

文章编号: 1000-0615(2019)04-0958-10

DOI: 10.11964/jfc.20180411259

基于栈式自编码BP神经网络预测水体亚硝态氮浓度模型

付泰然^{1,2}, 刘广鑫^{1,2}, 万全元², 吴霆³,
赵丽娟², 林鑫^{2*}, 杨灵^{3*}

(1. 华中农业大学水产学院, 湖北武汉 430070;

2. 仲恺农业工程学院动物科技学院, 广州市水产病害与水禽养殖重点实验室,

广东省水环境与水产品安全工程技术研究中心,

广东省水禽健康养殖重点实验室, 广东广州 510225;

3. 仲恺农业工程学院信息科学与技术学院, 广东广州 510225)

摘要: 亚硝态氮对于水产养殖动物具有毒性, 对于其含量的及时监控非常重要。基于光谱法和电极法设计的亚硝态氮传感器价格昂贵, 难以大面积推广, 因此急需研发一种能快速预测养殖水体亚硝态氮的模型。实验通过实验室构建的水质在线检测系统测定水体中温度、pH、溶解氧、氧化还原电位4个参数, 同时用 α -萘胺比色法测定水体中亚硝态氮的浓度, 从4种参数中选取与亚硝态氮浓度相关的参数作为预测模型的关联变量。水质参数数据及亚硝态氮浓度数据分别经预处理后作为原始数据用于SAE神经网络的训练, 训练方法采用无监督逐层贪婪训练法, 用学习到的特征监督训练SAE-BP神经网络, 利用反向传播算法(BP)优化模型。训练得到结构为4-5-4-3-1的SAE-BP神经网络模型, 建立的神经网络模型对实验数据预测的拟合优度 R^2 为0.95, 预测结果的均方根误差RMSEP为0.099 71。研究表明, 亚硝态氮预测模型可以较为精准地预测水体中亚硝态氮的浓度。本模型将为开发在线快速监测养殖水体亚硝态氮浓度提供新的思路。

关键词: 亚硝态氮; 栈式自编码; SAE-BP神经网络; 预测模型

中图分类号: S 949; X 832

文献标志码: A

随着水产养殖业规模的扩大, 水产养殖水体污染问题变得愈加严重, 养殖过程中残饵、排泄物等是养殖水体污染的主要来源, 也是导致养殖水体富营养化的主要因素^[1]。水质“三氮”是指氨态氮、亚硝态氮以及硝态氮3种氮源的存在形式, “三氮”含量变化是水体富营养化的指标之一。3种形态的氮可以互相转化, 其中氨态氮、亚硝态氮超标会对养殖动物产生毒害作用^[2-9]。因此实时并准确掌握水体中氨态氮以及亚硝态氮的含量变化及其变化趋势对高效养殖具有重要意义。目前测定水体中亚硝态氮浓度的方法有分光光度法、光谱法以及电极法, 其中分光光度法操作繁琐, 且数据存在一定的滞

后, 难以获得实时的浓度变化。基于电极法和光谱法设计的亚硝态氮传感器价格昂贵, 难以大规模推广, 因此需要探索其他可以实时监测养殖水体中亚硝态氮浓度的方法。影响水体中亚硝态氮浓度的变量有温度、pH值、溶解氧、氧化还原电位、各种脱氮固氮微生物等^[10-14], 而且这些变量间可相互影响。因此用常规方法来建立亚硝态氮浓度预测模型有一定难度。近年来深度学习的快速发展为解决这种复杂的多变量非线性问题提供了新思路。目前已有多种基于神经网络的水质参数预测模型, 如基于递归RBF神经网络的氨氮预测模型和基于BP神经网络的水体氨氮预测模型等^[15-18]。本实验提出一种栈

收稿日期: 2018-04-25 修回日期: 2018-09-26

资助项目: 广东省高等教育“创新强校工程”专项(KA170500G); 广州市民生科技攻关计划(201803020033, 201704020030)

通信作者: 林鑫, E-mail: linli@zhku.edu.cn; 杨灵, E-mail: yang98613@163.com

式自编码神经网络亚硝态氮预测模型,尤其是其无监督的学习机制,为亚硝态氮浓度的实时、准确预测提供了一种新的思路和方向。

1 材料与方法

1.1 建立预测模型流程

模型建立过程 ①数据采集。通过在线水质检测系统采集实时数据;②数据处理。将采集到的数据进行预处理,并挑选合适的变量和数量合适的样本;③辨识建模。挑选合适的模型,用采集到的数据进行训练、校正和测试,数据达到要求后投入使用。

预测模型变量选择 本实验所用水为广州市居民用的自来水,实验装置为本实验室搭建的恒温养殖系统,水桶由空气压缩机进行增氧,水桶中的水温由控温系统控制,水桶容积为300 L,高度为1.2 m。影响水体中亚硝态氮浓度的变量有温度、pH、氧化还原电位、各种脱氮固氮微生物数量等,亚硝态氮的浓度与这些变量存在一定关联,从建模原理的角度分析,如果没有选择某些关联变量,最终预测模型会出现一定误差。选择过多的变量则导致模型过拟合,在预测模型变量的选择上还需要兼顾变量是否检测方便且稳定。最终本实验选择温度、pH、溶解氧、氧化还原电位作为水体中亚硝态氮浓度的辅助变量。为了最大程度地排除脱氮固氮微生物的影响,在进行数据测定之前使用二氧化氯对实验用水进行消毒(使二氧化氯终浓度达到3 mg/L),曝气5 d,通过测定确认余氯已经被清除,而且这个浓度的二氧化氯几乎杀灭了所有的细菌种类。

