

文章编号: 1674-5566(2023)05-0943-10

DOI: 10.12024/jsou.20230604222

基于特征构造预处理与 TCN-BiGRU 的池塘溶解氧预测模型

张 铮¹, 高 森¹, 张泽扬²

(1. 上海海洋大学 工程学院, 上海 201306; 2. 上海城市电力发展有限公司, 上海 200123)

摘 要: 为实现对池塘溶解氧变化趋势的准确和可靠预测, 降低池塘养殖风险, 提出了一种基于特征构造预处理 (Feature construction, FC) 与时间卷积网络 (Temporal convolutional network, TCN)-双向门控循环单元 (Bidirectional gate recurrent unit, BiGRU) 的溶解氧预测模型 FC-TCN-BiGRU。通过对样本序列的统计特征、环境因子特征和季节特征进行构造, 挖掘变量间深层次的相关性; 采用 TCN 对构造特征序列进行多层卷积和降维处理, 在保留全局时序特征的同时去除冗余信息; 结合 BiGRU 对降维特征进行建模, 实现对溶解氧的准确预测。另外, 构建了沙猫群算法 (Sand cat swarm optimization, SCSO) 优化的非参数核密度估计 (Kernel density estimation, KDE) 对溶解氧预测误差的分布范围进行估计。实验结果表明, 模型的均方误差、平均绝对误差、均方根误差和决定系数分别为 0.027 5、0.143 2、0.165 8 和 0.94, 优于其他比较模型。同时, 区间估计能够有效覆盖溶解氧的波动范围, 量化预测过程中的不确定性。在池塘溶解氧的短期预测中, 本模型具有较好的准确性和鲁棒性, 对水产养殖的水体健康监测和养殖效率的提升具有一定的指导意义。

关键词: 池塘养殖; 溶解氧; 特征构造; 预测模型; 区间估计

中图分类号: TP 391; TP 23 **文献标志码:** A

溶解氧是集约化水产养殖过程中的主要限制因素之一: 缺氧环境会造成鱼类生长迟缓^[1], 免疫下降和死亡^[2]; 富氧环境则会增加鱼类气泡病的患病概率^[3]。由于池塘养殖环境开放, 溶解氧受到多种生态因素的共同影响^[4-5], 表现出非线性和时变性的特点。因此, 对溶解氧的变化趋势和范围进行准确预测成为降低养殖风险和实现水质精确调控的关键^[6-7]

目前, 针对溶解氧预测算法的研究可分为基于前馈神经网络模型和基于循环神经网络模型两类, 如: CAO 等^[8]采用经验模态分解对溶解氧序列进行时序划分, 利用正则化极限学习机 (RELM) 进行组合预测, 使得模型具有较好的泛化能力。REN 等^[9]提出一种遗传算法优化的模糊神经网络对鱼菜共生系统中的溶解氧进行预测, 克服了传统模型稳定性差的问题。然而以上这些前馈神经网络模型无法捕捉数据间的时间相关性, 以 LSTM 和 GRU 为代表的循环神经网络在

时间序列预测任务中表现出良好性能^[10], 已经广泛应用于水质预测建模。CAO 等^[11]利用 K 均值聚类对溶解氧序列进行聚类分簇, 通过主成分分析法对数据进行降维, 结合 GRU 对鲈鱼养殖水质进行预测, 但模型参数要根据主观经验进行调整, 具有一定盲目性。GUO 等^[1]在此基础上, 将路径查找算法加入模型, 实现对关键参数的自动寻优, 提高了模型的短期预测精度。

近年来, 为了提高预测算法在复杂环境下的鲁棒性, 数据分解、特征提取和注意力机制等技术开始应用于循环神经网络模型中。YU 等^[12]将原始水质数据分解为若干子序列后, 利用 c 均值聚类对子序列进行重组, 然后针对每个聚类序列建立 BiGRU 预测模型, 减少噪声对预测结果的干扰。陈英义等^[13]通过小波变换和卷积网络对数据进行降噪并挖掘变量间的潜在信息, 解决了传统溶解氧模型的数据冗余问题。CAO 等^[14]引入注意力机制增强了 GRU 对输入数据中重要信息

收稿日期: 2023-07-26 修回日期: 2023-09-07

基金项目: 上海市水产动物良种创制与绿色养殖协同创新中心项目 (2021 科技 02-12); 上海市崇明区农业科创项目 (2021CNKC-05-06)

作者简介: 张 铮 (1981—), 男, 副教授, 研究方向为渔业物联网, 水产养殖智能化装备。E-mail: z-zhang@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>

的关注程度,使得模型可以更好地处理长序列场景。

但现有的溶解氧预测模型多集中于预测算法的改进,忽略了结合实际养殖经验对原始数据的进一步深层挖掘。同时,区间估计具有量化不确定性、提供未来概率信息的能力,已经广泛应用于船舶俯仰预测^[15]、大气污染物质量浓度预测^[16-17]、电力负荷预测^[18]、建筑能耗预测^[19]等领域,但在溶解氧预测中,仍然是很大的空白。

针对以上两个问题,提出了基于FC-TCN-BiGRU (Feature Construction and Temporal Convolutional Network-Bidirectional Gate Recurrent Unit)的区间预测模型。输入特征是决定预测精度的核心要素^[20-21],通过特征构造可以提高原始样本集的代表能力,从而使得TCN-BiGRU能够更准确地捕捉数据之间的潜在关系和模式。沙猫群算法优化的非参数核密度估计方法(Sand Cat Swarm Optimization-Kernel density estimation, SCSO-KDE)可以对预测误差进行概率分布拟合,将确定性预测结果拓展到概率分布区间,以提高溶解氧预测模型的可靠性和稳定性。

