lee K, Density and entanglement-based clustering of sequence data, International Conference on Fuzzy Theory and Its Applications, iFUZZY 2015, 2016 January, pp. 40-43.

[14] K. M. Lee, Mining Generalized Fuzzy Quantitative Association Rules with Fuzzy Generalization Hierarchies,IFSA World Congress and 20th NAFIPS International Conference, 2001, 1 (1), pp. 2977-2982.