

基于 Bi-BPNNs 的船舶轨迹修复模型

李 越¹, 袁 智^{1,2}, 刘 奕²

(1. 湖北工程学院新技术学院, 湖北 孝感 432000; 2. 武汉理工大学航运学院, 湖北 武汉 430063)

[摘要] 为了解决船舶轨迹数据的异常和丢失问题, 辅助轨迹复原和情景推演, 提出一种双向学习模型, 用于修复船舶轨迹。从 AIS (Automatic Identification System) 报文中提取航行船舶的上下文轨迹特征向量, 改进基础 BP (Back Propagation) 神经网络的拓扑结构, 构建具有双向预测功能的 BP 神经网络 (Bi-BPNNs) 模型, 修复缺失的轨迹数据。使用长江干线航行船舶的真实轨迹数据对构建的模型进行验证和分析, 通过与基础 BP 网络和常用线性插值方法的对比, 证明 Bi-BPNNs 模型在船舶轨迹修复上具有更好的性能和效果。

[关键词] 船舶轨迹; 修复模型; Bi-BPNNs; 双向预测

[中图分类号] U 698.6

A Model of Ship Trajectory Data Repair Based on Bi-BPNNs

LI Yue¹, YUAN Zhi^{1,2}, LIU Yi²

(1. College of Technology, Hubei Engineering University, Xiaogan 432000, China;

2. School of Navigation, Wuhan University of Technology, Wuhan 430063, China)

Abstract: In order to solve the issues of trajectory data abnormality and losses, as well as to assist trajectory restoration and scenario deducing, a two-way learning method for ship trajectory data repair is proposed. Which can solve the problem of trajectory data abnormality and loss, and assist trajectory restoration and scenario inference, so as to ensure ship traffic flow analysis and marine accident investigation. Specifically, the context trajectory feature vectors of ships are extracted from AIS (automatic identification system) data, the topological structure of BP (Back Propagation) neural network is improved, and the Bi-directional BP neural network (Bi-BPNNs) model is constructed to repair the missing trajectory data. Finally, the real trajectory data of the ships sailing on the Yangtze River are collected for verification of the proposed model. What's more, compared with basic BP network and common linear interpolation methods, the Bi-BPNNs model proposed in this paper has better performance and effect on ship trajectory data repair.

Keywords: ship traffic flow; trajectory repair; Bi-BPNNs; bi-directional prediction

0 引言

日益繁忙的水上交通给船舶交通监管和事故预防带来了巨大压力^[1]。船舶轨迹数据因为记录了船舶航行过程中的实时位置和航线信息, 从而为水上船舶轨迹分析、交通流分析、船舶交通监管和水上事故预防提供了重要的信息^[2]。区域内船舶基于 GPS (global positioning system)、AIS (automatic i-

[收稿日期] 2020-07-20

[基金项目] 湖北省教育厅项目 (B2017504); 湖北工程学院新技术学院科学研究项目 (2017Hgxy19, 2019Hgxy06); 国家自然科学基金项目 (51709219)

[作者简介] 李越 (1991—), 女, 助教, 硕士, 从事生物信息处理、数据建模研究。通信作者: 袁智 (1988—), 男, 讲师, 博士生, 从事交通运输工程、航海智能技术研究。

dentification system) 设备不断地与他船和岸基以报文的形式发送和接收船舶数据, 这些数据包括船舶的静态特征、实时位置、速度、航线以及其他信息。但是, 由于受采集设备、信号传输以及环境等因素的影响, 原始采样的轨迹数据存在异常和错误等质量问题^[3], 从而造成船舶轨迹数据异常和部分缺失。因此, 有必要构建船舶轨迹修复模型来提升船舶轨迹数据的质量, 从而提高轨迹数据的利用效率和研究价值。针对船舶轨迹修复的研究, 国内外学者已经取得了一些成果。刘立群等^[4]通过三次样条插值法分别对船舶经度、纬度信息进行修复。秦红星等^[5]以线性插值为基础, 提出了一种船舶轨迹修复算法。李佳等^[6]利用粒子群优化算法和相似轨迹优化最小二乘支持向量机模型修复残缺轨迹数据。Liu 等^[7]基于支持向量机方法构建了船舶轨迹预测模型。Lv^[8]借助大数据和云计算技术对船舶轨迹进行了分析。现有这些方法多基于线性插值方法, 不能很好地把握整体轨迹的特征, 尤其不能很好地处理弯曲航段的轨迹。将船舶轨迹数据看作有序的时间序列, 通过分析异常或缺失轨迹的上下文信息, 构建轨迹修复模型, 可以实现船舶轨迹的修复。与传统的线性插值方法相比, 神经网络在轨迹预测建模方面有明显的优势, 可以根据原始轨迹数据的上下文信息双向学习整个船舶轨迹的变化规律, 从而保证轨迹修复的精度。BP (back propagation) 神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络, 有很强的非线性逼近功能, 在数据分析和预测方面有着广泛的应用^[9]。