1 数据来源与方法

为了保证研究结果的可信度和真实性,开发一套池塘环境监测系统,并在研究区域进行布设,以采集实际水质和气象参数,构建研究所需的数据集。

1.1 研究区域

选取上海市崇明区上海城市电力公司水产养殖基地(北纬31.58°,东经121.84°)鲈鱼塘进行实验。鱼塘面积约为1.2 hm²,宽度40 m,长度288 m,水深2.5 m,养殖品种为加州鲈鱼(*Micropterus salmoides*)。

1.2 环境监测系统架构及数据采集

采用课题组设计研发的基于物联网的低功耗池塘环境监测系统,实现池塘水体和周围气象环境的实时采集和监测。水质终端节点采用浮标作为承载装置,并通过太阳能系统实现设备的自供电,水质传感器位于水下0.6 m处,可采集溶解氧、pH、水温、氧化还原电位、电导率等5种水质数据;在池塘的东南侧距地面2 m高度处安装气象传感器,采集同期温湿度、气压、光照、风速、风向、太阳辐射和雨量等8种气象数据。在微控

制器中搭载RT-thread嵌入式实时操作系统,对数据采集、处理、上传过程和低功耗设计进行多任务划分,并利用其同步和通信机制对各任务进行合理调度和及时响应。

所有数据均通过基于MQTT协议的4G无线通信模块上传到数据库进行数据存储,并转发到物联网智慧云平台进行实时显示和统计(图1)。选取2023年4月11日—5月25日,共45 d的水质和气象数据,采集间隔为10 min,共计6 480条数据。将样本集按照7:3的比例进行训练集和测试集的划分,并将预测模型用于2023年5月26、27两日的溶解氧预测验证。

2 FC-TCN-BiGRU溶解氧预测模型构建

2.1 整体流程

基于特征构造预处理与TCN-BiGRU算法构建应用于池塘溶解氧预测的模型,主要包含数据获取及清洗、数据特征构造、数据预测和区间估计4部分,其流程如图2所示。

首先,通过低功耗池塘环境监测系统获取实验区水质和气象数据。利用线性插值、Hermit插值等方法对原始数据集进行清洗,以减少异常噪声数据的干扰。其次,根据实际的养殖经验对影响溶解氧较大的因素进行分析和特征构造(FC),挖掘它们与溶解氧深层相关性。采用归一化对分类样本集进行处理,避免因不同量纲对结果造成影响,并通过Pearson相关系数进行分析,选取影响溶解氧的强构造特征,为提高溶解氧预测模型的精度提供高质量输入集。最后,将训练集数据输入到TCN-BiGRU模型中进行训练,采用测试集样本对1 h(6 steps)后的溶解氧进行滚动预测并得到误差值。误差作为一种预测的不确定性因素,通过SCSO-KDE算法对其进行拟合,得到误差概率密度分布函数。在提高预测精度的同时给出可能的区间分布范围,保证预测模型的稳定性和可靠性。

2.2 数据特征构造

2.2.1 分类统计特征构造

溶解氧的含量因昼夜交替而呈现不同的变化模式^[22]。日间溶解氧主要通过池塘中的浮游植物、水中藻类等进行光合作用而产生,约占总溶解氧含量的80%以上^[23]。夜间缺乏光照,导致各种水生生物持续进行氧气消耗,从而使溶解氧减少。

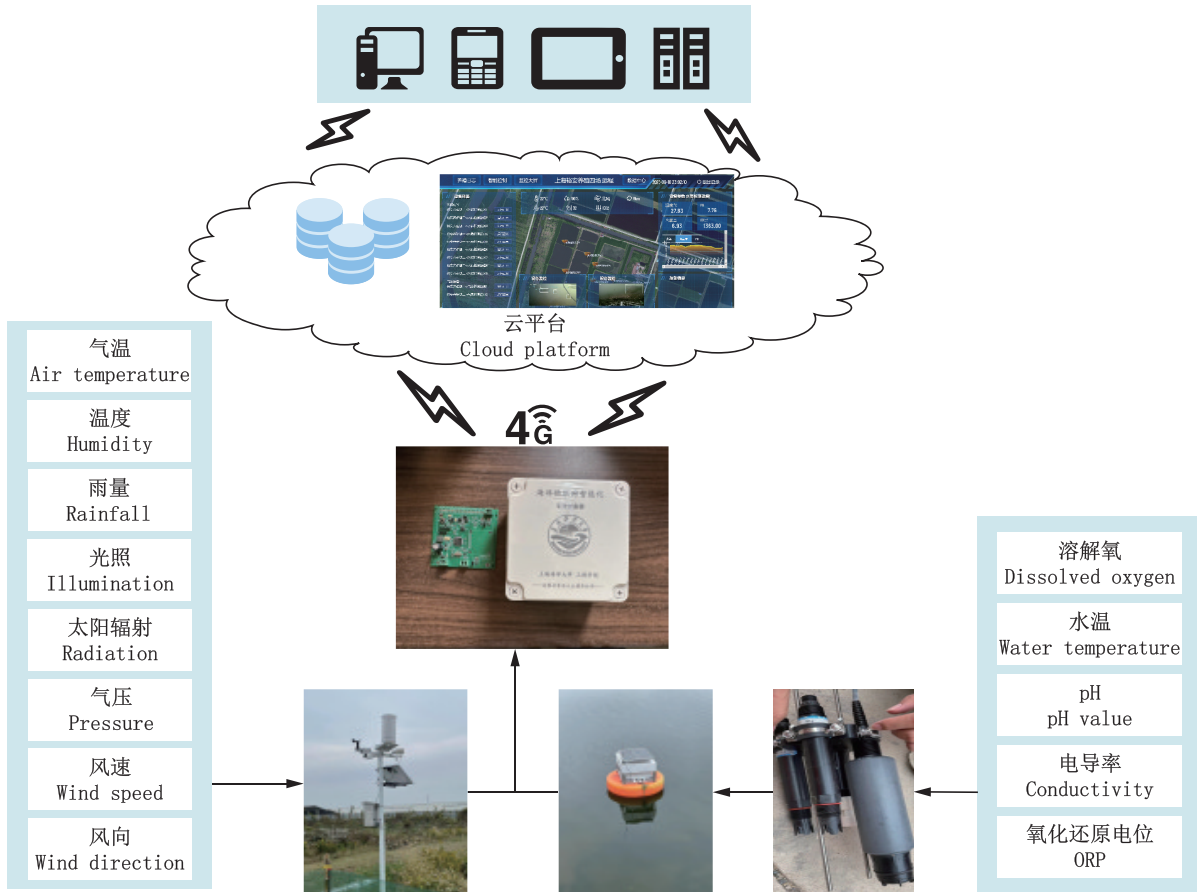


图1 基于物联网的低功耗池塘环境监测系统

Fig. 1 Low-power environmental monitoring system in pond based on internet of things

