



## 以优化的 BP 神经网络模型为基础研究山东近海口虾蛄空间分布与环境因子的关系

李鹏程<sup>1,2</sup>, 张崇良<sup>1,2\*</sup>, 任一平<sup>1,2,3</sup>, 徐宾铎<sup>1,2</sup>, 薛莹<sup>1,2</sup>

(1. 中国海洋大学水产学院, 山东 青岛 266003;

2. 海州湾渔业生态系统教育部野外观测研究站, 山东 青岛 266003;

3. 青岛海洋科学与技术试点国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266237)

**摘要:** BP 神经网络模型作为一种常用的机器学习方法, 被广泛应用于物种分布模型, 以解析生物分布与环境因子的关系。与传统回归模型相比, 该模型可以灵活处理变量间的非线性关系, 但其结构复杂, 在参数设置方面存在不确定性, 从而影响模型的预测与应用。根据 2016—2017 年山东近海口虾蛄渔业资源调查与环境数据, 利用 BP 神经网络模型构建口虾蛄资源分布模型, 同时利用数据分组处理算法 (group method of data handling, GMDH)、遗传算法 (genetic algorithm, GA) 和自适应算法 (adaptive algorithm) 分别对模型输入变量、初始权值和隐节点数目 3 方面进行优化, 构建 7 种不同组合优化模型。结果显示, 7 种模型的优化效果存在明显差异, 单方面和两方面组合优化模型预测性能基本保持一致; 而三方面共同优化其均方根误差与残差平方和分别为 0.35 和 1.94, 较初始模型的 0.52 和 2.40 更小, 且相关系数最大为 0.45, 表明模型优化效果最好。对比优化前后发现, 口虾蛄资源密度随纬度和底层盐度变化趋势基本保持一致, 而随底层温度的升高, 口虾蛄资源密度存在较大差异。此外, 最优模型较初始模型增加水深为关键环境因子, 对口虾蛄的资源密度具有重要影响。本研究进一步开发了 BP 神经网络模型参数优化的方法, 证明了参数优化对 BP 模型的预测性能具有重要影响, 模型优化对于分析口虾蛄资源密度与环境因子的关系具有重要意义。

**关键词:** 口虾蛄; 物种分布模型; 人工神经网络; 模型优化; 预测性能

**中图分类号:** S 932.5<sup>+</sup>; TP 183

**文献标志码:** A

物种分布模型 (species distribution models, SDMs) 是研究物种空间分布与环境因子关系的重要方法。在海洋渔业研究中, 应用较广泛的主要有广义线性模型 (generalized linear model, GLM)、广义可加模型 (generalized additive model, GAM) 和栖息地指数模型<sup>[1]</sup>等基于回归分析方法的模型。

近年来, 机器学习方法得到迅速发展, 在关键因子筛选、渔场预测<sup>[2-3]</sup>和复杂动态系统建模与预测问题<sup>[4]</sup>等方面具有明显优势。人工神经网络作为机器学习的重要方法, 于 20 世纪 80 年代迅速发展<sup>[5]</sup>, 是模拟动物神经细胞群结构和功能特性构成的一种信息处理系统<sup>[6]</sup>。其中, 多层前馈

收稿日期: 2020-07-16 修回日期: 2020-08-31

资助项目: 国家重点研发计划 (2018YFD0900906); 国家自然科学基金 (31802301)

第一作者: 李鹏程 (照片), 从事渔业资源评估与管理研究, E-mail: 17805450993m@sina.cn

通信作者: 张崇良, E-mail: zcl.0903@163.com



网络学习 (error back propagation, BP) 是最基础、应用最广泛的模型<sup>[7-8]</sup>, 由 Rumelhart 等<sup>[9]</sup> 提出, 具有较好的自学习性、自适应性、稳健性和泛化性<sup>[10]</sup> 等特点。近年来, 该模型被越来越多地应用于预测海洋鱼类的分布研究<sup>[11]</sup>。但 BP 模型仍存在一些不足: 如初始权值的选取是随机的、容易陷入局部极小值和网络结构选择的不确定性<sup>[12]</sup> 等问题。因此, BP 模型的优化受到越来越多的关注, 如张磊等<sup>[8]</sup> 就极小值问题产生的原因及各种改进方法进行总结, 比较分析其优缺点, 并对改进局部极小值问题进行综合研究; 李伟林等<sup>[7]</sup> 设计自适应调节算法对 BP 模型的隐藏层结构进行优化。相关研究更多关注模型单一优化方法的选择, 而对模型不同方法组合优化的研究较少, 同时缺乏确定的指标对模型优化前后性能进行比较和判断。

以山东近海优势种口虾蛄 (*Oratosquilla oratoria*) 为例, 通过不同优化算法, 对模型输入变量、初始权值和隐节点数目进行单一和组合优化, 并比较优化前后的模型性能, 选择优化效果最好的模型用于分析口虾蛄空间分布与环境因子之间的关系。本实验旨在为 BP 模型在渔业研究中的应用提供参考, 为口虾蛄适宜栖息地保护和渔业资源管理提供科学依据。

## 1 材料与方法

### 1.1 数据来源

口虾蛄数据来自 2016 年 10 月, 2017 年 1 月、5 月和 8 月山东近海的渔业资源底拖网调查, 调查海域为 119.3°~124.0°E、35°~37°N, 采用系统采样设置 63 个站位 (图 1)。口虾蛄数据按照拖

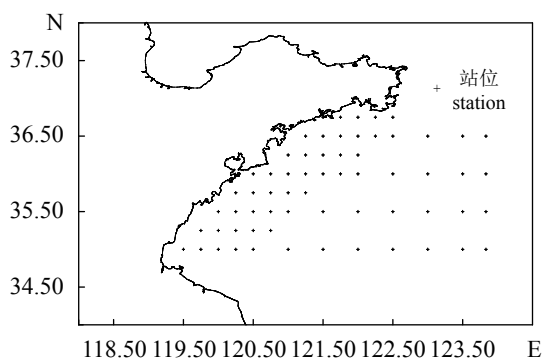


图 1 山东近海渔业资源与环境调查站位示意图

Fig. 1 Fishery resources and environment survey station in the coastal waters of Shandong

网 (1 h) 拖速 (2 kn) 进行标准化, 得到口虾蛄的资源量 (g/h)。样品处理和环境数据测定等均按照《海洋调查规范》(SC/T 9403—2012) 进行<sup>[13]</sup>。环境数据由同步使用的 CTD (型号为 CTD75M/1167) 获得, 包括水深、表层温度、底层温度、表层盐度和底层盐度。

