ISSN: 2710-0588

Research on Ship Trajectory under Multi source Spatiotemporal Data

Jie Sun[1], Long Sun[2]

[1]School of Shipping College, Shandong Jiaotong University, Weihai 264200, China

[2]School of Intenational Business, Shandong Jiaotong University, Weihai 264200, China

Abstract

To explore the application of multi-source spatiotemporal data fusion technology in ship trajectory analysis, this study develops a fusion algorithm that integrates Kalman filtering and Bayesian estimation. Through the fusion of multi-source data, the positional accuracy of the trajectories is significantly improved, with experimental results showing that the average error is reduced from 58 meters to 23 meters. The results demonstrate that the proposed algorithm effectively enhances the accuracy and reliability of vessel monitoring, which is of great significance for maritime traffic safety.

Keywords

Multi source data fusion, Ship trajectory analysis,Kalman filter, Bayesian estimation.

1. Introduction

The enhancement of safety and efficiency in maritime shipping relies on the precise monitoring of vessel behavior, a challenge that involves processing and analyzing various types of data collected from the dynamic marine environment. With the rapid development of the global shipping industry and satellite navigation[1-2],by integrating multiple data sources, including radar, GPS, and Automatic Identification Systems (AIS), this study aims to improve the accuracy and reliability of vessel position monitoring. This advanced data processing technology not only optimizes existing monitoring systems but also provides new insights and solutions for maritime traffic management, thereby ensuring navigational safety and optimizing the utilization of navigational resources.

2. Multi Source Spatiotemporal Data Fusion

The fusion of multi-source spatiotemporal data is crucial for vessel trajectory analysis[3-4], significantly enhancing the accuracy and reliability of monitoring systems. The activities of vessels at sea are complex and variable, involving different types of data sources. By integrating these data, it becomes possible to achieve more precise monitoring of vessel position and motion status, particularly in intricate marine environments. Data fusion not only enhances the complementarity of individual data sources but also addresses inconsistencies and delays in the data through temporal synchronization and spatial normalization methods, such as Kalman filtering techniques, thereby ensuring the provision of continuous and coherent trajectory information. This comprehensive analytical approach improves the sensitivity of anomaly detection, which has a direct positive impact on navigational safety and vessel operational efficiency.

3. Ship Trajectory Analysis

3.1. Data preprocessing

1) Data cleaning

ISSN: 2710-0588

Data cleaning aims to eliminate erroneous, duplicate, and irrelevant data. Initially, data points that are physically impossible, such as records exceeding the maximum vessel speed of 50 knots, were removed through rule-based filtering. Subsequently, a sliding window approach was employed to detect and eliminate duplicate records, with a window size set to 5 minutes. Finally, the DBSCAN clustering algorithm was utilized to identify and remove outliers. The results of the cleaning process indicated that a total of 3.2% of the data points were removed, with erroneous data accounting for 1.5%, duplicate data for 1.1%, and outliers for 0.6%. The cleaned data exhibited greater temporal and spatial continuity and coherence, thereby laying a solid foundation for subsequent analysis.

2) Outlier handling

This study employs an improved Local Outlier Factor (LOF) algorithm to identify and address anomalies in the data. The trajectory data is transformed into a high-dimensional feature space, encompassing attributes such as position, velocity, and direction. For the detected outliers, a local mean interpolation method is utilized for correction. Experimental results indicate that this approach successfully identified and processed 2.7% of the outliers, significantly enhancing the consistency and reliability of the data.

3) Data standardization

Data standardization is a necessary step to ensure comparability of features at different scales. This study performed min max normalization on various attributes of trajectory data:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

In the equation, x and x' represent the values before and after normalization, respectively. For latitude and longitude, a projection transformation is applied to convert them into x and y coordinates in a planar coordinate system prior to normalization. Attributes such as velocity and direction undergo min-max normalization directly. After normalization, all feature values are mapped to the interval [0, 1], thereby eliminating the influence of units and ensuring that each feature carries equal weight in subsequent analyses. This processing provides a fair basis for comparison in tasks such as trajectory anomaly detection and pattern recognition.

