

Research Article

Ship Trajectory Reconstruction from AIS Sensory Data via Data Quality Control and Prediction

Xinqiang Chen^{1,2}, Jun Ling¹, Yongsheng Yang¹, Hailin Zheng³, Pengwen Xiong⁴, Octavian Postolache⁵, and Yong Xiong⁶

¹Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

²Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China

³Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China

⁴School of Information Engineering, Nanchang University, Nanchang 330031, China

⁵ISCTE-Instituto Universitario de Lisboa and Instituto de Telecomunicacoes, Lisbon 1649-026, Portugal

⁶Hunan Lianzhi Technology Co., Ltd., Changsha 410217, China

Correspondence should be addressed to Pengwen Xiong; steven.xpw@ncu.edu.cn

Received 28 May 2020; Revised 9 July 2020; Accepted 20 July 2020; Published 17 August 2020

Academic Editor: Javier Martinez Torres

Copyright © 2020 Xinqiang Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate ship trajectory plays an important role for maritime traffic control and management, and ship trajectory prediction with Automatic Identification System (AIS) data has attracted considerable research attentions in maritime traffic community. The raw AIS data may be contaminated by noises, which limits its usage in maritime traffic management applications in real world. To address the issue, we proposed an ensemble ship trajectory reconstruction framework combining data quality control procedure and prediction module. More specifically, the proposed framework implemented the data quality control procedure in three steps: trajectory separation, data denoising, and normalization. In greater detail, the data quality control procedure firstly identified outliers from the raw ship AIS data sample, which were further cleansed with the moving average model. Then, the denoised data were normalized into evenly distributed data series (in terms of time interval). After that, the proposed framework predicted ship trajectory with the artificial neural network. We verified the proposed model performance with two ship trajectories downloaded from public accessible AIS data base.

1. Introduction

Maritime transportation occupies over 90% of global trade in terms of goods delivering volume. Enhancing traffic safety attracts huge attention considering that maritime traffic incident can cause significant loss of human life, navigation environment damage, etc. [1, 2]. To avoid potential maritime accidents, various maritime surveillance data are collected for the purpose of navigation environment awareness, which provides accurate early-warning information to maritime traffic participants [3]. The AIS data involves meaningful spatial-temporal maritime traffic information which supports various navigation operation decisions. More specifically, the AIS data is a popular data source for analyzing ship trajectory variation tendency. Note that AIS is a type of self-

reporting system originally designed for preventing potential accident, which is a mandatory facility for cargo ships (i.e., ship with gross tonnage larger than 300) [4–8]. Moreover, fishing boats with length longer than 15 m are required to install AIS equipment in the European Union Member States [9–12].

The AIS equipment transmits the static and kinematic ship information (e.g., ship type, call sign, speed, latitude, longitude, heading, Maritime Mobile Service Identity (MMSI), etc.) at a variable refresh rate. More specifically, the AIS system broadcasts the ship information ranging from several minutes to two seconds based on the ship travelling speeds (i.e., the AIS system updates its data at lower frequency under larger maneuvering speed). In that manner, ship (equipped with AIS facility) position can be obtained in

real time in coastal area. Moreover, the large-scale AIS datasets have been stored at regional or national data centers, which can be accessible by users in request (note that users may need to pay for accessing the AIS data). Previous studies suggested that AIS data quality imposes significant influence on the maritime traffic safety analysis, and thus improving AIS data quality has become an active topic in the maritime community [13, 14].

AIS data anomaly removal studies involve unsupervised clustering method and neural network based and statistical models [15, 16]. Liu et al. proposed an adaptive Douglas-Peucker framework to suppress AIS data outliers in the manner of data compression [17]. Deng introduced a Markov based model to explore ship movement patterns, which were further used to identify the abnormal AIS data samples [18]. Zhang et al. proposed a hierarchical density-based spatial clustering of applications with noise based model to cluster and denoise the raw AIS trajectories [19]. Rong et al. cleansed the raw AIS data in the lateral and longitudinal dimensions with a novel probability trajectory prediction model [20]. The neural network relevant models have shown many successes in tackling the AIS trajectory denoising and prediction tasks [21]. Hoque and Sharma applied long short-term memory neural network to forecast ship trajectories, which were employed to suppress the AIS data anomaly [22]. Kim and Lee proposed a novel deep neural network model to remove AIS outliers and thus predict both medium- and long-term ship trajectory variation tendencies [23]. Similar researches can be found in [8, 24–28].

We aim to propose a novel AIS denoising and prediction framework with the support of data quality control procedure. Our main contributions can be summarized as follows: (1) we cleansed the raw AIS data with the steps of trajectory separation, outlier removal, and data normalization; (2) we predicted ship trajectory via the denoised AIS data with the artificial neural network (ANN); (3) we testified the proposed framework performance on two ship trajectories. The study can help maritime traffic participants forecast accurate ship trajectories and thus take early-warning measurements to enhance maritime traffic efficiency and safety. The remainder of the paper is organized as follows. We introduce the data source used in our study in Section 2. After that, the methodology details about the AIS data denoising are illustrated in Section 3, and then the ANN model used for predicting ship trajectories is presented. The experimental results are shown in Section 4. Section 5 briefly concludes the study and illustrates future work.

