

基于改进 LSTM 的航运物流路径轨迹修复研究

白响恩¹, 方明权¹, 徐笑锋^{1*}, 肖英杰¹, 吴永明²

(1. 上海海事大学 商船学院, 上海 201306; 2. 宁波引航站, 浙江 宁波 315000)

摘要: **目的** 提高航运物流路径轨迹信息的挖掘精度和安全效率。**方法** 对宁波舟山港条帚门水域窄口航道船舶的类型、数量、长度进行统计分析, 对货运船 AIS 物流路径轨迹异常进行识别与修复。考虑船舶实际航行的环境因素, 提出一种新的数据纠偏方法。考虑船舶属性与环境因素, 以通航宽度和三自由度运动学的转向能力识别异常数据, 然后运用 Bi-LSTM 法对筛选后航运物流路径进行轨迹修复。**结果** 所提筛选方法不需要轨迹聚类或建立额外的模型进行判别, 筛选数据量占总数量的 34.26%, 修复后的 AIS 货船物轨迹数据量在原有基础上提升了 115.34%。**结论** 使用文中方法可以有效纠偏和修复异常航运物流路径轨迹数据, 为航运物流轨迹数据挖掘提供一定的基础方法。

关键词: 物流路径; 货运船舶; AIS 轨迹; 船舶特性; Bi-LSTM 修复

中图分类号: U675.7 文献标识码: A 文章编号: 1001-3563(2023)17-0152-08

DOI: 10.19554/j.cnki.1001-3563.2023.17.018

Ship Logistics Path Trajectory Repair Based on Improved LSTM

BAI Xiang-en¹, FANG Ming-quan¹, XU Xiao-feng^{1*}, XIAO Ying-jie¹, WU Yong-ming²

(1. Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China;

2. Ningbo Pilot Station, Zhejiang Ningbo 315000, China)

ABSTRACT: The work aims to improve the mining accuracy and safety efficiency of cargo logistics path trajectory information. In this study, the type, quantity and length of ships in the narrow channel of the strip Tiaozhoumen waters of Ningbo Zhoushan Port were statistically analyzed for the identification and repair of AIS logistics path trajectory abnormal points of cargo ships. Considering the environmental factors of actual ship navigation, a new data correction method was proposed. Considering the ship attributes and environmental factors, the abnormal data were identified with the navigation span and steering capability of the three-degree-of-freedom kinematics, and finally the Bi-LSTM method was applied to repair the screened shipping logistics path trajectory. The proposed screening method did not require trajectory clustering or building additional models for discrimination, the screened data accounted for 34.26% of the total quantity, and the repaired AIS cargo ship trajectory data were improved by 115.34% on the original basis. The method can effectively correct and repair abnormal shipping logistics path trajectory data, and provide some basic methods for shipping logistics trajectory data mining.

KEY WORDS: logistics path; cargo ship; AIS trajectory; ship characteristics; Bi-LSTM repair

海上运输是我国主要运输方式和经济发展的重
要来源, 海运运输促进了全球范围的贸易和物流发

展。对船舶物流路径轨迹分析能得出诸多影响运输
效率的因素, 其中货运物流船舶 AIS 轨迹信息有誤

是造成航运事故的重要原因之一^[1]。智能物流航运研究是世界范围内的热门话题, 可视化与数据挖掘技术的发展, 促使这项工作通过大数据分析来调查货运物流效率等问题。宁波舟山港条帚门航道是缓解交通拥堵、增加货运量的重要物流航道, 对提升货运物流效率起到重要作用, 是主要中型货船货运的主要物流航道。物流路径航迹的分析是 AIS 数据研究领域的热点, AIS 提供的信息(经度、纬度、航速、航向等)是物流运输经济、安全、效率的重要信息来源, 充分挖掘数据中的信息能有效提高海上物流运输作业安全性、效率性, 保护海洋自然生态环境。国内外专家^[2-4]对 AIS 数据进行去噪与修复后对货运物流轨迹和港口吞吐量进行了深度挖掘, 实现了可视化分析, 提高了物流效率与安全。Yang 等^[5]利用长江上海段 4 923 艘船舶的 AIS 数据估计船舶速度与拥堵程度之间的相关性, 分析了不同港口市场的航运物流量, 为测试其他航运渠道的速度限制的合理性提供了有效的安全保障。

原始 AIS 货运物流轨迹数据冗余, 在处理中既要保留原有船舶物流路径特征属性, 也要清除异常数据。AIS 异常数据的检测方法有基于统计学与数学原理的方法^[6-7]。这类专家以统计学结果和经验为判断依据, 设置固定距离作为半径范围。因此, 这类方法精确度不高且只能在分秒短时间内有效果。相似性度量的方法^[8-9]起初是设置来判断形状相似性的。海陆交通有较大差别, 即使在同一条航道内, 因为环境水流和船舶自身的船舶领域等因素形成 2 条相似的轨迹的概率也很低, 甚至完全不同, 其适用性和准确性仍需要思考。聚类与压缩的方式^[10-11]往往需要寻找代轨迹, 建立数据库等工作, 轨迹聚类会造成轨迹形状固定化, 可能不利于后续轨迹的修复和数据的挖掘。国外^[12-13]学者认为风浪流等环境因素会影响到货运轨迹, 在窄口段水域 AIS 货运物流轨迹往往存在异常和缺失 2 种情况。因此, 上述的方法只是关注明显的固定信息错误, 或将时间切割成均匀的数据集来处理, 他们很少去考虑时间和空间属性问题。多数学者往往只针对数据缺失的问题进行修复, 少有学者对物流信息的纠偏进行深入研究, 重复与错误的信息不仅不利于数据的挖掘, 也会影响机器学习算法的深度学习与预测的精确度。

