

基于PCA-LSTM神经网络的凡纳滨对虾养殖水质预测

习文双, 江敏, 吴昊, 潘璠, 唐燕

Prediction of water quality in *Litopenaeus vannamei* aquaculture ponds based on the PCA-LSTM neural network model

XI Wenshuang, JIANG Min, WU Hao, PAN Fan, TANG Yan

在线阅读 View online: <https://doi.org/10.12024/jsou.20220903980>

您可能感兴趣的其他文章

Articles you may be interested in

凡纳滨对虾池塘养殖过程中水质与虾虹彩病毒病发生的相关性

Correlation between water quality and shrimp hemocyte iridescent virus disease occurrence of *Litopenaeus vannamei* in ponds

上海海洋大学学报. 2020, 29(5): 641 <https://doi.org/10.12024/jsou.20200302951>

暗纹东方立体种养模式和单养模式的水质变化及养殖能效比较

Comparative analysis of water quality changes and cultural efficiencies between multistory culture model and traditional monoculture model of obscure puffer (*Takifugu obscurus*)

上海海洋大学学报. 2021, 30(5): 789 <https://doi.org/10.12024/jsou.20200603090>

淡水集中连片池塘与养殖尾水处理系统的综合水质评价

Comprehensive water quality evaluation of freshwater concentrated continuous ponds and aquaculture tail water treatment system

上海海洋大学学报. 2022, 31(1): 170 <https://doi.org/10.12024/jsou.20200402987>

过硫酸氢钾干预下对虾养殖系统中水质指标和菌群结构变化分析

Effects of potassium monopersulfate on water environment index and microbial community structure of *Litopenaeus vannamei* in pond culture system

上海海洋大学学报. 2022, (2): 452 <https://doi.org/10.12024/jsou.20210203287>

养殖密度对硝化型生物絮团系统中凡纳滨对虾生长和水质的影响

Effect of different stocking density on growth performance of *Litopenaeus vannamei* and water quality in nitrifying bio-floc system

上海海洋大学学报. 2020, 29(1): 27 <https://doi.org/10.12024/jsou.20180402300>

文章编号: 1674-5566(2023)01-0108-10

DOI:10.12024/jso.20220903980

基于 PCA-LSTM 神经网络的凡纳滨对虾养殖水质预测

刁文双¹, 江敏^{1,2}, 吴昊³, 潘璠¹, 唐燕¹

(1. 上海海洋大学 海洋生态与环境学院, 上海 201306; 2. 上海海洋大学 海洋水域环境生态上海高校工程研究中心, 上海 201306; 3. 上海海洋大学 水产与生命学院, 上海 201306)

摘要: 基于上海奉贤区 2 个水产养殖合作社 2014—2018 年和 2021 年的检测数据, 选取水温 (T)、溶解氧 (DO)、高锰酸盐指数 (I_{Mn})、总磷 (TP)、总氮 (TN)、氨氮 (TAN)、亚硝酸盐氮 (NO_2^- -N)、硝酸盐氮 (NO_3^- -N) 共 8 个水质指标进行研究, 提出了基于主成分分析 (Principal component analysis, PCA) 和长短时记忆神经网络 (Long short-term memory neural network, LSTM) 的预测模型。首先采用主成分分析法对数据进行特征提取和降维, 选取高锰酸盐指数 (I_{Mn}) 和氨氮 (TAN) 作为水质预测指标, 建立基于 PCA 法的 LSTM 模型; 接着采用 PCA-LSTM 模型对不同养殖塘的水质进行预测; 最后, 将其与单一 LSTM 模型进行对比以验证模型的优劣。结果表明: PCA-LSTM 模型可用于凡纳滨对虾养殖池塘水中 I_{Mn} 和 TAN 的预测, 预测结果优于单一 LSTM 模型。

关键词: 凡纳滨对虾; 养殖水质预测; 长短时记忆神经网络; 主成分分析

中图分类号: S 968.22 **文献标志码:** A

凡纳滨对虾 (*Litopenaeus vannamei*) 作为世界上养殖最广泛的对虾品种之一, 具有适应性强、生长快、肉质鲜美等特点^[1]。近些年随着养殖产业的不断发展, 一些细菌或病毒性病害频繁爆发^[2]、养殖水质恶化, 极大程度地影响了对虾的生长繁殖。因此有必要加强养殖水质监测^[3], 提前预测水质状况, 以便及时采取水质调控措施, 有效降低养殖风险, 提高养殖效益。对于实际养殖场而言, 极端天气如台风、暴雨等会造成养殖池塘水质突变, 从而影响预测效果。同时养殖池塘会受到人为活动的干预, 水质情况更为复杂多变, 也为养殖水质的预测增加一定的难度。

时间序列预测是对历史数据进行挖掘和分析, 并据此预测未来变化的一种方法。传统的时间序列预测主要有向量自回归 (Vector autoregressive model, VAR) 和自回归积分移动平均模型 (Autoregressive integrated moving average model, ARIMA) 等^[4], 它们通常假设变量之间存

在线性相关性, 这些必要的假设限制了模型的应用, 计算的复杂度随着变量数量的增加呈指数增长^[5]。近年来, 深度学习技术逐步兴起, 而同一时刻多个水质因子的含量会相互影响^[6], 因此提出了一系列基于深度学习的多元时间序列预测方法。例如循环神经网络 (Recurrent neural networks, RNN)、长短时记忆神经网络 (Long short-term memory neural network, LSTM) 等, 它们可用于处理序列波动变化较大并存在长期趋势的时间序列, 而 LSTM 有效地解决了 RNN 训练时存在梯度消失和梯度爆炸等问题, 大大提高了网络的准确性, 已广泛应用于各领域, 例如天气预报^[7]、交通预测^[8]、水质预测^[9]等。曹守启等^[10]以河蟹养殖溶解氧为研究对象, 建立 K-means 聚类和改进粒子群优化的 LSTM 模型, 和单一 LSTM 模型相比, 预测结果更接近真实值。孙龙清等^[11]为提高河蟹集约化养殖池塘中溶解氧含量的预测精度, 提出一种基于改进的天牛须搜索算法

收稿日期: 2022-09-16 修回日期: 2022-11-03

基金项目: 上海市虾类产业技术体系建设项目 (沪农科产字[2022]第 5 号); 广东省重点领域研发计划项目 (2020B0202010009); 上海市科技兴农项目 (沪农科创字[2019]第 3-5 号)

