

A Functional Model of AIS Data Fusion

Yongming Wang^{1(✉)} and Lin Wu²

¹ Dalian Maritime University, Dalian 116018, China
trancomm@163.com

² Institute of Computing Technology, Chinese Academy of Science,
Beijing 100190, China

Abstract. In recent years, maritime situational awareness based on the fusion of AIS (Automatic Identification System) data has attracted more and more researchers. However, the diversity of terms and methodologies hinders the understanding and communication among them. Besides, AIS data was used mainly for traffic pattern discovery and abnormal vessel detection, more values await discovery. To overcome these two problems, this paper proposes a functional model of AIS data fusion. This model may provide a common frame of reference for discussions of AIS data fusion as well as a checklist for functions a maritime situational awareness system should provide. Based on this model, this paper introduces our works.

Keywords: Maritime situation awareness · Automatic Identification System · AIS · Data fusion · Functional model · JDL model

1 Introduction

The AIS [1] is a self-reporting system installed on ships used to exchange kinematic (dynamic) and identity (static) information etc. with other nearby ships, base stations and satellites. The initial purpose of AIS is collision avoidance. It has interfaces for position (GNSS), heading (compass) and rate of turn (gyrocompass or ROT Indicator) sensors. Other information like navigation status, Maritime Mobile Service Identity (MMSI) number, name of ship, type of ship, destination, Estimated Time of Arrival (ETA), etc. is required to input manually. Ships of 300 gross tons and upwards on international voyages, 500 tons and upwards for cargos not in international waters and passenger vessels are required to fit an AIS transceiver [2]. The time interval of receiving AIS messages from a ship is supposed to be no more than 3 min when within the range of shore-based stations (typically less than 60 miles) or on the order of hours otherwise. Global AIS data collected from shore-based stations and satellites is now available on the Internet, through service providers (e.g., China Transport Telecommunications & Information Center (CTTIC)). However, as we will show in this paper, AIS data is rather dirty, full of noises and errors.

As a main source of vessels' near real time information, AIS data has been studied by growing number of researchers for traffic pattern discovery [3, 5, 6]. In their research, vessel objects, waypoint objects and route objects are discovered and updated by fusing AIS data. Based on discovered patterns, abnormal vessels were detected by

comparing vessels' behaviors with normal patterns [9]. For a more comprehensive maritime situational awareness, detailed types of entities should be identified, such as fishing areas [4], ports, anchorages, etc.

Diverse terms and methodologies about AIS data fusion exist in literature, making the communication and reuse among the community difficult. For example, different terms were used to represent the assessment of traffic routes, such as "analysis of motion patterns" [5], "learning for vessel trajectories" [6], "learning of maritime traffic patterns" [7], "extraction of knowledge" [8], "traffic route extraction" [3], "vessel pattern knowledge discovery" [9], "traffic knowledge discovery" [10], "vessel track information mining" [11], etc. These terms stem from the field of machine learning and data mining, which aim at solving the problem of knowledge or pattern discovery. On the other hand, data fusion is about the assessment of entities of interest, including their states, relationships among them and impacts. Patterns discovered by machine learning or data mining is a source of information in data fusion, which is often referred to as "models [12]". We study the problem of maritime situational awareness in the framework of data fusion rather than machine learning or data mining in this paper, and adopt those widely-used terms in fusion community.

To provide a common frame of reference as well as to explore more potential applications of AIS data, we propose a functional model of AIS data fusion in this paper. It describes what analysis functions or processes need to be performed [13] in a maritime situational awareness system. Another category of model is process model (e.g., Boyd's Observe-Orient-Decide-Act model), which describes how analysis is accomplished [13]. A functional model is useful in system engineering by "providing visualization of a framework for partitioning and relating functions and serving as a checklist for functions a system should provide" [12].

The JDL (Joint Directors of Laboratories) data fusion model [15] is a well-accepted functional model in fusion community. Our model is an instantiation of its revised version [14], and it focuses on Level 0 to Level 3. Based on this model, this paper introduces our work.

The remainder of the paper is organized as follows. Section 2 provides a detailed description of the proposed functional model. Section 3 summaries our existing works under the functional model. Section 4 draws the conclusion and future work.