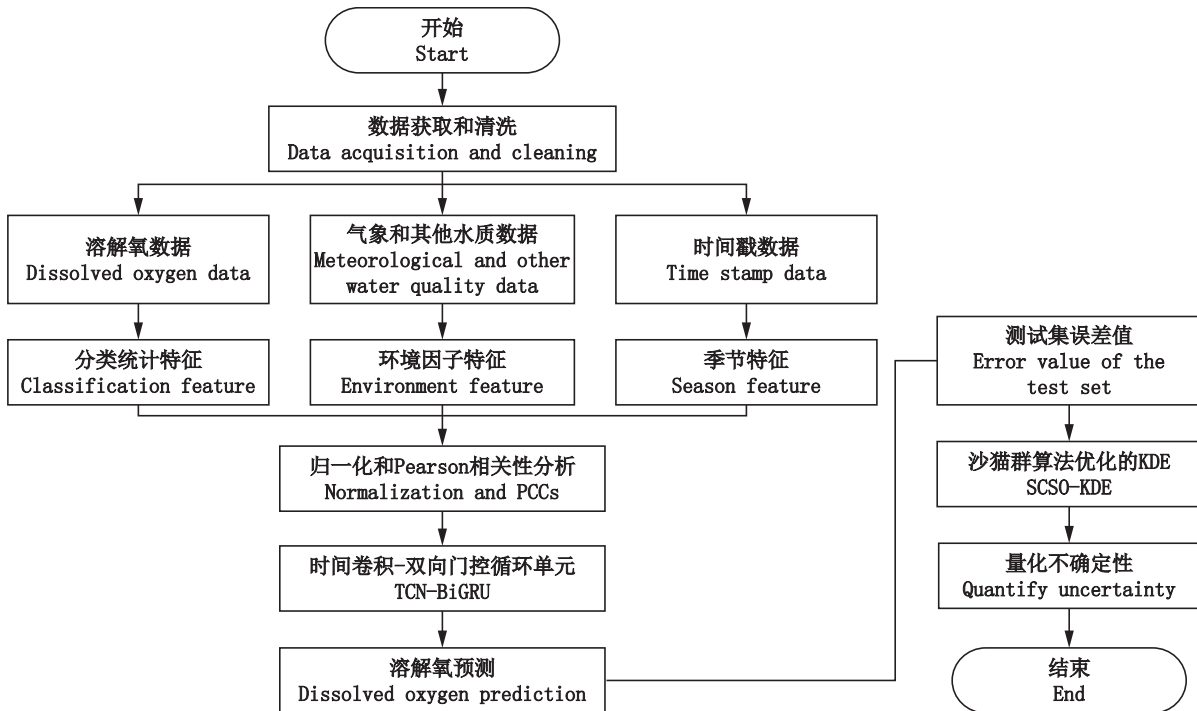


图2 总体框架流程图

Fig. 2 Overall framework flowchart

考虑到昼夜变化以及气象因素对光照时长的影响,根据光照强度对溶解氧数据进行了动态划分。当光照强度大于0 Lx时,将其归类为日间样本;当光照强度等于0 Lx时,将其归类为夜间样本。

2.2.2 环境因子特征构造

光照和太阳辐射强度是影响日间溶解氧的关键因素,而在相关性分析时由于未经数据特征构造无法被选作高质量影响因子^[4,8]。考虑到当前时刻的溶解氧含量实际是一段时期的累积结果,采用式(1)计算光照和太阳辐射的累积量并作为新的构造特征。

$$C(t) = \sum_{i=0}^t a_i \tag{1}$$

式中: a_i 为光照或太阳辐射在*i*时刻的采集值, $C(t)$ 为*t*时间段内的累积量。

2.2.3 季节特征构造

池塘鲈鱼的养殖时间主要集中在每年的4月到10月,其耗氧量随生长而逐渐增加。为帮助预

测模型识别不同的养殖时间节点,引入儒略日特征^[24]作为模型输入。考虑到鲈鱼苗投放时间的变动范围较大,而对于北半球中纬度地区而言,一年中最冷的时间相对固定(1月中旬左右),因此将1月15日的儒略日值置为-1,一年内的任一时间可采用式(2)计算:

$$I = -\cos \frac{2\pi(J - 15)}{365} \tag{2}$$

式中: J 为儒略日; I 为儒略日特征值。冬季时, I 接近-1;夏季时, I 接近1。

2.3 时间卷积-双向门控循环单元(TCN-BiGRU)

时间卷积网络是由BAI等^[25]提出的一种用于处理时间序列的卷积神经网络,其主要结构为包含扩张因果卷积的残差块(图3)。因果卷积可以保证输出结果仅依赖于过去的输入信息,从而有效避免未来信息的泄漏;扩张卷积允许在卷积时对输入进行间隔采样,在扩大感受野的同时解决多元时间序列信息的提取问题。