预测模型数据采集 亚硝态氮浓度的测定采用 α -萘胺比色法, α -萘胺与亚硝态氮偶合反应,在酸性溶液中生成紫红色染料,产物最大吸收峰在543 nm波长处。通过测定最终反应液在543 nm处的吸光度,与标准曲线比对,可求得此水样的亚硝态氮浓度。自来水中初始亚硝态氮浓度很低(0.02~0.05 mg/L),此浓度已接近 α -萘胺比色法测定浓度范围的下限,为了提高亚硝态氮测定的准确性,每一组水样中添加1 g亚硝酸钠,实验过程中保持每组水样总体积为150 L,并确保每组实验亚硝态氮起始浓度相似。为获得泛化性能较好的预测模型,训练模型使用的

数据间需要存在一定差异,因此本实验通过分别调节温度或pH值的方式,获得不同温度或pH下亚硝态氮的浓度数据,其中温度组的温度设置为20、25和30 °C,初始pH值为原始值8;pH组的初始pH值设置为5、6、7、8、9,温度设置为30 °C(表1)。必须说明的是,这些温度和pH值是人为理想的设定,实际上是通过压缩机增氧,向水体注入含有一定温度的空气。空气中的二氧化碳等气体会导致水体pH的波动;增氧过程气体对水体的搅动也会导致水体热量的变化,所以实际上无论温度还是pH值在传感器在线监测过程中所记录的数值都存在一定波动,在线系统可以及时精确记录这些微小变化,克服了人工测定的滞后性,为建模提供了大量原始数据。

表1 实验组设置

Tab. 1 Experimental group setting

编号 experimental group no.	pH	温度/°C temperature
1	5	30
2	6	30
3	7	30
4	8	30
5	9	30
6	8	20
7	8	25
8	8	30

开始实验后每30 min用移液器收集水样,为避免取样时间与测定数据不同步,取样后马上移入比色管中进行反应,每5 mL水样分别加入0.1 mL磺胺盐酸溶液(10 g/L)和0.1 mL盐酸萘乙二胺溶液(1 g/L),反应5 min后在543 nm波长处测定水样的吸光度。将测得的吸光度代入 α -萘胺比色法测定亚硝态氮标准曲线: $y=1.9073x-0.0378$ ($R^2=0.9916$ 、自变量 x 为亚硝态氮浓度,因变量 y 为吸光度)得到对应的亚硝态氮浓度,最终得到包含取样时间以及亚硝态氮浓度的原始数据。温度、pH、溶解氧、氧化还原电位4种辅助变量的数据通过实验室的在线水质监测系统测定。

水质智能监测平台包括前端监测设备和云服务器,前端监测设备通过传感器采集水体的理化参数,采集频率为15 s 1次(每15 s采集1组数据,1 min采集4次数据,经预处理剔除异常数据

后对每分钟采集的数据取平均值并储存于服务器), 并完成模数转换以及数据格式的转换。数据粗糙集约简、数据归一化等预处理过程由云服务器完成。值得说明的是, 一方面本实验在线系统可以每15 s采集每一个实验组4个水质指标(温度、pH、溶解氧、氧化还原电位)1次, 共有8个实验组, 可以预计在6 h内产生了大量的在线数据, 具体计算: $6\text{ h} \times 60\text{ min/h} \times 4\text{ 次/min}$ (每15 s采集1次) $\times 4\text{ 个水质指标} \times 8\text{ 个实验组} = 46\ 080$ 。也就是说在6 h的实验过程中, 在线系统记录了4万多个水质指标。另一方面, 人工比色法测定亚硝酸盐浓度最快也是30 min 1次, 为了验证预测数值和人工实际测定数值是否一致, 在线测定的4个水质指标的时间点和水质采集时间点一致, 根据亚硝态氮取样时间点, 在智能监测平台服务器中找到对应时间点的4种水质参数, 最后, 全部数据经预处理后, 将全部114组数据随机分成训练集和预测集2组, 分组方法采用多折交互验证法, 将全部114组数据均分成4份, 每次训练时选用其中的3份作为训练集, 剩下的1份作为预测集用作验证, 将上述过程重复10次, 通过这样设置训练集与预测集可以得到比较稳定的模型评价结果。

1.2 算法描述

自动编码器 自编码算法是一种无监督学习算法, 使用反向传播算法使得输出值等于输入值。即: $y(i) = x(i)$; $x(i) \in \mathbb{R}^n$; $h_{w,b}(x) \approx x$, 式中 w 是网络层之间的连接权重矩阵, b 是加在神经元上的偏置向量。自编码神经网络尝试逼近一个恒等函数, 使得输出接近于输入。如图1所示, 图中的“+1”表示偏置项系数。激活选择 sigmoid 函数:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (1)$$

对于包含有 m 个样本的数据集, 因自动编码器目标是使网络输出结果 $h_{w,b}(x)$ 与给定的目标值 y 的差距最小化。代价函数定义:

$$J(w, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{w,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{S_i} \sum_{j=1}^{S_{i+1}} (W_{ji}^{(l)})^2 \quad (2)$$

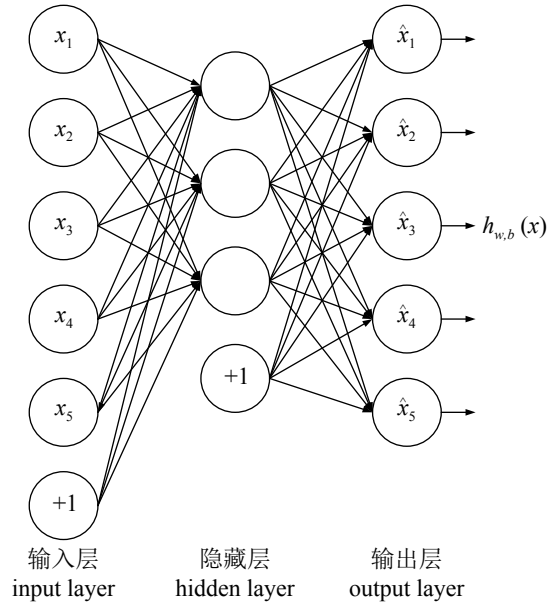


图1 自动编码器
Fig. 1 Autoencoder

式中, λ 为权重衰减参数, 公式的前一项为均方差项, 第二项为权重衰减项, 其意义为减少权重的幅度, 防止模型过拟合, 自动编码器的训练即为调节权值矩阵 W 与偏置向量 b 优化代价函数的过程。参数更新的表达式:

