

胡晶晶, 罗永明, 张纲强, 等. 基于邻域粗糙集的气象因子选择在虾塘水温预测中的应用[J]. 江苏农业学报, 2023, 39(3): 732-740.
doi:10.3969/j.issn.1000-4440.2023.03.014

基于邻域粗糙集的气象因子选择在虾塘水温预测中的应用

胡晶晶¹, 罗永明², 张纲强³, 匡昭敏², 谢映², 曾行吉⁴

(1. 广西民族大学电子信息学院, 广西 南宁 530006; 2. 广西壮族自治区气象科学研究所, 广西 南宁 530022; 3. 广西民族大学人工智能学院, 广西 南宁 530006; 4. 广西壮族自治区气象信息中心, 广西 南宁 530022)

摘要: 基于邻域粗糙集对影响虾塘水温变化的气象因子进行选择, 并选取模型预测虾塘水温, 为南美白对虾养殖趋利避害提供科学参考。首先, 将平均气温、最高气温、最低气温、降水量、气压、2 min 风速、10 min 风速和瞬时风速等 8 个气象因子组合输入 SFNN 模型(单隐层前馈神经网络模型)、高斯回归模型和岭回归模型进行虾塘水温预测, 选取预测效果最好的 SFNN 模型为本研究预测模型。然后, 运用邻域粗糙集和熵理论, 考虑气象因子和虾塘水温之间的相关性、冗余性和交互性, 选出影响虾塘水温变化的主要气象因子。最后, 利用选出的主要气象因子和 SFNN 模型实现虾塘水温预测。将基于邻域粗糙集选出的 5 个气象因子组合与 8 个气象因子组合, 以及 8 个气象单因子分别输入 SFNN 模型, 预测结果表明: 邻域粗糙集选出的 5 个气象因子组合预测结果最好, 其预测均方根误差、均方误差、平均绝对误差最小, 分别为 1.121 1、1.256 9 和 0.893 8, 决定系数(R^2) 为 0.791 6; 在气象单因子中, 气压对虾塘水温的预测结果较好。因此, 基于邻域粗糙集选出的 5 个气象因子组合, 通过 SFNN 模型进行虾塘水温预测结果最好, 此方法在南美白对虾养殖趋利避害、防灾减灾中具有一定的实用价值。

关键词: 虾塘水温; 气象因子; 邻域粗糙集; SFNN 模型

中图分类号: S966.12 **文献标识码:** A **文章编号:** 1000-4440(2023)03-0732-09

Application of meteorological factor selection based on neighborhood rough set in shrimp pond water temperature prediction

HU Jing-jing¹, LUO Yong-ming², ZHANG Gang-qiang³, KUANG Zhao-min², XIE Ying², ZENG Xing-ji⁴

(1. College of Electronic Information, Guangxi Minzu University, Nanning 530006, China; 2. Guangxi Zhuang Autonomous Region Institute of Meteorological Sciences, Nanning 530022, China; 3. School of Artificial Intelligence, Guangxi Minzu University, Nanning 530006, China; 4. Guangxi Zhuang Autonomous Region Meteorological Information Center, Nanning 530022, China)

Abstract: Based on the neighborhood rough set, the meteorological factors affecting the change of shrimp pond water temperature were selected. A model was selected to predict the water temperature of shrimp pond, which provided a scientific

reference for shrimp culture to seek advantages and avoid disadvantages. Firstly, eight meteorological factors, including average temperature, highest temperature, lowest temperature, amount of precipitation, barometric pressure, two-minute wind speed, ten-minute wind speed, and instantaneous wind speed, were combined into the single hidden layer feed forward neural network (SFNN) model, Gaussian regression model, and bridge regression model to predict the water temperature of shrimp ponds. SFNN model with the best prediction effect was selected as the pre-

收稿日期: 2022-07-21

基金项目: 广西壮族自治区自然科学基金项目(2020GXNSFAA2-38046); 广西壮族自治区气象局气象科研计划重点项目(桂气科 2020Z03); 广西民族大学软件工程重点实验室项目(2022-18XJSY-03)

作者简介: 胡晶晶(1996-), 女, 江苏淮安人, 硕士研究生, 主要从事粗糙集理论、数据挖掘与知识发现研究。(E-mail) jingjingu_star304@163.com

通讯作者: 罗永明, (E-mail) mingyongluo858@163.com

diction model. Then considering the correlation, redundancy and interaction between meteorological factors and shrimp pond water temperature, the main meteorological factors affecting the shrimp pond water temperature were selected by using neighborhood rough set and entropy theory. Finally, the water temperature of the shrimp pond was predicted by using the selected meteorological factors and SFNN model. Combination of five meteorological factors selected based on neighborhood rough set, the combination of eight meteorological factors, and eight single meteorological factors were input into the SFNN model, respectively. The prediction results showed that, the combination of five meteorological factors selected by neighborhood rough set had the best prediction results, and its root mean square error, mean square error and mean absolute error were the smallest, which were 1.121 1, 1.256 9 and 0.893 8, respectively. And the determination coefficient was 0.791 6. Among single meteorological factors, atmospheric pressure predicted best on shrimp pond water temperature. Therefore, the combination of five meteorological factors selected based on the neighborhood rough set had the best prediction in shrimp pond water temperature by the SFNN model. This method had practical value in seeking advantages and avoiding disadvantages, and disaster prevention and reduction in prawn breeding.