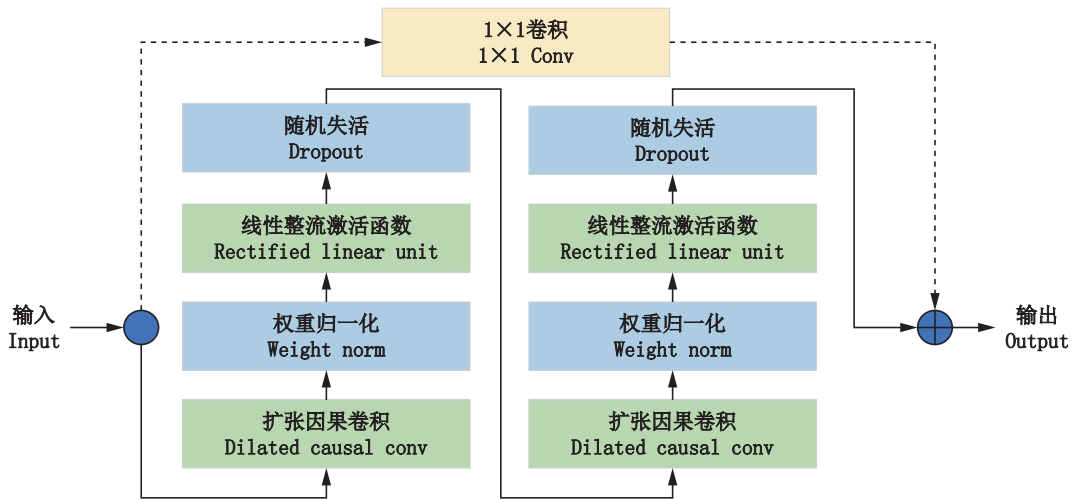


图3 TCN残差块

Fig. 3 Residual block in TCN

结合TCN和BiGRU的优势,构建溶解氧序列的预测模型,其架构如图4所示。首先,通过堆叠3层的TCN残差模块以获取更大范围的输入序列感受野并对特征进行提取和降维,同时避免出现梯度爆炸和梯度消失等问题每个残差块具有相同的内核大小*k*,其扩张因子*D*分别为1、2、4。BiGRU获取到TCN处理后的数据序列,它将正反两个方向的GRU层连接起来,一个按时间步从前往后(正向)处理输入序列,另一个反向处理。通

过这种方式,BiGRU可以更加完整地探索时序的依赖关系,获取上下文关联。最后,经全连接层将高维特征映射为最终预测结果。

2.4 溶解氧的区间估计

2.4.1 沙猫群优化算法

沙猫群优化算法是受沙猫搜索和狩猎行为启发而提出的一种生物启发式优化策略^[26]。在*d*维优化问题中,种群个数为*N*,每只个体代表问题解的1×*d*维数组 $X_i = (X_{i1}, X_{i2}, X_{i3}, \dots, X_{id})$ 。沙猫群算法的寻

优过程如公式(3)所示,分为搜索阶段和捕食阶段,通过自适应参数 R 的取值范围实现两阶段的过渡。

$$\vec{X}(t+1) = \begin{cases} \vec{r} \times (\vec{x}_b(t) - rand(0, 1) \times \vec{x}_c(t)), & |R| > 1 \\ \vec{x}_b(t) - \vec{r} \times \vec{x}_{rnd} \times \cos(\theta) & , |R| \leq 1 \end{cases} \quad (3)$$

式中: \vec{r} 模拟沙猫听觉感知的个体差异。 $rand(0,$

$1)$ 为 $0\sim 1$ 的随机数; t 为迭代次数; R 为自适应参数;当 $|R| > 1$ 时,沙猫处于搜索阶段,根据最佳位置 \vec{x}_b 、当前位置 \vec{x}_c 和 \vec{r} 更新到下一时刻的最佳位置 $\vec{X}(t+1)$ 。当 $|R| \leq 1$ 时,进入捕食猎物阶段, \vec{x}_{rnd} 是根据 \vec{x}_b 和 \vec{x}_c 所生成的随机位置, θ 的取值为 $0\sim 2\pi$,使沙猫可以沿不同的角度移动,避免陷入局部最优。经多次迭代,获取到问题空间的最优解。

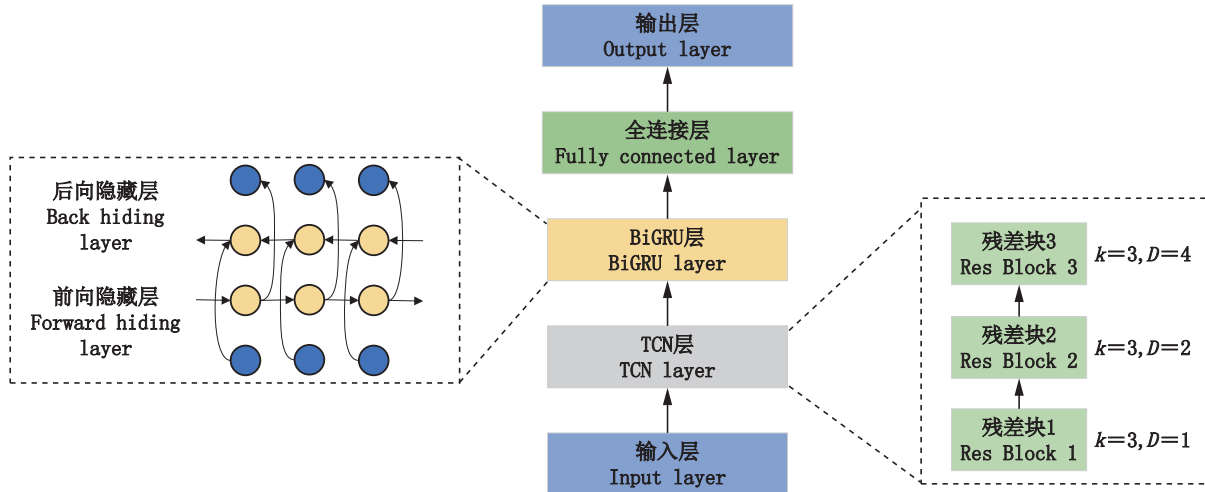


图4 TCN-BiGRU 架构
Fig. 4 Architecture of TCN-BiGRU

2.4.2 SCSO-KDE 进行区间估计

KDE 是一种有效估计预测误差潜在分布的非参数方法。与参数估计不同,KDE 能够直接提供概率密度函数,通过使用给定的样本数据特征来估计预测区间的上下界,而无须事先假设样本数据的分布。对于一组溶解氧数据预测误差集 $\{x_1, x_2, \dots, x_i\}$,则 x_i 的概率密度函数为