$$W = W - \alpha \frac{\partial J(w, b)}{\partial w} \quad (3)$$

$$b = b - \alpha \frac{\partial J(w, b)}{\partial b} \quad (4)$$

式中, α 为初始化参数学习率。

栈式自动编码器 栈式自动编码网络是由多层自编码器组成的神经网络, 栈式自编码神经网络的编程方式是按照从前到后的顺序依次执行每一层的自编码器, 前一层自编码器的输出作为后一层自编码器的输入, 同理栈式自编码器的解码过程就是反向按顺序执行每一项自动编码器(图2)。

本栈式自编码网络采用无监督逐层贪婪训练算法进行预训练(图3), 每次只训练一个隐含层, 此层自编码器优化后再开始训练下一层, 直至训练完最后一个隐藏层, 通过无监督逐层训练, 获得神经网络中所有层最佳的初始权重, 再将最后一个隐藏层的最高阶特征输入到BP神经网络进行有监督训练, 训练方式为梯度下降法, 用有监督训练的方式对整体网络的

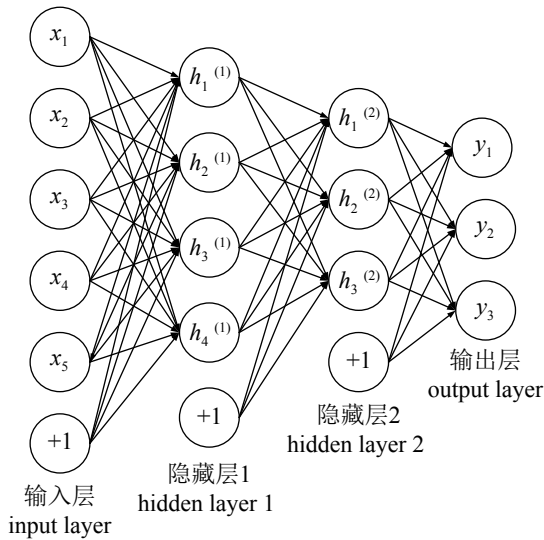


图 2 栈式自动编码器

Fig. 2 Stacked autoencoder

权重进行微调, 栈式自动编码器采用了无监督学习特征的方式, 而BP神经网络是一种监督式

的学习算法, 二者结合后的模型能够对数据进行有效预测, 从而得到所需的水质参数信息。

1.3 预测模型建立

神经网络具有自主学习的功能且适合解决复杂系统中多变量非线性问题。目前栈式神经网络已经应用于很多领域, 但是在应用过程中训练样本数量和质量、参数的选择、数据的处理以及算法的结构等不同均会影响到神经网络模型的性能, 基于以上讨论, 本实验提出一种基于栈式神经网络的亚硝态氮预测模型。

模型构建方法: 在线水质检测系统连续采集到的水质参数数据(温度、pH、溶解氧、氧化还原电位)以及对应的亚硝态氮浓度数据经预处理后训练神经网络模型, 其中亚硝态氮浓度为不同时间点下通过分光光度法所测得, 而这些时间点的水质参数数据则通过在线水质检测系统获得(图4)。

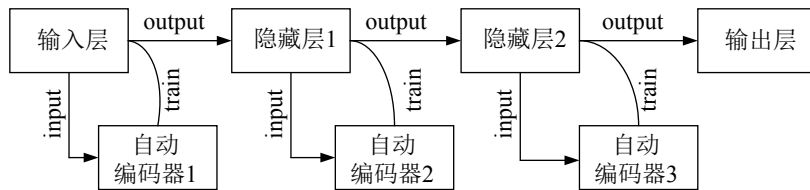


图 3 栈式自动编码器预训练模式图

Fig. 3 Training mode of stacked autoencoder

预测模型的性能依赖于水质参数测量数据的准确性。由于数据采集过程中时常产生误差, 如果使用带有较大误差的数据进行建模, 会使预测模型的性能下降。因此需要对测量数据进行预处理, 将异常数据进行过滤。而为了保证之后步骤中数据处理的方便以及程序运行速率, 还需进行粗糙集约简和数据的归一化处理。温度、pH等影响因子具有不同的物理意义和不同的量纲及数量级, 因此在训练网络之前需要对数据进行归一化处理, 本实验采用离差标准化法, 即最大最小标准化法进行数据归一化处理:

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

式中, x 为原始数据, \hat{x} 为标准化处理后的数据, x_{\max} 和 x_{\min} 分别为该组数据中的最大值和最小值。经过标准化处理后, 全部数据均处于[0, 1]

内, 有利于网络的训练。全部数据预处理后随机分为2组: 训练样本集和测试样本集, 接着将训练样本传入深度网络进行无监督贪婪训练, 训练完毕后用测试样本对深度网络可靠性进行验证。

1.4 模型参数优化

目前栈式自编码器结构的选取尚无完善的理论依据, 因此根据数据与应用的方式不同调整参数的设置, 本研究通过多次实验确定了如下参数设置, 学习速率可以认为权重的更新速率, 学习速率过大时表现为代价函数曲线震荡, 而学习速率过小时, 代价函数收敛的速率过慢; 迭代次数即为对模型进行优化的次数, 当误差达到预设值时训练自动结束; 确定栈式神经网络的层数以及网络隐藏层节点数对于网络模型最终的预测效果十分重要。网络层数过多会导致模型性能下降, 导致过拟合, 而层数过少会导致模型与真实函数的拟合性不佳, 合理的网