2. Data

The U.S. Marine Energy Administration and National Oceanic and Atmospheric Administration provides large-scale AIS data, which benefits many AIS relevant studies due to its public accessibility (<https://marinecadastre.gov/ais/>) [29, 30]. The original AIS dataset includes both kinematic and static information for the ship, which contains MMSI, Co-ordinated Universal Time (UTC), latitude, longitude, speed over ground (SOG), heading, course over ground (COG), timestamp, call sign, and so forth. We collect the AIS data

from the Gulf of Mexico with latitude ranging from 18°N to 31°N, and the longitude falls in the interval [75°W, 100°W]. The minimum time interval for sampling the AIS data is 1 s, and the maximum value is 500 s. We collect 12813 AIS data samples on April 11, 2017, from the above-mentioned database (see Figure 1). Following the international standard for representation of latitude and longitude, the south latitude (west longitude) is denoted as a negative number, while north latitude (east longitude) is presented with a positive number.

3. Methodology

The raw AIS data may contain different types of outliers due to instable signal transmission rate, data transmission congestion, etc. It is important to suppress such data anomalies for the purpose of exploiting reliable maritime traffic kinematic information from the AIS dataset. To address the issue, we firstly implement the data quality control procedure to remove the trajectory outliers and then predict the trajectory with artificial neural network. The schematic overview for the proposed framework is shown in Figure 2.

3.1. Data Quality Control for Suppressing AIS Outliers. Ship trajectory data (i.e., AIS data) is stored in the database via data delivering/receiving timestamp, and thus we need to aggregate trajectory data (from a single ship) before conducting ship trajectory analysis relevant researches. For the purpose of thoroughly removing anomalous AIS samples, we implement the data quality control with steps of ship trajectory extraction, data cleansing, and data formatting (i.e., AIS time interval normalization).

3.1.1. Ship Trajectory Extraction. The ship trajectory extraction can be divided into separating trajectories from different ships and removing discontinuous ship trajectories. It is noted that the ship can be uniquely identified by the MMSI, which is thus applied to separate AIS data samples from different ships. Then, the raw AIS samples are sorted by timestamp in an ascending manner. It is found that an AIS sample may be recorded in database for several times. To address the issue, the repetition samples are removed to avoid being further processed when the constraints in equation (1) are satisfied. The outputs from the above step are the raw ship trajectories. We find that several time intervals between neighboring samples are very large (e.g., four hours), indicating that many AIS data are lost. Such AIS data discontinuity imposes big challenge for analyzing ship kinematic moving state in detail. To overcome the disadvantage, we divide the raw ship trajectory into different segments when the time interval between neighboring samples exceeds a threshold (see equation (2)):

$$\begin{cases} T_a = T_b, \\ \text{Lat}_a = \text{Lat}_b, \\ \text{Lon}_a = \text{Lon}_b, \end{cases} \quad (1)$$

$$T_i > T_{th}, \quad (2)$$

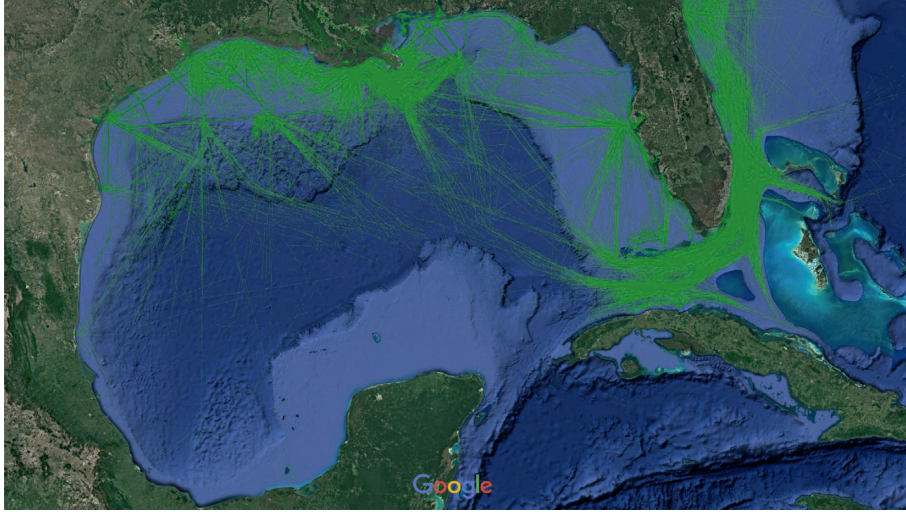


FIGURE 1: Raw AIS data collected from Gulf of Mexico.

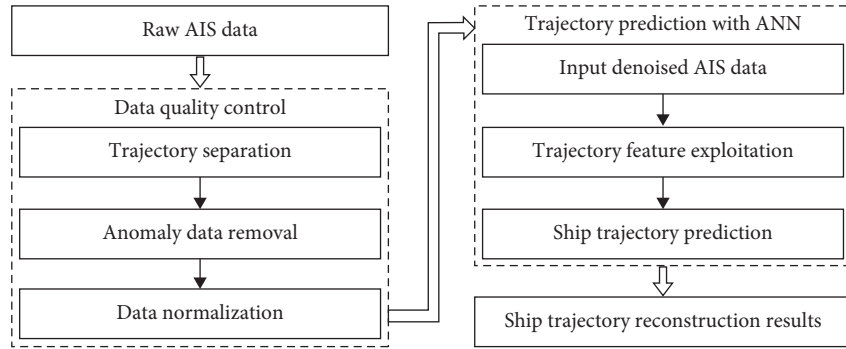


FIGURE 2: Overview for the proposed AIS trajectory reconstruction framework.

where T_a and T_b are the timestamps from two AIS records. The ship positions at timestamp T_a are denoted as Lat_a and Lon_a , respectively. Lat_b and Lon_b are the counterparts at timestamp T_b . T_i is the time interval between neighboring samples. T_{th} is the threshold, which is set to 4 hours by default.