Key words: shrimp pond water temperature; meteorological factors; neighborhood rough set; single hidden layer feed forward neural network (SFNN) model

近年来,广西水产养殖面积不断增大,南美白对虾因其个体大、生长快、营养需求低、抗病能力强等优点^[1],已经成为广西对虾养殖的一个主要养殖品种,并成为养殖户收入增长的主要来源。南美白对虾又名凡纳滨对虾(简称:对虾),其生长发育与虾塘的水环境密切相关^[2],其中水温起伏变化直接影响其生长发育,或引发疾病,甚至导致死亡。由于广西南美白对虾基本是露天海水养殖,观测水温仪器易被腐蚀,成本高,无法在生产中大量应用。然而,虾塘水温与气象条件息息相关^[3-4]。因此,构建虾塘水温预测模型,通过气象因子实现水温预测,应用于生产,对广西南美白对虾养殖趋利避害、防灾减灾具有重要的现实意义。国内外关于水温预测研究多集中在海水^[5-6]、水库^[7-8]及河流^[9-10]中。邓爱娟等^[9]利用逐步回归法对养殖鱼塘水温进行了预测。韦耀国等^[10]分析了气温、相对湿度、地温、太阳辐射等因素,建立了基于BP神经网络的南水北调中线水温预报模型。施珮等^[11]基于EMD-IGA-SELM方法预测了养殖池水温。陈英义等^[12]针对池塘水温小样本,提出了基于GA-BP神经网络预测养殖塘短期水温。柳海涛等^[13]根据水文与气象数据,基于人工神经网络建立了鱼类产卵场水温预测模型。朱森林等^[14]根据高斯过程回归建立了密苏里河水温预测模型。此外,还有许多学者以气温作为影响水温的主要气象因子,提出水温预测模型^[15-18]。由于粗糙集理论^[19]在处理信息不确定性方面的巨大优势,Hu等^[20]将邻域的概念引入粗糙集理论,研究并提出了邻域粗糙集,并进行数据挖掘,近年来在群体决

策^[21]、属性选择^[22-23]、不确定性推理^[24]等领域得到了成功应用。同时,结合邻域粗糙集与熵理论的属性选择方法在信息熵^[25]、条件信息熵^[26]、互信息^[27]和自信息^[28]等方面都取得了显著进展。前人在利用气象因子预测水温方面已经开展了大量有价值的研究,但现有虾塘水温预测模型没有充分考虑气象因子之间相互作用对虾塘水温变化的影响,且缺乏对气象因子进行选择的有效工具。为了弥补这一方面研究的缺失,邻域粗糙集能有效刻画影响虾塘水温变化的关键气象因子,为气象因子选择提供了新的视角。本文选取了平均气温、最高气温、最低气温、降水量、气压、2 min 风速、10 min 风速和瞬时风速共8个气象因子,度量各个气象因子对虾塘水温变化的不确定性,对8个气象因子进行选择,得到影响虾塘水温预测的关键气象因子,分析其对虾塘水温的预测效果,实现虾塘水温预测功能,为南美白对虾养殖过程中趋利避害提供科学参考。

1 材料和方法

1.1 研究区域

广西北部湾沿海地处亚热带季风气候区,气候温和,降水充沛,具有得天独厚地理优势,是农业农村部指定的重点南美白对虾产业带,2020年广西南美白对虾产量 3.3×10^5 t,占中国总产量27.67%,产量排名中国第2。

1.2 数据来源与试验环境

选取广西钦州市犀牛脚镇沙角村的一个露天南美白对虾养殖水塘作为水温数据的采集点(东经

108.84°, 北纬 21.65°), 采集点位于广西南美白对虾产区较中心位置。采用美国 YSI ProDSS 多参数水质分析仪, 每隔 1 min 采集水温、pH 值、溶解氧等数据, 共采集数据 61 676 条。采集时间: 2021 年 10 月 13 日至 11 月 25 日, 采集数据期间有 3 次明显降温过程。借鉴前人研究经验^[3-4, 18], 选取平均气温、最高气温、最低气温、降水量、气压、2 min 风速、10 min 风速和瞬时风速共 8 个气象因子作为影响虾塘水温的因素, 其中气象数据来自离观测点最近的合浦西场国家自动气象观测站, 东经 108.95°, 北纬 21.14°, 观测得到的 8 个气象因子的小时数据共 1 004 条。根据时间, 将虾塘观测的水温与气象数据匹配成小时数据样本集。

本研究使用深度学习框架 Tensorflow2.0 和 Python3.6 搭建试验环境。试验数据集在所有模型中划分为训练集和测试集, 划分比例为 9:1, 以 1 h 为单位的 100 条测试数据来预测虾塘水温的每小时变化。