### 1.2 模型优化方法

本研究在传统 BP 模型的基础上通过对输入变量、初始权值和隐节点数目这 3 个方面对模型进行优化, 并评估不同模型的优化效果。

**输入变量选择** 数据分组处理算法 (group method of data handling, GMDH) 通过不完全归纳法<sup>[14]</sup> 在筛选输入变量方面具有独到的优点, 能较好地处理变量选择问题<sup>[15-17]</sup>。该方法由 Ivakhnenko 等<sup>[4]</sup> 提出, 用于预报海洋和河流中的鱼群。与传统 BP 模型相比, GMDH 算法具有明确的解析式、建模过程自组织控制, 不需要任何初始假设等特点, 通过“遗传-变异-选择”的原则<sup>[18]</sup> 选出最优复杂度模型。

Bates 等<sup>[19]</sup> 于 1969 年提出模型组合方法提高模型预测的准确性。因此, 根据传统 BP 模型对非线性系统的建模具有很好的优越性, 但输入变量难以确定的问题, 本研究将 BP 模型与 GMDH 算法相结合, 以提高模型的性能。GMDH 算法通过 Matlab R2018b(9.5.0) 软件中 maGMDH 实现。

**初始权值确定** 传统 BP 模型初始值选取是随机的, 容易导致学习效率低、收敛速度慢和局部极小值等问题<sup>[20-21]</sup>。因此, 初始权值的选择对模型性能具有至关重要的作用。遗传算法 (genetic algorithm, GA) 具有良好的寻优能力, 通过模拟生物进化过程而形成全局优化搜索算法<sup>[22]</sup>, 广泛应用于 BP 模型初始权值的优化问题<sup>[23-26]</sup>。其遵循“适者生存和优胜劣汰”的原理<sup>[27]</sup>, 选择进化好的个体作为最优解, 从而提高模型的性能。

因此, 本研究先采用 GA 算法对初始权值进行优化, 定位出较好的搜索空间。然后利用 BP 模型在局部搜索出最优解<sup>[28]</sup> 以避免陷入局部极小值问题, 从而对 BP 模型性能进行优化。GA 算法通过 R3.5.2(Eggshell Igloo, 2018) 软件中的 genalg 包实现。

**网络结构优化** 在传统 BP 模型中, 网络结构优化的问题是国内外的研究热点<sup>[29]</sup>。本研究中其输出层由问题本身决定, 输入层通过 GMDH

算法选择, 而隐含层数和隐节点个数的确定往往缺乏依据。Hornik 等<sup>[30]</sup>研究表明, 一个隐含层的神经网络, 隐节点足够多就可以任意精度地逼近一个非线性函数。因此, 本研究在此基础上, 通过选择合适的隐节点数目提升模型性能。

本研究通过自适应算法 (adaptive algorithm) 确定隐节点个数<sup>[28-29]</sup>。该方法是在网络学习过程中, 根据环境要求自适应地学习和调整结构, 得到最适隐节点数的网络结构。其通过竞争决定神经元数目的增减<sup>[29]</sup>, 从而优化模型性能。自适应算法通过以下方式实现:

$$\text{样本发散度: } S_i = \frac{1}{n} \sum_{p=1}^n O_{ip}^2 - \bar{O}_i^2$$

同层隐节点  $i, j$  的相关系数:

$$R_{ij} = \frac{\frac{1}{n} \sum_{p=1}^n O_{ip} * O_{jp} - \bar{O}_i * \bar{O}_j}{\sqrt{S_i} * \sqrt{S_j}}$$

式中,  $O_{ip}, O_{jp}$  是隐节点  $i, j$  在学习第  $p$  个样本时的输出,  $\bar{O}_i, \bar{O}_j$  是  $i, j$  在学习完  $n$  个样本后的平均输出,  $n$  为学习样本总数,  $S_i$  表示隐节点对网络的训练作用,  $R_{ij}$  表示隐节点  $i, j$  的相关连程度。相关研究表明,  $S_i < (0.01 \sim 0.001)$  即可被删除;  $R_{ij} > (0.8 \sim 0.9)$  即说明  $i, j$  功能重复, 需要合并<sup>[31-32]</sup>。

### 1.3 BP 模型构建与评估

本研究选取经度、纬度、季节 3 个时空因子, 以及水深、表层温度、底层温度、表层盐度、底层盐度等共 8 个因子作为解释变量。在模型构建之前, 采用方差膨胀因子 (variance expansion factor, VIF)<sup>[33]</sup> 进行多重共线性检验。一般原则下  $\sqrt{VIF} > 2$ , 表明存在多重共线性问题<sup>[1]</sup>, 建模前予以删除。传统 BP 模型采用逐步回归法<sup>[34]</sup> 进行因子筛选和拟合, 运用累积方差解释率 (variance interpretation rate) 和残差平方和 (residual sum of squares, SSE) 判断模型的拟合效果, 残差平方和越小, 累积方差解释率越高, 模型的拟合效果越好。当模型累积方差解释率和残差平方和不再变化, 模型即构建结束。

本研究采用 GA 算法、GMDH 算法和自适应算法对模型 3 方面进行优化, 构成 7 种不同的组合优化 BP 模型 (表 1), 其中 Model 1 为传统 BP 模型, Model 2~4 为单方面优化模型, Model 5~

7 为两方面组合优化模型, 而 Model 8 为三方面共同优化模型。

表 1 不同优化组合的 BP 模型

Tab. 1 BP models with different combinations of optimizing methods

模型 model	优化方法 optimization		
	遗传算法 genetic algorithm	数据分组处理算法 group method of data handling	自适应算法 adaptive algorithm
Model 1			
Model 2	+		
Model 3		+	
Model 4			+
Model 5	+	+	
Model 6	+		+
Model 7		+	+
Model 8	+	+	+

注: +表示在 Model 1 的基础上选择该种算法对其进行优化  
Notes: + means selecting this algorithm to optimize it based on Model 1