3.1.2. Removal of AIS Data Anomaly. After obtaining the AIS data in the above step, we implement the anomaly data denoising procedure to remove AIS data noises. Typical AIS data outliers are summarized as follows: (a) The longitude and (or) latitude: it is far beyond the reasonable range. We collect the AIS data in the Gulf of Mexico, and the longitude (latitude) is supposed to fall in $75^\circ W$ ($18^\circ N$) and $100^\circ W$ ($31^\circ N$). The ship trajectory will be considered as outliers when the ship spatial data (i.e., latitude and longitude) exceed the range. Moreover, sudden longitude (latitude) variation is another type of typical outlier, and we employ the moving average method to correct the data outliers (see equation (3)). (b) Abnormal velocity data: after manually checking the raw AIS data, we find several ship speed samples are very high (i.e., larger than 30 knots). It is less likely for a ship travelling in inland waterways at such speed for the purpose of ensuring maritime traffic safety. (c) Ship

course outlier: ship may change its moving direction in coastal areas to avoid maritime traffic collision. But large ship course variation is not permitted in real world. We average the neighboring ship courses to remove such data outlier. Given ship headings C_1 , C_2 , and C_3 from three neighboring AIS trajectory samples (with timestamps t_1 , t_2 , and t_3), we consider C_2 as the outlier when the condition in equation (4) is satisfied. The ship heading C_2 is updated as C'_2 with equation (5):

$$\left\{ \begin{array}{l} lat_i = \frac{lat_{(i-1)} + lat_{(i+1)}}{2} |lat_i - lat_{(i-1)}| \\ > lat_{th} \text{ and } |lat_i - lat_{(i+1)}| > lat_{th}, \\ lon_i = \frac{lon_{(i-1)} + lon_{(i+1)}}{2} |lon_i - lon_{(i-1)}| \\ > lon_{th} \text{ and } |lon_i - lon_{(i+1)}| > lon_{th}. \end{array} \right. \quad (3)$$

$$\left\{ \begin{array}{l} |C_2 - C_1| > C_{th}, \\ |C_2 - C_3| > C_{th}, \\ |C_1 - C_3| \leq C_{th}, \end{array} \right. \quad (4)$$

$$\begin{cases} C'_2 = C_1 + \frac{C_1 - C_3}{t_{13}} \times t_{12}, & C_1 > C_3, \\ C'_2 = C_3 + \frac{C_3 - C_1}{t_{13}} \times t_{12}, & C_1 < C_3. \end{cases} \quad (5)$$

Here, ship latitude and longitude at timestamp i are lat_i and lon_i , respectively. The rule is applicable to $\text{lat}_{(i-1)}$, $\text{lat}_{(i+1)}$, $\text{lon}_{(i-1)}$, and $\text{lon}_{(i+1)}$. lat_{th} and lon_{th} are the thresholds for the latitude and longitude, respectively. C_{th} is the ship heading variation threshold. The parameter t_{13} is time difference between timestamps t_1 and t_3 , and t_{12} is the time interval between t_1 and t_2 .

3.1.3. AIS Data Normalization. We can obtain noise-free AIS dataset after implementing the above two steps. It is found that time interval may vary from different data samples, which hinders ship trajectory reconstruction model from accurately extracting AIS intrinsic patterns. In that manner, it is difficult to predict ship trajectory in real-world applications. To address the issue, we employ the cubic spline interpolation and moving average models to normalize the AIS data series. Given three noise-free AIS trajectories A_1 , A_2 , and A_3 , we label and store the AIS samples A_1 and A_3 assuming that one of the following conditions is met (see equations (6) to (9)). Moreover, the AIS sample A_2 is denoted as flag data when the constraints in equation (10) are satisfied. We normalize the ship trajectory samples between A_1 and A_2 with the cubic spline interpolation, and, for more details, we suggest the reader to refer to [31]. The ship AIS data between A_2 and A_3 is normalized with the moving average model, and details can be found in [26].

Note that appropriate time interval is crucial for ship trajectory analysis due to the fact that large time interval can lead to ship kinematic information loss, and smaller time interval may introduce trivial ship moving patterns. After carefully exploiting time interval distributions via the collected AIS data samples (see Figure 3), we find the majority of time for interval samples is 60 s, which is set as default value in our study without further specifications:

$$\frac{d_{12}}{d_{23}} > d_{t1}, \quad (6)$$

$$t_{12} > t_{t1}, \quad (7)$$

$$|v_1 - v_2| > v_{\text{th}}, \quad (8)$$

$$d_{12} < d_{t2}, \quad (9)$$

$$\begin{cases} t_{23} < t_{t2}, \\ d_{12} > d_{t3}, \\ d_{23} > d_{t4}, \end{cases} \quad (10)$$

where $d_{12} \cdot (d_{23})$ is the ship displacement between positions $A_1 \cdot (A_2)$ and $A_2 \cdot (A_3)$ (A_3). Parameter $t_{12} \cdot (t_{23})$ is time cost for the ship travelling from position $A_1 \cdot (A_2)$ to position $A_2 \cdot (A_3)$. The speed recorded in the trajectory sample

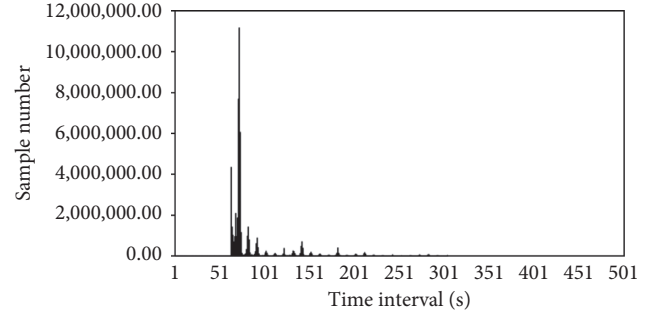


FIGURE 3: Time interval distribution for the collected AIS samples.