作者简介: 刁文双 (1999—), 女, 硕士研究生, 研究方向为环境化学。E-mail: M200401154@st.shou.edu.cn

通信作者: 江敏, E-mail: mjiang@shou.edu.cn

版权所有 © 《上海海洋大学学报》编辑部 (CC BY-NC-ND 4.0)

Copyright © Editorial Office of Journal of Shanghai Ocean University (CC BY-NC-ND 4.0)

<http://www.shhydx.com>

(Improved beetle antennae search algorithm, IBAS) 和 LSTM 相结合的模型,该模型 MSE、RMSE、MAE 分别为 0.644、0.803、0.531,预测精度较单一 LSTM 模型有较大提升。陈英义等^[12]以淡水养殖的杂交鲮鱼为研究对象,建立 PCA-LSTM 模型预测溶解氧,模型 MSE、RMSE、MAE 分别为 0.274、0.089 和 0.147,均优于单一 LSTM 模型。以上模型的预测效果较好,说明时间序列预测同样适用于养殖水质。

在水产养殖领域,时间序列预测模型大多聚焦于溶解氧的预测,对于其他影响养殖活动的重要水质因子,例如高锰酸盐指数(I_{Mn})、氨氮(TAN)、总氮(TN)、总磷(TP)等的预测研究则较少。同时水质变化作用机理复杂,一种模型预测精度存在局限性,因此本文以凡纳滨对虾养殖池塘为研究对象,采用主成分分析法降维,构建 PCA-LSTM 水质组合预测模型,为今后凡纳滨对虾养殖水质预测提供参考。

1 材料与方法

1.1 数据来源

实验样本来源于上海市奉贤区的 2 个水产养殖合作社,分别记为 A 和 B。为满足研究对于数据的要求,每个养殖场选取固定的凡纳滨对虾土池养殖塘,试验期间每 5 天对养殖池塘进行水质检测,检测指标包括水温(T)、溶解氧(DO)、高锰酸盐指数(I_{Mn})、总磷(TP)、总氮(TN)、氨氮(TAN)、亚硝酸盐氮(NO_2^- -N)、硝酸盐氮(NO_3^- -N)等。其中水温(T)、溶解氧(DO)采用美国 YSLProDO 溶氧仪测定, I_{Mn} 、TP 分别采用酸性高锰酸钾法(GB 11892—89)、钼酸铵分光光度法(GB 11893—89)测定,TN、TAN、 NO_2^- -N、 NO_3^- -N 分别采用碱性过硫酸钾消解紫外分光光度法(GB11894—89)、纳氏试剂分光光度法(GB 7479—87)、分子吸收分光光度法(GB 7493—87)、紫外分光光度法(HJ/T 346—2007)进行测定。

1.2 数据预处理

1.2.1 缺失数据处理

由于原始数据存在少部分缺失,需要对空值进行填补。对于前后时间间隔较小的缺失数据,选择采用前后 5 天同一时刻的平均值进行填补;而对于时间跨度较大的缺失数据,选择天气

类型相近的时间进行补全^[13]。本研究中,原始数据的缺失量大约为 5%,异常值数据占比约为 1%。

1.2.2 异常数据处理

对于常规水质而言,前后相邻的检测数据一般情况下不会发生急剧的突变,因此若检测数据存在异常值,可采用拉依达准则(3σ 准则)进行筛选,其方法原理如下:

设样本数据为 X_1, X_2, \dots, X_n , 平均值为 \bar{X} , 偏差为 $v_i = X_i - \bar{X} (i = 1, 2, \dots, n)$, 按照 Bessel 公式计算其标准偏差(σ), 公式如下:

$$\sigma = [\sum v_i^2 / (n - 1)]^{1/2} \quad (1)$$

若某一样本数据 X_i 的偏差(v_i)满足,则认为异常数据,需要将其采用平均值代替进行平滑处理^[14]。

对于极易受到外界影响的养殖水质而言,暴雨等气候因素往往会使水质参数发生剧变,因此对于拉依达准则后发现的异常数据,将进一步参考当时的气候状况,酌情考虑是否剔除。

1.3 主成分分析法

主成分分析法(Principal components analysis, PCA)是一种降维的统计方法,它的原理是通过少数几个主成分来揭示多个变量间的内部结构,以替代原始变量,从而减少数据^[15]。而每个主成分都是原始变量的线性组合,且互不相关,可以有效地简化复杂问题^[16]。对于建模而言,通过 PCA 选择的主成分变量作为网络的输入,相关系数高的指标作为网络的输出,简化了网络的结构,提高了模型预测效果^[17]。具体算法步骤如下:(1)为消除量纲影响,对原始数据采用标准化处理,得到标准化矩阵。(2)进行 KMO 检验和 Bartlett 检验,确定数据是否适合进行 PCA 分析。(3)通过标准化矩阵计算相关系数矩阵。(4)求相关系数矩阵的特征根和特征向量,并计算贡献率和累计贡献率。(5)获得主成分矩阵。得到确定后的多个主成分,同时计算各主成分得分。

1.4 基本 LSTM 网络模型

长短时记忆神经网络(LSTM)能够深入挖掘时间序列中的固有规律^[18]。它的结构如图 1 所示。其中包括输入门 i_t , 遗忘门 f_t 和输出门 o_t , 存储单元 C_t 用来存储当前时刻历史信息。 x_t 是当前时间步的输入, C_{t-1} 是上一个时间步的单位状态, h_{t-1} 是上一个时间步的隐藏状态, C_t 是当前

时间步的单位状态, h_t 是当前时间步的隐藏状态。LSTM 单元的运行过程分为 3 个阶段, 第一阶段: 遗忘门决定了遗忘的程度, C_{t-1} 被遗忘, 如方程式(2)所示; 第二阶段: 输入门确定要记忆的信息 x_t 和 h_{t-1} , 用于更新 C_{t-1} 到 C_t , 如方程式(3)和(4)所示; 第三阶段: 输出门决定 h_t , 如方程式(5)和(6)所示。

$$f_t = \sigma(W_{f_x} \times x_t + W_{f_h} \times h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(W_{i_x} \times x_t + W_{i_h} \times h_{t-1} + b_i) \otimes \tanh(W_{c_x} \times x_t + W_{c_h} \times h_{t-1} + b_c) \quad (3)$$

$$c_t = f_t \otimes C_{t-1} + i_t \quad (4)$$

$$o_t = \sigma(W_{o_x} \times x_t + W_{o_h} \times h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

