International Core Journal of Engineering Volume 11 Issue 4, 2025
ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0057

Improving the Accuracy of AIS Data Missing Value Imputation Using an Enhanced Interpolation Method

Qian Gao, Shaoyi Guo

School of Shipping, Shandong Jiaotong University, Weihai 264200, China

Abstract

During navigation, vessels continuously transmit Automatic Identification System (AIS) data, which contains a wealth of information. However, raw AIS data is often disorganized and may contain anomalies or missing values, making it difficult to directly track navigation trajectories. To make these data usable and ensure their integrity while reducing subsequent trajectory prediction errors, it is essential to preprocess the raw AIS data. This involves extracting and constructing a dataset of vessel navigation trajectories using the latitude and longitude of vessels entering and leaving ports, as well as the latitude, longitude, speed, and navigation status of AIS trajectory points. Subsequently, the extracted navigation trajectory data is preprocessed to remove anomalies. Finally, an enhanced hybrid interpolation method is employed to impute missing trajectory data, thereby improving the accuracy of interpolation.

Keywords

Enhanced Hybrid Interpolation Method; Interpolation; AIS Data Preprocessing.

1. Introduction

With the increasing number of vessels equipped with AIS systems, the mechanism for sharing information among ships has become more sophisticated, leading to a rapid increase in the volume of AIS data. AIS systems can collect millions of AIS data points from vessels, which contain rich maritime navigation characteristics. Through in-depth analysis, valuable information can be extracted[1]. However, due to limitations in AIS system equipment and data collection environments, raw AIS data often has varying degrees of missing values. Untreated missing data can lead to significant loss of important information. Therefore, it is crucial to repair these missing data to ensure data integrity. Zhou et al[2]. proposed a method for identifying erroneous AIS data and used an improved cubic spline interpolation method to repair AIS data. Liu et al[3]. proposed a trajectory repair method combining cubic spline interpolation and Vondrak filtering. This method first uses Vondrak filtering to smooth the trajectory data and then employs cubic spline interpolation to repair missing trajectory data. Experimental results showed that this method significantly improved the repair of missing trajectories. Guo et al[4] . proposed a trajectory repair method based on ship kinematics, which analyzes the navigation time area of the trajectory and combines ship motion characteristics to repair and reconstruct the vessel's trajectory. Experimental validation demonstrated that this method performed well in specific waters. Although previous studies have made progress in handling missing vessel trajectories, current methods typically use a uniform repair strategy for different parts of the same trajectory. This approach often leads to significant repair errors when dealing with trajectories with large variations. Therefore, this paper proposes a new hybrid interpolation repair method that segments the trajectory based on its variation type and applies different interpolation techniques to different types of trajectory segments to improve the accuracy of interpolation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0057

2. Data

2.1 Data Cleaning

AIS data includes three main categories of information: dynamic data, static data, and navigation data. Dynamic data primarily involves key information such as the latitude, longitude, heading, and speed of the vessel. Static data includes detailed information such as the Maritime Mobile Service Identify (MMSI), flag state, and deadweight tonnage. Navigation data focuses on critical indicators such as draft depth and navigation status.

Raw AIS data sometimes contains outliers, and the purpose of data cleaning is to identify and correct these anomalies to enhance the overall quality of the data. This provides a solid foundation for subsequent clustering analysis and predictive modeling, thereby reducing prediction errors. In this study, the following cleaning steps were performed for data with obvious errors:

Sort the AIS data by MMSI number to ensure that data for the same MMSI is arranged in chronological order.

1) Remove records with MMSI numbers that are not nine digits long.

2) Delete duplicate data.

3) Exclude records with ship lengths less than 90 meters.

After removing erroneous AIS data, although the correct AIS sequence was obtained, some data may still be missing and need to be repaired to obtain complete and accurate data.

2.2 Data Missing

AIS data missing refers to the situation where the time interval between adjacent trajectory points is long and the latitude and longitude change significantly. The causes of trajectory data missing are mainly as follows:AIS transmission is unreliable, and data loss can occur during transmission.AIS equipment failure or intentional shutdown of the AIS device, resulting in no AIS data being sent for a period of time.During AIS data cleaning, some data is removed due to anomalies, leading to missing AIS data.