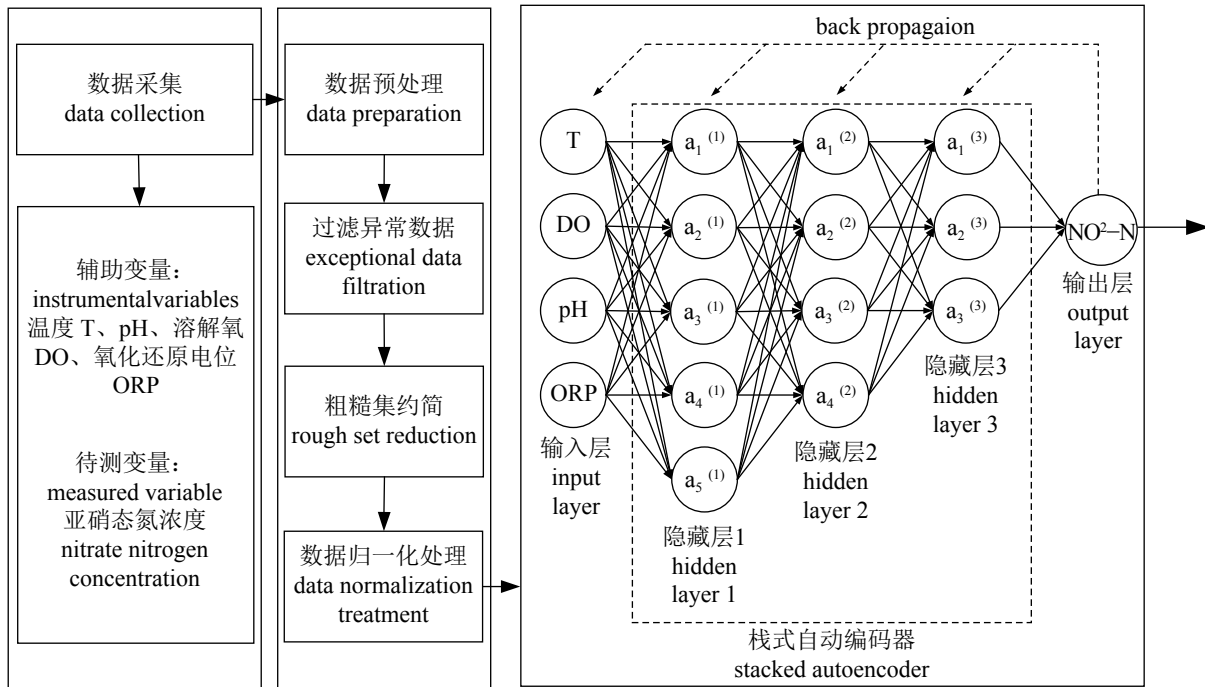


图 4 基于栈式自动编码器的亚硝态氮预测模型

Fig. 4 Prediction model based on stacked autoencoder

络层数会让更高层学习到更高阶更抽象的特征，隐藏层中的节点数较少会导致学习的不充分，而过多的节点会导致学习负荷过大，会增加训练的时间。本研究通过实验最终确定网络隐藏层层数为3，各隐藏层节点数分别为5、4、3，在这个设置下，预测模型的准确率最高。因为本模型输入神经元层数为4，第一隐藏层层数为5，因此需要对隐藏层神经元加入稀疏性限制，经多次实验确定稀疏参数 ρ 和稀疏惩罚项参数 β 的值(表2)。

表 2 预测模型参数设置

Tab. 2 Parameters set for the prediction model

参数 parameter	值 value
预训练学习速率 pre-training learning rate	0.1
微调学习速率 fine tuning learning rate	0.1
预训练最大迭代次数 the maximum number of iterations of pre-training	400
微调最大迭代次数 the maximum number of iterations of fine tuning	300
稀疏参数 sparse parameter	0.5
稀疏惩罚项参数 sparsity penalty term parameter	3
权重衰减参数 weight decay parameter	300

2 结果

2.1 水质参数与亚硝态氮浓度相关性

通过分析亚硝态氮在不同pH值和温度值下的测定值，研究了随着pH值升高，氧化还原电位和亚硝态氮浓度的变化规律；以及随着温度升高，亚硝态氮浓度的变化规律。不同初始pH条件下各组氧化还原电位ORP的变化情况，水体中pH值与氧化还原电位呈反比(图5)。各组均添加1 g的亚硝态钠30 min后，初始pH=6和初始pH=7的实验组中亚硝态氮的浓度显著低于其他实验组，而初始pH=5、pH=8和pH=9的实验组中各亚硝态氮浓度无显著差异(图6)。在不同温度下亚硝态氮浓度变化情况，30 °C时，各取样时间点的亚硝态氮浓度均显著高于25 °C与20 °C 2组，而在25 °C与20 °C组中各取样时间点的亚硝态氮浓度无显著差异(图7)。

2.2 SAE-BP神经网络模型预测结果

图8和图9分别给出了通过温度、pH、溶解氧、氧化还原电位实时预测亚硝态氮模型的训练以及预测结果，其中训练集样本 R^2 值为0.900 36；预测集样本 R^2 值为0.952 08,说明经训练的亚硝态氮预测模型的相关性较好，且拟合度高。图8和