本文对船舶原始轨迹数据进行分析, 引入 BP 神经网络方法, 构建 Bi-BPNNs 预测模型, 从前后两个方向对异常或缺失的船舶轨迹进行循环修复, 从而解决船舶轨迹数据丢失问题。在此基础上, 基于采集的真实船舶轨迹数据进行实例研究, 并与基础 BP 网络模型和常用线性插值方法进行对比分析, 从而验证本文提出模型的有效性和实用性。

1 基于 BP 神经网络的船舶轨迹双向修复模型

船舶轨迹可看作是船舶航行时间和当前位置构成的序列, 相邻的轨迹点之间有一定的联系。基于 BP 神经网络构建轨迹修复模型, 可以学习历史轨迹序列之间的变化特征来预测未来时间的轨迹, 从而实现缺失轨迹的修复^[10]。

BP 神经网络有三个要素: 网络拓扑结构、传递函数和学习算法。一个简单 BP 神经网络的拓扑结构包括输入层、隐含层和输出层。其中, 隐含层又有单隐含层和多隐含层结构, 隐含层越多, 网络越复杂, 学习的能力也越强, 但是训练的时间过长, 会导致模型过拟合。BP 网络的传递函数是非线性变换函数 Sigmoid 函数, 其特点是函数本身及其导数都是连续的。BP 神经网络的学习算法为 δ 算法, 输入层、隐含层和输出层的权值调整公式均由学习率、误差信号和输入信号三个因素决定, 输入层误差信号直接反应了网络的输出误差, 隐含层的误差信号从输出层开始逐层反传过来得到。

船舶轨迹数据中的轨迹点不仅与上文轨迹有关系, 还与下文轨迹有关系。如果只用上文轨迹进行学习, 那么修复的轨迹点越多, 误差越大。如果增加对下文轨迹信息的学习, 两边同时修复缺失轨迹, 将有效减少误差, 保证轨迹修复结果的可靠性和准确性。然而, BP 神经网络是一种单向学习的网络, 想要实现上下文轨迹的双向学习, 需要改进其网络结构。因此, 本文基于 BP 神经网络设计了双向预测模型 Bi-BPNNs, 其结构如图 1 所示。输入层除了包含缺失轨迹的上文的特征向量, 还增加了下文的特征向量, 按时间顺序翻转之后再传输给网络的隐含层, 在输出层的时候

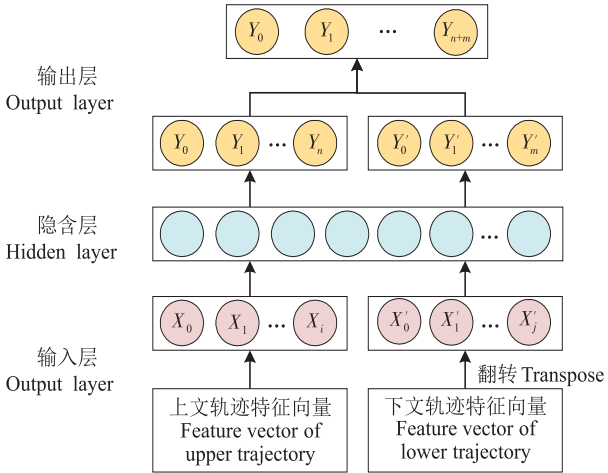


图 1 修复船舶轨迹的 Bi-BPNNs 网络结构
Fig.1 Structure of Bi-BPNNs network for ship trajectory repair

跟上文预测的正向轨迹结合, 最终输出完整的轨迹修复结果。

2 模型构建与参数设置

2.1 模型输入

从采集的船舶 AIS 数据选取时间、经度、纬度、航向、航速为船舶轨迹特征向量, 作为模型的输入变量。输入变量除了包含缺失轨迹的上文信息, 即前一时间段的轨迹, 还包括后一时间段的下文轨迹。因为船舶轨迹特征向量是有序的时间序列, 所以下文的轨迹信息在输入隐含层之前还需要进行翻转, 以便网络学习到下一时间段中有更多的轨迹特征, 从而提高缺失部分轨迹的预测精度。