研究通过交叉验证的方法评价模型的预测效果。交叉验证随机抽取样本数据的 75% 作为训练样本用于模型的训练和拟合, 剩余 25% 数据进行模型预测。模型重复运行 1000 次, 采用均方根误差 (root mean square error, RMSE)、相关系数 (correlation coefficient, COR) 和残差平方和 (residual sum of squares, SSE) 的平均值, 作为判断模型预测性能指标, 比较不同组合优化模型预测性能。其中 SSE 表示真实值与预测值之差的平方和, 衡量模型的拟合程度。其值越小表示模型的拟合效果越好<sup>[35]</sup>, 反之越差。计算公式:

$$SSE = \sum_{i=1}^n (P_i - O_i)^2$$

式中,  $P_i$  表示真实值,  $O_i$  表示预测值,  $n$  为预测值个数。

相关系数 (COR) 用于研究变量之间的线性相关程度, 本研究中描述预测值与真实值之间的相关性, 反映模型的预测效果, 其绝对值越接近于 1 说明相关程度越好<sup>[36]</sup>, 即预测效果越好, 反之较差。其公式<sup>[36]</sup>:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

式中,  $\text{Cov}(X, Y)$  为  $X, Y$  的协方差,  $\text{Var}[X]$  为  $X$

的方差,  $\text{Var}[Y]$ 为  $Y$  的方差。

RMSE 作为衡量模型预测准确性的指数, 能够反映模型预测值的离散程度<sup>[37]</sup>, 其值越小, 表示模型预测的准确性越高, 描述的实验数据更精确。其计算公式:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}$$

式中,  $P_i$ 为真实值,  $O_i$ 为预测值,  $n$  为预测值个数。

## 2 结果

### 2.1 模型拟合结果

通过 VIF 检验表明, 底层温度、底层盐度、表层温度、表层盐度、水深、季节、经度和纬度的 VIF 值分别为 2.48、1.96、3.82、1.72、1.71、2.29、2.66 和 2.96, 各因子间多重共线性不显著, 均可作为解释变量加入模型。

Model 1 构建结果表明, 纬度、底层温度和底层盐度为关键解释变量, 与口虾蛄资源密度存在密切关系, 其累积方差解释率为 64.43%。其中纬度的贡献率最大, 为 39.39%, 其次为底层温度和底层盐度, 贡献率分别为 17.12% 和 7.92% (表 2)。

表 2 Model 1 的拟合结果及各因子的重要性

Tab. 2 Fitting results of Model 1 and importance of each factor

模型 model	加入的 因子 added factors	累积方差 解释率/% variance interpretation rate	贡献率/% importance	残差 平方和 residual sum of squares
Model 1	N	39.39	39.39	2.94
	SBT	56.51	17.12	2.55
	SBS	64.43	7.92	2.40

注: N表示纬度, SBT表示底层温度, SBS表示底层盐度, 下同  
Notes: N is latitude, SBT is the bottom temperature, SBS is the bottom salinity, the same below

Model 2~8 在 Model 1 基础上对输入变量、初始权值和隐节点数目进行优化, 得到不同组合优化模型结果 (表 3)。其中, Model 2、4 和 6 的解释变量与 Model 1 一致, 为纬度、底层温度和底层盐度。GMDH 算法优化的 Model 3、5、7 和 8 较 Model 1 增加水深作为关键解释变量。隐节点数目在不同优化算法下存在差异, 未进行自适应算法优化的 Model 2、3 和 5 最优隐节点

数目为 3; 在未进行 GMDH 算法下采用自适应算法优化的 Model 4 和 6 最优隐节点数目为 4; 在 GMDH 算法条件下增加自适应算法优化的 Model 7 和 8 最优隐节点数目为 5。

表 3 不同组合优化 BP 模型构建结果

Tab. 3 Results of BP model construction of different combinations optimized

模型 model	加入的因子 added factors	初始权值 initial weights	隐节点数目 number of hidden nodes
Model 2	N+SBT+SBS	0.001 6	3
Model 3	N+SBT+SBS+depth	—	3
Model 4	N+SBT+SBS	—	4
Model 5	N+SBT+SBS+depth	0.001 6	3
Model 6	N+SBT+SBS	0.001 6	4
Model 7	N+SBT+SBS+depth	—	5
Model 8	N+SBT+SBS+depth	0.001 6	5

注: depth表示水深, —表示初始权值是随机的

Notes: depth represents the depth of water, — means that the initial weight is random

### 2.2 模型的预测性能

交叉验证得到 1000 次不同组合优化模型的均方根误差、相关系数和残差平方和等指标平均值 (表 4)。除 Model 4 和 6, 其他模型均方根误差均小于 Model 1, 其中 Model 3、5 和 8 均方根误差更小, 同时其相关系数优于 Model 1, 表明 Model 3、5 和 8 预测性能优于 Model 1, 其中 Model 8 相关系数最大, 模型预测性能最好; Model 2 和 7 的相关系数与 Model 1 相比基本保持不变, 模型预测性能与 Model 1 相差不大; Model 4 和 6

表 4 不同模型优化的结果

Tab. 4 Optimization results of different models

模型 model	均方根误差 RMSE	相关系数 COR	残差平方和 SSE
Model 1	0.52	0.38	2.40
Model 2	0.45	0.38	2.39
Model 3	0.36	0.39	2.25
Model 4	0.54	0.35	2.25
Model 5	0.37	0.38	2.24
Model 6	0.54	0.35	2.26
Model 7	0.45	0.36	1.96
Model 8	0.35	0.45	1.94



的相关系数明显低于 Model 1, 模型的预测性能较差。

### 2.3 分布与关键环境因子的关系

通过 BP 模型分析口虾空间分布与关键环境因子的关系。Model 1 和 8 结果表明, 口虾资源密度与纬度(北纬)和底层盐度关系基本保持一致, 口虾资源密度随纬度的升高呈逐渐上升趋势; 口虾在底层盐度时资源密度预测的偏离值较大, 但整体上随底层盐度增加呈先上升后下降趋势(图 2, 图 3)。

Model 1 和 8 中, 口虾资源密度与底层温度的关系存在明显差异, Model 1 口虾资源密度随底层温度升高先增加后保持不变, Model 8 呈现波动上升后下降的趋势; 另外, Model 8 增加水深变量为关键环境因子, 口虾资源密度随水深增加呈先上升后下降的趋势。

## 3 讨论

实验采用不同方法对传统 BP 模型 3 个方面进行优化, 通过均方根误差、相关系数和残差平方和综合比较分析模型优化效果, 结合山东近海口虾渔业资源调查与环境数据对模型预测性能进行评估。研究表明, GMDH 算法单独优化和 GA、GMDH 算法组合优化均能有效提升 BP 模型预测性能, 且三方面共同优化模型预测效果最好。本实验能够为 BP 神经网络模型的应用提供有价值的参考, 同时为山东近海口虾空间分布与各环境因子关系的研究提供更为合适、准确的模型, 从而为山东近海口虾的渔业资源管理提供科学指导。