A_1 is v_1 , and the rule is applicable to v_2 . d_{t1} , t_{t1} , v_{th} , d_{t2} , t_{t2} , d_{t3} , and d_{t4} are the thresholds, with the default settings being 5, 360, 5, 10, 160, 30, and 30.

3.2. Ship Trajectory Prediction with the ANN Model. The artificial neural network model has shown great successfulness in many roadway traffic flow prediction applications, which demonstrates its potential in ship trajectory prediction task. The main advantages of the back-propagation (BP) neural network are strong nonlinear curve fitting capability, low complexity, and self-learning ability, which can easily identify and predict ship trajectory variation tendency. Moreover, the ANN model can output ship trajectory prediction results in real-time manner due to the low computational cost, which can provide instant maritime traffic information for tackling time-demanding maritime tasks. Based on the above reasons, we employ the ANN model to predict AIS trajectories. The ANN model exploits intrinsic relationship between input training (and testing) data and the output samples with the human-like information perception rule.

For the given g th neuron node, we denote by A_j ($j = 1, 2, \dots, J$) the input ship AIS trajectory. w_j ($j = 1, 2, \dots, J$) represents the weight for each ship trajectory. Based on that, the input AIS data for the g th neuron node is obtained by equation (11). Note that the hidden layer in a BP neural network plays the role of extracting ship travelling patterns from the AIS data. With the help of transfer function, the BP neural network can learn the nonlinearity patterns among the input ship AIS data samples, and the sigmoid transfer function used in our study is shown in equation (12). The BP network measures difference between the predicted AIS trajectories and ground-truth data, which is returned back to network to adjust the model structure and neuron settings for the purpose of obtaining optimal ship trajectory prediction results:

$$N_{\text{in}}^g = \sum_{j=1}^J w_j^* A_j, \quad (11)$$

$$f(s_g^j) = \frac{1}{(1 + \exp(-s_g^j))}, \quad (12)$$

where N_{in}^g is the input for the g th neuron node and s_g^j is the state of the g th neuron of the hidden layer with the j th AIS sample.

3.3. Evaluation Metrics. To quantify the ship trajectory prediction performance, we compare the predicted AIS data with ground truth data with typical statistical measurements. Following the rule in previous studies [26, 27], we employ the root mean square error (RMSE), mean absolute error (MAE), Frechet distance (FD), and average Euclidean distance (AED) to measure the prediction goodness. For any given ship trajectories, the prediction accuracy is quantified with the above-mentioned statistical indicators (see equations (13) to (16)). The smaller RMSE, MAE, FD, and AED indicate more accurate ship trajectory prediction accuracy, and vice versa. Note that both the RMSE and MAE indicators are implemented to quantify the ship trajectory prediction accuracy in terms of longitude and latitude, respectively:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - g_i)^2}, \quad (13)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |p_i - g_i|, \quad (14)$$

$$\text{FD} = \max_{i \in [1, n]} \sqrt{(p_i(x) - g_i(x))^2 + (p_i(y) - g_i(y))^2}, \quad (15)$$

$$\text{AED} = \frac{\sum_{i=1}^n \sqrt{(p_i(x) - g_i(x))^2 + (p_i(y) - g_i(y))^2}}{n}, \quad (16)$$

where n is the number of AIS data samples. p_i is the i th predicted AIS data sample, and g_i is the i th ground truth AIS data samples. The parameters $p_i(x)$ and $p_i(y)$ are the latitude and longitude for the i th predicted AIS data sample, and $g_i(x)$ and $g_i(y)$ are the counterparts for the i th ground truth AIS sample.

4. Experiment

For the purpose of evaluating framework performance, we have collected two typical ship trajectories (i.e., two groups of AIS data samples) from the observed navigation region. The AIS data for the ship with MMSI No. 357234000 and No. 367715380 were collected from the above-mentioned data base, which were denoted as Case 1 and Case 2, respectively. The ship trajectory for Case 1 was sampled from 24 January, 2017, to 25 January, 2017. The data samples for Case 2 were collected on 6 January, 2017. The framework was implemented on Windows10 OS with 16 GB RAM and 4 GHz CPU. We employed Python (3.5 version) to perform the data quality control and prediction procedure on the ship trajectory data.