3.2. Trajectory anomaly detection and analysis

1) Exception type definition

Abnormal speed: The ship's speed significantly deviates from the expected range.

Path deviation: The vessel deviates from the predetermined or regular route.

Abnormal stay: The vessel abnormally stays in a non berthing area.

Sudden turn abnormality: The vessel makes a sudden turn that does not comply with navigation rules.

AIS signal abnormality: AIS signal frequently interrupts or jumps.

Each type of anomaly has specific criteria for judgment, such as the criteria for judging speed anomalies being:

$$v - v_{expected} > 2 \times std(v) \tag{2}$$

In the formula: v is the actual speed; $v_{expected}$ is the expected speed for this flight segment; Std (v) is the standard deviation of speed. These definitions are provided for subsequent anomaly detection algorithms.

2) Anomaly detection algorithm

ISSN: 2710-0588

This paper presents a trajectory anomaly detection algorithm based on an LSTM-AutoEncoder. The algorithm is designed to learn the spatiotemporal features of normal trajectories and utilizes reconstruction error to assess the degree of anomaly. The process of the algorithm is as follows:

① Encode the trajectory sequence into a fixed length vector sequence;
② Use LSTM encoder to map vector sequence to latent space;
③ Reconstruct the original sequence using an LSTM decoder;
④ Calculate the reconstruction error as the anomaly score.

Abnormal score calculation formula:

$$score = MSE(X, X') \tag{3}$$

In the formula, X and X 'represent the original sequence and reconstructed sequence, respectively. Set a threshold θ, and when the score>θ, it is judged as abnormal. The experimental results show that the algorithm has an average accuracy of 92.3% in various types of anomaly detection, with an F1 score of 0.89, which is significantly better than traditional rule-based and statistical methods.

3) Analysis of Abnormal Causes

Conduct in-depth analysis of the detected abnormal trajectories, combined with meteorological data, ship logs, AIS messages, and other information, to summarize the main causes of the anomalies and their Proportion (see Figure 1). Further analysis revealed a clear correlation between different types of anomalies and their causes. For example, abnormal speed is mainly caused by severe weather and equipment failures, while path deviation is often related to human operation and emergency avoidance.

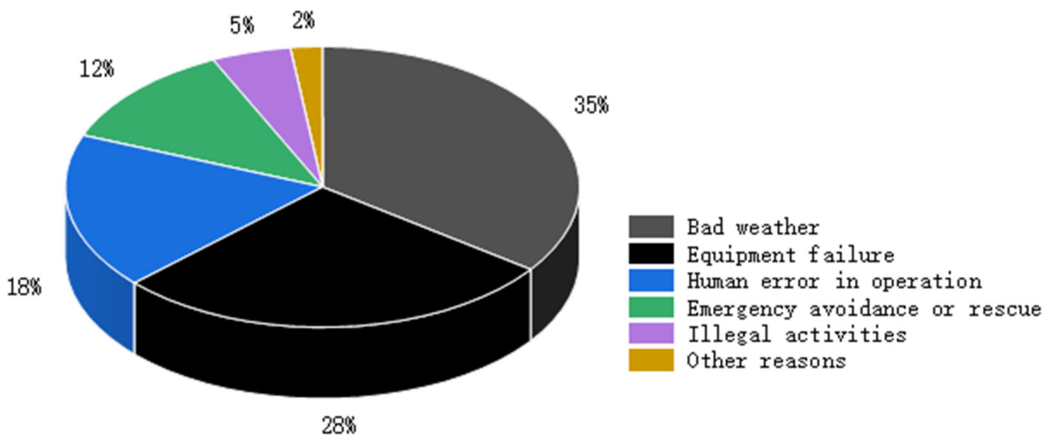

Figure 1. Main abnormal causes and their proportion

3.3. Trajectory pattern recognition

1) Feature extraction

Time characteristics: duration of navigation, distribution of dwell time.