修复原始货运轨迹可以帮助工作人员寻找影响船舶物流的诸多信息, 发现船舶转向、变速的原因, 同时也可以利用这些信息进行后续的工作。

航运物流轨迹具有连续性, 数据前后关联紧密, 在时间和空间上属于连续的过程。常用的方法主要包括插值法^[14-15]、改进的插值法^[16]及 K 邻近法等^[17]。这些方法依据历史条件对短时间内少量数据缺失修复精度较高, 但对长时间数据的修复精度低, 甚至无法进行轨迹修复。还有学者用基于统计学与数学原理方法进行修复, 例如高斯过程、双线性编码器等^[18-19]。

虽然弥补了长时间无法修复的问题, 但建立模型复杂, 需要确定的参数分量众多, 且难以修复速度、船首向等空间以外的动态数据。

深度学习中的神经网络在近年来尤为流行, 在船舶轨迹的重塑与预测方面拥有优秀的效果。国外有学者运用改进后的神经网络对轨迹进行矫正以消除大量冗余数据, 并将数据集导入循环神经网络(Recurrent Neural Network, RNN)等网络中进行训练修复, 结果表明不管是去噪还是修复都比传统方法更好^[20-21]。与其他神经网络相比, 长短期记忆网络(Long Short Term Memory, LSTM)集的方法在船舶轨迹修复方面表现更佳^[22-23]。机器学习是通过已有数据训练来识别已定义问题的属性和模式的方法^[24]。由于船舶的航行参数具有不确定因素, 通过数学模型拟合和统计学中计算出阈值判别的方法难以适应航行物流货运轨迹。通过机器学习方法来训练和学习数据特征, 是未来处理类似大数据的趋势。

针对上述方法的不足和窄口段水域 AIS 货运物流轨迹异常和缺失的 2 种情况, 本文考虑众多学者没有涉及的实际环境因素和船舶本身属性, 来解决 AIS 货运物流轨迹中除空间距离以外的速度和首向角等数据的纠偏和修复轨迹, 为物流货运数据深度挖掘、船舶路径预测等提供了基础保障。

1 AIS 数据

依据国际海事组织的规定, 绝大部分船都必须安装自动识别系统(Automatic Identification System, AIS)。AIS 通过 GPS 以及基站设施等设备能够持续发送船舶的信息。国外研究者使用了 AIS 数据追踪货运物流轨迹信息, 该数据提供了港口之间和移动的单只船舶的精细物流路径轨迹信息^[25-26], 对其深度挖掘并以此构建网络, 找到适合该水域船舶物流运输最短路径, 以此解决运输效率和安全问题。

1.1 AIS 数据组成

船舶 MMSI 号(9 位)、船名、船舶类型、船宽、船长等信息, 这类固定信息往往禁止修改, 这类信息称之为静态信息。

船舶航行的具体位置(船舶的经纬度坐标表示)、航速、航向(包括航迹向、船首向)等。这些数据主要通过船上设备和基站设施获取, 这类信息称之为动态信息。当 AIS 设备不连接到基站设施时, 或当船舶信号出错器被干扰时, 发送的动态数据可能会出错。因此 AIS 动态数据错误率较高, 也是影响后续物流数据挖掘等工作的难点之一。

1.2 异常点的分类

1) 基本信息错误指静态数据中的错误信息, 例如 MMSI 号错误, 编号数字不是规定的 9 位, 经纬度

超出范围,速度不合理等。

2) 时间不完整性指一定时间内数据的缺失。根据最新的国际电信联盟通信标准^[27]的规定,AIS发送的时间间隔与船舶类型和航速相关。对于A类船,AIS报文发送的间隔应不大于10 s;对于B类船,发送间隔不应超过30 s。

3) 空间逻辑异常点是考虑了船舶本身行为的不合逻辑点等。对于异常转向点,本文结合实际环境因素,考虑三自由度的坐标系运动原理,运用船舶摆动直径的方法进行判别。

2 基本理论与方法

2.1 轨迹预处理

预处理是利用AIS的第1步,大多数AIS数据记录到了小数点后7位,利用Haversine公式计算2点之间的精确距离,结合实际经纬度的表达式式(1)。

$$h = \sin^2((x_2 - x_1)/2) + \cos(x_1) \cos(x_2) \sin^2((y_2 - y_1)/2) \quad (1)$$

$$d = 2r \arcsin(\sqrt{h}) \quad (2)$$

式中: y 、 x 分别为经度和纬度; r 为地球半径, r 取6 371 m。经过haversine计算的2点之间距离可以精确到米。2点之间的精确矢量距离是下一步筛选与去噪工作的基础。

2.2 异常漂移点识别

异常漂移点可以指空间上的异常,也可以指船舶航行过程中不符合规范操作的异常,例如,船舶间的矢量距离太远或是船舶不在规定内的通航航道内航行。往往前者识别精度较低,特别是在窄口段水域。而航迹带宽度是船舶安全航行操纵的必要宽度,常受侧风、波浪和水流等影响,通常用来设计航道模拟轨迹。很少学者考虑了船舶本身的特性与环境条件。与道路交通不同,大多数学者在处理AIS数据的时候都没有考虑这类参数。环境对船舶轨迹有一定影响,因此本文考虑宁波舟山本地的实际风速、水流等实际因素来处理AIS数据,修复流程如图1所示。