4.1. Ship Trajectory Reconstruction on Case 1. We first presented the ship trajectory reconstruction results on Case 1 (i.e., the ship with MMSI No. 357234000) and then verified the model performance on the AIS data from the ship with MMSI No. 367715380. The spatial-temporal ship trajectory distribution shown in Figure 4 indicated

that the ship was travelling back and forth in small area considering that both ship longitudes and latitudes varied in small range. But several obvious outliers were found in the raw ship trajectory data. More specifically, the anomalous ship positions from several AIS data samples were far away from their neighbors, which showed unreasonable ship displacement. After carefully checking the raw AIS data, we found that average ship moving speed was quite slow (i.e., smaller than 4 knots). The main reason is that the ship is a special survey seismic vessel which was towing on the water surface at a large area (e.g., towing a dozen of sensors connected by hydrophone streamer cables). Moreover, the ship's instant speed reached 20 knots when the ship position was considered as obvious outlier (e.g., abnormal ship latitude and/or longitude). It can be inferred that the ship finished the task in the current coastal region and thus speeded up to another sea area. Besides, the abnormal ship longitude positions were different from those of the latitude counterparts (see Figures 4(a) and 4(b), respectively). In that way, anomaly ship trajectory sample was observed when latitude (or longitude) was interfered by neighboring ship positions.

The denoised ship trajectory showed that abnormal data samples were successfully removed by the proposed framework considering that no outliers were observed in the spatial-temporal trajectory distributions. The denoised ship latitude and longitude distributions shown in Figures 5(a) and 5(b) confirmed the above analysis. It is observed that the denoised ship longitude varied from -89.6° to -89.8° , and the latitude data varied from 27.8° to 28° . It can be inferred that the ship travelled in an area with a radius about 2 km, and thus the ship was indeed in mooring state. We observed that ship trajectory samples were not evenly distributed considering that many data discontinuities are found in Figure 5. To alleviate such discontinuity, we have normalized the ship trajectory samples which are shown in Figure 6. The raw denoised ship trajectories were interpolated into evenly distributed data series, and discontinuous data samples were successfully removed (see Figures 6(a) and 6(b), respectively).

The ship trajectory reconstruction results were further evaluated by ship trajectory prediction accuracy, which can be found in Table 1 (note that our proposed framework is denoted as DANN). We have implemented another popular trajectory prediction model (i.e., the long short-term memory model (abbreviated as LSTM)) [32] for the purpose of prediction performance comparison. From the perspective of MAE, the longitude error of our proposed framework was approximately one-tenth to that of LSTM, which is 3.94×10^{-3} . The latitude MAE obtained by our model is 2.07×10^{-3} , which is about 1% to that of the LSTM counterpart. The RMSE indicators for the longitude and latitude obtained by the proposed framework were both 5.41×10^{-4} , which showed similar variation tendency to those of MAE. Both the FD and AED indicators demonstrated distance between predicted and ground truth data samples, which showed similar tendency to those of MAE and RMSE.

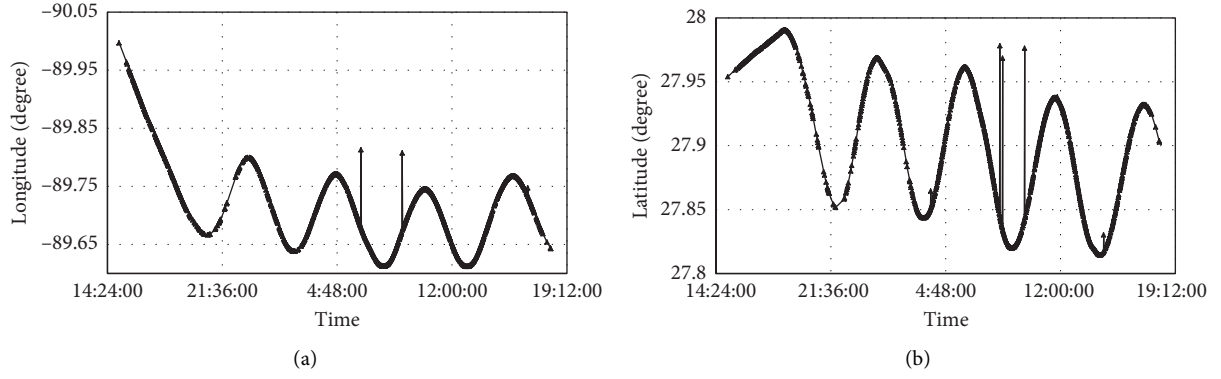


FIGURE 4: Raw ship trajectory distributions via AIS data for Case 1. (a) Raw longitude data distribution. (b) Raw latitude data distribution.

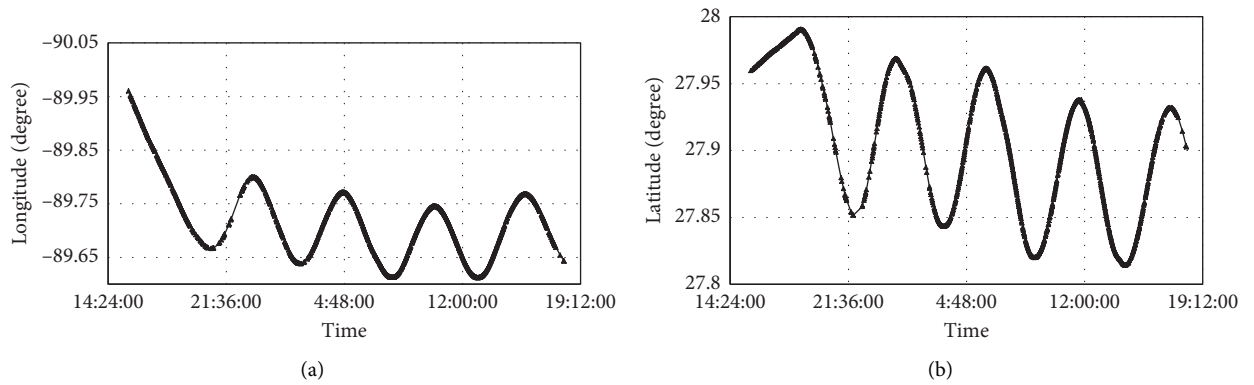


FIGURE 5: Denoised ship trajectory distributions via AIS data for Case 1. (a) Denoised longitude data distribution. (b) Denoised latitude data distribution.