For trajectories with missing data, if the missing time exceeds one-sixth of the total trajectory time, that segment of the trajectory is discarded and not used as experimental data. If it does not exceed one-sixth, interpolation repair is performed. The time interval between trajectory data points is set as a threshold. When it exceeds the set threshold, interpolation repair is needed. The repaired data is interpolated with one missing value every 15 minutes. If it is less than 15 minutes but more than 8 minutes, one value is interpolated. If it is less than 15 minutes and not more than 8 minutes, no value is interpolated.

3. Improved Interpolation Method

Given that trajectories typically cover vast areas and have varied shapes, a single interpolation method often fails to accurately correct errors. Therefore, this paper proposes a segmented composite interpolation repair strategy based on trajectory type. This strategy divides trajectories into two types: straight-line navigation and curved navigation. Specifically, if the absolute value of the heading difference between each of the three consecutive trajectory points before and after the interpolation point exceeds 8 degrees, it is classified as a curved trajectory; otherwise, it is considered a straight-line trajectory.

3.1 Curved Trajectory Repair

When trajectory data is severely missing, these data points are not used for experimental purposes. However, for trajectories with only short segments of missing data, cubic spline interpolation has been proven to be effective in repairing these small segments of curved trajectories. Therefore, we choose to use cubic spline interpolation to repair these missing curved trajectories.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0057

The core of cubic spline interpolation is a solution process that involves solving a three-bend moment equation system to obtain a set of curve functions, ultimately generating a smooth curve that passes precisely through a series of given shape value points[5].

If a function S(x) satisfies the following conditions:

$\mathcal{S}(x)$is a polynomial of degree not exceeding three on each subinterval $[x_i - 1, x_i]$(where $i = 1,2,\ldots,n$),

$\mathcal{S}(x)$is a twice continuously differentiable function on the interval [a,b].

$\mathcal{S}(x)$satisfies $\mathcal{S}(x_i) = y_i$ at each node.

If the dataset $[\ t_0, t_k\ ]$has nodes that require repair. $t_0 = x_1 < x_2 <, \ldots, < x_k = t_k$, x_k The corresponding function p_k,The solution method is as follows:

Input the k interpolation nodes of the data set. $t_0 = x_1 < x_2 <, \ldots, < x_k = t_k$For the function values$p_1, p_2, \ldots, p_k$, when the adjustment condition is $p_1'' = p_k''$=0, the interpolation points x_0 are sought.

Compute:

$h_i = x_i - x_{i-1}(i = 1,2,\ldots,k-1)$

$u_i = \frac{h_{i-1}}{h_{i-1}+h_i}$, $d_i = 6\left(\frac{p_{i+1}-p_i}{h_i} - \frac{p_i - p_{i-1}}{h_{i-1}}\right)\frac{1}{h_{i-1}+h_i}$, $(i = 1,2,\cdots,k-1)$

$\alpha_1 = \frac{6}{h_1}\left(\frac{p_2 - p_1}{h_1} - p_1'\right)$, $\alpha_k = \frac{6}{h_{k-1}}\left(p_k' - \frac{p_k - p_{k-1}}{p_{k-1}}\right)$.

Solve the system of equations using the Thomas algorithm (also known as the tridiagonal matrix algorithm).

Output $\mathcal{S}(x_i)$ for each interval.

Outputx_0Interval$[x_i - 1, x_i]$,compute the interpolation p_0.