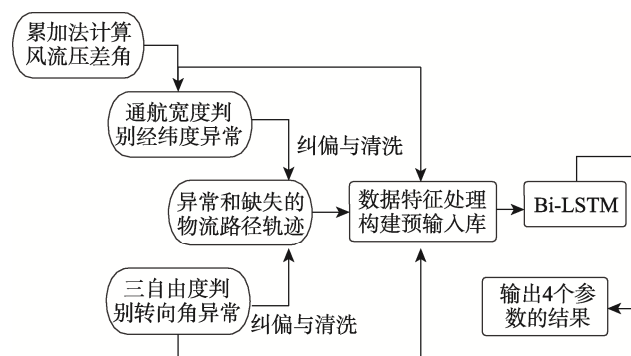


图1 修复流程
Fig.1 Flowchart of repair

2.3 通航设计宽度

通航设计宽度受到环境因素以及船舶本身性质的影响,其计算过程涵盖了环境水流、风速、船舶属性等。富裕宽度指船舶航行时,在两艘会船的船舶(队)之间和船舶(队)与航道岸线之间所必须具有的安全距离之和是船舶航行安全的保障基础。在航道的宽度设计过程之中除了需要考虑船舶的漂移倍数和风流压差偏角之外,还需要考虑船舶的航行速度对宽度设计参数的影响。随着当前船舶速度和风浪流速度的增加,其对航道宽度设计参数的影响就不能再被忽略^[28]。经过文献阅读与宁波舟山实地调研,对于特定的船舶,因其计算过程考虑了诸多影响轨迹变化的因素,所以其宽度在理论上适合作为空间阈值的标准。单航道宽度按式(3)——(4)计算。

$$W = A + 2c \quad (3)$$

$$A = n(L \sin \gamma + B) \quad (4)$$

式中: W 为单航道通航宽度, m; A 为航迹带宽度, m; c 为船舶与航道边线间的富裕宽度, m; n 为船舶漂移倍数; L 为设计的船长, m; γ 为风流压差偏角, ($^\circ$); B 为设计的船宽, m。

2.4 累加法的风压偏差角计算

风流压差角指一定风等级下船舶在航行过程中航迹向与航首向之间的夹角。它是由风压差角和流压差角的矢量和计算而成,风中航迹线与真航向线的夹角叫作风压差角,简称风压差。船舶左舷受风,流中航迹线与真航向线的夹角叫作流压差角,简称流压差,用 β 表示。如图2所示,风流压差角受诸多因素影响,其大小随着风力、船速、船型、吃水等情况的不同而变化。风压差角和流压差角可用式(5)——(7)计算。

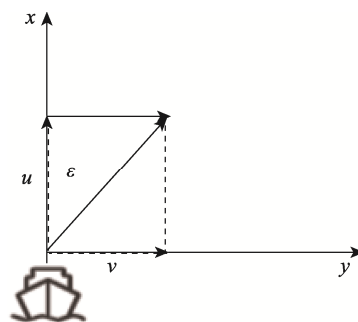


图2 风流压差角示意图
Fig.2 Drift and leeway angle

$$\tan \epsilon = \frac{u}{v} = \frac{v_w}{v_s} \quad (5)$$

$$\epsilon = \arctan \left(\frac{v_w}{v_s} \right) \quad (6)$$

$$\gamma = \alpha + \beta \quad (7)$$

式中: v_s 为船舶横向漂移速度, m/s; v_w 为船舶纵向漂移速度, m/s; α 为风压差角; β 为流压差角; γ 为风流压差角; ε 为速度夹角。

风流压差偏角在恒定的流场下可取固定值, 但是实际情况下风流是变动的, 船舶也是运动的, 在航行过程中船舶会左右摆动, 因此风流压差偏角也会变化。经过实地调研与大量资料文献考察发现因为舟山地区实际的流场复杂多变, 与设置固定值的前提条件有较大差异。本文采用累加的方法计算特定船舶在一定轨迹内的风流压差角。

γ 累加法为在初始时刻的风流压差角的基础上不断累加。例如, 假设得到一组风流压差角的集合为 $\{g, h, f, d\}$, 则 γ 的计算式为 $\gamma=g+|h-g|+|f-h|+|d-f|$ 。本文规定以偏航迹向右边的为正值、左边的为负值来计算。经数据统计, 货运船经窄口段水域的航速基本超过 6 n mile/h, 甚至达到 10 n mile/h 以上, 因此, 船舶与航道底边线间的富裕宽度 c 的取值按照不同船型取对应大于 6 n mile/h 的计算值。考虑最大流速的情况下, 涨潮流向为 $305^{\circ} \sim 310^{\circ}$, 落潮流向为 $112^{\circ} \sim 120^{\circ}$, 计算得最大横流为 0.9 m/s。因此, 船舶漂移倍数 n 的取值可按照 $0.75 < p \leq 1.00$ 这一档进行取值, 即 n 取 1.45。其中, 船舶漂移倍数和取值范围以及船舶与航道底边线的富裕宽度取值如表 1、2 所示, 其中 B 为船舶的宽度。

表 1 船舶漂移倍数 n 和风力等级关系
Tab.1 Relationship between ship drifting multiplier n and wind scale

风力等级	n 船舶漂移倍数
$p \leq 0.1$	1.81
$0.10 < p \leq 0.25$	1.75
$0.25 < p \leq 0.5$	1.69
$0.5 < p \leq 0.75$	1.59
$0.75 < p \leq 1.00$	1.45