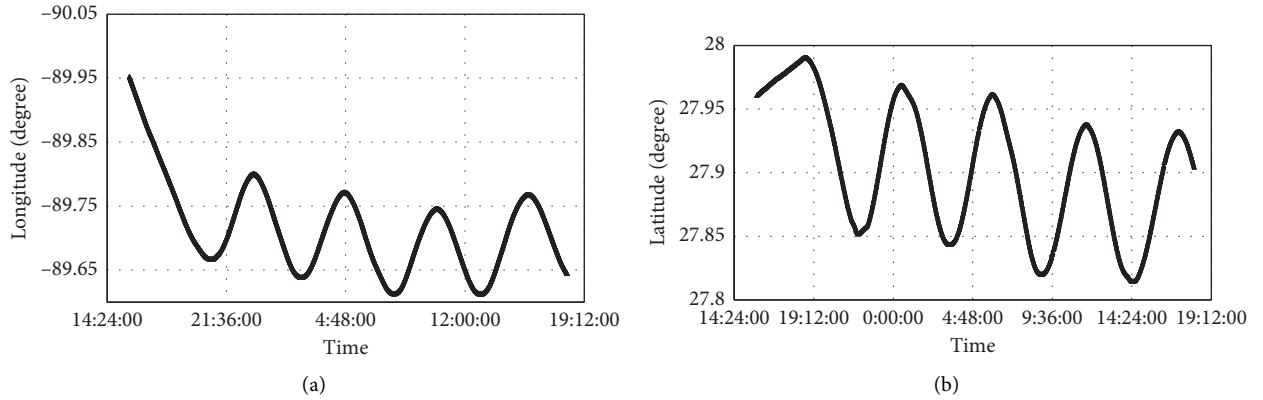


FIGURE 6: Normalized ship trajectory distributions via AIS data for Case 1. (a) Normalized longitude data distribution. (b) Normalized latitude data distribution.

TABLE 1: Ship trajectory prediction performance results for Case 1.

	DANN	LSTM
MAE (LON)*	3.94×10^{-3}	2.00×10^{-2}
MAE (LAT)	2.07×10^{-3}	2.49×10^{-1}
RMSE (LON)	5.41×10^{-4}	6.69×10^{-4}
RMSE (LAT)	5.41×10^{-4}	1.72×10^{-2}
FD	9.83×10^{-2}	3.05×10^{-1}
AED	5.16×10^{-3}	2.51×10^{-1}

*Note. MAE (LON) is the MAE value for longitude data, and the rule is applicable to MAE (LAT), RMSE (LON), and RMSE (LAT).

4.2. Ship Trajectory Reconstruction on Case 2. The model performance of the proposed framework was validated on another trajectory with ship MMSI No. 367715380. We applied the same procedure to improve the AIS data quality, and thus ship trajectory prediction was further implemented. We observed several sudden variations in both latitude and longitude data samples (see Figures 7(a) and 7(b), respectively). Moreover, the maximum distance between neighboring ship latitudes was over 1000 km, which is quite impossible in real world. Figures 7(c) and