1.3 数据预处理

采用美国 YSI ProDSS 多参数水质分析仪采集水温数据, 试验过程中, 仪器运行相当稳定, 无缺失值和异常值; 采用的气象观测站数据属于气象业务数据, 质量稳定可靠, 8 个气象因子数据仅存在极少的缺失值, 针对个别缺失的气象小时数据, 在数据样本集中进行了删除。为了减少不同维度数据之间由于量纲造成的差异, 加速模型拟合速度, 使用最大最小归一化方法将数据映射到 [0, 1] 范围内。归一化公式如下:

$$x_{\text{new}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

其中, x_{new} 为归一化后的值, x_{\min} 为最小值, x_{\max} 为最大值。

1.4 模型构建

1.4.1 单隐层前馈神经网络 单隐层前馈神经网络 (Single hidden layer feedforward neural network, SFNN) 模型^[29]是神经网络模型中最简单的拓扑结构, 由输入层、隐藏层和输出层构成。由图 1 可知 SFNN 的拓扑结构: 其各神经元分层排列, 每个神经元与前一层的神经元相连, 接受前一层的输出, 并输出给下一层。神经网络的能力主要取决于隐藏层, 包括隐藏层节点数和所使用的激活函数类型。其核心原理是通过激活函数将输入层数据进行非线性组合生成非线性决策边界, 从而提高网络的学习能力。

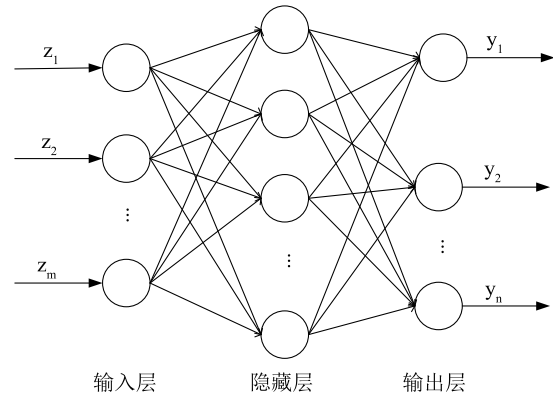


图1 SFNN(单隐层前馈神经网络)拓扑结构

Fig.1 Topological structure of single hidden layer feed forward neural network (SFNN)

假设 SFNN 有 m 个输入节点, t 个隐藏层节点, n 个输出层节点。以一个二分类数据集为例子, 给定一个训练样本集 $\{(z_{\varepsilon}, y_{\varepsilon})\}_{\varepsilon=1}^{\lambda}$, 其中 $z_{\varepsilon} = (z_{\varepsilon 1}, z_{\varepsilon 2}, z_{\varepsilon 3}, \dots, z_{\varepsilon m})^T$ 是第 ε 个 m 维实例向量, $y_{\varepsilon} \in \{0, 1\}$ 是第 ε 个实例对应的标签。设置输入层和隐藏层之间的连接权重 $w^{(1)} = [w_1^{(1)}, w_2^{(1)}, \dots, w_t^{(1)}]^T$, $w_{\beta}^{(1)} = [w_{\beta 1}^{(1)}, w_{\beta 2}^{(1)}, \dots, w_{\beta m}^{(1)}], \beta = 1, 2, \dots, t$ 。输入层和隐藏层之间的偏置为 $b^{(1)} = [b_1^{(1)}, b_2^{(1)}, \dots, b_t^{(1)}]^T$, $b_{\beta}^{(1)} = [b_{\beta 1}^{(1)}, b_{\beta 2}^{(1)}, \dots, b_{\beta m}^{(1)}], \beta = 1, 2, \dots, t$ 。隐藏层和输出层之间的连接权重 $w^{(2)} = [w_1^{(2)}, w_2^{(2)}]^T$, 其中 $w_h^{(2)} = [w_{h1}^{(2)}, w_{h2}^{(2)}, \dots, w_{ht}^{(2)}], h = 1, 2$ 。隐藏层和输出层之间的偏置 $b^{(2)} = [b_1^{(2)}, b_2^{(2)}]^T$, 其中 $b_h^{(2)} = [b_{h1}^{(2)}, b_{h2}^{(2)}, \dots, b_{hd}^{(2)}], h = 1, 2$ 。以 $w^{(l)}$ 为例, $l=1$ 与 $l=2$ 分别表示从输入层到隐藏层以及从隐藏层到输出层的连接权重。假设隐藏层的输出矩阵 $G[w^{(1)} \cdot Z + b^{(1)}]$ 为:

$$\begin{cases} g_1[w_1^{(1)} \cdot z_1 + b_1^{(1)}] & \cdots & g_1[w_1^{(1)} \cdot z_{\lambda} + b_1^{(1)}] \\ \vdots & & \vdots \\ g_t[w_t^{(1)} \cdot z_1 + b_t^{(1)}] & \cdots & g_t[w_t^{(1)} \cdot z_{\lambda} + b_t^{(1)}] \end{cases}$$