2 Functional Model of AIS Data Fusion

In this section, we describe the functional model from level 0 to level 3, see Fig. 1. This model is based on the revised version [14] of the JDL fusion model [15]. The original JDL model centers on military scenes and "suffered from an unclear partitioning scheme [14]". The revised version is more general and "provides a clear and useful partitioning while adhering as much as possible to current usage across the data fusion community [14]".

Our model is entity-oriented and entities can be vessels, traffic route segments, traffic stops, fishing areas, anchorages, etc. These entities' aggregates or relations can also be regarded as an entity. The "products of processes at each level are estimates of some existing or predicted aspects of reality [15]", and they are stored in database.

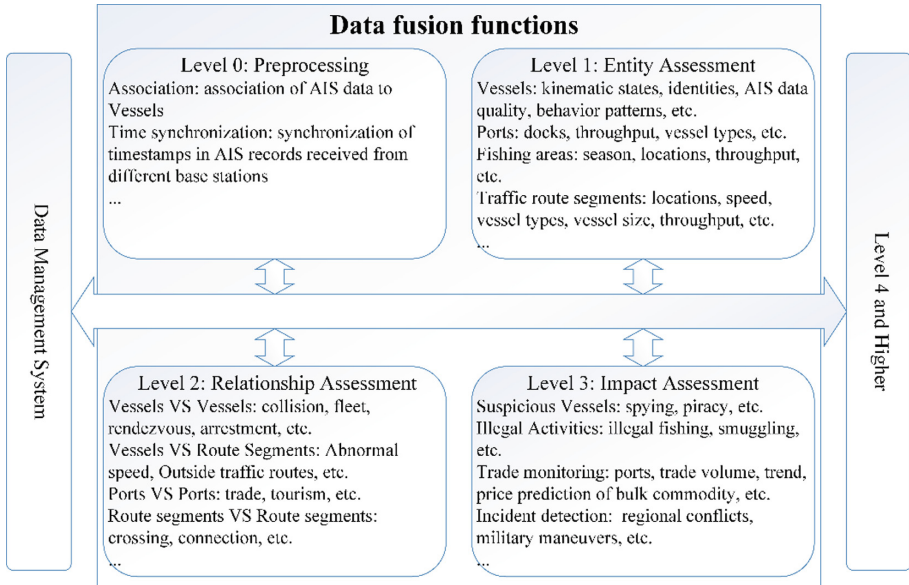


Fig. 1. Proposed functional model of AIS data fusion, which focuses on Level 0 to Level 3. AIS data as well as products of fusion is managed by the data management system on the left of this figure. AIS data

To highlight the data fusion functions in this figure, we put data sources, support databases and fusion databases into the data management system together. According to CTTIC [16], it is gathering about 10 billion records (around 1 TB) each year. So the data management system should be able to handle AIS data on the order of TB.

As level 4 (process assessment) and higher processes (user refinement) are not specifically fusion functions, they are not discussed in this paper.

2.1 Level 0: Preprocessing

AIS data is prone to be highly erroneous and noisy due to the deficiencies of the system and improper use of it. Although this problem was revealed [17, 18], limited work [19, 20] has been done on the preprocessing of AIS data.

As AIS was designed for collision avoidance, many issues may arise when applied for real-time surveillance. For example, an AIS message only contains the seconds part (6 bits) of the time that it was broadcast and it is up to receivers to affix the full time stamp. A comparison was made between the seconds part of the original time stamp and affixed one, and significant differences were found [17]. Besides, as clocks are not synchronized, a ship will jump on the map when its AIS messages are received from different receivers [20].

In addition to flaws in the design of AIS, improper use of it makes the situation even worse. For instance, the MMSI of each ship is supposed to be unique, and it's commonly used [9] to identify and track vessels. But it's not rare that a MMSI is shared

by different vessels [19], as this field is input manually and not reliable. To identify each vessel uniquely, we need an internal ID. For each AIS message, its association with an existed or new ship must be determined first. This problem can be solved by tracking vessels in real-time or off-line [4, 10, 19].