2.2 模型参数

轨迹修复模型的重要参数: 1) 时间步长过短会影响整体的学习效果, 太长则泛化能力不好; 2) 隐含层神经元数目对网络的性能影响较大, 也决定了网络的复杂度, 是网络模型重要的参数; 3) 训练函数可通过调整权值和阈值来减小整体误差的算法; 4) 传递函数是网络的重要组成部分, 而且必须是连续可微的。这些参数将通过实际轨迹数据的训练和测试实验进行选择和优化。

2.3 模型评估

通过训练值、预测值、真实值的 RMSE (root mean square error, 均方根误差) 和 MAE (mean absolute error, 平均绝对误差) 对构建的双向修复模型的轨迹修复结果进行评估和优化。同时, 为验证构建模型的性能, 还将其与常见的插值方法^[11]和基础的 BP 网络^[12]进行对比, 并通过计算 RMSE 和 MAE 进行对比分析, 计算方法:

$$\text{RMSE} = \left(\sum_{t=1}^T (y_t - \hat{y}_t)^2 / T \right)^{1/2}; \quad \text{MAE} = \sum_{t=1}^T |y_t - \hat{y}_t| / T.$$

式中: t 表示时间变量; T 表示轨迹变量的长度; y_t 和 \hat{y}_t 分别表示轨迹变量的真实值和预测值。

3 实验及结果

本实验使用的船舶轨迹数据来源于航行在长江干线的一艘散货船, 航行轨迹是长江干线江苏航段浏海沙水道, 区域范围为 $31^{\circ}55'53''\text{N}$, $120^{\circ}29'49''\text{E}$ 至 $32^{\circ}8'38''\text{N}$, $120^{\circ}46'1''\text{E}$, 航行时间为 2018 年 10 月 13 日。原始的采样数据包含日期、时间、船舶呼号、航速、航向、经度、纬度, 部分数据如表 1 所示。

表 1 部分原始采样数据
Tab. 1 Part of the original sampled data

日期 Date	时间 Time	船舶呼号 Call Sign	航速 SOG /kn	航向 COG /($^{\circ}$)	经度 Longitude /($^{\circ}\text{E}$)	纬度 Latitude /($^{\circ}\text{N}$)
2018/10/13	17:34:53	100900517	8.258	186.26	120.7667	32.0160
2018/10/13	17:35:51	100900517	8.084	185.89	120.7635	32.0157
2018/10/13	17:35:51	100900517	8.084	185.89	120.7635	32.0157
2018/10/13	17:36:49	100900517	8.106	180.36	120.7603	32.0153
⋮	⋮	⋮	⋮	⋮	⋮	⋮

3.1 轨迹预处理

除了设备误差和网络传输造成的轨迹数据缺失外, 采集到的原始数据还存在一些异常和噪声。例如, 船舶的经度或纬度数据范围不在采样的航行区域范围, 部分航速明显偏小或偏大。为此, 设计了轨迹数据预处理方法: 先将收集的原始数据按采样时间进行排序, 删除重复的数据; 接着对排完序的数据进行航段划分, 以便在更小的范围内进行判断, 避免下一步错误处理或清洗过度; 然后在航段内进行遍历, 分别判断经度、纬度、航速是否在正常范围之内, 如果超出范围就被认为是异常数据被清

洗掉，最终得到干净的轨迹数据集，预处理之后得到的船舶轨迹如图 2 所示。

3.2 船舶轨迹修复

为验证所构建的 Bi-BNNs 轨迹预测模型的性能，选取修复难度大的弯曲段轨迹进行实验，设置 {航速，航向，经度，纬度} 为修复模型的输入特征向量，并随机选取样本数据的 80% 作为训练集，剩下的 20% 为测试集。实验开始，将模型的参数神经元数量、时间步长、训练函数、传递函数分成 4 组进行测试，通过计算，比较模型的训练误差和测试误差，对各参数进行逐步优化，如表 2 所示。

从表 2 中的第 1 组和第 2 组实验中可以看到，当训练函数为“trainlm”、传递函数为“purelin”时，经度和纬度的训练误差最小；在第 3 组实验中，时间步长为 4 的误差比时间步长为 2 和 5 时的要小；同理，在第 4 组试验中，神经元数量为 12 时，模型的预测效果最佳。因此，最终设定轨迹修复模型的参数如下：神经元数量为 12；时间步长为 4；训练函数为“trainlm”；传递函数为“purelin”。这里，模型参数时间步长设置为 4，也就是说用 4 个历史轨迹数据点预测未来一个轨迹数据点，然后依次循环直至得到完整的轨迹预测。