表 2 船舶与航道边线间的富裕宽度 c
Tab.2 Superwidth c between ship and bottom edge of channel

船舶类型	航速/(n mile · h ⁻¹)	c /m
集装箱货运船	≤ 6	0.5B
	> 6	0.75B
散货船	≤ 6	0.75B
	> 6	B
油船或其他船	≤ 6	B
	> 6	1.50B

2.5 异常转向点识别

船舶的转向角指船舶运动过程中摆动的幅度大

小, 以 i 个点与 $i+1$ 个点之间的船舶航向角之差计算。船舶坐标系统的运动学是用于描述船舶在空间中的任意运动。与地面交通不一样的是, 船舶的运动具有空间性。Zhou 等^[29]利用船舶坐标体系的运动学构建 Unity 虚拟图像数据集, 使用卡尔曼滤波实时跟踪船舶坐标信息, 结果表明定位跟踪精确, 六自由度坐标体系能够良好展示船舶的运动过程, 能够精准描述船舶的运动状态。

在船舶运动和控制领域, 航向角及航行轨迹的改变是关注的重点, 即船舶在水平方向内的运动。对大多数船舶来说, 垂荡、纵摇和横摇运动对水平面内的运动影响甚小, 且不太容易计算与观察。本文主要针对 AIS 中的动态数据, 则三自由度的船舶运动问题可用式 (8) 表示。

$$\begin{cases} \dot{x}_o = u \cos \psi - v \sin \psi \\ \dot{y}_o = u \sin \psi + v \cos \psi \\ \dot{\psi} = r \end{cases} \quad (8)$$

式中: $\dot{x}_o$ 、 $\dot{y}_o$ 为惯性坐标系内位置向量的导数; u 、 v 为附体坐标系内的速度向量; r 为角速度向量; $\dot{\psi}$ 为欧拉角姿态向量的导数。

转向角速度 r 为船首向 ψ 关于时间的导数, 因此在水平面运动内可通过摆动直径 d 来测量。船舶设计的最大摆动直径与船舶的长度有关, 也会受环境水流等其他因素的影响。为了避免分析情况复杂, 本文只考虑船舶本身的运动特性。 d 通常是船长的 2~4 倍, 即 $d=k \times l$, k 取 4。假设船舶速度和船长分别为 v 、 l , 则转向速率 $r=360v/(k\pi l)$, 船舶从 i 个点点到 $i+1$ 个点的最大转向角应为该时间上的积分, 见式 (9)。

$$w_{\max} = \int_i^{i+1} \frac{360v}{k\pi l} dt \quad (9)$$

船舶从 i 到 $i+1$ 个点的转向角范围应为:

$$|C_{i+1} - C_i| \leq w_{\max} \int_i^{i+1} \frac{360v}{k\pi l} dt \quad (10)$$

式中: C_i 为 i 个点的首向角; 速度 v 可以取 $i+1$ 和 i 中 2 个速度值的最大值, 原因是充分考虑船舶的摆动动力, 使转向角范围在高风浪流环境下合理, 保留更多船舶动态数据; t 为 2 个点对应时间的差值。

2.6 Bi-LSTM

Bi-LSTM 模型输入过程为双向, 解决非线性问题方面优于 LSTM 网络, 能够更有效地拟合数据集。因此, 它可以用于序列分析, 并提供更准确的预测, 既保留了较长的依赖关系的优点, 又能结合时序性数据的特点, 还考虑了前后数据变化带来的影响^[30]。其输入结构和流程可用式 (11) — (13) 和图 3 表示。

$$h_t = \sigma(w_1 x_t + w_2 h_{t-1}) \quad (11)$$

$$h'_t = \sigma(w_3 x_t + w_4 h_{t+1}) \quad (12)$$

$$O_t = \phi_h(w_5 h_t + w_6 h_t') \quad (13)$$

式中: h_t 为前向传播层处理单元在当前时刻 t 的输出; h_t' 为反向层处理单元的输出; O_t 为输入层的值; X_t 为当前时刻的输出变量; ϕ_h 为输出层权重; w_i 为权重系数参与每层的计算; σ 激活函数。

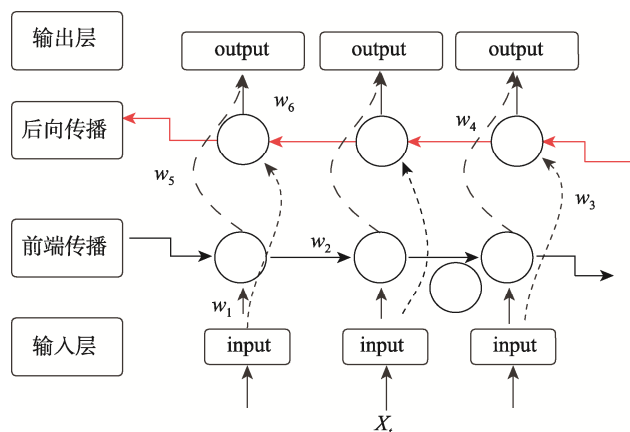


图 3 Bi-Lstm 神经网络
Fig.3 Bi-Lstm neural network

输入的数据是四维向量, 将数据转换为基于特征矩阵的行, 机器学习之前需要对数据进行特征工程的处理, 将样本向量转换为单位向量, 防止数据之间差距过大的影响。因此, 将数据归一化处理, 保证神经网络公平识别和深度学习数据, 其转换方式如下:

$$X_M = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (11)$$

式中: X 为原始数据; X_M 为归一化后数据; $X_{\min}$ 为数据最小值; $X_{\max}$ 为数据最大值。

本文采取机器学习中经典的 8.2 学习方法。其中训练集为前 80% 组数据, 测试集为后 20% 组数据。 $S(t) = (\text{经度}, \text{纬度}, \text{速度}, \text{航首向})$ 为训练目标, 将 AIS 数据中的经度、纬度、速度、航首向 4 个参数带入模型中, 步长设置为 4。其在 t 时刻的航行行为特征 $S(t)$ 表示为 $S(t) = \{x, y, v, c\}$, x, y, v, c 代表去噪后的纬度、经度、速度以及航向角。将训练样本 $S(t-3)$ 、 $S(t-2)$ 、 $S(t-1)$ 、 $S(t)$ 发送到输入层, 随后输出 $S(t+1)$, 用损失函数计算实际向量预测偏移之间的误差。参数调整一般包括批处理大小、优化器、训练轮次以及损失函数等。通过损失值最小化来调整权重参数, 以提高模型的预测精度。损失函数表示如下:

$$L_{\text{oss}} = \frac{\sum_{i=1}^n |x_i^o - x_i^p|}{n} \quad (12)$$

式中: n 为测试集长度; x_i^o 为观测值; x_i^p 为预测值。

2.6 基于距离的相似度量

为了验证本文方法在异常数据纠偏和补足修复上的优势, 利用轨迹聚类中常用的 Hausdorff 距离判

别法进行对比实验分析, 对物流轨迹的异常进行识别和纠偏。其原理为用每个轨迹点之间距离最小值中的最大值来计算。设 2 条船的轨迹分别用轨迹点集合 $A(a \in A)$ 和集合 $B(b \in B)$ 表示, 则 2 条轨迹的 Hausdorff 距离计算式为:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (13)$$

式中: $h(A, B)$ 为从船舶轨迹点集合 A 到船舶轨迹点集合 B 的单向 Hausdorff 距离。

3 实验分析

3.1 数据库

为验证上述所提出方法效果, 数据库选择宁波舟山条帚门航道及其周边水域, 时间范围从 2021 年 6 月 1 日 8 时 46 分 37 秒至 2021 年 6 月 30 日 23 时 59 分 22 秒。实验空间区域位于桃花岛与虾峙岛之间, 实验设备配置为 i5-12400, NVIDIA GeForce-RTX3060, 训练次数为 1 000 次, 总共对 628 511 条轨迹数据进行分析。

根据图 5 可知, 6 月一共有 674 艘船舶经过该区域, 其中货船占比最多, 达到了 362 艘, 占比为 53.71%。其次是集装箱船有 108 艘。因为该航道主要是浙江省舟山群岛虾峙岛与六横岛之间用于运输货物和缓解主流航道交通的重要航道。在 362 艘货船中, 船长为 100~185 m 占比最多, 总共有 228 艘。船长为 185 m 以下的货船总占比达到了 85.28%, 如表 3 所示。船长为 185 m 以下的货船是该窄口段航道的主要船舶。

为了验证本文方法, 依据上述统计结果, 选取该水域数量占比最大最具代表性船舶货船进行实验, 日期选择为 6 月 1 日到 6 月 3 日, 选取船长为 185 m 以下的货船来进行实验。该长度的货船占比最大, 且速度在航行过程中较快, 在此复杂窄口段水域极易造成数据出错和缺失。经过统计该时间段内一共有 57 243 条数据。1 000 t 左右的货船多为 A 类船舶且多为 150 m 以下的货船。根据固定标准和实际情况设置时间间隔在 10 s 以上、3 min 以下的为异常数据, 不考虑人为关闭 AIS 设备等其他情况, 只针对航行过程中的船舶。经过简单预处理与静态异常数据的筛选剔除后, 所有货船的 AIS 物流路径轨迹如图 4 所示。

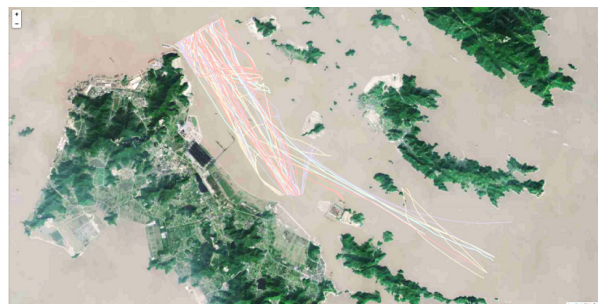


图 4 货船物流 AIS 轨迹图
Fig.4 Logistics AIS trajectory chart of cargo ships

表 3 船舶类型数量以及货船长度数量分布 Tab.3 Distribution of the number of vessel types and the length of cargo ships			
类型以及数量	占比/%	长度/m	货船长度数量以及占比
货船 360 艘	53	>250	19 (5%)
客船 15 艘	2	185 ~ 250	34 (10%)
集装箱船 108 艘	16	100 ~ 185	228 (63%)
其他船 56 艘	8	100<	81 (22%)
邮轮 84 艘	13	>250	19 (5%)