$$f(x) = \frac{1}{Nb} \sum_{i=1}^N K\left(\frac{x-x_i}{b}\right) \quad (4)$$

式中: N 为溶解氧误差样本的长度; K 选定为鲁棒性较好的高斯核函数; b 为带宽,过窄会导致估计呈现高方差(过拟合),过宽会导致高偏差(欠拟合)。通过沙猫群优化算法对 KDE 的带宽进行自适应选取,采用目标优化函数公式(5)用于定量描述概率密度函数(Probability density function, PDF)与溶解氧误差分布的拟合程度。

$$\min F = \frac{\sum_{i=1}^n (\tilde{y}_i - L_i)^2}{n} \quad (5)$$

式中: n 为误差序列频率直方图的个数; \tilde{y}_i 为第 i 个直方图的中心点处所对应的概率密度值; L_i 为

第 i 个直方图的高度值。 F 为适应度优化指标, F 越小,概率密度函数的拟合程度越好。

对于给定的置信度 $1 - \alpha$,根据溶解氧的预测结果和 PDF,得到其置信区间(公式 6)。

$$\left[P_i + \tilde{F}\left(\frac{\alpha}{2}\right), P_i + \tilde{F}\left(1 - \frac{\alpha}{2}\right) \right] \quad (6)$$

式中: P_i 为溶解氧预测值; α 为显著性水平; \tilde{F} 为 PDF 积分后的反函数。

3 结果与讨论

为了验证所提溶解氧预测模型的预测效果,设计了相关实验对特征构造、单点预测和区间估计等 3 个关键部分进行模型的性能评价和比较。在这一过程中,我们详细分析了各个实验的结果,并对其出现原因和机制进行了归纳和讨论。

3.1 实验设置

实验基于 Python、PyTorch 深度学习库和 MATLAB 完成。硬件配置包括 Windows 10 操作系统、运行频率为 3.20 GHz 的 Intel (R) Core (TM) i7-8700 CPU、32GB 内存和 GeForce RTX

3060Ti GPU。设置 FC-TCN-BiGRU 模型中卷积核尺寸为 3,膨胀系数 $D=1,2,4$,TCN 的 Dropout 概率设为 0.1;BiGRU 的堆叠次数为 2,迭代次数 $T=500$ 。沙猫种群个数为 $N=50$,迭代次数 $t=300$,灵敏度范围 $s_M \in [0, 2]$ 。

3.2 构造特征的有效性验证

采用 Person 相关系数比较原始各环境因子

特征和构造特征与溶解氧的相关性,其中相关系数高于 0.4 的特征如表 1 和表 2 所示。经过数据构造,日间 pH 相关性提高了 13.5%,光照累积量和太阳辐射累积量成为高相关特征,其与溶解氧相关性均高于 0.5,夜间大气压强相关性提高了 16.3%,其余各特征也有与原始样本相当的相关性。

表 1 溶解氧与原始环境因子的 Pearson 相关系数

Tab. 1 Pearson correlation coefficient among dissolved oxygen and original environmental factors

	pH	大气压强 Pressure	湿度 Humidity	电导率 Conductivity
溶解氧 Dissolved oxygen	0.59	0.43	-0.51	-0.41

表 2 溶解氧与构造特征的 Pearson 相关系数

Tab. 2 Pearson correlation coefficient among dissolved oxygen and constructed features

溶解氧 Dissolved oxygen	pH	大气压强 Pressure	湿度 Humidity	电导率 Conductivity	太阳辐射累积量 Gross radiation intensity	光照累积量 Gross illumination
日间 Day	0.67	0.41	-0.53	-0.43	0.54	0.54
夜间 Night	0.56	0.50	-0.48	-0.43	-	-

选取 pH、湿度、大气压强、电导率、季节构造特征和溶解氧的 10 个历史值,共计 15 个特征作为夜间的模型输入;在以上特征的基础上,选取太阳辐射累积量和光照累积量,共计 17 个特征作为日间的模型输入。

3.3 溶解氧预测模型的对比分析

为了验证提出的 FC-TCN-BiGRU 模型的预

测准确性,将其与极限学习机(ELM)、门控循环单元模型(GRU)、长短时神经网络模型(LSTM)和 TCN-BiGRU 等目前流行的时间序列预测模型进行比较。采用平均绝对误差(MAE)、均方误差(MSE)、均方根误差(RMSE)和决定系数(R^2)对实验结果进行评价。