### 3.1 关键环境因子的效应差异

通过优化 BP 模型后进行关键环境因子筛选, 与 Model 1 增加水深为关键环境因子对口虾资源密度有密切关系。口虾资源密度与纬度、底层盐度的关系优化前后基本保持不变, 但资源密度在底层盐度时预测的偏离值较大, 可能与口虾为广盐性<sup>[38]</sup>种类有关, 对盐度具有较强的适应性。也可能是由于调查站位设计原因, 导致盐度数据主要分布在高盐度附近, 低盐条件的站位较少造成的。Model 8 较 Model 1 在与底层温度的关系方面存在明显差异, 可能与温度过高使口虾耗氧率和能量消耗增加, 过低

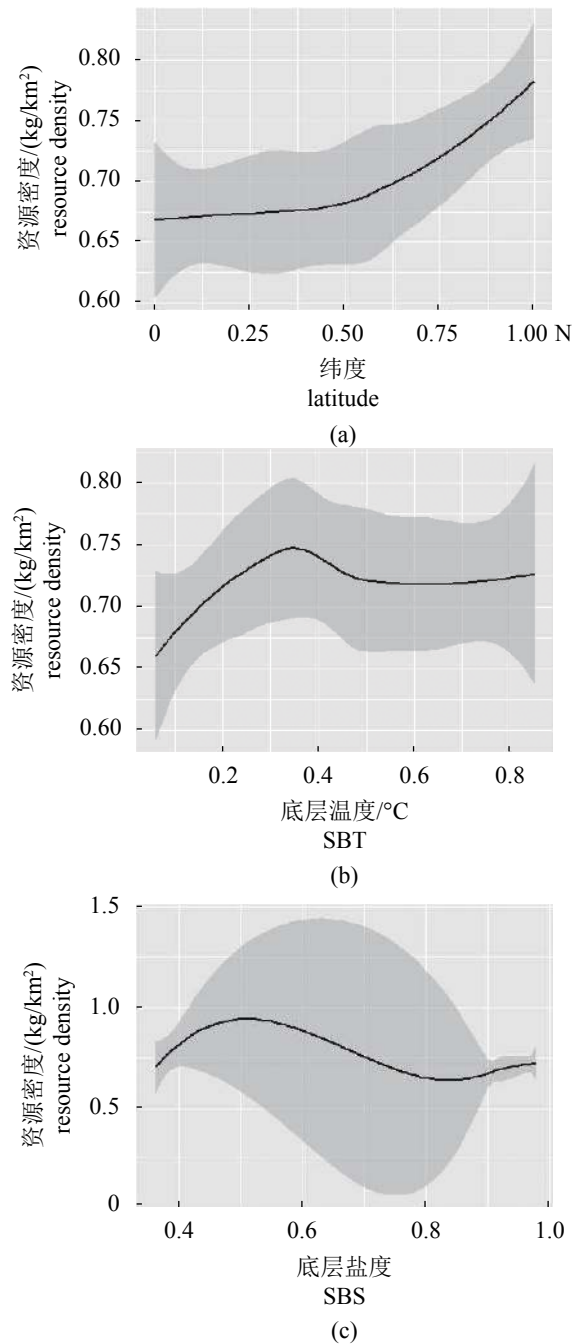


图 2 Model 1 各影响因子与口虾资源密度之间的关系

为加快网络的学习训练速度, 对数据进行归一化处理, 通过  $[x-\min(x)]/[\max(x)-\min(x)]$  实现, (a) 资源密度与纬度的关系, (b) 资源密度与底层温度的关系, (c) 资源密度与底层盐度的关系, 下同

Fig. 2 Relationship between the impact factors of Model 1 and the resource density of *O. oratoria*

In order to speed up the learning and training speed of the network, the data is normalized by  $[x-\min(x)]/[\max(x)-\min(x)]$ , (a) the relationship between resource density and latitude, (b) the relationship between resource density and bottom temperature, (c) the relationship between resource density and bottom salinity, the same below