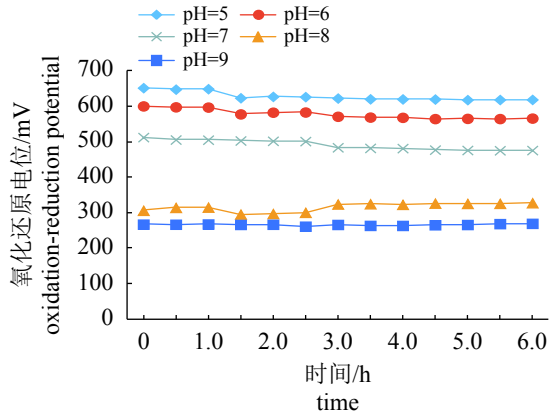


图 5 不同初始pH值对氧化还原电位的影响
 Fig. 5 Effects of different initial pH values on oxidation-reduction potential

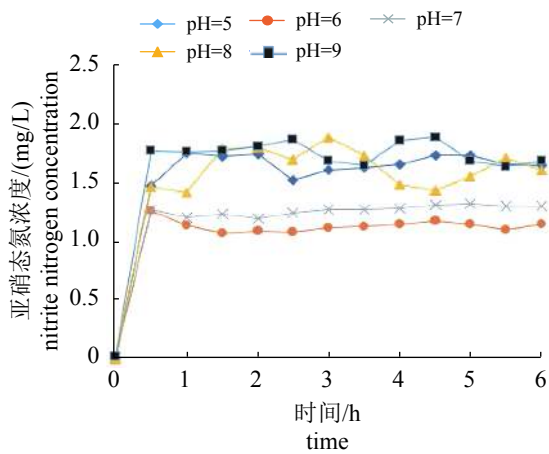


图 6 不同初始pH值对亚硝态氮浓度的影响
 Fig. 6 Effects of different initial pH values on nitrite nitrogen concentration

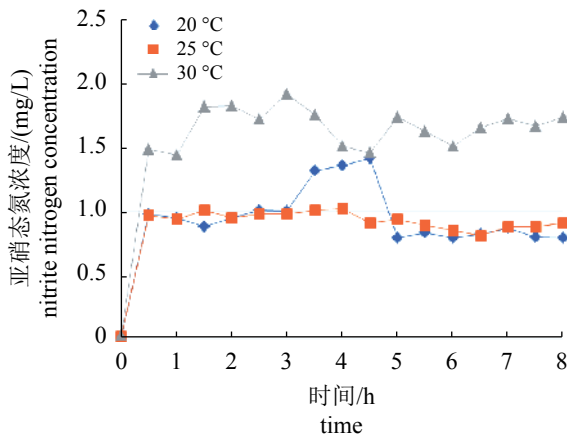


图 7 不同温度对亚硝态氮浓度的影响
 Fig. 7 Effects of different temperature on nitrite nitrogen concentration

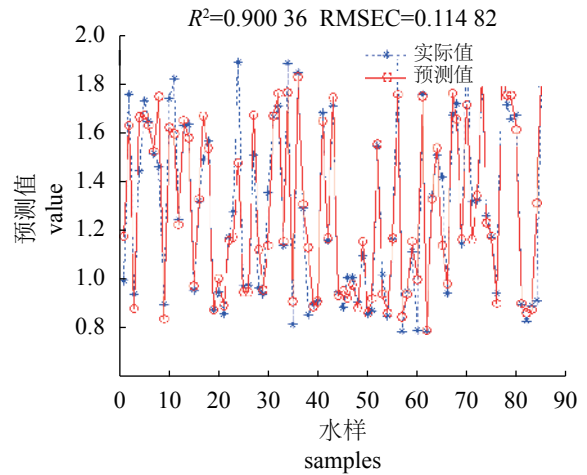


图 8 通过温度、pH、溶解氧、氧化还原电位实时预测水体亚硝态氮浓度模型的训练集预测结果

Fig. 8 Training sets prediction results of the real-time nitrite nitrogen concentrations using temperature, pH, DO and ORP as variables

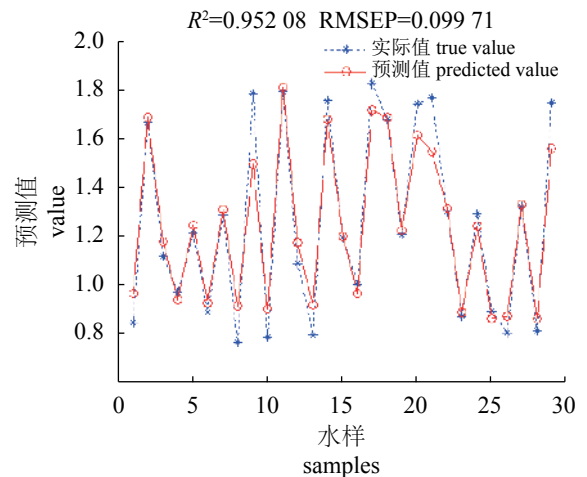


图 9 通过温度、pH、溶解氧、氧化还原电位实时预测水体亚硝态氮浓度模型的预测集预测结果

Fig. 9 Prediction results of the real-time nitrite nitrogen concentrations using temperature, pH, DO and ORP as variables

图9同样给出了预测模型训练以及测试的误差值, 此处为均方根误差:

$$RMSE = \sqrt{\frac{1}{2T} \sum_{t=1}^T [d(t) - y(t)]^2} \quad (6)$$

式中, T 为样本数, $d(t)$ 是期望输出值, $y(t)$ 是实际输出值。最终得到预测模型的训练误差RMSEC为0.114 82, 预测误差RMSEP为0.099 71。理想的模型应该具有低的RMSEC和RMSEP值, 此外

RMSEC和RMSEP值也不应该有很大的差异。因此可判断本预测模型性能良好,对实验数据的预测效果优异。

3 讨论

3.1 不同预测模型比较

目前已有多种预测亚硝态氮的模型被提出,如基于NAR神经网络的亚硝态氮预测模型^[19]、基于弹性BP神经网络的预测模型^[20]。其中基于NAR神经网络预测模型其均方根误差为0.006 7,小于本研究,但 R^2 值低于本模型0.95,因NAR神经网络预测的是亚硝态氮的未来变化趋势,而本研究目的是获得实时亚硝态氮浓度数据,在各自实时数据预测的精度上本模型优于NAR神经网络模型。