2.2 Level 1: Entity Assessment

In maritime situational awareness systems, entities of interest are mainly vessels [21], ports [5, 27], fishing areas [4], traffic route segments [7, 8, 10], turning points and junctions [7, 8], etc. When the area of interest is bounded, entry and exit points [10] are also important entities.

The state vector of a vessel includes position, speed, direction, etc. Sequential states of a vessel constitute its tracks, which can be partitioned into moving segments and stops. These segments and stops can be further classified into traffic route segments, traffic stops, fishing areas, ports, anchorages, etc. The detection and assessment of traffic route segments is a basic problem in Level 1 [3, 9, 10, 22].

2.3 Level 2: Relationship Assessment

Level 2 process is usually referred to as “situation refinement [15]” or “situation assessment [14]”. It has been pointed out that situations can be represented as sets of relations [14], so we refer to Level 2 process as relationship assessment in this paper for clarity.

Relations of interest for maritime situational awareness include but are not limited to:

- vessels VS vessels

Maritime authorities are deeply concerned about the relationship among vessels [26], because illegal activities and incidents can be detected by relation assessment. For example, we can discover smuggles by identifying vessels’ rendezvous at sea. And encounter of vessels may indicate critical incidents like collision, piracy or arrestment, etc.

- vessels VS route segments

Original AIS data is made up of point sets. After traffic route segments have been extracted in level 1 process, we can represent the track of each vessel as series of segments. This can remove redundancy in raw AIS data and compress it by about two orders. It was modeled as a classification problem in previous works [9], and solved by maximizing the posterior probability that a vessel is sailing over a certain route.

As discovered routes are characterized by spatial, temporal and other attribute-related features (e.g., type and size) of vessels sailing on them [3, 9], and features of vessels include their historical route patterns, we can detect anomalies when a vessel’s attributes are not compatible with the route segment, or when it is not compatible with its historical patterns.

- ports VS ports

There is limited work on analyzing the relationship between ports. An example was identifying patterns of ships' transition among ports by AIS data mining using the software R [23]. Based on discovered patterns, they estimated destination ports.

- route segments VS route segments

Extracted route segments are separate lines [22]. We need to know their relations of connection [8], in order to form the global route map.

2.4 Level 3: Impact Assessment

Level 3 process combines products of level 1 and level 2 to assess impacts. Functions provided at this level include predicting vessels' positions, assessing the intention and threat of vessels and detecting anomalies.

As time intervals between messages of the same ship varies from seconds to hours [17] due to limited coverage and bandwidth of receivers, it is necessary to predict positions for situational awareness [5, 9, 17, 19].

Anomaly detection and risk assessment are usually based on the assessment of relationship among entities. Most existing literature only focused on detecting anomalous ships [8, 9, 11, 24–27].

3 Our Existing Works

3.1 Level 0: Preprocessing

We have performed association on global AIS data of 34 months (from August 2012 to May 2015) based on spatial and temporal proximity [28]. When two records with the same MMSI were more than 100 km away and the speed reckoned was more than 120 knots, then they were regarded as from different ships [28]. The results of pre-processing showed that, among 1,998,200 distinct MMSIs recorded, 1,607,664 were suspicious: the average number of AIS messages broadcast by each of them was only 11.3. The remaining 390,536 MMSIs were used by 491,346 vessels, 43,034 messages had been received from one vessel on average.

3.2 Level 1: Entity Assessment

We proposed a vessel trajectory partitioning method based on hierarchical fusion of position data reported by AIS, aiming at improving preciseness and processing speed [29]. Our method consists of two steps: position point fusion and sub-trajectory fusion. The key idea of our method is to describe trajectories by hierarchical concepts. In the first step, trajectories are partitioned by grouping positions in raw AIS records into sub-trajectories represented as straight line segments. In the second step, we aggregate successive sub-trajectories into more abstract concepts: route segments and stops. We applied our method on a data set containing 473963 AIS records from 10 vehicle carriers along the coast of China. Algorithms were implemented by python 2.7.3 on a

notebook with Intel Core2 Duo CPU T6600 @2.20 GHz, 2 GB RAM. The average execution time of our two level partitioning on each vessel is around 2.30 s and 0.22 s respectively. After partitioning, we got 250 route segments and 16 stops. Experimental results showed that our method split routes and identified stops precisely at the computational complexity of $O(n)$.