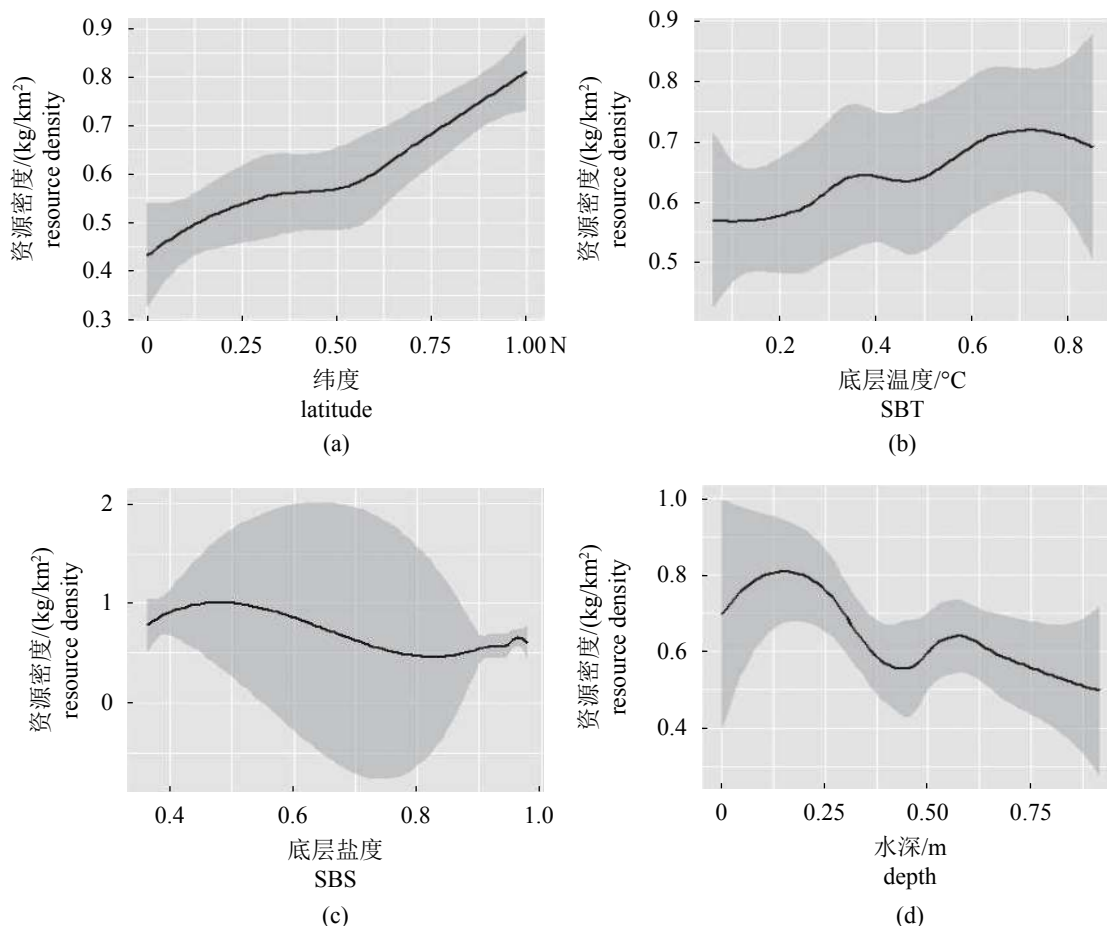


图3 Model 8 各影响因子与口虾蛄资源密度之间的关系

Fig. 3 Relationship between the impact factors of Model 8 and the resource density of *O. oratoria*

会抑制其器官组织的活动性，影响其生长发育有关。徐海龙等<sup>[39]</sup>研究表明，口虾蛄最适温度为16~24℃，调查海域底层温度为6.26~27.12℃，口虾蛄资源密度随底层温度升高呈先上升后下降的趋势更符合以往研究。同时，可能与Model 8对实验数据拟合效果最好，能更好地反映真实情况有关。另外，Model 8中的增加水深与口虾蛄资源密度存在密切关系，其资源密度随水深增加呈先上升后下降的趋势。该结果可能与口虾蛄的最适水深为30m以浅海域<sup>[40]</sup>有关，在30m范围内，随水深增加，口虾蛄资源密度呈上升趋势，超过最适水深时口虾蛄资源密度逐渐降低。另一方面，可能与口虾蛄的生活史特征有关。相关研究表明，口虾蛄资源量季节间存在明显差异，资源量由高到低依次为夏季、春季、秋季、冬季<sup>[40]</sup>。同时，口虾蛄季节间在深浅水区作短距离洄游<sup>[40]</sup>，春、夏季在近岸浅水区觅食和产卵，秋、冬季逐渐向深水区越冬迁徙。因此，

造成口虾蛄资源密度随水深增加先上升后下降。此外，不同季节口虾蛄的空间分布和资源量存在差异，但本研究模型构建中未选择季节作为关键解释因子，可能与季节效应被其他环境因子代替有关，季节作为一个综合性因素，其实际效应也体现在水温等因子的变化上。

### 3.2 单方面优化效果

本研究表明，采用3种方法分别优化对Model 1的影响存在较大差异。GA算法优化，模型的预测性能和拟合效果相对于Model 1基本保持不变，其模型预测的准确性略有提升。该结果与穆阿华等<sup>[41]</sup>于2005年提出的GA算法和BP算法相结合，发现的GA算法能够以较快的速度减小搜索空间，且不易陷入局部极小值问题等结果不相符，可能与GA算法优化得到的初始权值与Model 1随机选取的相差不大有关；GMDH算法优化对模型输入变量的选择提供更

为合适的方法, 2006 年蒋薇等<sup>[18]</sup>提出采用 GMDH 算法对 BP 算法进行改进, 指出 GMDH 算法能够有效优化网络结构。优化后的模型预测性能、拟合效果和模型预测的准确性均有明显提升, 与蒋薇等研究相一致。可能与 GMDH 算法从模型结构方面进行优化, 筛选出更为合适的输入变量有关。该算法增加水深变量作为影响口虾资源密度的关键环境因子, 对网络结构进行优化, 进而提升了模型的预测性能。水深变量的增加可能与口虾底栖生活有关, 在一定程度上验证了 GMDH 算法对变量选择的优势<sup>[14]</sup>, 同时模型预测的准确性提升说明与模型变量选择的正确性有关。自适应算法与 GMDH 算法优化模型均是对模型结构进行优化, 适应算法优化, 模型的拟合效果在单方优化中最高, 但模型预测性能和预测的准确性均明显下降, 可能是自适应算法优化的隐节点数目 4 相比经验选择的数目 3, 在一定程度上增加了模型的复杂度, 偏离了数据的真实情况, 出现过度拟合现象, 对模型预测性能造成影响。

### 3.3 多重优化效果

Model 5~7 对 Model 1 进行两方面组合优化, 在 GMDH 算法和自适应算法的单方面优化基础上增加 GA 算法进行组合优化, 增加 GA 算法优化前后模型整体性能基本保持不变。该结果与李松等<sup>[42]</sup>和陈文等<sup>[43]</sup>采用 GA 算法能有效提升模型性能的结论不一致, 但在一定程度上验证了 GA 算法优化前后, 模型预测性能和拟合效果基本保持不变的结论; 在 GA 算法和自适应算法单方面优化的条件下, 增加 GMDH 算法优化后, 模型的拟合程度、预测效果和模型预测的准确性均有明显提升。该结果可能与建模过程中模型结构和变量的自动筛选保证了系统的收敛速度, 避免模型的过拟合和拟合不足有关, 与陈建平<sup>[44]</sup>和李牡丹等<sup>[45]</sup>研究结果相一致。同时可能与 GMDH 算法能够在很大程度上使模型更加接近实际情况, 并收敛于全局最优有关; 而在 GA 算法和 GMDH 算法单方面优化的基础上, 增加自适应算法组合优化后, 模型的预测性能略有降低, 整体性能也出现下降。该结论与刘维群等<sup>[46]</sup>研究结果有所差异, 可能是隐节点数目的增加使口虾资源密度与环境因子之间的关系复杂化, 与调查得到的真实关系有所差别,

从而降低模型的性能。该结果与单方面采用自适应算法优化降低模型性能的结果相一致, 一定程度上验证了结果的准确性。同时, 3 种方法共同优化在所有组合模型中预测性能最好, 但模型预测的准确性与 GMDH 算法单方面优化及 GA 算法和 GMDH 算法两方面组合优化基本一致, 说明在模型预测的准确性方面没有明显优化, 这可能与模型之间存在的确定优化关系有关。