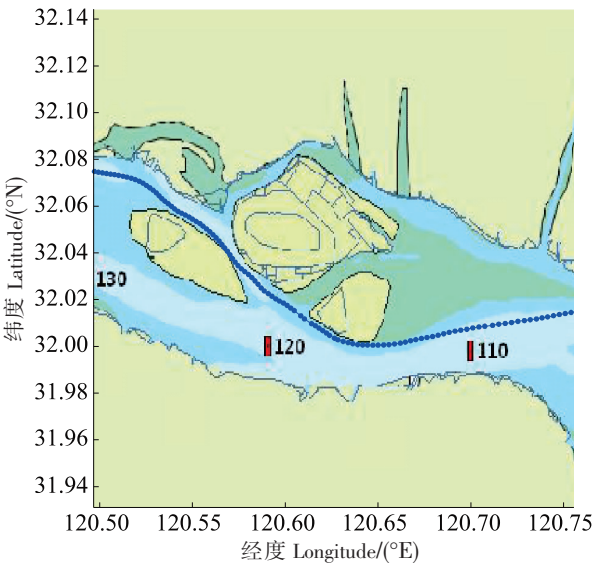


图 2 预处理后的船舶轨迹
Fig.2 Ship trajectory after preprocessing

表 2 Bi-BPNNs 模型的不同参数设置实验

Tab. 2 Some training experiments with different parameter settings for Bi-BPNNs model

序号 NO.	神经元数量 Neurons	时间步长 Time step	训练函数 Training function	传递函数 Transfer function	RMSE($\times 10^{-2}$)/($^{\circ}$)	
					经度 Longitude	纬度 Latitude
1	10	3	traingdx	purelin	0.186 0	0.166 4
	10	3	traingd	purelin	1.120 7	1.268 8
	10	3	traingdm	purelin	4.288 9	8.304 3
	10	3	trainrp	purelin	0.128 5	0.152 3
	10	3	traingcb	purelin	0.116 8	0.140 7
	10	3	trainlm	purelin	0.031 4	0.053 6
	10	3	trainseg	purelin	0.184 5	0.178 3
	10	3	trainlm	poslin	0.109 9	0.089 5
2	10	3	trainlm	logsig	0.740 3	0.178 3
	10	3	trainlm	tansig	0.721 5	0.132 1
	10	3	trainlm	radbas	1.056 3	0.218 6
	10	3	trainlm	satlins	0.896 9	0.148 7
	10	3	trainlm	purelin	0.150 3	0.104 5
3	10	2	trainlm	purelin	0.150 3	0.104 5
	10	4	trainlm	purelin	0.018 1	0.047 3
	10	5	trainlm	purelin	0.028 4	0.048 6
4	6	4	trainlm	purelin	0.031 4	0.061 4
	8	4	trainlm	purelin	0.022 9	0.050 2
	12	4	trainlm	purelin	0.017 6	0.042 9
	14	4	trainlm	purelin	0.027 8	0.085 5

根据确定的双向预测模型, 分别利用上下文的船舶轨迹点双向循环预测 9 个轨迹点, 得到共含有 18 个预测轨迹点的轨迹, 如图 3 所示。图 3 中, 蓝色点表示真实测得的轨迹, 红色点表示预测的轨迹, 其中图 3a) 是包含预测信息的轨迹, 图 3b) 是弯曲航段真实轨迹与预测轨迹的对比。图 4 展示了 Bi-BNNs 模型的船舶轨迹修复性能。