3.2 异常点的删除

依据前面设置的方法,结果筛选对比图如表 3 所示。6 月 1 日—3 日中在所设定的条件船舶下,AIS 所有数据有 57 243 个点,其中不包括船舶停止等无效点。表 3 中静态异常数据点共有 386 个,占总数据量的 0.67%。本文方法筛选出的空间异常点为 10 790 个,比传统相似性度量方法多了 1 830 个点,而基于距离的相似度量法无法纠偏船舶的异常转向点。本文方法纠偏的异常转向点共有 3 961 个,占比达 8.52%,占比相对较少,此占比是相对于距离筛选点后的数据,及在 46 453 个点基础上进一步筛选。因为转向角不考虑空间距离,这样筛选可以提高精度以及避免重复筛选。距离筛选点占比较大,但也不足 20%。可能原因是该水域较小,窄口段流速较大,且边界有船舶经过,造成少许船舶只有几个或是几十个数据点被记录在内,而且也有原始下载的 AIS 数据因为项目要求可能已经进行了预处理的情况。整体筛选出 14 751 个点,占原始数据约为 27%,基于轨迹聚类中的距离相似性度量法筛选数据的占比为 15.60%。

表 3 异常数据筛选结果对比 Tab.3 Comparison of abnormal data screening results					
方法	静态数 据数量	空间过滤 点数量	空间过滤 点占比/%	转向过滤 点数量	转向过滤 点占比/%
通航宽度与 三自由度	386	10 790	18.85	3 961	8.52
相似度量法	386	8 960	15.60		

3.3 数据的修复

选取 MMSI 编号为 412110936 的货船,该船的基本信息如表 4 所示。选择时间跨度为 6 月 1 日 23: 27: 25 到 6 月 2 日 00: 01: 05,一共 178 个点。根据前面基础理论,运用累加法算出该船舶在该时间段内的风流压差偏角 $\gamma=10.156^{\circ}$,此风压差偏角只考虑船舶航行过程中的角。异常静态点有 5 个,该船的理论通行航宽度 $W=110.5$ m,超越该半径范围外的空间

点为异常点。异常距离点为 40 个,异常转向点为 16 个,总计筛选点为 61 个,占整体数据的比例为 34.26%。

3.4 修复后结果图以及分析

修复的结果分别从经纬度、速度、船首向以及经纬度的损失方差进行分析。Bi-LSTM 重塑的点数为 203,RNN 重塑的点数为 168,比原始数据还要少,Bi-LSTM 的经纬度损失方差为 0.004 9,而 RNN 的为 0.015 3,比改进后的算法高 312.22%。图 5 为修复前的轨迹图,其轨迹点存在漂移异常点(第 31 个点前后)以及缺失点(点数 112 前后),此处漂移点指超越航道设计宽度半径的点而非曲线上的欧氏距离空间点,欧式距离点往往在 200~300 m 以外。对于宽航道水域,因为水流缓慢和航道宽度等因素,可能导致其半径范围会更大。修复后的轨迹点如图 6 所示,不存在异常漂移的现象且所有点均在理论设计航道中,在图 6 中 Bi-LSTM 修复的结果中,点 30 到点 45 有一定密集情况,是因为后面 4 个点间隔时间在 3 s 左右,导致机器学习输出点更为密集。修复后的总数据

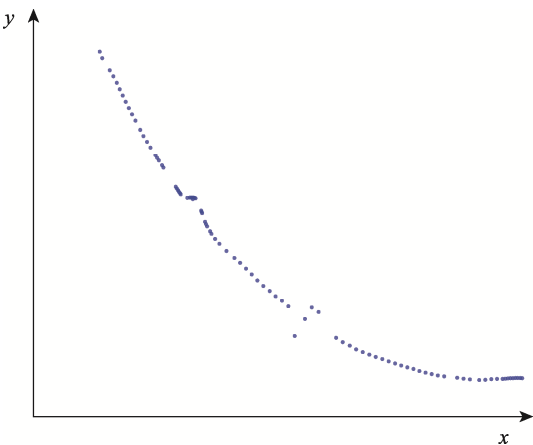


图 5 原始轨迹
Fig.5 Original trajectory

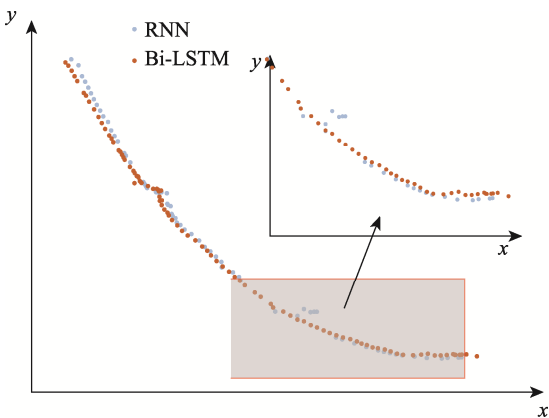


图 6 修复后轨迹
Fig.6 Trajectory after repair

量达 203 个点,在原有数据上提升了 115.34%,连续性更强,拟合度更高,满足 AIS 报文规律。RNN 的修复结果起初效果良好,但在缺失部分出现了少许不合逻辑的轨迹点,其原因是 RNN 存在梯度问题,在长时间修复结果上表现不如 Bi-LSTM 模型,在此期间损失率也最高,因此整体损失均方差达 0.015 3。由图 7 可知,RNN 修复的船首向拐点多,最小船首向为 271.3985° ,与前一个点的船首向相差 0.365° 左右,比整体相差率高了 36%。修复后最小转向角与前一个转向角相差率在合理范围内,且符合船舶运动模型。整体曲线更加平滑,更能反映船舶航行真实状态。修复速度对比图 8 可知,修复后的速度均在 2 n mile/h 以上,且在速度为 6~9 n mile/h 时,Bi-LSTM 修复后的曲线更加平滑,且不存在 i 到 $i+1$ 个点之间速度差过大的情况。表明长时间的物流路径修复中,Bi-LSTM 更加稳定,其修复后的参数能够真实反映船舶航行状态。