综上所述, 在一种方法优化模型的基础上增加不同优化方法, 与模型性能的提升之间并非简单的累加关系, 其可能存在复杂的不确定性。

## 4 结语

本研究讨论了 BP 神经网络模型的优化方法, 并提出可参考的优化策略, 为其他类型神经网络模型的发展提供参考。但需要注意的是, BP 神经网络模型的不确定性并不局限于本研究所探讨的 3 个方面, 同时模型效果也受到研究目标物种的生物学性状、生态系统结构特征以及数据质量<sup>[47]</sup>等的影响。此外, 人工神经网络模型作为一种黑箱工具, 虽在预测能力方面表现优秀, 但缺乏明确的解析关系。在未来模型优化和评估的研究中, 应更注重模型结果的生物学和生态学解释, 在确保数据准确性的同时获取长时间序列数据, 以提高模型质量, 为海洋渔业资源管理和保护提供依据。

### 参考文献 (References):

- [1] 栾静, 张崇良, 徐宾铎, 等. 海州湾双斑螯栖息分布特征与环境因子的关系[J]. 水产学报, 2018, 42(6): 889-901.  
Luan J, Zhang C L, Xu B D, et al. Relationship between catch distribution of Portunid crab (*Charybdis bimaculata*) and environmental factors based on three species distribution models in Haizhou Bay[J]. Journal of Fisheries of China, 2018, 42(6): 889-901 (in Chinese).
- [2] 陈雪忠, 樊伟, 崔雪森, 等. 基于随机森林的印度洋长鳍金枪鱼渔场预报[J]. 海洋学报, 2013, 35(1): 158-164.  
Chen X Z, Fan W, Cui X S, et al. Fishing ground forecasting of *Thunnus alalung* in Indian ocean based on random forest[J]. Acta Oceanologica Sinica, 2013, 35(1): 158-164 (in Chinese).
- [3] 杨胜龙, 张禹, 张衡, 等. 不同模型在渔业 CPUE 标准化中的比较分析[J]. 农业工程学报, 2015, 31(21): 259-

- 264.
- Yang S L, Zhang Y, Zhang H, *et al.* Comparison and analysis of different model algorithms for CPUE standardization in fishery[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(21): 259-264 (in Chinese).
- [ 4 ] Ivakhenko A G, Savchenko E A, Ivakhenko G A. Problems of future GMDH algorithms development[J]. *Systems Analysis Modelling Simulation*, 2003, 43(10): 1301-1309.
- [ 5 ] 王坤. BP在数码管字识别中的应用[J]. *中国数据通信*, 2004, 6(11): 88-91.
- Wang K. Application of BP to nixie tube- number's recognition[J]. *China Data Communications*, 2004, 6(11): 88-91 (in Chinese).
- [ 6 ] 谭显胜, 周铁军. BP算法改进方法的研究进展[J]. *怀化学院学报*, 2006, 25(2): 126-130.
- Tan X S, Zhou T J. Methods to improve BP neural network[J]. *Journal of Huaihua University*, 2006, 25(2): 126-130 (in Chinese).
- [ 7 ] 李伟林, 廖恩红. 自适应调节的隐节点神经网络结构优化算法[J]. *计算机工程与设计*, 2017, 38(6): 1664-1667, 1685.
- Li W L, Liao E H. Adjustable hidden node algorithm for optimal neural network architecture[J]. *Computer Engineering and Design*, 2017, 38(6): 1664-1667, 1685 (in Chinese).
- [ 8 ] 张磊, 胡春, 钱锋. BP算法局部极小问题改进的研究进展[J]. *工业控制计算机*, 2004, 17(9): 33-34, 50.
- Zhang L, Hu C, Qian F. An overview of the improved methods for solving the problem of local Minimum in the BP algorithm[J]. *Industrial Control Computer*, 2004, 17(9): 33-34, 50 (in Chinese).
- [ 9 ] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [10] 墨蒙, 赵龙章, 龚媛雯, 等. 基于遗传算法优化的BP神经网络研究应用[J]. *现代电子技术*, 2018, 41(9): 41-44.
- Mo M, Zhao L Z, Gong A W, *et al.* Research and application of BP neural network based on genetic algorithm optimization[J]. *Modern Electronics Technique*, 2018, 41(9): 41-44 (in Chinese).
- [11] 高晓. 连续型 Schaefer 产量模型和 BP 网络模型在渔业中的应用研究 [D]. 武汉: 华中农业大学, 2009.
- Gao X. The study between continuous Schaefer-form production model and BP network model in fishery[D]. Wuhan: Huazhong Agricultural University, 2009 (in Chinese).
- [12] 鲁娟娟, 陈红. BP神经网络的研究进展[J]. *控制工程*, 2006, 13(5): 449-451, 456.
- Lu J J, Chen H. Researching development on BP neural networks[J]. *Control Engineering of China*, 2006, 13(5): 449-451, 456 (in Chinese).
- [13] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 12763.6-2007 海洋调查规范 第 6 部分: 海洋生物调查 [S]. 北京: 中国标准出版社, 2008.
- General Administration of Quality Supervision, Inspection and Quarantine of the people's Republic of China, Standardization Administration of China. GB/T 12763.6-2007 Specifications for oceanographic survey—part 6: marine biological survey[S]. Beijing: Standards Press of China, 2008 (in Chinese).
- [14] 邹昊飞, 夏国平, 杨方廷. 基于自组织算法的改进型 GAANN 预测模型[J]. *中国管理科学*, 2005, 13(6): 75-80.
- Zou H F, Xia G P, Yang F T. Forecasting model using a hybrid GMDH and improved arithmetic of neural network based on genetic algorithm[J]. *Chinese Journal of Management Science*, 2005, 13(6): 75-80 (in Chinese).
- [15] 贺昌政, 吕建平. 自组织数据挖掘理论与经济系统的复杂性研究[J]. *系统工程理论与实践*, 2001, 21(12): 1-5, 35.
- He C Z, Lv J P. Study of self-organizing data mining theory and the complexity of economic systems[J]. *Systems Engineering-Theory & Practice*, 2001, 21(12): 1-5, 35 (in Chinese).
- [16] 李晓峰. 基于自组织方法的人工神经网络经济预测模型的建立[J]. *预测*, 2002, 21(6): 64-66, 63.
- Li X F. The establishment of economy forecasting model based on GMDH and artificial neural network[J]. *Forecasting*, 2002, 21(6): 64-66, 63 (in Chinese).
- [17] 何跃, 鲍爱根, 贺昌政. 自组织建模方法和GDP增长模型研究[J]. *中国管理科学*, 2004, 12(2): 139-142.
- He Y, Bao A G, He C Z. Self-organizing modeling methods and a model study on the growth of GDP[J]. *Chinese Journal of Management Science*, 2004, 12(2): 139-142 (in Chinese).