式中: σ 为 Sigmoid 激活函数; \tanh 为双曲正切函数; \otimes 表示元素乘法运算; W_{f_x} 、 W_{i_x} 、 W_{c_x} 、 W_{o_x} 表示 x_t 的权重参数, W_{f_h} 、 W_{i_h} 、 W_{o_h} 表示 h_{t-1} 的权重参数; b_i 、 b_o 、 b_f 分别为 3 个控制门的偏置项^[19]。LSTM 神经元运用 3 个门结构控制信息的输入和输出, 让信息选择性地传输^[20-21]。

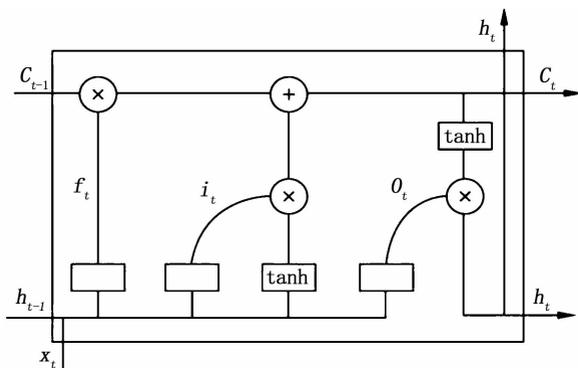


图 1 LSTM 单元结构

Fig. 1 Structure of LSTM cell

1.5 模型评价指标

单一的评价指标并不能完全反映模型的性能, 选择不同的评价指标可以对模型进行更加客观综合的评价^[22]。本文选用的预测模型性能评价指标有平均绝对误差 (MAE)、均方根误差 (RMSE)、平均绝对百分误差 (MAPE) 和相对误差 (Er)^[12]。MAE 和 RMSE 用于判定预测值与真实值之间的偏差, MAPE 主要描述预测值偏离真实值的百分比, 同时这 3 个评价指标的数值越小意味着模型预测效果越好。Er 用于计算预测值和真实值的偏差率, 从而反映测量的可信程度。

各评价指标的计算公式如下所示:

$$E_{\text{MAE}} = \frac{1}{m} \sum_{i=1}^m |y_i - y'_i| \quad (7)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - y'_i)^2} \quad (8)$$

$$E_{\text{MAPE}} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - y'_i}{y_i} \right| \quad (9)$$

$$E_r = \frac{y_i - y'_i}{y_i} \quad (10)$$

式中: m 为样本数; y_i 为第 i 时刻的真实值; y'_i 为第 i 时刻的预测值。

1.6 数据分析方法

采用 Microsoft Excel 2010 进行数据处理; 采用 IBM SPSS 26 软件对各水质指标进行主成分分析; 采用 MATLAB 2018a 软件建立长短时记忆神经网络模型; 采用 OriginPro 2022b 进行绘图。

2 结果与分析

2.1 水质检测结果

各水质因子在 2014—2018 年 A 养殖场和 2021 年 B 养殖场整个养殖周期的变化情况见表 1。其中 2014—2018 年合作社 A 的 1 号塘记为 A-1, 2018 年合作社 A 的 2 号塘记为 A-2, 2021 年合作社 B 的 1 号塘记为 B-1。

2.2 PCA-LSTM 模型的建立

2.2.1 模型预测指标的确定

养殖水质因子间存在复杂的耦合关系, 难以确定神经网络的最佳输入, 若均作为神经网络的输入变量, 会导致模型结构相对复杂, 难以得到理想的效果^[23], 因此本研究采用 PCA 选择的主成分变量作为 LSTM 网络的输入, 降低模型输入参数的维度, 消除了变量之间的冗余性和相关性。根据 1.3 节描述的 PCA 提取主成分的方法, 对 A-1 塘 2014—2018 年 8 项水质指标参数进行 PCA 主成分提取, 计算特征值与贡献率, 结果见表 2。

按照特征值大于 1 的原则, 提取 4 个主成分作为网络输入 (前 4 个主成分累计贡献率约为 74%, 表明其能够反映原始数据的绝大部分信息)。通过 PCA 分析将网络的输入由 8 维降低为 4 维, 优化了网络的输入, 得到的主成分变量分别命名为 F1、F2、F3 和 F4。同时计算得到主成分矩阵 (表 3)。

表 1 数据样本统计分析

Tab. 1 Statistical analysis of the data samples

水质因子 Water factors	A-1		A-2		B-1	
	平均值 ± 标准差 Mean ± SD	范围 Range	平均值 ± 标准差 Mean ± SD	范围 Range	平均值 ± 标准差 Mean ± SD	范围 Range
T/°C	27.500 ± 3.600	18.200 ~ 34.900	27.500 ± 4.300	18.700 ~ 35.400	26.800 ± 3.400	18.400 ~ 31.200
DO/(mg/L)	7.650 ± 2.200	2.420 ~ 15.710	7.700 ± 2.400	3.590 ~ 12.230	9.620 ± 1.580	6.600 ~ 15.400
I _{Mn} /(mg/L)	9.370 ± 3.500	2.320 ~ 19.600	10.910 ± 2.780	5.840 ~ 17.170	6.460 ± 1.170	3.950 ~ 8.000
TP/(mg/L)	0.498 ± 0.491	0.010 ~ 2.300	0.430 ± 0.171	0.256 ~ 1.061	0.261 ± 0.099	0.041 ~ 0.447
TN/(mg/L)	3.050 ± 1.535	0.600 ~ 9.790	4.733 ± 3.243	2.094 ~ 16.710	2.833 ± 1.033	1.100 ~ 4.300
NO ₂ ⁻ -N/(mg/L)	0.090 ± 0.133	0.001 ~ 0.750	0.057 ± 0.068	0.001 ~ 0.280	0.068 ± 0.076	0.003 ~ 0.339
NO ₃ ⁻ -N/(mg/L)	0.605 ± 0.373	0.080 ~ 2.250	1.206 ± 0.516	0.255 ~ 6.619	0.239 ± 0.159	0.016 ~ 0.537
TAN/(mg/L)	0.829 ± 0.619	0.120 ~ 3.650	1.206 ± 0.516	0.489 ~ 2.239	0.537 ± 0.337	0.070 ~ 1.348

注:A-1 为 2014—2018 的检测数据(共 175 组),A-2 为 2018 年的检测数据(共 35 组),B-1 为 2021 年的检测数据(共 35 组)。

Notes:A-1 is the sense data from 2014 to 2018(A total of 175 groups),A-2 is the sense data in 2018(A total of 35 groups),B-1 is the sense data in 2021(A total of 35 groups).