基于弹性BP神经网络的预测模型所得到的均方根误差RMSE为0.331 7,基于弹性BP神经网络的模型使用pH值、溶解氧作为辅助参数。在水中影响亚硝酸盐含量的因素很多,亚硝酸盐在水中具有还原性,容易被溶解在水中的各种氧化剂氧化成硝酸盐,水体氧化性能的强弱可以一定程度上由亚硝酸盐的相对浓度体现,而亚硝酸盐浓度的高低也会对水体的氧化性能产生影响,氧化还原电位ORP是水体氧化性能的宏观体现;温度对水体中各种生化反应也有影响。本实验选择了温度、pH、溶解氧和氧化还原电位4种参数对亚硝态氮浓度进行预测,预测模型的均方根误差为0.099 7,选取不同的参数建模,可能是2种模型对各自数据预测精确度不同的原因。

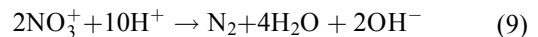
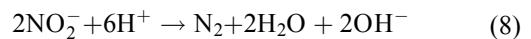
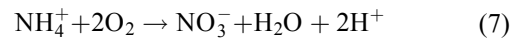
BP神经网络理论上可以以任意精度逼近非线性且连续的函数,但是传统的BP神经网络难以获得良好的初始权值和初始阈值,这有可能导致弹性BP神经网络的误差较高。本研究通过无监督训练栈式自动编码器,获得相较于传统BP网络更好的权重分配,且同样具有对非线性连续函数的逼近能力。

3.2 模型的水质参数选取讨论

一个好的预测模型应包含有足够全面的关联变量,以及关联变量之间或关联变量与待预测参数之间的相关性。本研究阐述了pH值与氧化还原电位之间的相关性、pH值与亚硝态氮浓度的相关性以及温度与亚硝态氮浓度之间的

相关性。

pH值是氢离子浓度指数,是溶液中氢离子活性的衡量标准。水体的氧化还原电位可以反映整个系统氧化还原性能的综合指标,养殖水体中存在多种氧化还原电对如 $\text{Fe}^{3+}/\text{Fe}^{2+}$ 、 $\text{NO}_2^-/\text{NO}_3^-$ 、 $\text{NH}_4^+/\text{NO}_2^-$ 和 $\text{SO}_4^{2-}/\text{H}_2\text{S}$,还有溶解氧($\text{O}_2/\text{H}_2\text{O}$)等,不同氧化还原电对的相对含量不同以及各种氧化还原电对中反应进程的不同都会对水体ORP产生影响,氧化还原电位的电位越高,表示系统整体的氧化性越强。养殖水中 $\text{NH}_4^+-\text{NO}_2^--\text{NO}_3^-$ 在相互转化的过程中常伴随pH的转变: NH_4^+ 经过亚硝化、硝化作用和水体中氧化物质的氧化最终生成 NO_3^- ;而 NO_3^- 和 NO_2^- 在反硝化细菌或某些还原剂的作用下还原为 N_2 的总反应方程式:



由公式(7)可知,在氨氮氧化、硝化的阶段会引起pH下降,由式(8)、(9)可推得在反硝化、还原的阶段pH会上升。因此可以认为pH值的变化可在一定程度上反映水体中氮氧化(硝化)、还原(反硝化)的进程。

温度升高,单位体积内活化分子增加,发生化学反应的有效碰撞增大,宏观表现为反应速率增加。亚硝酸盐可以与氯胺反应生成强致癌物质亚硝胺,不同温度下反应产物的生成量不同,25℃时亚硝胺的生成量显著高于低温5℃时的生成量^[21]。水体中温度还与微生物生长代谢有着密切关系,不同的微生物有不同的最适生长温度,在相同温度下硝化细菌与反硝化细菌的代谢活力有差异,这导致了在不同温度下水体中亚硝态氮的积累量不同,如在15~35℃区间内亚硝态氮积累率随温度升高而增加且 N_2O 的释放量也越高^[22]。

综上所述,pH、氧化还原电位以及温度3种水质参数与亚硝态氮均存在一定的相关性,虽然未讨论溶解氧与亚硝态氮的关联,但是,溶解氧在水体氧化还原电位中占了很大权重,因此本研究选择pH、氧化还原电位、溶解氧以及温度进行建模。在养殖水体中还有许多难以测定的因素影响到亚硝态氮的浓度,比如各种硝化细菌及反硝化细菌等,实时获得这些因素

的可量化数据是未来进一步提升预测模型性能的关键。

3.3 模型的优化讨论

本研究发现, 在所有实验组亚硝态氮起始浓度都相同的条件下, 实际测定的亚硝态氮浓度的波动较大, 这可能是因为本研究所用水样为居民用水, 且测定前使用二氧化氯消毒, 水体的复杂程度要低于养殖池中的水体, 没有了池塘中各种微生物以及缓冲因子对无机氮的缓冲作用, 水体中亚硝态氮随着各种水质参数的变化趋势更为明显。因此本研究下一步工作将评估并优化此模型的泛化性能。

4 结论

本研究应用了基于SAE-BP神经网络的建模方法, 通过选择温度、pH、溶解氧、氧化还原电位建立了水体亚硝态氮浓度的预测模型, 可以较为精确地预测实验室条件下的亚硝态氮浓度。预测模型的 R^2 为0.95, 均方根误差RMSEP为0.099 71。本研究为建立更完善的亚硝态氮浓度预测模型提供了理论支持, 由于条件限制, 本研究没有验证在水质可能发生变化且结构更复杂的水体中此模型的预测效果, 而对于面向养殖水体的亚硝态氮预测模型应具有一定的推广能力, 这将作为本研究需要进一步探讨的重点。根据BP神经网络的建模原理可知, 更多有意义的水质参数可以提升整体模型的精确度, 随着今后水体中亚硝态氮相关研究趋于完善, 选择更多与亚硝态氮浓度相关联的可测量参数, 基于SAE-BP神经网络的建模方法可以更为准确地预测水体中亚硝态氮的浓度。