- [18] 蒋薇, 何跃, 魏仕强. 基于GMDH方法和BP神经网络的经济预警模型[J]. *现代管理科学*, 2006(10): 58-60.  
Jiang W, He Y, Wei S Q. Economic early warning model based on GMDH method and BP neural network[J]. *Modern Management Science*, 2006(10): 58-60 (in Chinese).
- [19] Bates J M, Granger C W J. The combination of forecasts[J]. *Journal of the Operational Research Society*, 1969, 20(4): 451-468.
- [20] 齐银峰, 谭荣建. 基于改进粒子群优化算法的BP神经网络在大坝变形分析中的应用[J]. *水利水电技术*, 2017, 48(2): 118-124.  
Qi Y F, Tan R J. Application of improved particle swarm optimization algorithm-based BP neural network to dam deformation analysis[J]. *Water Resources and Hydro-power Engineering*, 2017, 48(2): 118-124 (in Chinese).
- [21] 洪露, 马长山, 谢宗安. 基于遗传算法的神经网络权值优化[J]. *贵州工业大学学报(自然科学版)*, 2003, 32(6): 48-51.  
Hong L, Ma C S, Xie Z A. Optimization of weights in neural networks based on genetic algorithm[J]. *Journal of Guizhou University of Technology (Natural Science Edition)*, 2003, 32(6): 48-51 (in Chinese).
- [22] 吴旻, 陈长明, 史哲, 等. 遗传算法及其改进研究[J]. *湖北广播电视大学学报*, 2011, 31(9): 157-158.  
Wu M, Chen C M, Shi Z, *et al.* Research on genetic algorithm and its improvement[J]. *Journal of Hubei TV University*, 2011, 31(9): 157-158 (in Chinese).
- [23] 孙海峰, 沈颖, 王亚楠. 基于遗传算法优化BP神经网络的接触电阻预测[J]. *电测与仪表*, 2019, 56(5): 77-83.  
Sun H F, Shen Y, Wang Y N. Prediction of contact resistance based on optimized BP neural network of genetic algorithm[J]. *Electrical Measurement & Instrumentation*, 2019, 56(5): 77-83 (in Chinese).
- [24] 车国鹏, 刘永红. 遗传算法优化BP神经网络的交通流参数预测[J]. *综合运输*, 2018, 40(6): 64-67, 108.  
Che G P, Liu Y H. A prediction of traffic flow parameters based on BP neural network optimized by genetic algorithm[J]. *China Transportation Review*, 2018, 40(6): 64-67, 108 (in Chinese).
- [25] 韩旭, 王蒙. 基于遗传算法优化BP神经网络的故障电弧识别[J]. *测控技术*, 2016, 35(12): 21-25, 29.  
Han X, Wang M. Arc fault identification based on BP neural network optimized by genetic algorithm[J]. *Measurement & Control Technology*, 2016, 35(12): 21-25, 29 (in Chinese).
- [26] 乔姗姗. 基于遗传算法优化的 BP 神经网络在建筑工程投标报价中应用的研究 [D]. 扬州: 扬州大学, 2012.  
Qiao S S. Research based on genetic algorithm optimization on the application of BP neural networks in tender offer for construction projects[D]. Yangzhou: Yangzhou University, 2012 (in Chinese).
- [27] 陈郁明. 遗传算法在自动卷组中的应用 [D]. 贵阳: 贵州大学, 2008.  
Chen Y M. Application of genetic algorithm in automatic test paper generation[D]. Guiyang: Guizhou University, 2008 (in Chinese).
- [28] 李晓峰, 刘光中. 人工神经网络BP算法的改进及其应用[J]. *四川大学学报(工程科学版)*, 2000, 32(2): 105-109.  
Li X F, Liu G Z. The improvement of BP algorithm and its application[J]. *Journal of Sichuan University (Engineering Science Edition)*, 2000, 32(2): 105-109 (in Chinese).
- [29] 范佳妮, 王振雷, 钱锋. BP人工神经网络隐层结构设计的研究进展[J]. *控制工程*, 2005, 12(S1): 109-113.  
Fan J N, Wang Z L, Qian F. Research progress structural design of hidden layer in BP artificial neural networks[J]. *Control Engineering of China*, 2005, 12(S1): 109-113 (in Chinese).
- [30] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. *Neural Networks*, 1989, 2(5): 359-366.
- [31] 李晓峰. 动态全参数自调整BP神经网络预测模型的建立[J]. *预测*, 2001, 20(3): 69-71.  
Li X F. The establishment of forecasting model based on BP neural network of self-adjusted all parameters[J]. *Forecasting*, 2001, 20(3): 69-71 (in Chinese).
- [32] 李晓峰, 徐玖平, 王荫清, 等. BP人工神经网络自适应学习算法的建立及其应用[J]. *系统工程理论与实践*, 2004, 24(5): 1-8.  
Li X F, Xu J P, Wang Y Q, *et al.* The establishment of self-adapting algorithm of BP neural network and its application[J]. *Systems Engineering-Theory & Practice*, 2004, 24(5): 1-8 (in Chinese).
- [33] Kabacoff R I. R in action: data analysis and graphics with R[M]. Greenwich: Manning Publications, 2011: 1-474.