表 2 主成分分析

Tab. 2 Principal component analysis

成分 Component	特征值 Eigenvalue	初始贡献率 Initial contribution/%	累积贡献率 Cumulative contribution/%
1	2.168	27.104	27.104
2	1.365	17.060	44.164
3	1.305	16.313	60.477
4	1.064	13.298	73.774
5	0.657	8.210	81.984
6	0.588	7.350	89.334
7	0.484	6.053	95.387
8	0.369	4.613	100.000

表 3 主成分系数矩阵

Tab. 3 Coefficient matrix of the principal components

项目 Items	成分 1	成分 2	成分 3	成分 4
	Component 1	Component 2	Component 3	Component 4
T/°C	-0.183	-0.236	0.392	0.420
DO/(mg/L)	-0.187	0.464	-0.019	-0.385
TP/(mg/L)	-0.049	0.555	0.258	0.284
TN/(mg/L)	0.361	0.184	0.018	0.280
NO ₂ ⁻ -N/(mg/L)	0.280	-0.004	0.137	-0.579
NO ₃ ⁻ -N/(mg/L)	0.240	0.176	0.456	0.043
TAN/(mg/L)	0.350	-0.133	-0.152	0.167
I _{Mn} /(mg/L)	0.038	0.268	-0.543	0.303

马真^[24]以凡纳滨对虾为研究对象,得到对养殖水质影响较大的因子是 DO、水温、氨氮、pH、亚硝酸盐氮和 I_{Mn};潘璠等^[25]对凡纳滨对虾养殖池塘的浮游植物群落结构变化进行研究,结果表明养殖池塘中的浮游植物优势类群(硅藻门、蓝藻门、绿藻门)都与 TN、TP、I_{Mn}极显著正相关;唐燕等^[26]研究表明受人为活动干扰的凡纳滨对虾养殖塘后生浮游动物群落结构特征变化主要受到

TP 和 I_{Mn}的影响;水产养殖尾水污染物排放标准中规定的指标为悬浮物、pH、高锰酸盐指数、总磷和总氮。因此,就凡纳滨对虾养殖池塘而言,认为 TN、TP、I_{Mn}对养殖过程的影响较大。结合本研究中凡纳滨对虾实际养殖情况看,总磷的含量基本符合排放标准;总氮在养殖后期含量高,超过二级排放标准 5 mg/L,而总氮中的氨氮是对虾养殖过程中始终存在的有害物质,也是对虾的排泄物、残饵、粪便以及动植物尸体等含氮有机物分解的最终产物。综合考量,本实验最终选取 I_{Mn}、TAN 作为本研究的预测指标。

选取 A-1 塘作为数据样本,统一划分 80% (即 140 组数据)作为训练集,20% (即 35 组数据)作为测试集。根据上述主成分分析结果,得到主成分变量 F1、F2、F3 与 F4,并与 I_{Mn}和 TAN 的数据合并,作为 PCA-LSTM 模型的输入变量,分别预测下一时刻的 I_{Mn}和 TAN。PCA-LSTM 网络的输入层节点数为 5,输出层节点数为 1。

2.2.2 模型的超参数优化

分别采用 0.1、0.05、0.01、0.005、0.001 共 5 组学习率对训练集进行训练,其中迭代次数设定为 250,隐含层设定为 32 层,对比分析训练集学习率大小对模型训练效果的影响。在确定 I_{Mn}和 TAN 的最优学习率后,分别采用 50、100、150、200、250、300、350 共 7 组不同迭代次数对训练集进行训练,其中隐含层设定仍为 32 层,对比分析训练集迭代次数对模型训练效果的影响。上述确定学习率和迭代次数后,分别采用 2、4、8、16、32、64、128、256 个隐含层对训练集进行训练,对比隐含层数大小对模型训练效果的影响。以上

优化过程均采用网络训练过程的 RMSE、LOSS 进行判断, RMSE、LOSS 越小代表模型效果越好。不同学习率、迭代次数、隐含层的 RMSE、LOSS 对

比如图 2 所示, 最终确定 I_{Mn} 和 TAN 的最佳学习率均为 0.05; I_{Mn} 和 TAN 的最佳迭代次数均为 300; I_{Mn} 和 TAN 的最佳隐含层分别为 64 和 32。