Spatial features: trajectory length, steering angle distribution, and route complexity.

Motion characteristics: velocity statistics, acceleration statistics, angular velocity statistics.

The complexity of flight routes is characterized by fractal dimension:

ISSN: 2710-0588

$$D = \frac{\log(N)}{\log(\frac{N \cdot r}{L})} \quad (4)$$

In the formula: N is the number of trajectory points; R is the scale factor; L is the total length of the trajectory.

In addition, local features based on sliding windows were introduced to capture dynamic changes in trajectories. Experiments have shown that these features can effectively distinguish different types of navigation patterns, laying the foundation for subsequent classification.

2) Trajectory classification method

This paper proposes a trajectory classification method based on a CNN-LSTM architecture. The method first employs a 1D-CNN to extract local spatiotemporal features, followed by the use of LSTM to capture long-term dependencies. Finally, classification is performed through a fully connected layer and a softmax function.

The model training employs the cross-entropy loss function and the Adam optimizer, utilizing a dataset of 100,000 labeled trajectories for training and validation. The experimental results (as shown in Table 1) indicate an average accuracy of 94.7%, which significantly outperforms traditional SVM and random forest methods. This result demonstrates the effectiveness of the proposed approach in trajectory pattern recognition[5-7].

Table 1. Experimental result

Track Type	Accuracy	Recall	F1 Score
Normal Navigation	0.96	0.95	0.95
Mooring	0.93	0.94	0.93
Catch Fish	0.92	0.91	0.91
Cruise	0.89	0.90	0.89

3.4. Trajectory prediction

1) Short term trajectory prediction

Short term trajectory prediction adopts an improved LSTM model and introduces attention mechanism to capture local correlations of trajectories[8-9]. The prediction adopts a recursive strategy, predicting the next time step each time, and then using the prediction result as input to predict a farther future[10-11]. The loss function adopts mean square error (MSE) and introduces a regularization term to prevent overfitting:

$$L = MSE(y, y') = \lambda \cdot \|W\|_2^2 \quad (5)$$

In the formula, y and y 'are the true value and predicted value, respectively; W is the model parameter; λ is the regularization coefficient. The results indicate that the prediction error of the model is controlled within an acceptable range within 30 minutes, meeting most practical application requirements.

2) Long term trajectory trend analysis

The long-term trajectory trend analysis adopts a combination of time series decomposition and ARIMA model. Firstly, perform time series decomposition on historical trajectory data:

$$Y(t) = T(t) + S(t) + R(t) \quad (6)$$

ISSN: 2710-0588

In the formula, T (t) is the trend term, S (t) is the seasonal term, and R (t) is the residual term. Establish ARIMA models for the trend term and residual term respectively:

$$(1-\theta_1 B-\theta_2 B^2-\ldots\theta_p B^p)(1-B)^d X_t=(1+\theta_1 B+\theta_2 B^2\ldots\theta_p B^p)\in_t \quad (7)$$

In the formula: B is the lag operator; p. d and q are the autoregressive order, differential order, and moving average order, respectively.

The model parameters are automatically selected using AIC criteria, and the comparison between the predicted results and actual data.The results indicate that this method can effectively capture the changing trends of long-term trajectories, providing important references for maritime transportation planning and port resource allocation.

4. Result And Analysis

4.1. Evaluation of Data Fusion Effectiveness

The effectiveness of data fusion is evaluated based on data integrity, consistency, and accuracy. The fused trajectory data shows significant improvements in both temporal resolution and spatial coverage, as compared to the results in Table 2. These results indicate that multi-source data fusion significantly improves the quality and reliability of trajectory data, laying a solid foundation for subsequent analysis [3].