- [34] Wood S N. Stable and efficient multiple smoothing parameter estimation for generalized additive models[J]. *Journal of the American Statistical Association*, 2004, 99(467): 673-686.
- [35] 曾韦霖, 马文军, 刘涛, 等. 构建气温-死亡关系模型中温度指标的选择[J]. *中华预防医学杂志*, 2012, 46(10): 946-951.  
Zeng W L, Ma W J, Liu T, *et al.* What temperature index is the best predictor for the impact of temperature on mortality[J]. *Chinese Journal of Preventive Medicine*, 2012, 46(10): 946-951 (in Chinese).
- [36] 盛骤, 谢式千, 潘承毅. 概率论与数理统计 [M]. 第四版. 北京: 高等教育出版社, 2009: 104-106.  
Sheng Z, Xie S Q, Pan C Y. Probability and statistics[M]. Fourth Edition. Beijing: Higher Education Press, 2008: 104-106 (in Chinese).
- [37] Hyndman R J, Koehler A B. Another look at measures of forecast accuracy[J]. *International Journal of Forecasting*, 2006, 22(4): 679-688.
- [38] 王春琳, 徐善良, 梅文骥, 等. 口虾蛄的生物学基本特征[J]. *浙江水产学院学报*, 1996, 15(1): 60-62.  
Wang C L, Xu S L, Mei W X, *et al.* A biological basic character of *Oratosquilla oratoria*[J]. *Journal of Zhejiang College of Fisheries*, 1996, 15(1): 60-62 (in Chinese).
- [39] 徐海龙, 刘海映, 林月娇. 温度和盐度对口虾蛄呼吸的影响[J]. *水产科学*, 2008, 27(9): 443-446.  
Xu H L, Liu H Y, Lin Y J. Effect of temperature and salinity on respiration of mantis shrimp (*Oratosquilla oratoria*)[J]. *Fisheries Science*, 2008, 27(9): 443-446 (in Chinese).
- [40] 李明坤, 徐宾铎, 薛莹, 等. 山东南部近海口虾蛄空间分布特征及其季节变化[J]. *水产学报*, 2019, 43(8): 1749-1758.  
Li M K, Xu B D, Xue Y, *et al.* Spatial distribution characteristics and seasonal variation of *Oratosquilla oratoria* in the southern coastal waters of Shandong Province[J]. *Journal of Fisheries of China*, 2019, 43(8): 1749-1758 (in Chinese).
- [41] 穆阿华, 周绍磊, 刘青志, 等. 利用遗传算法改进BP学习算法[J]. *计算机仿真*, 2005, 22(2): 150-151, 166.  
Mu A H, Zhou S L, Liu Q Z, *et al.* Using genetic algorithm to improve BP training algorithm[J]. *Computer Simulation*, 2005, 22(2): 150-151, 166 (in Chinese).
- [42] 李松, 刘力军, 解永乐. 遗传算法优化BP神经网络的短时交通流混沌预测[J]. *控制与决策*, 2011, 26(10): 1581-1585.  
Li S, Liu L J, Xie Y L. Chaotic prediction for short-term traffic flow of optimized BP neural network based on genetic algorithm[J]. *Control and Decision*, 2011, 26(10): 1581-1585 (in Chinese).
- [43] 陈文, 庞琳娜. GABP神经网络在交通流预测中的应用研究[J]. *微计算机信息*, 2009, 25(14): 245-247.  
Chen W, Pang L N. The research of the application of GABP neural network in traffic flow prediction[J]. *Control & Automation*, 2009, 25(14): 245-247 (in Chinese).
- [44] 陈建平, 杨宜民, 张会章, 等. 一种基于GMDH模型的神经网络学习算法[J]. *云南大学学报(自然科学版)*, 2008, 30(6): 569-574.  
Chen J P, Yang Y M, Zhang H Z, *et al.* A neural network learning algorithm based on GMDH model[J]. *Journal of Yunnan University (Natural Science Edition)*, 2008, 30(6): 569-574 (in Chinese).
- [45] 李牡丹, 王印松. 基于灰色GMDH网络组合模型的风速预测[J]. *可再生能源*, 2017, 35(4): 522-527.  
Li M D, Wang Y S. Wind speed prediction based on grey GMDH network combined model[J]. *Renewable Energy Resources*, 2017, 35(4): 522-527 (in Chinese).
- [46] 刘维群, 李元臣. BP网络中隐含层节点优化的研究[J]. *交通与计算机*, 2005, 23(2): 83-86.  
Liu W Q, Li Y C. Optimization of hidden layer units of BP neural network[J]. *Computer and Communications*, 2005, 23(2): 83-86 (in Chinese).
- [47] 王晨. BP神经网络在中压加氢裂化装置多方面预测中的应用研究[J]. *石油炼制与化工*, 2018, 49(7): 95-99.  
Wang C. Application in multi-aspect prediction in hydrocracking unit by BP neural network[J]. *Petroleum Processing and Petrochemicals*, 2018, 49(7): 95-99 (in Chinese).

## Relationship between spatial distribution of *Oratosquilla oratoria* and environmental factors in Shandong offshore based on optimized BP neural network model analysis

LI Pengcheng<sup>1,2</sup>, ZHANG Chongliang<sup>1,2\*</sup>, REN Yiping<sup>1,2,3</sup>, XU Binduo<sup>1,2</sup>, XUE Ying<sup>1,2</sup>

(1. College of Fisheries, Ocean University of China, Qingdao 266003, China;

2. Field Observation and Research Station of Haizhou Bay Fishery Ecosystem, Ministry of Education, Qingdao 266003, China;

3. Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China)

**Abstract:** As a common machine learning method, BP neural network model is widely used in species distribution models to analyze the relationship between biological distribution and environmental factors. Compared with the traditional regression models, this model can flexibly deal with the nonlinear relationship between variables. However, there are substantial uncertainties in parameter setting as a result of its complex structure, which may affect the prediction and application of this model. This study considered approaches to optimize the model parameters, including the group method of data handling, genetic algorithm and adaptive algorithm, to improve initial weights and the number of hidden nodes of the model, respectively. Seven combinations of optimized BP models were implemented based on the survey data obtained from fishery resources and environment in Shandong offshore between 2016 and 2017. Our results showed that there were significant differences in the predictive performance of the seven optimization models. The predictive performance of the one-way and two-way optimization models was approximately the same. The root mean square error and the square of residual error were 0.35 and 1.94 respectively, which were smaller than the initial model's 0.52 and 2.40, and the maximum correlation coefficient was 0.45, indicating that the optimization effect of the model was the best. After the comparison and optimization, it was found that the resource density of *Oratosquilla oratoria* was basically different with the increase of bottom salinity while the resource density of *O. oratoria* was significantly different with the increase of bottom temperature. In addition, the increase of water depth in the optimal model compared with the initial model was a key environmental factor, which had an important effect on the resource density of *O. oratoria*. In this study, the parameter optimization method of the BP neural network model was further developed, which proved that the parameter optimization had important effect on the prediction performance of the BP model, and the model optimization was of great significance for the analysis of the relationship between resource density and environmental factors.

**Key words:** *Oratosquilla oratoria*; species distribution model; artificial neural network; model optimization; predictive performance

**Corresponding author:** ZHANG Chongliang. E-mail: zcl.0903@163.com

**Funding projects:** National Key Research and Development Program (2018YFD0900906); National Natural Science Foundation of China (31802301)