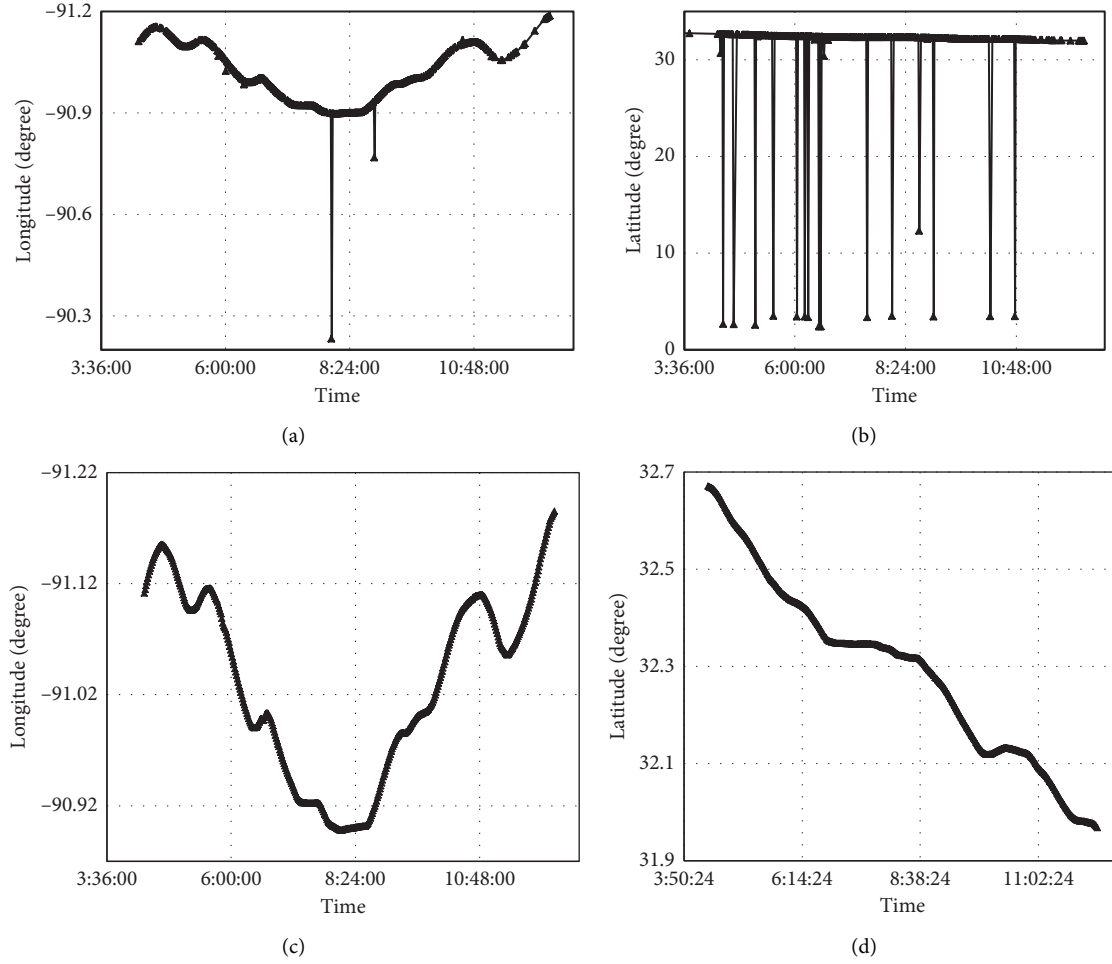


FIGURE 7: Ship trajectory distributions via AIS data for Case 2. (a) Raw longitude data distribution. (b) Raw latitude data distribution. (c) Normalized longitude data distribution. (d) Normalized latitude data distribution.

TABLE 2: Ship trajectory prediction performance results for Case 2.

	DANN	LSTM
MAE (LON)	2.23×10^{-3}	1.64×10^{-1}
MAE (LAT)	2.60×10^{-3}	1.94×10^{-1}
RMSE (LON)	8.10×10^{-4}	1.17×10^{-2}
RMSE (LAT)	8.14×10^{-4}	2.35×10^{-2}
FD	3.46×10^{-2}	4.40×10^{-1}
AED	4.25×10^{-3}	2.82×10^{-1}

7(d) demonstrated the AIS data after implementing the data quality control procedure. It is demonstrated that ship trajectory outliers were successfully removed by our proposed framework. The ship trajectory prediction results indicated that our proposed framework obtained higher accuracy compared to the LSTM model. For instance, the MAE indicators for the DANN latitude and longitude were 2.23×10^{-3} and 2.60×10^{-3} , respectively, which were both approximately one-tenth to those of the LSTM counterparts (see Table 2). The RMSE, FD, and AED indicators showed similar variation tendency to those of the MAE.

5. Conclusion

It is not easy to obtain accurate ship trajectory information via historical AIS data due to unpredicted noises. We proposed a novel framework by integrating steps of data denoising, data normalization, and trajectory prediction. The proposed framework firstly identified different ship trajectories via time interval between neighboring data samples, which is the first substep in the data quality control procedure of the proposed framework. Then, the data outliers in the raw AIS data were determined with a group of constraints, which were further corrected by the moving average method. After that, the denoised AIS data were normalized into data samples for the purpose of ship trajectory analysis applications. Then, we predicted ship trajectory with the ANN model for the purpose of further evaluating model performance. The experiments were implemented on two ship trajectories (i.e., typical outliers were observed in the raw data). The statistical results showed that our proposed framework can successfully remove abnormal AIS data outliers and obtained satisfying ship trajectory prediction performance (i.e., the average MAE,

RMSE, FD, and AED are 2.71×10^{-3} , 6.77×10^{-4} , 6.65×10^{-2} , and 4.71×10^{-3}).

We can expand our work by conducting further studies in the following aspects. First, we applied our proposed framework to cleanse and predict ship trajectories on the AIS data from two special-purpose ships, which is more challenging due to the irregular and unpredicted spatial-temporal movements. In future, we can employ the proposed framework to denoise and forecast ship trajectories for general-purpose merchant ships (e.g., oil tankers and container ships) to further testify the model performance. Second, it is noted that the ANN module in the proposed framework may suffer from the overfitting disadvantage, which may degrade the model performance. We can employ additional bio-inspired models to further enhance the model prediction accuracy. Third, we can obtain more holistic model performance by comparing it against other popular ship trajectory prediction models. Fourth, we can test the model robustness on the AIS data collected under more complicated navigation environment interferences (e.g., ship sailing at narrow and busy channels). Last but not least, we can implement maritime situation awareness task (e.g., ship behavior analysis and prediction) by exploiting the obtained historical AIS data.

Data Availability

All data generated or analyzed during this study can be obtained from the corresponding author upon request by email.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Conceptualization was performed by Xinqiang Chen and Jun Ling; methodology was contributed by Xinqiang Chen, Jun Ling, and Yongsheng Yang; original draft preparation was done by Xinqiang Chen, Jun Ling, Hailin Zheng, and Yong Xiong; review and editing were carried out by Yongsheng Yang, Pengwen Xiong, Octavian Postolache, and Yong Xiong; funding acquisition was done by Yongsheng Yang and Pengwen Xiong.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (51709167, 51579143, and 61663027) and Shanghai Committee of Science and Technology, China (18040501700, 1829501100, and 17595810300).