To describe shipping density of each area, we defined vessel and traffic density [28]. The definition of vessel density in a region was taken as the expected number of vessels per unit area at any time, and traffic density as the average number of vessels crossing this region per unit area per unit time. We calculated vessel and traffic density using a grid-based method. The traffic density in 2014 is shown in Fig. 2.

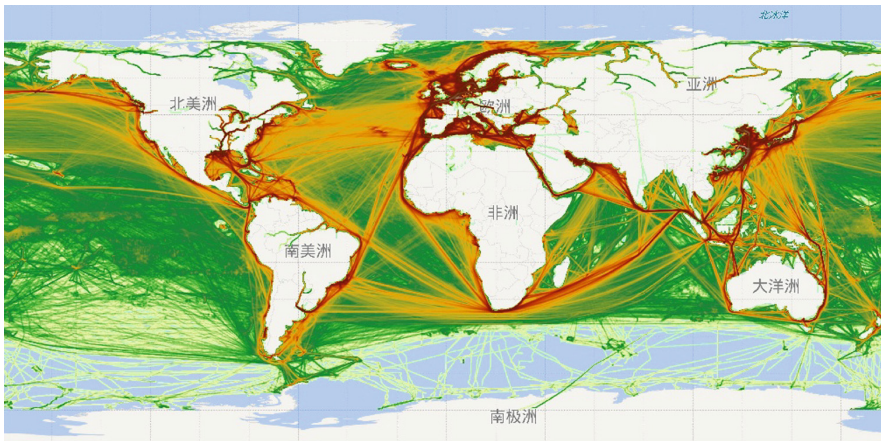


Fig. 2. Global traffic density in 2014, at the spatial resolution of 10 min longitude by 10 min latitude

3.3 Level 2: Relationship Assessment

We are developing an online system detecting vessels' rendezvous. In the system, we divide the earth into grids of the size 1° latitude by 1° longitude, and put ships into them according to their locations. Rendezvous detection is performed every 10 min by calculating distance among ships inside the same grid and surrounding grids.

We have mined ships' mooring positions base on AIS data, including position, speed and status. Each position was represented as a grid of 0.6 s longitude by 0.6 s latitude, and the membership that each grid belonged to a berth was calculated separately according to tracks located in this grid using fuzzy inference. After that, we clustered mooring positions into berths using DBSCAN, taking multiple attributes of positions into consideration in the function of distance. An example of a clustered berth is demonstrated in Fig. 3.