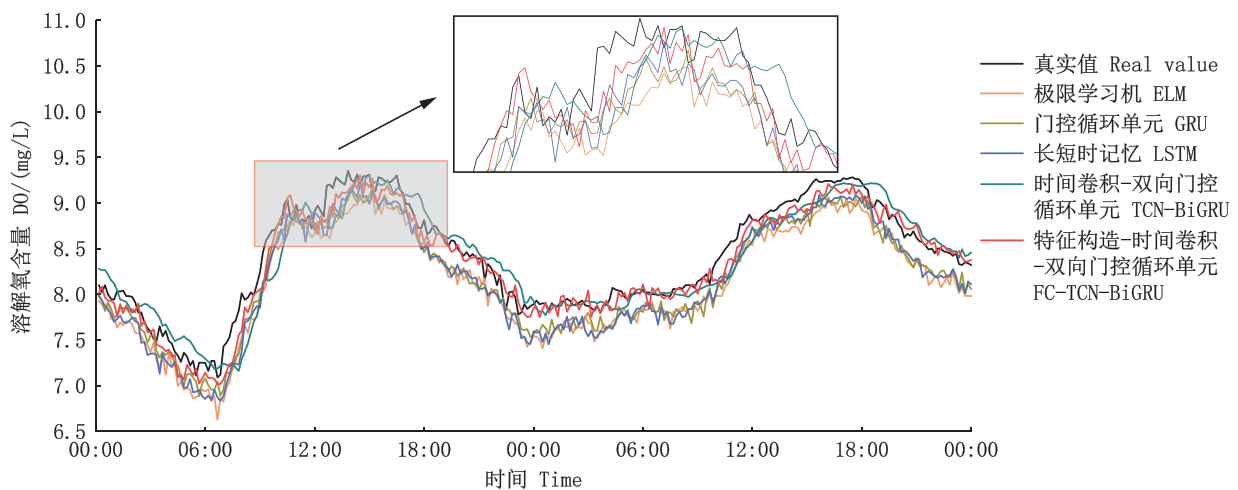


图 5 各模型的预测性能对比
Fig. 5 Comparison of prediction performance of different models

通过实际建模训练,以预测 2023 年 5 月 26、27 日的溶解氧进行结果验证,各模型的预测性能

对比折线图如图 5 所示。可以看出,ELM 作为一种前馈神经网络无法准确捕捉数据间的时序特

征,拟合真实值的效果最差。LSTM和GRU中的门控机制允许模型选择性地接收或忽略输入信息,因此在早晨溶解氧数据变化出现突变拐点的时间段(05:00—07:00)依然有较好的预测效果。组合模型TCN-BiGRU可以结合不同算法的优势,在提取到更大感受野范围特征信息的同时有效学习到数据间的长期依赖关系,降低了过拟合的风险。图5中的局部放大图为26日12:00—18:00的溶解氧变化曲线。在此时间段内,池塘浮游植物、水中藻类的光合作用受光照强度的影响,使得溶解氧在17:00左右达到峰值后逐渐下降;同时,由于温度、气压和人为等因素的复合作用,溶解氧在这一时间段波动变化。在预测前根据昼夜模式划分样本集并构造高质量特征作为TCN-BiGRU模型输入,可以有效提升复杂环境下溶解氧的预测精度,获得最优的预测结果。

表3为不同模型的各项指标计算结果的对比,本文所提出的FC-TCN-BiGRU模型 MSE、MAE、RMSE、 R^2 分别为 0.027 5、0.143 2、0.165 8 和 0.940 0。相较于未经过特征构造的 TCN-

BiGRU 模型分别提高了 27.8%、8.3%、15.1% 和 3.3%。同时,与经典的 LSTM、GRU 和 ELM 相比,所提算法的 RMSE 分别提高了 21.2%、31.9% 和 44.3%。

3.4 溶解氧的区间估计性能验证

为分析 SCSO-KDE 的优化程度,采用不同的带宽取值对误差直方图进行拟合,效果如图6所示,可知,SCSO算法以公式(5)为目标函数,通过不断改变沙猫的位置逐渐向最优结果逼近。经过 SCSO 优化后的概率密度曲线可以准确地反映出误差数据的集中分布范围,具有良好的平滑度和精度。

表3 不同模型的预测结果指标对比
Tab. 3 Comparison of prediction results of different models

模型 Models	均方误差 MSE	平均绝对误差 MAE	均方根误差 RMSE	决定系数 R^2
ELM	0.088 7	0.281 1	0.297 8	0.74
GRU	0.059 3	0.224 6	0.243 5	0.83
LSTM	0.044 3	0.187 3	0.210 5	0.87
TCN-BiGRU	0.038 1	0.156 1	0.195 2	0.91
FC-TCN-BiGRU	0.027 5	0.143 2	0.165 8	0.94

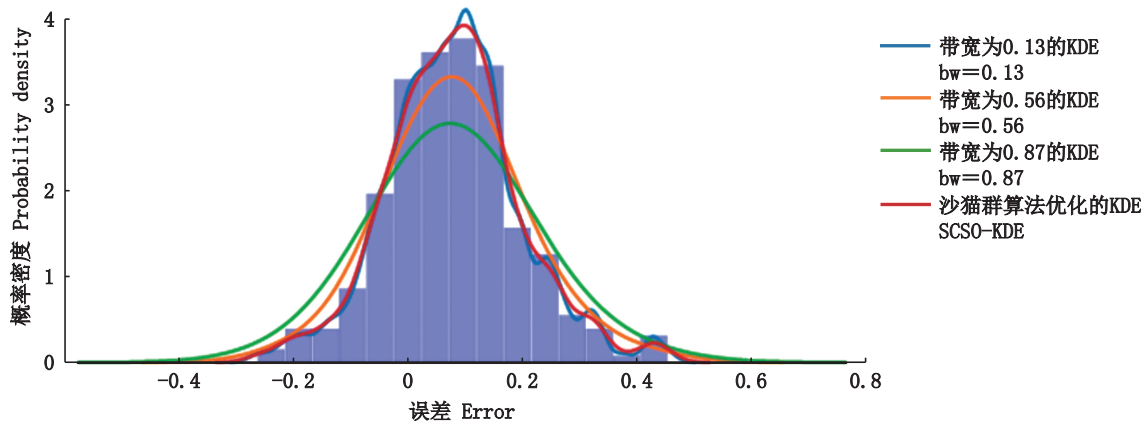


图6 不同带宽下的误差拟合曲线
Fig. 6 Probability density curves with different bandwidths

为验证所提区间估计算法的可靠和准确性,本文采用预测区间覆盖率(Prediction interval coverage probability, PICP)和预测区间归一化平均宽度(Prediction interval normalized average width, PINAW)两种常用准则分别评价区间覆盖和区间宽度。PICP反映了实际溶解氧采集值落在预测区间上下界的概率,用于评估区间预测的可靠性;PINAW反映了预测的准确性,在保证预测区间可靠性的同时获取更窄的区间宽度。