其中, $Z = (z_1, z_2, \dots, z_{\lambda})$, 隐藏层激活函数 $G(\cdot) = [g_1(\cdot), g_2(\cdot), \dots, g_t(\cdot)]^T$, 最常用的激活函数包括 Sigmoid 和 ReLU。假设输出层的输出为 $\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{\lambda})$, 那么 $\bar{y} = w^{(2)} \cdot G[w^{(1)} \cdot X + b^{(1)}] + b^{(2)} = (\frac{\bar{y}_{11} \cdots \bar{y}_{1\lambda}}{\bar{y}_{21} \cdots \bar{y}_{2\lambda}})$, 使用 Sigmoid 函数给出预测结果, 关于第 h 维第 ε 个输出向量 $\bar{y}_{h\varepsilon}$, 有 $\hat{y}_{h\varepsilon} = \frac{\exp(\bar{y}_{h\varepsilon})}{\sum_{h=1}^2 \exp(\bar{y}_{h\varepsilon})}, \varepsilon = 1, 2, \dots, \lambda$, 因此, SFNN 的最终预测结果为:

$$\hat{y} = \begin{bmatrix} \frac{\exp(\bar{y}_{11})}{\sum_{h=1}^2 \exp(\bar{y}_{h1})} & \cdots & \frac{\exp(\bar{y}_{1d})}{\sum_{h=1}^2 \exp(\bar{y}_{hd})} \\ \frac{\exp(\bar{y}_{21})}{\sum_{h=1}^2 \exp(\bar{y}_{h1})} & \cdots & \frac{\exp(\bar{y}_{2d})}{\sum_{h=1}^2 \exp(\bar{y}_{hd})} \\ \vdots & \ddots & \vdots \\ \frac{\exp(\bar{y}_{\lambda 1})}{\sum_{h=1}^2 \exp(\bar{y}_{h1})} & \cdots & \frac{\exp(\bar{y}_{\lambda d})}{\sum_{h=1}^2 \exp(\bar{y}_{hd})} \end{bmatrix} = \begin{pmatrix} \hat{y}_{11} & \cdots & \hat{y}_{1\lambda} \\ \hat{y}_{21} & \cdots & \hat{y}_{2\lambda} \\ \vdots & \ddots & \vdots \\ \hat{y}_{\lambda 1} & \cdots & \hat{y}_{\lambda \lambda} \end{pmatrix}$$

1.4.2 SFNN 模型参数设置 由于神经网络模型可以使用一个包含足够多节点的隐藏层以任意精度逼近任意复杂度的连续函数,为了降低模型复杂度,SFNN 模型设置了一层隐藏层,如何设置隐藏层节点的数目相较经验参数调优、网格搜索、随机搜索等参数调优方法,贝叶斯优化^[30]是基于贝叶斯定理和高斯过程的参数优化方法,近年来被广泛用于机器学习模型的参数调优。

本研究借助 Tensorflow2.0 的 Keras Tuner 库中提供的贝叶斯优化器对 SFNN 模型中的隐藏层节点数和学习率进行贝叶斯参数寻优,优化的参数约束条件为:隐藏层节点数为 [5, 500],学习率为 [e-4, e-1]。SFNN 模型的参数设置为:隐藏层数为 1 层,隐藏层节点数为 445,学习率为 e-3,使用均方误差作为损失函数,ReLU 作为激活函数,采用优化器 Adam 使模型收敛更快,使用概率为 0.1 的 Dropout 以减少过拟合。

1.4.3 模型评价指标 本研究选取均方根误差 (RMSE)、均方误差 (MSE)、平均绝对误差 (MAE) 和决定系数 (R^2) 作为评价指标,其中 RMSE、MSE 和 MAE 表示模型对于真实值拟合的误差,值越小表示结果越准确; R^2 越接近 1 表示模型效果越好。相关评价指标具体如下:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m [y_i - f(x_i)]^2} \quad (2)$$

$$MSE = \frac{1}{m} \sum_{i=1}^m [y_i - f(x_i)]^2 \quad (3)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m [|y_i - f(x_i)|] \quad (4)$$

$$R^2 = 1 - \frac{MSE[y_i, f(x_i)]}{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2} \quad (5)$$

其中, y_i 为真实值, $f(x_i)$ 为模型预测值, \bar{y} 为 y_i 的均值, m 为样本总数。

1.4.4 SFNN 模型验证的对比模型选择 为了验证 SFNN 模型的有效性,将高斯回归模型^[14]和岭回归模型^[31]作为对比预测模型。高斯回归模型是基于完整的贝叶斯学习算法,其适用于处理高维数、样本

小和非线性等复杂的回归问题,广泛应用于水温、水质等预测。岭回归模型具有参数少,运行效率高特点,是简化的支持向量机回归算法,拟合效果好。因此,采用高斯回归模型和岭回归模型作为本研究的对比模型。