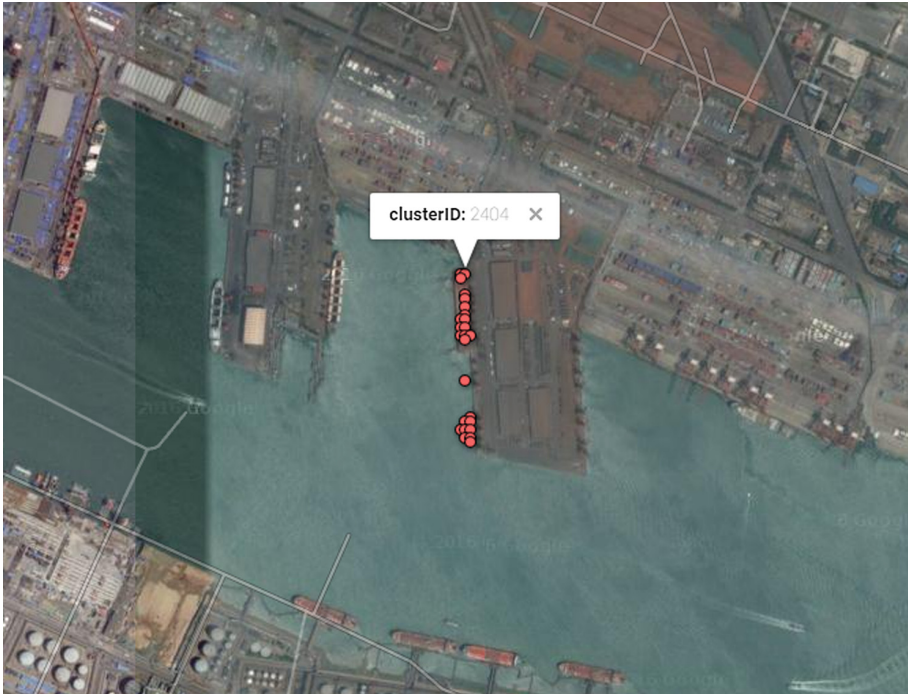


Fig. 3. A clusters of mooring points, belonging to a berth

3.4 Level 3: Impact Assessment

Based on vessel and traffic density maps described in Sect. 3.1, we are developing an online system detecting two types of anomalies: areas with vessel number deviating from expected and ships sailing in unusual routes. The former anomaly is detected by comparing current number of vessels with patterns in vessel density maps. The latter anomaly is detected by comparing each vessel's route with traffic density maps.

4 Conclusion and Future Work

We proposed a functional model of AIS data fusion in this paper. We hope that this model can provide a common frame of reference for maritime situational awareness based on AIS data fusion. This model is an instantiation of the well-accepted JDL model and serves as a checklist for functions a maritime situational awareness system should provide. Based on this model, this paper introduced our works.

Our future work would be to achieve more comprehensive global maritime situational awareness based on AIS data fusion. In the entity assessment step, we will detect and estimate more kinds of entities. Previous work mainly focused on stops and routes. To achieve global maritime situational awareness, we would consider more types of entities: vessels, traffic route segments, traffic stops, fishing areas, anchorages and ports,

etc. With those entities, we can assess more interesting kinds of relations in level 2 process, such as fleet identification, which has not been studied yet as far as we know. Although anomaly detection is a hot topic in level 3 process, most efforts were made on identifying ships whose tracks are not compatible with traffic patterns. In our future work, those ships deviating from their historical patterns will also be marked as anomaly. Taking more types of entities into consideration, we plan to detect more kinds of anomaly based on the analysis of their states and relations among them. Possible anomalies include birth and death of fishing areas and anchorages, changes of traffic route segments, etc. Apart from anomaly detection, we will also perform other level 3 processes such as route planning, trade monitoring, piracy detection, etc.

References

1. Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band. International Telecommunications Union (ITU) Recommendation ITU-RM, pp. 1371–1374 (2010)
2. Safety of life at sea (solas) convention chapter V, regulation 19
3. Pallotta, G., Vespe, M., Bryant, K.: Traffic route extraction and anomaly detection from AIS data. In: Proceedings of the International COST MOVE Workshop on Moving Objects at Sea, Brest, France (2013)
4. Mazzarella, F., Vespe, M., Damalas, D., et al.: Discovering vessel activities at sea using AIS data: mapping of fishing footprints. In: Proceedings of the 16th International Conference on Information Fusion, pp. 1–7 (2014)
5. Ristic, B., Scala, B.L., Morelande, M., et al.: Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In: Proceedings of the 11th International Conference on Information Fusion, pp. 1–7 (2008)
6. Vries, G.K.D., Someren, M.: Machine learning for vessel trajectories using compression, alignments and domain knowledge. *Expert Syst. Appl.* **39**(18), 13426–13439 (2012)
7. Vespe, M., Visentini, I., Bryan, K., et al.: Unsupervised learning of maritime traffic patterns for anomaly detection. In: Proceedings of the 9th IET Data Fusion & Target Tracking Conference: Algorithms & Applications, pp. 1–5 (2012)
8. Guillarme, N.L., Lerouvreurt, X.: Unsupervised extraction of knowledge from S-AIS data for maritime situational awareness. In: Proceedings of the 16th International Conference on Information Fusion, pp. 2025–2032 (2013)
9. Pallotta, G., Vespe, M., Bryant, K.: Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. *Entropy* **15**(6), 2218–2245 (2013)
10. Pallotta, G., Vespe, M., Bryant, K.: Traffic knowledge discovery from AIS data. In: Proceedings of the 16th International Conference on Information Fusion, pp. 1996–2003 (2013)
11. Deng, F., Guo, S., Deng, Y., et al.: Vessel track information mining using AIS data. In: Proceedings of the International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI), pp. 1–6 (2014)
12. Kessler, O., White, F.: Data Fusion Perspectives and Its Role in Information Processing. In: Handbook of Multisensor Data Fusion, pp. 15–43 (2009)
13. Antony, R.: Data fusion automation: a top-down perspective. In: Handbook of Multisensor Data Fusion, pp. 137–163 (2009)