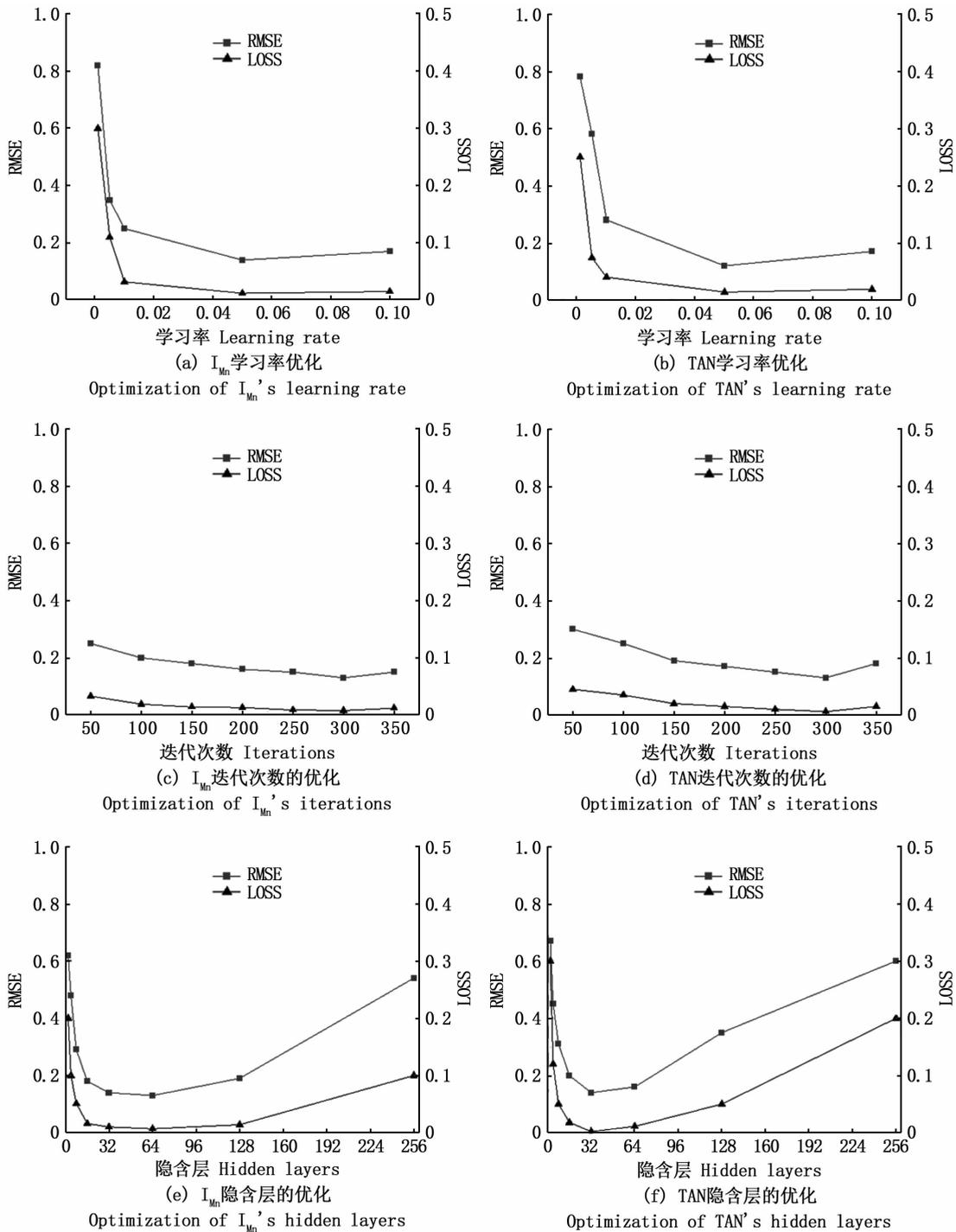


图 2 I_{Mn} 和 TAN 的超参数优化

Fig. 2 Hyperparameter optimization of the I_{Mn} and TAN

2.2.3 模型预测结果

优化模型超参数后,将测试集(即 2018 年 A-1 塘)的数据代入模型,进一步调整模型参数,得到最优模型。图 3 比较了 2018 年 A-1 塘 I_{Mn} 、TAN 实测值与预测值的变化趋势,并给出预测值的 95% 置信区间。从图中可以看出, I_{Mn} 和 TAN

的预测值和真实值大体上趋于一致,仅有个别点位的真实值超出了预测值 95% 的置信区间,模型预测效果较好。因此,认为该模型对于凡纳滨对虾养殖水体的水质预测方面的应用具有一定的参考价值。

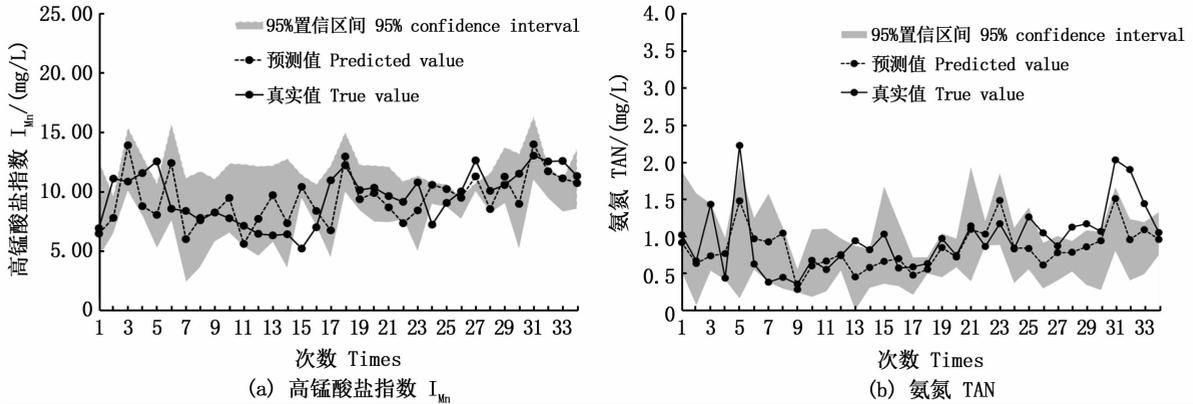


图 3 2018 年 A-1 塘 PCA-LSTM 模型的预测结果
Fig. 3 Prediction results of the PCA-LSTM model in pond A-1 in 2018

为了更加直观地展示预测值和真实值的偏差,采用公式(10)计算得到的相对误差进行表征。从图 4(a)中可以看出, I_{Mn} 的整体误差相对较小,仅第 5 次和第 18 次的相对误差较大,预测值比真实值分别低了 4.66 mg/L 和 5.98 mg/L; 而 TAN 亦是如此,个别点位有一定误差,但总体上趋于一致,主要是第 23 次和第 32 次的预测值比真实值低,分别低了 0.69 mg/L、0.98 mg/L。这可能是暴雨和台风天气后,养殖水质突变,导

致预测结果和真实值存在偏离。另外,该养殖场采用的是露天土池养殖,养殖过程中饲料投喂量、养殖密度和对虾生长状况等也会对养殖水质产生一定影响,但若将这些影响因素加入,模型体系就过于庞大,而水质在很大程度上也是这些因素综合呈现的结果,所以本研究未将这些因素纳入模型的输入参数,这可能也是模型预测存在一定偏差的原因。

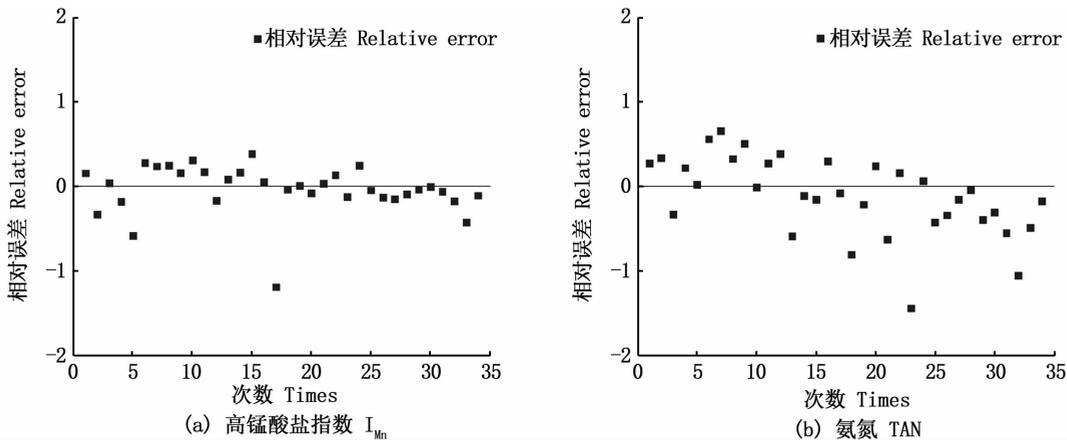


图 4 2018 年 A-1 塘 PCA-LSTM 模型的相对误差
Fig. 4 Relative error of the PCA-LSTM model in pond A-1 in 2018