1.5 气象因子与虾塘水温的不确定性度量

目前大多数水温预测模型仅考虑气象因子与虾塘水温的相关性,常常忽略了它们之间的其他关系。从图 1 可见,本研究利用气象因子模拟虾塘水温时,发现它们之间除了存在相关性外,还存在冗余性和交互性。因此,基于邻域粗糙集与熵理论,使用邻域互信息和邻域条件互信息度量它们之间的不确定性。

定义 1^[20], 设 $U = \{x_1, x_2, \dots, x_i\}$ 是一个有限对象集, $NDIS = (U, C \cup \{d\}, \vartheta)$ 表示一个邻域决策信息系统, $C \cap \{d\} = \emptyset$, $B \subseteq C$, $C = \{a_1, a_2, \dots, a_j\}$ 为条件属性集, $\{d\}$ 为决策属性集, $\forall a \in C \cup \{d\}$, $(U, C \cup \{d\}, \vartheta)$ 确定一个信息函数 $a: U \rightarrow V_a$, 其中 V_a 是属性 a 的信息函数值的集合, $\vartheta > 0$ 为邻域半径参数。样本之间的相似性使用距离表示为:

$$D_B(x_i, x_k) = \left[\sum_{a=a_1}^{a_j} |a(x_i) - a(x_k)|^p \right]^{1/p} \quad (6)$$

其中, $a(x_i)$, $a(x_k)$ 分别表示样本 x_i, x_k 在属性 a 下的值。样本 x_i 在属性子集 B 下的邻域类表示为:

$$[x_i]_B^\vartheta = \{x_i | x_i \in U, D_B(x_i, x_k) \leq \vartheta\} \quad (7)$$

其中, $[x_i]_B^\vartheta$ 表示样本 x_i 在 B 上的邻域类, 邻域大小取决于邻域半径参数 ϑ 。

定义 2^[26], 给定一个 $NDIS = (U, C \cup \{d\}, \vartheta)$, $B \subseteq C \cup \{d\}$, 在属性子集 B 中的邻域熵定义为:

$$NI(B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \lg \left(\frac{|[x_i]_B^\vartheta|}{|U|} \right) \quad (8)$$

其中, $|U|$ 表示样本数量。

(1) 相关性: 即气象因子对虾塘水温是有影响的。利用邻域互信息来度量气象因子与虾塘水温之间的相关性:

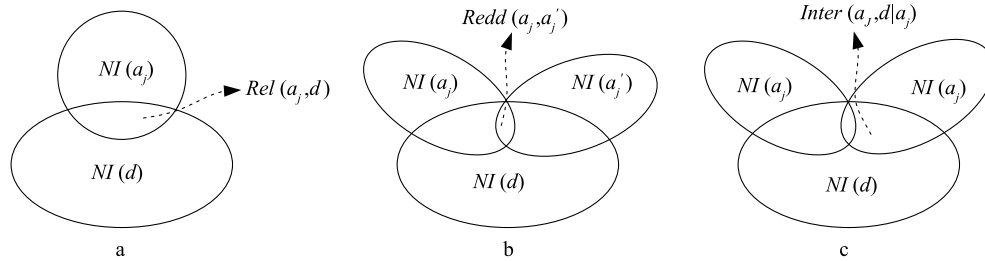
定义 3^[32], 给定一个 $NDIS = (U, C \cup \{d\}, \vartheta)$, $a_j \in C$, $\{d\}$ = 虾塘水温, C 为气象因子组成的条件属性集, 气象因子和虾塘水温之间的相关性定义为:

$$Rel(a_j, d) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \lg \left(\frac{|[x_i]_{a_j}^\vartheta| \cdot |[x_i]_d^\vartheta|}{|[x_i]_{a_j}^\vartheta \cap [x_i]_d^\vartheta| \cdot |U|} \right) \quad (9)$$

其中, $[x_i]_{a_j}^\vartheta$ 表示样本 x_i 在 a_j 上的邻域类,

$[x_i]_d^\vartheta$ 表示样本 x_i 在 d 上的邻域类, $[x_i]_{a_j}^\vartheta \cap [x_i]_d^\vartheta$ 表示样本 x_i 在 $a_j \cap d$ 上的邻域类。对 $\forall a_1, a_2 \in C$, 若 a_2 相比 a_1 与虾塘水温更相关, 那么 $Rel(a_2, d) >$

$Rel(a_1, d) > 0$, 说明 a_2 对虾塘水温的影响比 a_1 更大。如图 2(a) 所示, 相交区域是气象因子 a_j 对虾塘水温的影响大小。



a: 相关性; b: 冗余性; c: 交互性。

图 2 气象因子与虾塘水温关系

Fig.2 Relationship between meteorological factors and shrimp pond water temperature

(2) 冗余性: 即气象因子对虾塘水温没有影响或产生了相同影响。利用邻域互信息来度量气象因子之间的冗余性:

定义 4^[32], 给定一个 $NDIS = (U, C \cup \{d\}, \vartheta)$, $a_j \in C, a_j' \in Reduct$, $Reduct$ 是已经选择的气象因子, $\{d\}$ = 虾塘水温, C 为气象因子组成的条件属性集, 气象因子之间的冗余性定义为:

$$Redd(a_j, a_j') = -\frac{1}{|U|} \sum_{i=1}^{|U|} \lg \left(\frac{|[x_i]_{a_j}^\vartheta| \cdot |[x_i]_{a_j'}^\vartheta|}{|[x_i]_{a_j}^\vartheta \cap [x_i]_{a_j'}^\vartheta| \cdot |U|} \right) \quad (10)$$

由图 2(b) 可知, $Redd(a_j, a_j')$ 度量了气象因子

$$Inter(a_j, d|a_j) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \lg \left(\frac{|[x_i]_{a_j}^\vartheta \cap [x_i]_d^\vartheta| \cdot |[x_i]_{a_j'}^\vartheta \cap [x_i]_d^\vartheta|}{|[x_i]_{a_j}^\vartheta \cap [x_i]_{a_j'}^\vartheta \cap [x_i]_d^\vartheta| \cdot |[x_i]_{a_j}^\vartheta|} \right) \quad (11)$$

该式解释了单个气象因子对虾塘水温影响较小, 但当气象因子之间相互作用时, 它们可能对虾塘水温产生更大的影响。由图 2(c) 可见, $Inter(a_j, d|a_j)$ 度量了气象因子相互作用对虾塘水温的影响大小。

1.6 气象因子选择

评价函数用于衡量气象因子对虾塘水温的影响

$$mfNRS = \argmax(a_j) = Rel(a_j, d) - \frac{1}{|Reduct|} \sum_{a_j' \in Reduct} Redd(a_j, a_j') + \frac{1}{|C-Reduct|-1} \sum_{a_j \in C-Reduct-a_j'} Inter(a_j, d|a_j) \quad (12)$$

其中, $Rel(a_j, d)$ 度量了气象因子与虾塘水温之间的相关性, $Redd(a_j, a_j')$ 度量了气象因子之间的冗余性, $Inter(a_j, d|a_j)$ 度量了气象因子之间的交互性。mfNRS 综合度量了气象因子与虾塘水温的相关性、冗余性和交互性, 可以根据评价函数 mfNRS 选择出影响虾塘水温的气象因子。

a_j 与 a_j' 对虾塘水温产生了相同影响的大小。

(3) 交互性: 即与相关性和冗余性不同, 强调气象因子之间相互作用对虾塘水温的影响。邻域条件互信息是在已知一个气象因子的条件下, 度量另外一个气象因子对虾塘水温的影响大小。利用邻域条件互信息来度量气象因子之间的交互性:

定义 5^[32], 给定一个 $NDIS = (U, C \cup \{d\}, \vartheta)$, $a_j \in C, a_j \in C-Reduct-\{a_j'\}$, $Reduct$ 是已经选择的气象因子, $\{d\}$ = 虾塘水温, C 为气象因子组成的条件属性集, 气象因子之间的交互性定义为:

大小, 基本思想是: 气象因子与虾塘水温之间的相关性越强、冗余性越小、气象因子之间交互性越强, 对虾塘水温预测能力就越好。本研究设计了一个基于邻域粗糙集的气象因子评价函数如下:

基于邻域粗糙集的气象因子选择算法:

输入: $NDIS = (U, C \cup \{d\}, \vartheta)$;

输出: 一个选择气象因子选择子集

/* 初始化 */

气象因子选择子集 $Reduct \leftarrow \emptyset$;

1) For each $a_j \in C$ do

2) $\text{ComputeRel}(a_j, d)$; /* 根据定义 3 计算气象因子与水温之间的相关性 */

3) End for

4) $\text{Reduct} \leftarrow \{a_j' \} \leftarrow$; /* 选择相关性最大的气象因子 */

5) $C \leftarrow C - \{a_j'\}$;

6) For each $a_j \in C$ do

7) For each $a_j' \in \text{Reduct}$ do

8) $\text{Compute Redd}(a_j, a_j')$; /* 根据定义 4 计算气象因子与水温的冗余性 */

9) For each $a_j \in C - \text{Reduct} - \{a_j'\}$ do

10) $\text{Compute Inter}(a_j, d | a_j)$; /* 根据定义 5 计算气象因子与水温的交互性 */

11) Compute mfnRS ; /* 根据公式 (12) 计算气象因子重要性 */

12) End for

13) Return Reduct ;

2 结果

2.1 SFNN 模型预测结果

为了分析不同模型对虾塘水温预测的效果,将 8 个气象因子组合作为高斯回归模型、SFNN 模型和岭回归模型的输入,进行虾塘水温预测。结果由表 1 和图 3 可知, SFNN 模型预测效果最好, $RMSE$ 、 MSE 和 MAE 最小, 分别为 1.326 1、1.758 6 和 1.028 8; 决定系数 (R^2) 最大为 0.708 4, 比高斯回归模型和岭回归模型分别高出 0.314 2 和 0.258 6。整体来看, SFNN 模型预测虾塘水温效果好于其他两种预测模型。因此, 本试验选择 SFNN 模型作为虾塘水温预测模型。