References

- [1] R. Zhen, Y. Jin, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier," *Journal of Navigation*, vol. 70, no. 3, pp. 648–670, 2017.
- [2] J. Xue, P. H. A. J. M. Van Gelder, G. Reniers, E. Papadimitriou, and C. Wu, "Multi-attribute decision-making method for prioritizing maritime traffic safety influencing factors of autonomous ships' maneuvering decisions using grey and fuzzy theories," *Safety Science*, vol. 120, pp. 323–340, 2019.
- [3] Y. Zhou, W. Daamen, T. Vellinga, and S. Hoogendoorn, "Review of maritime traffic models from vessel behavior modeling perspective," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 323–345, 2019.
- [4] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust ship tracking via multi-view learning and sparse representation," *Journal of Navigation*, vol. 72, no. 1, pp. 176–192, 2019.
- [5] X. Chen, X. Xu, Y. Yang, H. Wu, J. Tang, and J. Zhao, "Augmented ship tracking under occlusion conditions from maritime surveillance videos," *IEEE Access*, vol. 8, no. 1, pp. 42884–42897, 2020.
- [6] Z. Fang, L. Jian-yu, T. Jin-jun, W. Xiao, and G. Fei, "Identifying activities and trips with GPS data," *IET Intelligent Transport Systems*, vol. 12, no. 8, pp. 884–890, 2018.
- [7] X. Chen, Y. Yang, S. Wang et al., "Ship type recognition via a coarse-to-fine cascaded convolution neural network," *Journal of Navigation*, vol. 73, no. 4, pp. 813–832, 2020.
- [8] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, "High-resolution vehicle trajectory extraction and denoising from aerial videos," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.
- [9] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, 2013.
- [10] X. Chen, L. Qi, Y. Yang et al., "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.
- [11] X. Chen, "Robust visual ship tracking with an ensemble framework via multi-view learning and wavelet filter," *Sensors (Basel)*, vol. 20, no. 3, 2020.
- [12] Y. Zou, X. Zhong, J. Tang et al., "A copula-based approach for accommodating the underreporting effect in wildlife–vehicle crash analysis," *Sustainability*, vol. 11, no. 2, p. 418, 2019.
- [13] R. J. Bye and P. G. Almklov, "Normalization of maritime accident data using AIS," *Marine Policy*, vol. 109, Article ID 103675, 2019.
- [14] C. Iphar, C. Ray, and A. Napoli, "Data integrity assessment for maritime anomaly detection," *Expert Systems with Applications*, vol. 147, Article ID 113219, 2020.
- [15] L. Zhao and G. Shi, "Maritime anomaly detection using density-based clustering and recurrent neural network," *Journal of Navigation*, vol. 72, no. 4, pp. 894–916, 2019.
- [16] J. Tang, F. Gao, F. Liu, and X. Chen, "A denoising scheme-based traffic flow prediction model: combination of ensemble empirical mode decomposition and fuzzy C-means neural network," *IEEE Access*, vol. 8, pp. 11546–11559, 2020.
- [17] J. Liu, H. Li, Z. Yang, K. Wu, Y. Liu, and R. W. Liu, "Adaptive douglas-peucker algorithm with automatic thresholding for AIS-based vessel trajectory compression," *IEEE Access*, vol. 7, pp. 150677–150692, 2019.
- [18] F. Deng, "Vessel track information mining using AIS data," in *Proceedings of the International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, Beijing, China, September 2014.
- [19] S.-K. Zhang, G.-Y. Shi, Z.-J. Liu, Z.-W. Zhao, and Z.-L. Wu, "Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity," *Ocean Engineering*, vol. 155, pp. 240–250, 2018.

- [20] H. Rong, A. P. Teixeira, and C. Guedes Soares, "Ship trajectory uncertainty prediction based on a Gaussian Process model," *Ocean Engineering*, vol. 182, pp. 499–511, 2019.
- [21] L. Du, F. Goerlandt, and P. Kujala, "Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data," *Reliability Engineering & System Safety*, vol. 200, Article ID 106933, 2020.
- [22] X. Hoque and S. K. Sharma, "Ensembled deep learning approach for maritime anomaly detection system," in *Proceedings of the ICETIT 2019*, Springer International Publishing, Cham, Switzerland, 2020.
- [23] K. I. Kim and K. M. Lee, "Deep learning-based caution area traffic prediction with automatic identification system sensor data," *Sensors (Basel)*, vol. 18, no. 9, 2018.
- [24] M. Gao and G.-Y. Shi, "Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms," *Ocean Engineering*, vol. 205, p. 106919, 2020.
- [25] L. Zhang, Q. Meng, Z. Xiao, and X. Fu, "A novel ship trajectory reconstruction approach using AIS data," *Ocean Engineering*, vol. 159, pp. 165–174, 2018.
- [26] X. Chen, "Anomaly detection and cleaning of highway elevation data from google earth using ensemble empirical mode decomposition," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 5, Article ID 04018015, 2018.
- [27] X. Chen, "Traffic flow prediction at varied time scales via ensemble empirical mode decomposition and artificial neural network," *Sustainability*, vol. 12, no. 9, 2020.
- [28] R. J. Bye and A. L. Aalberg, "Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports," *Reliability Engineering & System Safety*, vol. 176, pp. 174–186, 2018.
- [29] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, 2018.
- [30] V. F. Arguedas, G. Pallotta, and M. Vespe, "Maritime Traffic Networks: from historical positioning data to unsupervised maritime traffic monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 722–732, 2017.
- [31] H. Behjat, Z. Dogan, D. Van De Ville, and L. Sornmo, "Domain-informed spline interpolation," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 3909–3921, 2019.
- [32] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38287–38296, 2019.