2.3 PCA-LSTM 模型与单一 LSTM 模型的预测效果比较

为了进一步验证 PCA-LSTM 模型的预测效果,本文采用单一 LSTM 模型进行比较。利用 A-1 塘 2014—2017 年的数据建立了单一 LSTM 模型,单一 LSTM 模型的输入节点数为 8,输出节点数为 1,其中网络结构设置、参数的优化以及验证集的预测均与 2.2 节一致,最终得到 I_{Mn} 和 TAN 的最优模型参数。 I_{Mn} 模型参数设置为学习率 0.1,迭代次数为 200,隐含层为 32 层;TAN 的模型参数设置为学习率 0.05,迭代次数为 200,隐含层为

64 层。

利用建立的 PCA-LSTM 和单一 LSTM 模型,对 2018 年 A-1 塘、A-2 塘和 2021 年 B-1 塘的 I_{Mn} 和 TAN 进行预测。为评价模型预测效果,采用公式(7)、(8)、(9)计算 MAE、RMSE 和 MAPE 值。两种模型预测结果的指标评价价值如图 5 所示。从图中可以看出,PCA-LSTM 模型的预测评价指标值均小于单一 LSTM 模型。结果表明,与单一 LSTM 预测模型相比,经过 PCA 法对数据降维后的 LSTM 模型的泛化能力更强,预测精度更高。

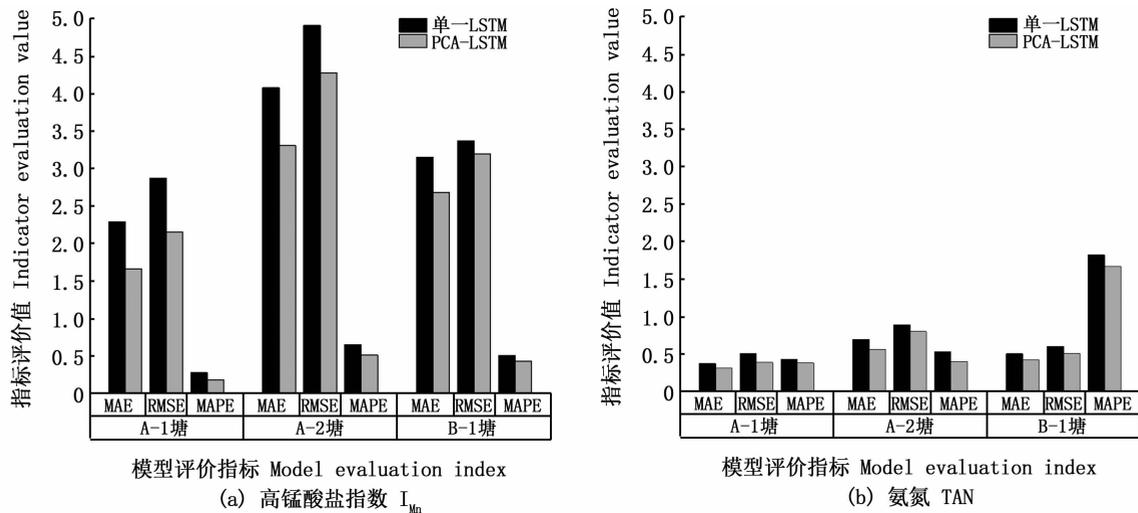


图 5 两种模型预测结果精度分析

Fig. 5 Forecast accuracy of two prediction models

3 讨论

3.1 PCA-LSTM 模型的优势

目前在水产养殖领域,为进一步对水质变化进行预判和预警,可采用长短时记忆神经网络预测未来时刻水质因子的变化,但预测难度会大大增加^[27]。而有研究表明采用 PCA 法可以有效提高模型的预测效果,张梦迪等^[13]采用 PCA 法筛选的水质预测指标,不仅可以很好地反映监测断面的整体水质状况,还可以简化网络结构,提高预测效果;袁红春等^[28]在中华绒螯蟹养殖水环境中,采用主成分分析法提取主成分变量,建立非线性有源自回归模型(Nonlinear autoregressive models with exogenous inputs, NARX)对氨氮进行预测,研究表明,利用 PCA 降维后的数据较原始数据预测效果更好。本研究结果表明,在进行有

较强非线性特征的养殖水质预测研究中,基于 PCA 的 LSTM 神经网络模型相较于单一的 LSTM 模型而言,预测精度较高,与前人的报道具有较好的一致性。PCA 法实现了模型输入的降维,显著减少了神经网络输入数据的维度,消除了变量之间的冗余性和相关性,提升了 LSTM 的预测精度,同时证实了先降维、后预测得到的模型预测效果更佳的说法^[29]。

3.2 影响模型预测效果的因素分析

众多研究表明,水质指标可以在很大程度上反映养殖过程和人为活动。金晶等^[30]以中华绒螯蟹养殖试验池塘为例,研究投喂量对水质变化的影响,发现高低饲料投喂量会导致水中溶解氧、氨氮质量浓度的上升或下降。刘忆瀚等^[31]研究结果表明微生态制剂对凡纳滨对虾的生长性能、酶活力以及养殖水质均有一定影响,对高锰

酸盐指数、氨氮、亚硝酸盐氮、总氮均有一定降解作用。本研究得到的 PCA-LSTM 模型整体预测效果不错,仍有个别时刻预测结果的偏差较大,原因可能有两点:第一,台风、暴雨等极端天气易使养殖塘的水质发生突变,从而影响模型的预测效果;第二,为了满足模型对于数据的要求,训练时的网络输入的数据是在原始数据的基础上进行一系列预处理,比如剔除异常值、填补空缺值等等,模型对于偏离正常变化趋势的水质指标的预测可能出现偏差。综上,后续在对虾养殖池塘水质预测中,可对模型进行以下改进:第一,加入气象因子、养殖密度和饲料投喂量等变量作为模型的输入;第二,提高采样频率,获得样本量大的数据进行建模;第三,进一步提高模型的预测效果。