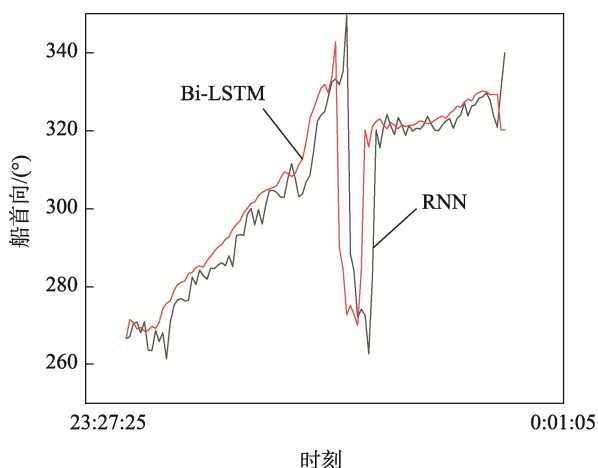


图 7 修复前后船首向
Fig.7 Ship heading angle diagram
before and after repair

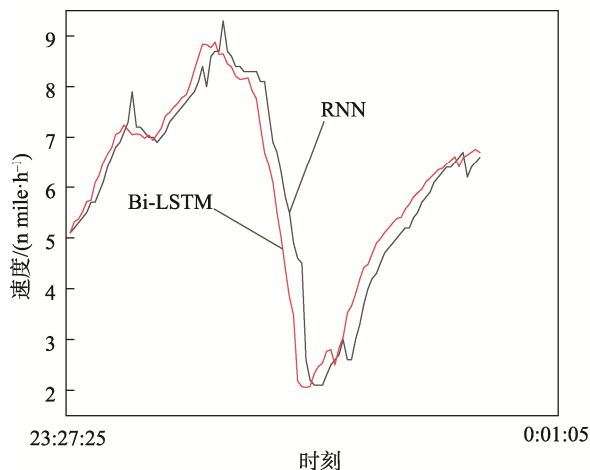


图 8 修复前后速度对比
Fig.8 Speed comparison diagram
before and after repair

4 结语

本文进行了大量文献阅读与宁波舟山港地区的实地调研,并对 57 243 条 AIS 数据分析,运用累加法计算出风流压差角偏角,从而获得通航设计宽度作为空间异常点的判别依据;根据水平方向船舶的三自由度摆动的公式计算出转向角范围,进行数据的清洗与去噪;最后运用 Bi-LSTM 法对 MMSI 号为 412110936 中缺失数据进行了修复,得出以下结论。

1) 此窄口段水域多为 100~185 m 的 A 类型货船,是货运物流运输的主要水域。水域风力和横流较大,船舶航行速度较快,大部分速度为 6~15 海里,原始物流路径轨迹信息显示大部分 AIS 物流路径轨迹信息存在缺失、漂移、异常等现象。

2) 本文所提出的筛选方法比传统基于聚类的相似性度量法精度更高,能够筛选空间距离以外的转向点数据,在 57 243 条数据中过滤了 15 137 条异常数据,占比为 26.443%。

3) 经度、纬度、船首向、速度 4 个参数的修复结果表明,经过 Bi-LSTM 处理之后比 RNN 处理之后的货运物流路径轨迹数据修复更加优异,轨迹更加连续平滑,更能反映船舶真实航行状态。本文方法对具有 AIS 数据异常的各种类型船皆可进行轨迹修复,可为后续聚类压缩工作中阈值设定提供参考范围,减少工作量,提高后续货运物流信息的精确度。

参考文献:

- [1] SUGRUE D, ADRUANES P. A Data Fusion Approach to Predict Shipping Efficiency for Bulk Carriers[J]. Transportation Research Part E: Logistics and Transportation Review, 2021, 149: 102326.
- [2] 唐经旺. 基于 AIS 轨迹数据挖掘的港口特征分析系统的研究与实现[D]. 北京: 北京邮电大学, 2021.
TANG Jing-wang. Research and Implementation of Port Feature Analysis System Based on AIS Trajectory Data Mining[D]. Beijing: Beijing University of Posts and Telecommunications, 2021.
- [3] WU L, XU Y, WANG Q, et al. Mapping Global Shipping Density from AIS Data[J]. The Journal of Navigation, 2017, 70(1): 67-81.
- [4] 贾磊. 基于海量 AIS 数据的内河航运信息可视化研究[D]. 武汉: 武汉理工大学, 2018.
JIA Lei. Research on Visualization of Inland Shipping Information Based on Massive AIS Data[D]. Wuhan: Wuhan University of Technology, 2018.
- [5] YANG Dong, WU Ling-xiao, WANG Shuai-an, et al. How Big Data Enriches Maritime Research a Critical Review of Automatic Identification System (AIS) Data Applications[J]. Transport Reviews, 2019, 39(3): 1-19.