14. Steinberg, A.N., Bowman, C.L.: Revisions to the JDL data fusion model. In: *Handbook of Multisensor Data Fusion*, pp. 45–67 (2009)
15. Hall, D.L., Llinas, J.: An introduction to multisensor data fusion. *Proc. IEEE* **85**(1), 6–23 (1997)
16. Website of AIS data services provided by CTTIC. http://www.myships.com/ShipLBS/index_en.html
17. Greidanus, H., Alvarez, M., Eriksen, T., et al.: Basin - wide maritime awareness from multi - source ship reporting data. *Trans. Nav. Int. J. Mar. Navig. Saf. Sea Transp.* **7**(2), 35–42 (2013)
18. Harati-mokhtari, A., Wall, A., Brooks, P., et al.: Automatic Identification System (AIS): A Human Factors Approach. http://217.26.101.136/middenlimburg/downloads/documenten/pdf/ais_human_factors.pdf
19. Mazzarella, F., Alessandrini, A., Greidanus, H., et al.: Data fusion for wide-area maritime surveillance. In: *Proceeding of COST MOVE Workshop on Moving Objects at Sea, Brest* (2013)
20. Pallotta, G., Horn, S., Braca, P., et al.: Context-enhanced vessel prediction based on Ornstein-Uhlenbeck processes using historical AIS traffic patterns: real-world experimental results. In: *Proceedings of 17th International Conference on Information Fusion*, pp. 1–7 (2014)
21. Falcon, R., Abielmona, R., Blasch, E.: Behavioral learning of vessel types with fuzzy-rough decision trees. In: *Proceedings of 17th International Conference on Information Fusion*, pp. 1–8 (2104)
22. Lee, J.G., Han, J., Whang, K.Y.: Trajectory clustering: a partition-and-group framework. In: *Proceedings of ACM SIGMOD International Conference on Management of data*, pp. 593–604 (2007)
23. Hadzagic, M., St-hilaire, M.O., Webb, S., et al.: Maritime traffic data mining using R. In: *Proceedings of 16th International Conference on Information Fusion*, pp. 2041–2048 (2013)
24. Balmat, J.F., Lafont, F., Maifret, R., et al.: A decision-making system to maritime risk assessment. *Ocean Eng.* **38**(1), 171–176 (2011)
25. Mascaro, S., Nicholso, A.E., Korb, K.B.: Anomaly detection in vessel tracks using Bayesian networks. *Int. J. Approx. Reas.* **55**(1), 84–98 (2014)
26. Lane, R.O., Nevell, D.A., Hayward, S.D., et al.: Maritime anomaly detection and threat assessment. In: *Proceedings of 13th International Conference on Information Fusion*, pp. 1–8 (2010)
27. Castaldo, F., Palmieri, F.A., Bastani, V., et al.: Abnormal vessel behavior detection in port areas based on Dynamic Bayesian Networks. In: *Proceedings of 17th International Conference on Information Fusion*, pp. 1–7 (2014)
28. Wu, L., Xu, X.J., Wang, Q., et al.: Mapping global shipping density from AIS Data. *J. Nav.* **70**, 67–81 (2016)
29. Wu, X.B., Wu, L., Xu, Y., et al.: Vessel trajectory partitioning based on hierarchical fusion of position data. In: *Proceedings of 18th International Conference on Information Fusion*, pp. 1230–1237 (2015)
30. Natale, F., Gibin, M., Alessandrini, A., et al.: Mapping fishing effort through AIS data. *PloS one* **10**(6), 1–16 (2015)