3.2 Straight-Line Trajectory Repair

For the repair of straight-line navigation trajectories, linear interpolation is effective. This paper employs linear interpolation for this purpose. Linear interpolation approximates the navigation trajectory formed by AIS data as a linear trajectory, similar to a linear function, with the interpolation function being a first-degree polynomial[6]. The missing segment of the trajectory sequence is connected by a straight line between the two data points before and after the missing segment. The missing values are fitted using the values of the two trajectory points before and after the missing segment.Let the missing data be $y(t)$, where t represents the time of the missing data, and $y(t)$ represents the missing data at time t. t_1 and t_2 represent the times before and after t, and $y(t_1)$ and $y(t_2)$ represent the corresponding data at these times. The data points $(t_1, y(t_1))$ and $(t_2, y(t_2))$ are linearly fitted, and the interpolated value at time t can be calculated using the following formula:

$$y_t = y(t_1) + \frac{y(t_2) - y(t_1)}{t_2 - t_1}(t - t_1)$$

4. Experimental Results and Analysis

To verify the proposed repair method, data from the AIS database of the vessel SOLSTICE (MMSI 256286000, vessel type B) was selected for the period from 10:00 to 22:00 on July 28, 2024, near 21.26977°N, 158.05162°W. A total of 219 data points were used. The original data contained some errors, which were corrected, resulting in 181 data points. The proposed method was then applied to improve the data through interpolation. The results are shown in the following figure.See figure 1and 2, Original Data and Interpolated Data.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0057

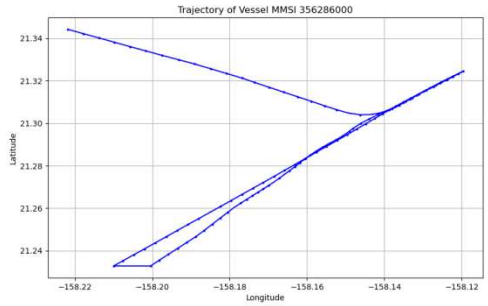

Figure 1. Original Data

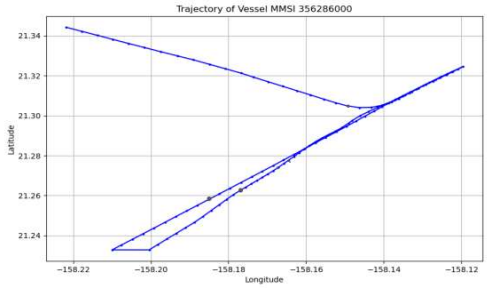

Figure 2. Interpolated Data.

5. Conclusion

This paper employs a combination of cubic spline interpolation and linear interpolation to repair data. Specifically, cubic spline interpolation is first used to smooth the data, generating a smooth curve that passes through all known data points. Experimental results demonstrate that this method effectively repairs the missing trajectory points of the vessel SOLSTICE (MMSI 256286000), with the repaired trajectory closely matching the actual trajectory. This not only validates the effectiveness of the hybrid interpolation method but also provides high-quality data support for subsequent vessel trajectory analysis and prediction. Future research will explore more types of interpolation methods and combine them with machine learning algorithms to further improve the accuracy and efficiency of AIS data repair, thereby providing stronger technical support for maritime traffic management and vessel navigation safety.

References

[1] Zhen R, Shao Z P, Pan J C. Research progress and prospects on ship behavior feature mining and prediction based on AIS data[J]. Journal of Earth Information Science, 2021, 23(12): 2111-2127.

[2] Zhou C, Liu W D, Wan G H, et al. Application of spline interpolation in AIS data repair[J]. China Water Transport. Channel Technology, 2018, (04): 76-79. DOI: 10.19412/j.cnki.42-1395/u.2018.04.016.

[3] Liu L Q, Wu C Z, Chu D F, et al. Research on ship trajectory repair based on Vondrak filtering and cubic spline interpolation[J]. Traffic Information and Safety, 2015, 33(04): 100-105.

[4] Guo S, Mou J, Chen L, et al. Improved Kinematic Interpolation for AIS Trajectory Reconstruction[J]. Ocean Engineering, 2021, 234(8).

[5] Zhou Y, Mu F, Hu J. Adaptive state updating particle filter tracking algorithm based on cubic spline interpolation[C]//2021 International Conference on Electronic Information Engineering and Computer Science (EIECS). IEEE, 2021: 484-488.

[6] Noor N M, Al Bakri Abdullah M M, Yahaya A S, et al. Comparison of linear interpolation method and mean method to replace the missing values in environmental data set[C]//Materials Science Forum. Trans Tech Publications Ltd, 2015, 803: 278-281.