表 1 不同模型虾塘水温预测误差

Table 1 Prediction error of shrimp pond water temperature by different models

模型	$RMSE$	MSE	MAE	R^2
高斯回归模型	1.911 5	3.654 0	1.470 2	0.394 2
岭回归模型	1.821 6	3.318 5	1.437 5	0.449 8
SFNN 模型	1.326 1	1.758 6	1.028 8	0.708 4

$RMSE$: 均方根误差; MSE : 均方误差; MAE : 平均绝对误差; R^2 : 决定系数。

2.2 不同气象单因子预测效果分析

为了探究利用气象单因子预测虾塘水温的效果, 将气象单因子: 平均气温、最高气温、最低气

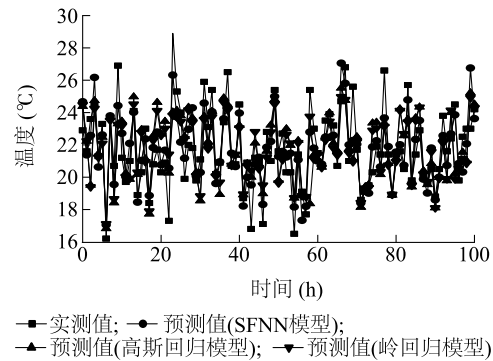


图 3 不同模型预测虾塘水温效果

Fig.3 Effects of different models on predicting shrimp pond water temperature

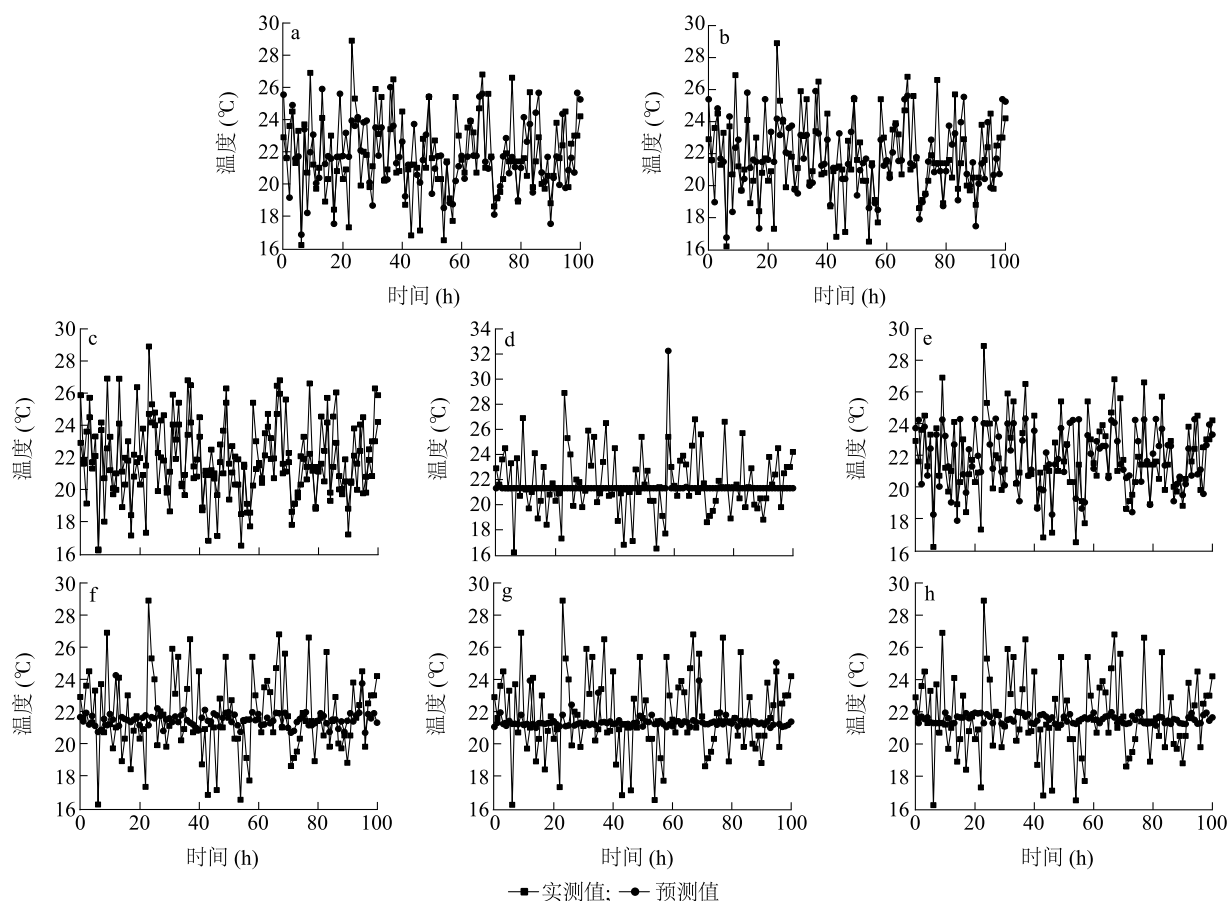
温、降水量、气压、2 min 风速、10 min 风速和瞬时风速等分别输入 SFNN 模型对虾塘水温进行预测。气象单因子对虾塘水温预测结果误差见表 2 所示, 其中, 利用气压预测虾塘水温与实测虾塘水温变化趋势拟合度较好, 详见图 4e, $RMSE$ 和 MSE 最小, 分别为 1.868 6 和 3.491 8; 利用最高气温、最低气温和平均气温预测虾塘水温的效果次之, 详见图 4a~图 4c, 处在同一水平; 利用 2 min 风速、10 min 风速和瞬时风速预测效果较差, 详见图 4f、图 4g 和图 4h, 处于同一水平; 降水量对虾塘水温预测效果最差, 详见图 4d, 在 57~60 h 时段内, 降水量对虾塘水温预测效果波动较大, 其余时段基本保持平稳, $RMSE$ 、 MSE 和 MAE 最大, 分别为 3.086 7、9.528 1 和 2.054 7。