参考文献:

- [1] Gowen R J, Bradbury N B. The ecological impact of salmonid farming in coastal waters: a review[J]. *Oceanography and Marine Biology Annual Review*, 1987, 25: 563-575.
- [2] 黄翔鹤, 李长玲, 郑莲, 等. 亚硝酸盐氮对凡纳滨对虾毒性和抗病相关因子影响[J]. *水生生物学报*, 2006, 30(4): 466-471.
Huang X H, Li C L, Zheng L, *et al.* The toxicity of $\text{NO}_2\text{-N}$ on *Litopenaeus vannamei* and effects of $\text{NO}_2\text{-N}$ on factors relating to the anti-disease ability[J]. *Acta Hydrobiologica Sinica*, 2006, 30(4): 466-471(in Chinese).
- [3] 王鸿泰, 胡德高. 池塘中亚硝酸盐对草鱼种的毒害及防治[J]. *水产学报*, 1989, 13(3): 207-214.
Wang H T, Hu D G. Toxicity of nitrite to grass carp (*Ctenopharyngodon idellus*) in ponds and its way of prevention[J]. *Journal of Fisheries of China*, 1989, 13(3): 207-214(in Chinese).
- [4] 张武肖, 孙盛明, 戈贤平, 等. 急性氨氮胁迫及毒后恢复对团头鲂幼鱼鳃、肝和肾组织结构的影响[J]. *水产学报*, 2015, 39(2): 233-244.
Zhang W X, Sun S M, Ge X P, *et al.* Acute effects of ammonia exposure on histopathology of gill, liver and kidney in juvenile *Megalobrama amblycephala* and the post-exposure recovery[J]. *Journal of Fisheries of China*, 2015, 39(2): 233-244(in Chinese).
- [5] 陈家长, 臧学磊, 胡庚东, 等. 氨氮胁迫下罗非鱼(GIFT *Oreochromis niloticus*)机体免疫力的变化及其对海豚链球菌易感性的影响[J]. *生态环境学报*, 2011, 20(4): 629-634.
Chen J Z, Zang X L, Hu G D, *et al.* The immune response of GIFT *Oreochromis niloticus* and its susceptibility to *Streptococcus iniae* under stress in different ammonia[J]. *Ecology and Environmental Sciences*, 2011, 20(4): 629-634(in Chinese).
- [6] 韩英, 张辉, 王琨. 亚硝态氮对鲤鱼种血液SOD及GSH-Px的影响[J]. *淡水渔业*, 2007, 37(1): 66-68.
Han Y, Zhang H, Wang K. Activity of SOD and GSH-Px in the blood of fingerling *Cyprinus carpio* under different nitrite concentration[J]. *Freshwater Fisheries*, 2007, 37(1): 66-68(in Chinese).
- [7] Dunier M, Siwicki A K, Demaël A. Effects of organophosphorus insecticides: effects of trichlorfon and dichlorvos on the immune response of carp (*Cyprinus carpio*): III. *In vitro* effects on lymphocyte proliferation and phagocytosis and *in vivo* effects on humoral response[J]. *Ecotoxicology and Environmental Safety*, 1991, 22(1): 78-87.
- [8] 洪美玲, 陈立侨, 孙新谨, 等. 亚硝酸盐急性胁迫对中华绒螯蟹幼体相关免疫指标和应激蛋白(HSP70)表达的影响[J]. *应用与环境生物学报*, 2011, 17(5): 688-693.
Hong M L, Chen L Q, Sun X J, *et al.* Effects of acute nitrite exposure on immunity indicators and HSP70 expression in Chinese mitten-hand crab (*Eriocheir sinensis*).