选取26、27日两天作为区间估计的时间段(图7)。当置信度分别为95%、90%和85%时,所提模型的预测分布区间能够较好地覆盖实际值,有效地量化了溶解氧变化的不确定性。同时,将所提模型与未经SCSO优化的KDE进行区间指标对比,结果如表4所示。

可以看出在90%和85%的置信度下,所提算法的PICP均有较高的区间覆盖度,分别为0.929和0.876。与未经SCSO优化的KDE相比,PICP分

别提高了1.98%、2.94%。在95%的置信水平下,虽然两者的区间覆盖率较为接近,但优化后的KDE具有更窄的区间宽度,其PINAW为0.291。

通过分析发现,KDE算法可以有效地拟合误差率直方图,同时采用SCSO对区间上下界进行微调,提高了区间估计的质量。

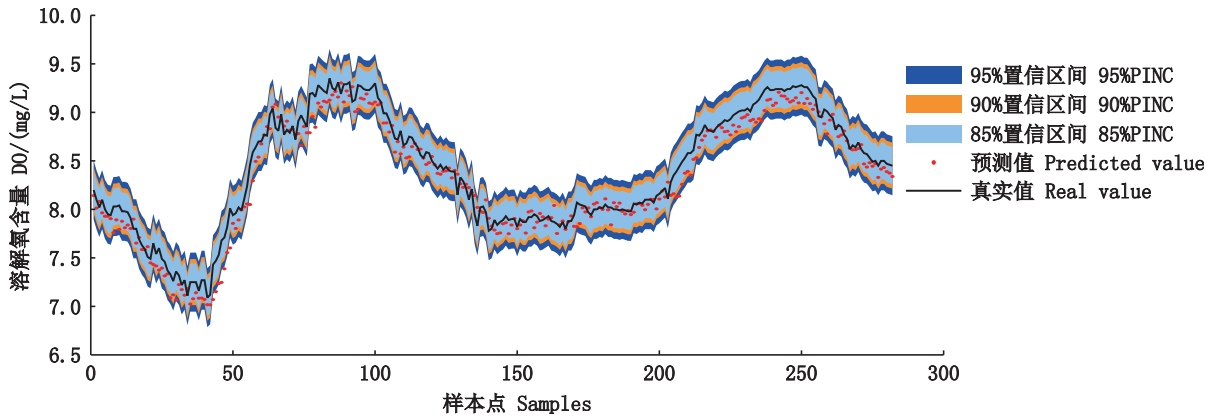


图7 不同置信水平下的概率预测区间

Fig. 7 Probability prediction intervals at different confidence levels

表4 区间估计评价指标
Tab. 4 Evaluation indices of interval estimation

模型 Models	置信度 PINC	预测区间覆盖率 PICP	预测区间归一化平均宽度 PINAW
FC-TCN-BiGRU-SCSO-KDE	95%	0.972	0.291
	90%	0.929	0.244
	85%	0.876	0.214
FC-TCN-BiGRU-KDE	95%	0.969	0.303
	90%	0.911	0.259
	85%	0.851	0.225

4 结论

本文提出一种基于特征构造预处理和TCN-BiGRU的溶解氧预测模型,用于预测池塘溶解氧1h后的变化趋势和分布范围,主要结论如下:

(1)通过特征构造方法对原始样本集进行数据预处理,分离出溶解氧变化的昼夜特征以及提高了输入特征间的相关性,克服传统预测模型输入质量低和信息冗余高的局限性。

(2)采用TCN对输入关键特征进行提取和降维,进而提高了BiGRU的准确性。将所提模型与ELM、LSTM、GRU和TCN-BiGRU的预测结果进行对比。实验结果表明,其具有更低MSE、MAE、RMSE和更高的 R^2 ,分别为0.027 5,0.143 2,0.165 8和0.940 0,在真实数据集中取得了较好的预测效果。

(3)利用沙猫群算法优化KDE的带宽,使得

模型在满足区间覆盖的前提下,具有更窄的区间宽度。与未优化的KDE相比,其90%置信度的PINAW降低了5.8%。在复杂养殖背景下,本模型具有更高的可靠性。

本模型在鲈鱼塘中进行了实验和验证,其中的数据特征构造思想、TCN-BiGRU的组合预测机制和区间估计方法,为溶解氧预测模型的构建提供一种新的思路,在水产养殖领域具有一定的应用推广价值。在今后的工作中,将在其他鱼种的池塘养殖领域和不同的养殖阶段进行研究和实验,进一步验证本模型的泛化性能。

参考文献:

[1] GUO J J, DONG J Q, ZHOU B, et al. A hybrid model for the prediction of dissolved oxygen in seabass farming [J]. Computers and Electronics in Agriculture, 2022, 198: 106971.

[2] ABDEL-TAWWAB M, MONIER M N, HOSEINIFAR S H, et al. Fish response to hypoxia stress: growth, physiological, and immunological biomarkers [J]. Fish Physiology and Biochemistry, 2019, 45(3): 997-1013.

[3] ZHOU X H, WANG J P, HUANG L, et al. Modelling and controlling dissolved oxygen in recirculating aquaculture systems based on mechanism analysis and an adaptive PID controller [J]. Computers and Electronics in Agriculture, 2022, 192: 106583.