4 结论

本文提出并构建了基于 PCA-LSTM 神经网络的凡纳滨对虾养殖水质预测模型,预测未来时刻的高锰酸盐指数和氨氮的变化趋势,可得出如下结论。

(1)采用 PCA 方法优化筛选 I_{Mn} 和 TAN 作为水质预测指标,同时得到的主成分变量能够较好地反映凡纳滨对虾养殖水质的整体情况,实现了模型输入数据的降维,消除了变量之间的冗余性和相关性。

(2)本文提出对 3 个不同养殖池塘的数据进行预测,结果表明 PCA-LSTM 模型预测效果均显著优于单一 LSTM 模型。PCA 方法对变量进行特征选择后确定模型输入,可使预测的精度得以提升。

参考文献:

- [1] HUANG H H, LIAO H M, LEI Y J, et al. Effects of different carbon sources on growth performance of *Litopenaeus vannamei* and water quality in the biofloc system in low salinity[J]. *Aquaculture*, 2022, 546: 737239.
- [2] LIAO G W, WU Q P, MO B H, et al. Intestinal morphology and microflora to *Vibrio alginolyticus* in Pacific white shrimp (*Litopenaeus vannamei*) [J]. *Fish & Shellfish Immunology*, 2022, 121: 437-445.
- [3] 倪蒙,原居林,储忝江,等. 环杭州湾围垦养殖区凡纳滨对虾(*Litopenaeus vannamei*)池塘水质特征及污染物排放研究[J]. *淡水渔业*, 2018, 48(2): 107-112.
- NI M, YUAN J L, CHU T J, et al. Water quality and

- pollutant discharge of *Litopenaeus vannamei* pond in the reclamation region around Hangzhou Bay [J]. *Freshwater Fisheries*, 2018, 48(2): 107-112.
- [4] MELESSE A M, AHMAD S, MCCLAIN M E, et al. Suspended sediment load prediction of river systems: An artificial neural network approach [J]. *Agricultural Water Management*, 2011, 98(5): 855-866.
- [5] CHEN Y W, DING F Q, ZHAI L B. Multi-scale temporal features extraction based graph convolutional network with attention for multivariate time series prediction [J]. *Expert Systems with Applications*, 2022, 200: 117011.
- [6] 刘晶晶,庄红,铁治欣,等. K-Similarity 降噪的 LSTM 神经网络水质多因子预测模型 [J]. *计算机系统应用*, 2019, 28(2): 226-232.
- LIU J J, ZHUANG H, TIE Z X, et al. Water quality multi-factor prediction model using LSTM neural network based on K-Similarity Noise reduction [J]. *Computer Systems & Applications*, 2019, 28(2): 226-232.
- [7] SOARES E, COSTA P, COSTA B, et al. Ensemble of evolving data clouds and fuzzy models for weather time series prediction [J]. *Applied Soft Computing*, 2018, 64: 445-453.
- [8] FANG W W, ZHOU W H, YAN J W, et al. Attention meets long short-term memory: A deep learning network for traffic flow forecasting [J]. *Physica A: Statistical Mechanics and its Applications*, 2022, 587: 126485.
- [9] 张贻婷,李天宏. 基于长短期记忆神经网络的河流水质预测研究 [J]. *环境科学与技术*, 2021, 44(8): 163-169.
- ZHANG Y T, LI T H. River water quality prediction based on long short-term memory neural network [J]. *Environmental Science & Technology*, 2021, 44(8): 163-169.
- [10] 曹守启,周礼馨,张铮. 采用改进长短期记忆神经网络的水产养殖溶解氧预测模型 [J]. *农业工程学报*, 2021, 37(14): 235-242.
- CAO S Q, ZHOU L X, ZHANG Z. Prediction model of dissolved oxygen in aquaculture based on improved long short-term memory neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37(14): 235-242.
- [11] 孙龙清,吴雨寒,孙希蓓,等. 基于 IBAS 和 LSTM 网络的池塘水溶解氧含量预测 [J]. *农业机械学报*, 2021, 52(s1): 252-260.
- SUN L Q, WU Y H, SUN X B, et al. Dissolved oxygen prediction model in ponds based on improved beetle antennae search and LSTM network [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(s1): 252-260.
- [12] 陈英义,程倩倩,方晓敏,等. 主成分分析和长短期记忆神经网络预测水产养殖水体溶解氧 [J]. *农业工程学报*, 2018, 34(17): 183-191.
- CHEN Y Y, CHENG Q Q, FANG X M, et al. Principal component analysis and long short-term memory neural