- [6] WANG Li-kun, LI Yang, WAN Zheng, et al. Use of AIS Data for Performance Evaluation of Ship Traffic with Speed Control[J]. *Ocean Engineering*, 2020, 204(3): 107259.
- [7] LAROCK T, Xu M, ELIASSI R T. A Path-Based Approach to Analyzing the Global Liner Shipping Network[J]. *EPJ Data Science*, 2022, 11(1): 18.
- [8] ZHANG Y, LIU A, LIU C, et al. A Track Initiation Algorithm Using Residual Threshold for Shore-Based Radar in Heavy Clutter Environments[J]. *Journal of Marine Science and Engineering*, 2020, 8(8): 614.
- [9] PARK J, JEONG J, PARK Y. Ship Trajectory Prediction Based on bi-LSTM Using Spectral-Clustered AIS Data[J]. *Journal of Marine Science and Engineering*, 2021, 9(9): 1037.
- [10] LIU Zi-hao, WU Zhao-lin, ZHENG Zhong-yi. A Novel Framework for Regional Collision Risk Identification Based on AIS Data[J]. *Applied Ocean Research*, 2019, 89: 261-272.
- [11] ZHOU Yang, DAAMEN W, VELLINGA T, et al. Ship Classification Based on Ship Behavior Clustering from AIS Data[J]. *Ocean Engineering*, 2019, 175: 176-187.
- [12] WANG Kai, YAN X, YUAN Yu-peng, et al. Dynamic Optimization of Ship Energy Efficiency Considering Time-Varying Environmental Factors[J]. *Transportation Research Part D Transport and Environment*, 2018, 62: 685-698.
- [13] TALAVERA A, AGUASCA R, GALVAN B, et al. Application of Dempster-Shafer Theory for the Quantification and Propagation of the Uncertainty Caused by the Use of AIS Data[J]. *Reliability Engineering & System Safety*, 2013, 111: 95-105.
- [14] GUO Shao-qing, MOU Jun-min, CHEN Lin-ying, et al. Improved Kinematic Interpolation for AIS Trajectory Reconstruction[J]. *Ocean Engineering*, 2021, 234(8): 109256.
- [15] 郭昊. 基于 AIS 数据的轨迹段聚类方法研究[D]. 南京: 南京信息工程大学, 2021: 56-66.
- GUO Hao. Research on the Clustering Method of Trajectory Segments Based on AIS Data[D]. Nanjing: Nanjing Information Engineering, 2021: 56-66.
- [16] NGUYEN V, IM N, LEE S M. The Interpolation Method for the Missing AIS Data of Ship[J]. *Journal of Navigation and Port Research*, 2015, 39(5): 377-384.
- [17] YAN Zhao-jin, XIAO Yi-jia, CHENG Liang, et al. Exploring AIS Data for Intelligent Maritime Routes Extraction[J]. *Applied Ocean Research*, 2020, 101: 102271.
- [18] RONG Hao, TEIXEIRA A P, GUEDES S C. Ship Trajectory Uncertainty Prediction Based on a Gaussian Process Model[J]. *Ocean Engineering*, 2019, 182: 499-511.
- [19] MURRAY B, PERERA L. A Dual Linear Autoencoder Approach for Vessel Trajectory Prediction Using Historical AIS Data[J]. *Ocean Engineering*, 2020, 209: 107478.
- [20] MURRAY B, PERERA L. An AIS-Based Deep Learning Framework for Regional Ship Behavior Prediction[J]. *Reliability Engineering & System Safety*, 2021, 215(5): 107819.
- [21] NGUYEN D, VADAINÉ R, HAJDUCH G, et al. GeoTrackNet-a Maritime Anomaly Detector Using Probabilistic Neural Network Representation of AIS Tracks and a Contrario Detection[C]// *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [22] YANG C H, WW C H, SHAO J C, et al. AIS-Based Intelligent Vessel Trajectory Prediction Using Bi-LSTM[J]. *IEEE Access*, 2022, 10: 24302-24315.
- [23] SUO Yong-feng, CHEN Wen-ke, CLARAMUNT C, et al. A Ship Trajectory Prediction Framework Based on a Recurrent Neural Network[J]. *Sensors*, 2020, 20(18): 5133.
- [24] LIU T, MA J. Ship Navigation Behavior Prediction Based on AIS Data[J]. *IEEE Access*, 2022, 10: 47997-48008.
- [25] DUCRUET C, ZAIDI F. Maritime Constellations: A Complex Network Approach to Shipping and Ports[J]. *Maritime Policy & Management*, 2012, 39(2): 151-168.
- [26] KOJAKU S, XU Meng-qiao, XIA Hao-xiang, et al. Multiscale Core-Periphery Structure in a Global Liner Shipping Network[J]. *Scientific Reports*, 2019, 9(1): 404.
- [27] YU Cao, ZHANG Wei, ZHU Yu-liang, et al. Impact of Trends in River Discharge and Ocean Tides on Water Level Dynamics in the Pearl River Delta[J]. *Coastal Engineering*, 2020, 157: 103634.
- [28] LEE W J, ROH M I, LEE H W, et al. Detection and Tracking for the Awareness of Surroundings of a Ship Based on Deep Learning[J]. *Journal of Computational Design and Engineering*, 2021, 8(5): 1407-1430.
- [29] ZHOU Yang, DAAMEN W, VELLINGA T, et al. Review of Maritime Traffic Models from Vessel Behavior Modeling Perspective[J]. *Transportation Research Part C Emerging Technologies*, 2019, 105: 323-345.
- [30] FERRANDIS J D Á, TRIANTAFYLLOU M S, CHRYSSOSTOMIDIS C, et al. Learning Functionals via LSTM Neural Networks for Predicting Vessel Dynamics in Extreme Sea States[J]. *Proceedings Mathematical, Physical, and Engineering Sciences*, 2021, 477(2245): 20190897.