[4] 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.
- [5] 曹守启,周礼馨,张铮. 采用改进长短时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.
- [6] HUANG J D, LIU S Y, HASSAN S G, et al. A hybrid model for short-term dissolved oxygen content prediction [J]. Computers and Electronics in Agriculture, 2021, 186: 106216.
- [7] NONG X Z, LAI C, CHEN L H, et al. Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China [J]. Ecological Indicators, 2023, 146: 109845.
- [8] CAO W J, HUAN J, LIU C, et al. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction [J]. Aquacultural Engineering, 2019, 84: 50-59.
- [9] REN Q, ZHANG L, WEI Y G, et al. A method for predicting dissolved oxygen in aquaculture water in an aquaponics system [J]. Computers and Electronics in Agriculture, 2018, 151: 384-391.
- [10] SHU T X, CHEN J H, BHARGAVA V K, et al. An energy-efficient dual prediction scheme using LMS filter and LSTM in wireless sensor networks for environment monitoring [J]. IEEE Internet of Things Journal, 2019, 6 (4): 6736-6747.
- [11] CAO X K, LIU Y R, WANG J P, et al. Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network [J]. Aquacultural Engineering, 2020, 91: 102122.
- [12] YU J W, KIM J S, LI X, et al. Water quality forecasting based on data decomposition, fuzzy clustering and deep learning neural network [J]. Environmental Pollution, 2022, 303: 119136.
- [13] 陈英义,方晓敏,梅思远,等. 基于WT-CNN-LSTM的溶解氧含量预测模型[J]. 农业机械学报, 2020, 51 (10): 284-291.
- CHEN Y Y, FANG X M, MEI S Y, et al. Dissolved oxygen prediction model based on WT-CNN-LSTM [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 284-291.
- [14] CAO X K, REN N, TIAN G L, et al. A three-dimensional prediction method of dissolved oxygen in pond culture based on Attention-GRU-GBRT [J]. Computers and Electronics in Agriculture, 2021, 181: 105955.
- [15] WEI Y Y, CHEN Z Z, ZHAO C, et al. Deterministic and probabilistic ship pitch prediction using a multi-predictor integration model based on hybrid data preprocessing, reinforcement learning and improved QRNN [J]. Advanced Engineering Informatics, 2022, 54: 101806.
- [16] YU Y, LI H T, SUN S L, et al. PM_{2.5} concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy [J]. Sustainable Cities and Society, 2022, 85: 104049.
- [17] JIANG F, ZHU Q N, TIAN T H. An ensemble interval prediction model with change point detection and interval perturbation-based adjustment strategy: A case study of air quality [J]. Expert Systems with Applications, 2023, 222: 119823.
- [18] WANG J Z, XING Q Y, ZENG B, et al. An ensemble forecasting system for short-term power load based on multi-objective optimizer and fuzzy granulation [J]. Applied Energy, 2022, 327: 120042.
- [19] LI Y, TONG Z M, TONG S G, et al. A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation [J]. Sustainable Cities and Society, 2022, 76: 103481.
- [20] 杨京渝,罗隆福,阳同光,等. 基于气象特征挖掘和改进深度学习模型的风电功率短期预测[J]. 电力自动化设备, 2023, 43(3): 110-116.
- YANG J Y, LUO L F, YANG T G, et al. Wind power short-term forecasting based on meteorological feature exploring and improved deep learning model [J]. Electric Power Automation Equipment, 2023, 43(3): 110-116.
- [21] VOUK B, GUID M, ROBNIK-ŠIKONJA M. Feature construction using explanations of individual predictions [J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105823.
- [22] 施珮,匡亮,王泉,等. 基于PC-RELM的养殖水体溶解氧数据流预测模型[J]. 农业工程学报, 2023, 39(7): 227-235.
- SHI P, KUANG L, WANG Q, et al. Data stream prediction model for dissolved oxygen in aquaculture water using PC-RELM [J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(7): 227-235.
- [23] YIN L, FU L J, WU H, et al. Modeling dissolved oxygen in a crab pond [J]. Ecological Modelling, 2021, 440: 109385.
- [24] 屈永霖,闻新宇,张慕琪,等. 使用人工神经网络改进2022年北京冬奥会数值天气预报后处理过程的算法研究[J]. 北京大学学报(自然科学版), 2022, 58(2): 210-220.

- QU Y L, WEN X Y, ZHANG M Q, et al. Develop an objective post-processing system with artificial neural network to improve numerical weather prediction for the Olympic winter games Beijing 2022[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2022, 58 (2) : 210-220.
- [25] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J/OL]. arXiv preprint arXiv: 1803.01271, 2018. (2018-03-04). <https://arxiv.org/abs/1803.01271>.
- [26] SEYYEDABBASI A, KIANI F. Sand cat swarm optimization: A nature-inspired algorithm to solve global optimization problems [J]. *Engineering with Computers*, 2023, 39(4): 2627-2651.

Prediction model of dissolved oxygen in pond based on feature construction pretreatment and TCN-BiGRU

ZHANG Zheng¹, GAO Sen¹, ZHANG Zeyang²

(1. College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai Urban Power Development Co., Ltd., Shanghai 200123, China)

Abstract: In order to achieve accurate and reliable prediction of the dissolved oxygen in ponds and mitigate aquaculture risks, we propose a predictive model based on Feature Construction (FC) pretreatment and Temporal Convolutional Network (TCN) coupled with Bidirectional Gate Recurrent Unit (BiGRU). By constructing statistical features, environmental factor features and seasonal features from the samples, deep-level correlations between variables are explored. Then, the structural feature sequences are processed using multiple layers of convolution and dimensionality reduction through TCN, while preserving the global temporal characteristics and removing redundant information. By integrating BiGRU to model the reduced features, accurate prediction of dissolved oxygen levels is achieved. Furthermore, the Sand Cat Swarm Optimization (SCSO) algorithm is employed to optimize the non-parametric Kernel Density Estimation (KDE) for estimating the distribution range of dissolved oxygen prediction errors. The experimental results indicate that the proposed model achieves superior performance compared to other comparative models, with respective values of 0.027 5 for Mean Squared Error (MSE), 0.143 2 for Mean Absolute Error (MAE), 0.165 8 for Root Mean Squared Error (RMSE), and 0.94 for the Coefficient of Determination (R^2). Meanwhile, the interval estimation effectively covers the fluctuation range of dissolved oxygen, thereby quantifying the uncertainty in the prediction process. In the short-term prediction of dissolved oxygen levels in ponds, this model demonstrates notable accuracy and robustness. It is instructive for water quality monitoring in aquaculture and the enhancement of farming efficiency.

Key words: pond culture; dissolved oxygen; feature construction; prediction model; interval estimation