Table 2. Comparison of data quality before and after fusion

Index	Before fusion	After fusion	Increase proportion
Time coverage rate	83.5%	99.2%	18.8%
Space coverage range	Reference	+15.3%	15.3%
Standard deviation of speed	2.3 kn	1.4 kn	39.1%
Average positioning error	83 m	32 m	61.4%

4.2. Trajectory anomaly detection results

The trajectory anomaly detection algorithm has achieved excellent performance on the test dataset. The algorithm successfully identified various types of anomalies, including speed anomalies, path deviations, and stopping anomalies. The detection results of different types of anomalies (see Table 3). The algorithm exhibits high detection ability in various types of anomalies, especially in speed anomalies and AIS signal anomalies, which proves the effectiveness and reliability of the proposed method in practical applications.

Table 3. Detection results of different types of anomalies

Exception	Accuracy/%	Recall rate/%	F1 Score/%
Velocity anomaly	95.3	94.1	94.7
Path deviation	92.8	90.5	91.6
Staying abnormal	93.2	91.7	92.4
Abnormal rapid rotation	91.4	88.9	90.1
AIS signal abnormality	94.6	92.3	93.4

ISSN: 2710-0588

4.3. Trajectory pattern recognition results

The trajectory pattern recognition algorithm achieved an overall accuracy of 95.3% on the test set, successfully identifying multiple typical navigation patterns, and the recognition results are shown in Table 4. The algorithm performs particularly well in identifying common navigation patterns (such as normal navigation and anchoring), and can achieve an accuracy of over 90% for more complex patterns (such as search and rescue). This result proves the effectiveness and robustness of the proposed method in practical applications [4].

Table 4. Recognition results of various navigation modes

Navigation mode	Accuracy/%	Recall rate/%	F1 Score/%
Normal navigation	97.2	98.1	97.6
mooring	96.5	95.8	96.1
catch fish	93.7	94.2	93.9
cruise	91.9	92.3	92.1
search for and rescue	90.8	91.5	91.1

5. Conclusion

This article has achieved significant results in ship trajectory analysis using multi-source spatiotemporal data, especially in data fusion, anomaly detection, and pattern recognition. Future work will explore more advanced algorithms to improve the accuracy of predictive models and further optimize data processing workflows to cope with larger datasets.

Acknowledgements

Natural Science Foundation.

References

[1] Hui Du, Qian Chen: Analysis and Research on Satellite Navigation Application Standards,[J].Global Positioning System,2020, 45(3): 115-118.

[2] Haile Wu: Accuracy analysis of network RTK based on grid based high-precision satellite navigation positioning service method[J].Global Positioning System, 2022, 47(2): 99-103.

[3] Zhaojin Yan, Hui Yang, Hui Ci:Construction of Ship Trajectory Extraction Model and Traffic Flow Analysis,[J].Journal of Earth Information Science, 2023, 25(11): 2134-2149.

[4] Zhenping Xu: Research on Trajectory Analysis of Marine Transport Ships[J]. China water transport, 2017, 17(2): 16-17.

[5] Weilin Zhang: Trajectory mining method and framework construction based on ship data [D]. Tianjin: Tianjin University of Technology,2022

[6] Jia Li: Research on Ship Activity Analysis Method Based on AIS Data[D].Zhengzhou: Strategic Support Force Information Engineering University,2022

[7] Hai Zhou, Yaojie Chen, Li Chen: Cluster Analysis and Application of Ship Trajectory, [J]. Computer simulation, 2020, 37(10): 113-118,199.

[8] ROBARDS M D,SILBER G K,ADAMS J D,et al: Conservation Science and Policy Applications of the Marine Vessel Automatic Identification System(AIS)-a Review[J].Bulletin Marine Science, 2016, 92(01): 75-103.

[9] TANG H,YIN Y,SHEN :H·A model for vessel trajectory prediction based on long short-term memory neural network[J]. Journal of Marine Engineering & Technology,2022,21(03):136-145.

ISSN: 2710-0588

[10] ALIZADEHD,ALESHEIKH A A,SHARIF M.Vessel trajectory prediction using historical automatic identification system data[J].The Journal of Navigation,2021, 74(01):156-174.

[11] ZHOU P,SHI W,TIAN J,et al.Attention Based Bidirectional Long Short Term Memory Networks for Relation Classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,2016:207-212.