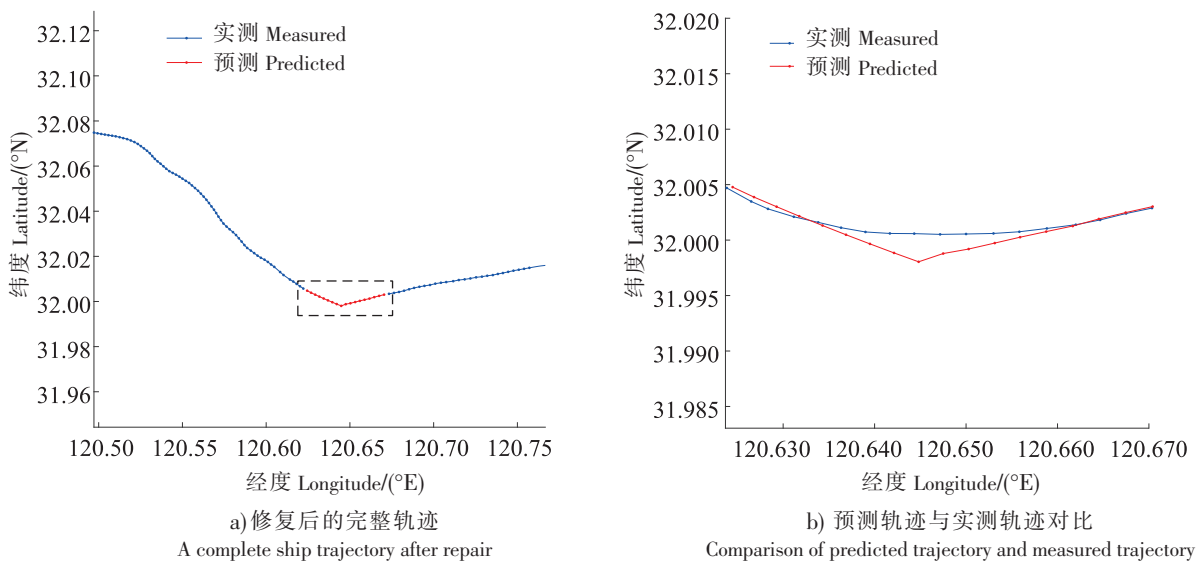


图 3 船舶轨迹修复结果
Fig.3 Results of ship trajectory repair

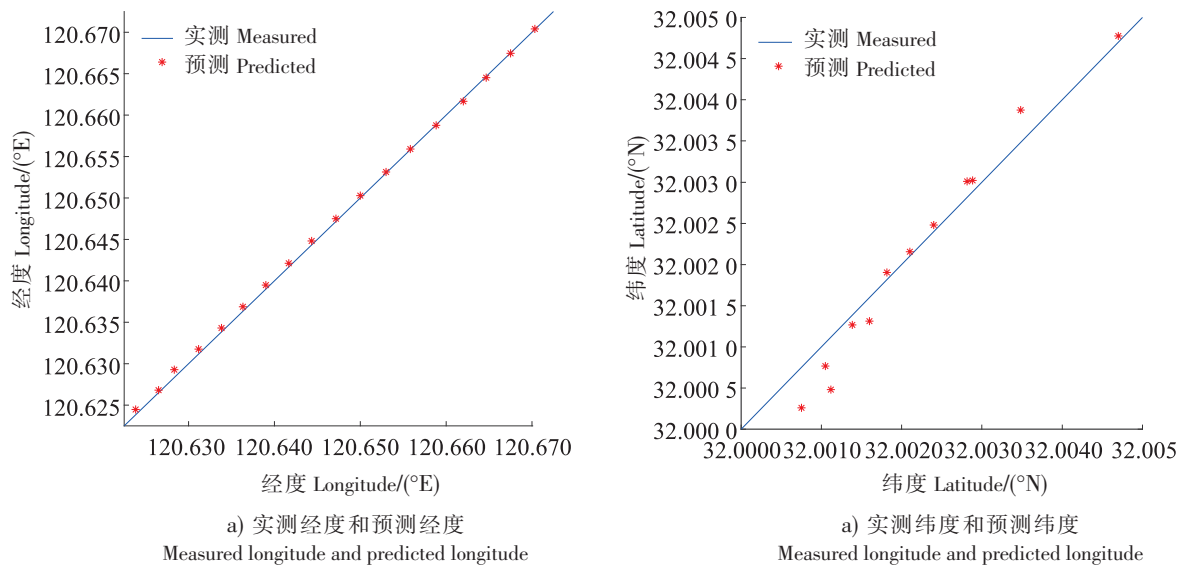


图 4 Bi-BNNs 模型的船舶轨迹修复性能
Fig.4 Performance of ship trajectory repair of Bi-BNNs model

此外, 为了进一步验证 Bi-BPNNs 模型的轨迹修复优越性, 基于相同的实验数据, 将其分别与常见的线性插值 (linear interpolation, LI)^[13]、逐段三次样条插值 (spline interpolation, SI)^[14]、立方插值 (pchip interpolation, PI)^[15]、和基础 BP 神经网络 (BPNNs) 模型进行对比, 并计算各方法修复轨迹数据的 RMSE 和 MAE, 得到 10 次实验结果的平均值和标准差, 如表 3 所示。从表 3 中可以看出 Bi-BPNNs 优于其他的方法和模型。