- sis)[J]. Chinese Journal of Applied & Environmental Biology, 2011, 17(5): 688-693(in Chinese).
- [9] 周鲜娇, 邱德全. 亚硝酸氮和副溶血弧菌对凡纳滨对虾部分免疫指标的影响[J]. 水生生态学杂志, 2009, 2(1): 49-53.
- Zhou X J, Qiu D Q. Effects of nitrite nitrogen and *Vibrio parahaemolyticus* on some immunity indicators of *Litopenaeus vannamei*[J]. Journal of Hydroecology, 2009, 2(1): 49-53(in Chinese).
- [10] 邓斌, 刘威, 李小明, 等. pH、ORP监控在亚硝酸型生物脱氮过程中的应用[J]. 环境科学与技术, 2007, 30(3): 97-99.
- Deng P, Liu W, Li X M, et al. Application of pH, ORP on-line monitor in shortcut biological nitrogen removal process[J]. Environmental Science & Technology, 2007, 30(3): 97-99(in Chinese).
- [11] Pagacova P, Blstakova A, Drtil M, et al. Continually measured orp and Ph signal for control of nitrogen removal[M]//Hlavinek P, Bonacci Q, Marsalek J, et al. Dangerous Pollutants (Xenobiotics) in Urban Water Cycle. Dordrecht: Springer, 2008: 287-296.
- [12] Bastian T, Brondum J. Do traditional measures of water quality in swimming pools and spas correspond with beneficial oxidation reduction potential?[J]. Public Health Reports, 2009, 124(2): 255-261.
- [13] Hanaki K, Wantawin C, Ohgaki S. Nitrification at low levels of dissolved oxygen with and without organic loading in a suspended-growth reactor[J]. Water Research, 1990, 24(3): 297-302.
- [14] Hellinga C, Schellen A A J C, Mulder J W, et al. The SHARON process: an innovative method for nitrogen removal from ammonium-rich waste water[J]. Water Science and Technology, 1998, 37(9): 135-142.
- [15] 杨琴, 谢淑云. BP神经网络在洞庭湖氨氮浓度预测中的应用[J]. 水资源与水工程学报, 2006, 17(1): 65-70.
- Yang Q, Xie S Y. Application of BP neural network into predicting $\text{NH}_3\text{-N}$ concentration of Dongting Lake[J]. Journal of Water Resources & Water Engineering, 2006, 17(1): 65-70(in Chinese).
- [16] 乔俊飞, 马士杰, 许进超. 基于递归RBF神经网络的出水氨氮预测研究[J]. 计算机与应用化学, 2017, 34(2): 145-151.
- Qiao J F, Ma S J, Xu J C. The water ammonia nitrogen prediction research based on a recursive RBF neural network[J]. Computers and Applied Chemistry, 2017, 34(2): 145-151(in Chinese).
- [17] 冯红利, 刘秀红, 杨庆, 等. 低溶解氧下氨氧化过程神经网络预测控制模型[J]. 中国环境科学, 2017, 37(1): 139-145.
- Feng H L, Liu X H, Yang Q, et al. Neural network prediction and control model for ammonia oxidizing process under low DO concentration[J]. China Environmental Science, 2017, 37(1): 139-145(in Chinese).
- [18] 管秋, 王万良, 徐新黎, 等. 基于神经网络的污水处理指标软测量研究[J]. 环境污染与防治, 2006, 28(2): 156-158.
- Guan Q, Wang W L, Xu X L, et al. Soft-sensing method development based on neural network for wastewater treatment[J]. Environmental Pollution and Control, 2006, 28(2): 156-158(in Chinese).
- [19] 汪翔, 何吉祥, 余磊, 等. 基于NAR神经网络对养殖水体亚硝酸盐预测模型的研究[J]. 渔业现代化, 2015, 42(4): 30-34.
- Wang X, He J X, She L, et al. Establishment of aquaculture water nitrite prediction model based on NAR neural network[J]. Fishery Modernization, 2015, 42(4): 30-34(in Chinese).
- [20] 林志贵, 姚芳琴, 冯林强, 等. 结合自适应遗传算法与弹性BP神经网络的亚硝酸盐预测模型[J]. 天津工业大学学报, 2015, 34(3): 67-72.
- Lin Z G, Yao F Q, Feng L Q, et al. Nitrite prediction model based on adaptive genetic algorithm and elastic BP neural network[J]. Journal of Tianjin Polytechnic University, 2015, 34(3): 67-72(in Chinese).
- [21] 杨磊. 水中亚硝胺类消毒副产物生成规律及其前质去除方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
- Yang L. Research on N-nitrosamines formation during disinfection and removal methods of their precursors[D]. Harbin: Harbin Institute of Technology, 2010. (in Chinese)
- [22] 郭宁, 张建, 孔强, 等. 温度对亚硝化及氧化亚氮释放的影响[J]. 环境工程学报, 2013, 7(4): 1308-1312.
- Guo N, Zhang J, Kong Q, et al. Effect of temperature on ammoxidation and nitrous-oxide emission in partial nitrification[J]. Chinese Journal of Environmental Engineering, 2013, 7(4): 1308-1312(in Chinese).

Establishment of a water nitrite nitrogen concentration prediction model based on stacked autoencoder-BP neural network

FU Tairan^{1,2}, LIU Guangxin^{1,2}, WAN Quanyuan², WU Ting³,
ZHAO Lijuan², LIN Li^{2*}, YANG Ling^{3*}

(1. College of Fisheries, Huazhong Agricultural University, Wuhan 430070, China;

2. Guangzhou Key Laboratory of Aquatic Animal Diseases and Waterfowl Breeding,

Guangdong Provincial Water Environment and Aquatic Products Security Engineering Technology Research Center,

Guangdong Provincial Key Laboratory of Waterfowl Healthy Breeding,

College of Animal Sciences and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China;

3. College of Information Science and Technology,

Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: Nitrite nitrogen is toxic to the aquatic animals. Monitoring the concentration of nitrite nitrogen is very critical for the culture of aquatic animals. Due to the high cost of the current commercial electrode sensor which is used to measure the concentration of nitrite nitrogen in water, this kind of sensor is very difficult to be popularized on a large scale. Therefore, it is an urgent need to develop another novel method to predict the concentration of nitrite nitrogen in water. In this paper, taking the advantage of the established online water monitoring system in our laboratory, the temperature, pH value, dissolved oxygen and oxidation-reduction potential were recorded from the water in tanks. Meanwhile, the actual concentration of nitrate nitrogen in water was measured using alpha-naphthalene colorimetric method. The data after pretreatment were used as the original data to be used for SAE neural network training. Thereafter, unsupervised greed training method was applied. The learnt characteristics were used for the supervision and training of BP neural network. The model was optimized using the back propagation (BP) algorithm. The prediction model R^2 of the nitrite nitrogen after training was 0.95, and root mean square error of the prediction (RMSEP) was 0.099 71, indicating that the model could accurately predict the concentrations of nitrate nitrogen in water. The established model will pave a new way for developing online system for monitoring the water nitrate nitrogen concentration in the future.

Key words: nitrate nitrogen; stacked autoencoder; SAE-BP neural network; prediction model

Corresponding author: YANG Ling. E-mail: linli@zhku.edu.cn; LIN Li. E-mail: yang98613@163.com

Funding projects: Innovation and Strong Universities Special Funds (KA170500G) from the Department of Education of Guangdong Province; Guangzhou Science and Technology Bureau (201803020033, 201704020030)