- network for predicting dissolved oxygen in water for aquaculture [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(17): 183-191.
- [13] 张梦迪, 徐庆, 刘振鸿, 等. 基于动态滑动窗口 BP 神经网络的水质时间序列预测 [J]. 环境工程技术学报, 2022, 12(3): 809-815.
ZHANG M D, XU Q, LIU Z H, et al. Prediction of water quality time series based on the dynamic sliding window BP neural network model [J]. Journal of Environmental Engineering Technology, 2022, 12(3): 809-815.
- [14] 刘双印. 基于计算智能的水产养殖水质预测预警方法研究 [D]. 北京: 中国农业大学, 2014.
LIU S Y. Prediction and early-warning of water quality in aquaculture based on computational intelligence [D]. Beijing: China Agricultural University, 2014.
- [15] BUZZELL G A, NIU Y B, AVIYENTE S, et al. A practical introduction to EEG time-frequency principal components analysis (TF-PCA) [J]. Developmental Cognitive Neuroscience, 2022, 55: 101114.
- [16] 于春蕾, 李梦悦, 尹伟石. 基于 PCA-BP 组合模型的股价预测分析 [J]. 长春理工大学学报 (自然科学版), 2021, 44(4): 125-130.
YU C L, LI M Y, YIN W S. Analysis of stock price forecast based on PCA-BP combination model [J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2021, 44(4): 125-130.
- [17] ASANTE-OKYERE S, SHEN C B, ZIGGAH Y Y, et al. Principal component analysis (PCA) based hybrid models for the accurate estimation of reservoir water saturation [J]. Computers & Geosciences, 2020, 145: 104555.
- [18] ZHI W, FENG D P, TSAI W P, et al. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? [J]. Environmental Science & Technology, 2021, 55(4): 2357-2368.
- [19] ZHANG Y T, LI C L, JIANG Y Q, et al. Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model [J]. Journal of Cleaner Production, 2022, 354: 131724.
- [20] 王海涛, 宋文, 王辉. 一种基于 LSTM 和 CNN 混合模型的文本分类方法 [J]. 小型微型计算机系统, 2020, 41(6): 1163-1168.
WANG H T, SONG W, WANG H. Text classification method based on hybrid model of LSTM and CNN [J]. Journal of Chinese Computer Systems, 2020, 41(6): 1163-1168.
- [21] 商艳红, 张静. 基于局部化双向 LSTM 和状态转移约束的养殖水质分类预测 [J]. 渔业现代化, 2019, 46(2): 28-34.
SHANG Y H, ZHANG J. Aquaculture water quality prediction based on local Bi-LSTM and state transformation constraint [J]. Fishery Modernization, 2019, 46(2): 28-34.
- [22] 陈能汪, 余镒琦, 陈纪新, 等. 人工神经网络模型在水质预警中的应用研究进展 [J]. 环境科学学报, 2021, 41(12): 4771-4782.
CHEN N W, YU Y Q, CHEN J, et al. Artificial neural network models for water quality early warning: A review [J]. Acta Scientiae Circumstantiae, 2021, 41(12): 4771-4782.
- [23] 姚启, 缪新颖. 基于主成分分析及 GA-LM 的水产养殖环境溶解氧和氨氮含量预测 [J]. 大连海洋大学学报, 2021, 36(5): 851-858.
YAO Q, MIAO X Y. Prediction of dissolved oxygen and ammonia nitrogen concentrations in aquaculture environment based on PCA and GA-LM [J]. Journal of Dalian Ocean University, 2021, 36(5): 851-858.
- [24] 马真. 凡纳滨对虾集约化养殖水质预警模型的研究 [D]. 青岛: 中国海洋大学, 2010.
MA Z. Study on pre-alarming water quality system for intensively cultured *Litopenaeus vannamei* [D]. Qingdao: Ocean University of China, 2010.
- [25] 潘璠, 江敏, 吴昊, 等. 凡纳滨对虾养殖池塘及外河道的浮游植物群落结构变化 [J/OL]. 水产学报: 1-13. (2022-08-24). <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=SCKX20220822000&uniplatform=NZKPT&language=gb>.
PAN F, JIANG M, WU H, et al. Changes of phytoplankton community structure in *Litopenaeus vannamei* culture ponds and outer river [J/OL]. Journal of Fisheries of China: 1-13. (2022-08-24). <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=SCKX20220822000&uniplatform=NZKPT&language=gb>.
- [26] 唐燕, 江敏, 吴昊, 等. 凡纳滨对虾养殖池塘及外河道中后生浮游动物群落结构差异及其影响因素 [J]. 上海海洋大学学报, 2022, 31(5): 1120-1135.
TANG Y, JIANG M, WU H, et al. Differences in metazooplankton community structure and its influencing factors in *Litopenaeus vannamei* aquaculture ponds and outer channels [J]. Journal of Shanghai Ocean University, 2021, 31(5): 1120-1135.
- [27] 李道亮, 刘畅. 人工智能在水产养殖中研究应用分析与未来展望 [J]. 智慧农业, 2020, 2(3): 1-20.
LI D L, LIU C. Recent advances and future outlook for artificial intelligence in aquaculture [J]. Smart Agriculture, 2020, 2(3): 1-20.
- [28] 袁红春, 赵彦涛, 刘金生. 基于 PCA-NARX 神经网络的氨氮预测 [J]. 大连海洋大学学报, 2018, 33(6): 808-813.
YUAN H C, ZHAO Y T, LIU J S. Ammonia nitrogen level forecasting based on PCA-NARX neural network [J]. Journal of Dalian Ocean University, 2018, 33(6): 808-813.
- [29] 张可, 崔乐. 基于 PCA-LSTM 模型的多元时间序列分类算法研究 [J]. 统计与决策, 2020, 36(15): 44-49.
ZHANG K, CUI L. Research on multivariate time series classification algorithm based on PCA-LSTM model [J].

- Statistics & Decision, 2020, 36(15): 44-49.
- [30] 金晶, 毛星, 戴红君, 等. 基于分布滞后非线性模型的蟹塘水质受投喂量影响的关系分析[J]. 江苏农业学报, 2022, 38(4): 994-1002.
- JIN J, MAO X, DAI H J, et al. Analysis of the relationship between water quality of crab pond and feeding amount of feedstuff based on distributed lag non-linear model [J]. Jiangsu Journal of Agricultural Sciences, 2022, 38(4): 994-1002.
- [31] 刘忆瀚, 蔺凌云, 尹文林, 等. 微生态制剂对凡纳滨对虾生长·酶活及养殖水质的影响[J]. 安徽农业科学, 2020, 48(15): 96-101, 104.
- LIU Y H, LIN L Y, YIN W L, et al. Effects of probiotics preparation on the growth, enzyme activities, aquatic water of *Litopenaeus vannamei* [J]. Journal of Anhui Agricultural Sciences, 2020, 48(15): 96-101, 104.

Prediction of water quality in *Litopenaeus vannamei* aquaculture ponds based on the PCA-LSTM neural network model

XI Wenshuang¹, JIANG Min^{1,2}, WU Hao³, PAN Fan¹, TANG Yan¹

(1. College of Marine Ecology and Environment, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai University Engineering Research Center for Water Environment Ecology, Shanghai Ocean University, Shanghai 201306, China; 3. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Based on the detection data of the two farms in Fengxian District of Shanghai from 2014 to 2018 and in 2021, 8 water quality indicators, including the water temperature (T), dissolved oxygen (DO), permanganate index (I_{Mn}), total phosphorus (TP), total nitrogen (TN), ammonia nitrogen (TAN), nitrite nitrogen (NO_2^- -N) and nitrate nitrogen (NO_3^- -N) were chosen to establish a prediction model based on principal component analysis (PCA) and long short-term memory (LSTM). Firstly, through principal component analysis which was used to reduce data feature extraction and dimension, I_{Mn} and TAN were determined to be the water quality prediction indexes to build a LSTM model based on the PCA analysis, then the PCA-LSTM model was used to predict the water quality of different ponds; Finally, comparison was carried out with a single LSTM model to verify the strengths and weaknesses of both models. The results show that the PCA-LSTM model can be used to predict I_{Mn} and TAN in *Litopenaeus vannamei* aquaculture ponds, and the prediction results are better than the single LSTM model.

Key words: *Litopenaeus vannamei*; aquaculture water quality prediction; long short-term memory neural network; principal component analysis