表 3 不同模型的船舶轨迹修复模型性能
Tab.3 Comparison of different models in ship trajectory repair

模型 Model	经度 Longitude		纬度 Latitude	
	RMSE	MAE	RMSE	MAE
LI	0.001 4 ± 0	0.001 0 ± 0	0.001 2 ± 0	0.001 1 ± 0
SI	0.012 6 ± 0	0.008 6 ± 0	0.083 1 ± 0	0.056 8 ± 0
PI	0.001 4 ± 0	0.001 0 ± 0	0.001 9 ± 0	0.001 2 ± 0
BPNNs	0.288 4 ± 0.003 5	0.008 4 ± 0.000 2	0.045 7 ± 0.003 0	0.001 4 ± 0.000 3
Bi-BPNNs	0.000 6 ± 0.000 1	0.000 5 ± 0.000 1	0.001 5 ± 0.000 4	0.001 0 ± 0.000 3

4 结论

本文针对船舶轨迹数据的缺失和异常问题，提出了 Bi-BPNNs 模型，利用对上下文的历史轨迹信息特征的学习，实现对缺失轨迹的双向循环预测，从而提升轨迹修复的精度和效率。弯曲航段的实例验证结果表明：经度修复的误差 RMSE 和 MAE 分别只有 0.000 6 和 0.000 5；纬度修复的误差 RMSE 和 MAE 分别只有 0.001 5 和 0.001 0。与其他多种插值方法和基础 BP 神经网络相比，本文提出的 Bi-BPNNs 在船舶轨迹修复上具有明显的性能优势。

[参 考 文 献]

[1] 刘长俭. 交通强国背景下中国水运发展战略思考 [J]. 中国水运, 2019(12): 14-15.

[2] 曾铮. 船舶航行轨迹分析与应用系统 [J]. 舰船科学技术, 2020, 42(2): 64-66.

[3] 周世波, 徐维祥. 船舶轨迹异常检测方法研究进展 [J]. 电子测量与仪器学报, 2017, 31(3): 329-337.

[4] 刘立群, 吴超仲, 褚端峰, 等. 基于 Vondrak 滤波和三次样条插值的船舶轨迹修复研究 [J]. 交通信息与安全, 2015, 33(4): 100-105.

[5] 秦红星, 杨茜. 改进线性插值的船舶轨迹修复迭代算法 [J]. 计算机辅助设计与图形学学报, 2019, 31(10): 1759-1767.

[6] 李佳, 初秀民, 刘兴龙, 等. 内河船舶缺失轨迹修复方法 [J]. 哈尔滨工程大学学报, 2019, 40(1): 67-73.

[7] LIU J, SHI G, ZHU K. Vessel trajectory prediction model based on AIS sensor data and adaptive chaos differential evolution support vector regression (ACDE-SVR) [J]. Applied Sciences, 2019, 9(15): 2983.

[8] LV S. Construction of marine ship automatic identification system data mining platform based on big data [J]. Journal of Intelligent & Fuzzy Systems. 2020, 38(2): 1249-55.

[9] 姜桂艳, 常安德, 牛世峰, 等. 基于 BP 神经网络的交通数据序列动态可预测性分析方法 [J]. 北京工业大学学报, 2011, 37(7): 1019-1026.

[10] 冯玉芳, 卢厚清, 殷宏, 等. 基于 BP 神经网络的故障诊断模型研究 [J]. 计算机工程与应用, 2019, 55(6): 24-30.

[11] 周鹏, 陈金海, 林乾锋, 等. 基于 AIS 的轨迹插值方法 [J]. 集美大学学报（自然科学版）, 2018, 23(6): 443-447.

[12] 凡甲甲, 祁云嵩, 葛霓琳. 基于改进 BP 神经网络的船舶轨迹识别方法 [J]. 计算机工程与设计, 2019, 40(12): 3639-3644.

[13] 秦红星, 杨茜. 改进线性插值的船舶轨迹修复迭代算法 [J]. 计算机辅助设计与图形学学报, 2019, 31(10): 1759-1767.

[14] 田利芹, 贾杰. 三次样条插值在船舶轨迹修复中的研究 [J]. 舰船科学技术, 2016, 38(6): 79-81.

[15] 罗天洪, 马力. 基于过路径点的三次多项式插值函数的仿蜘蛛机器人足路径规划 [J]. 机械传动, 2017, 41(7): 1-6.

(责任编辑 陈 敏 英文审校 郑青榕)