表 2 不同气象因子预测虾塘水温的误差

Table 2 Prediction error of shrimp pond water temperature based on different meteorological factors

气象因子	$RMSE$	MSE	MAE	R^2
平均气温	2.018 3	4.073 8	1.557 8	0.324 6
最高气温	1.957 3	3.831 0	1.507 9	0.364 9
最低气温	2.079 3	4.323 5	1.640 2	0.283 2
降水量	3.086 7	9.528 1	2.054 7	-0.579 4
气压	1.868 6	3.491 8	1.535 3	0.421 1
2 min 风速	2.431 6	5.913 0	1.896 7	0.019 7
10 min 风速	2.441 9	5.963 0	1.913 7	0.011 5
瞬时风速	2.488 9	6.194 7	1.991 0	-0.026 9
8 个气象因子组合	1.323 0	1.750 3	1.029 7	0.708 4
5 个气象因子组合	1.121 1	1.256 9	0.893 8	0.791 6

$RMSE$ 、 MSE 、 MAE 、 R^2 见表 1 注。



a: 平均气温; b: 最高气温; c: 最低气温; d: 降水量; e: 气压; f: 2 min 风速; g: 10 min 风速; h: 瞬时风速。

图 4 不同气象单因子预测虾塘水温效果

Fig.4 Effects of different single meteorological factors on predicting shrimp pond water temperature

2.3 不同气象因子组合预测效果分析

本研究设计的基于邻域粗糙集的气象因子选择算法,充分考虑了气象因子与虾塘水温的相关性、冗余性和交互性,选择的气象因子组合为最低气温、最高气温、气压、降水量和 10 min 风速等 5 个气象因子,从表 2 和图 5 可知,同样是用 SFNN 模型对虾塘水温进行预测,选择的 5 个气象因子组合预测效果最好,比 8 个气象因子组合预测误差都小, $RMSE$ 、 MSE 和 MAE 分别小了 0.201 9、0.493 4 和 0.135 9。

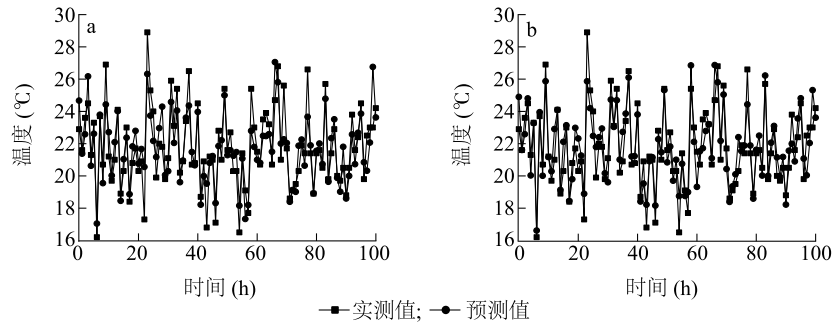
另外,8 个气象因子组合预测效果又好于气象单因子的预测效果。从 $RMSE$ 、 MSE 和 MAE 上看,降水量和 10 min 风速分别作为气象单因子预测水温的效果并不理想,但是在选择的 5 个气象因子组合中,降水量和 10 min 风速与其他气象因子相互作用后,表现出了良好的预测效果,这充分证明了气象因子之间相互作用对虾塘水温预测是有影响的。

3 讨论

邓爱娟等^[9]使用逐步回归法建立了洪湖地区春夏季鱼塘水温预报,研究水温与气温的相关性,分析了水温与当日以及前 1~3 d 的气温关系显著,选取当日、前 1~3 d 气象因子平均气温、最高气温和最低气温作为水温的影响因子,模型预测效果在 5 月较好,7 月较差。朱森林等^[14]利用机器学习高斯回归建立了密苏里河水温-气温模型,并与线性、非线性和随机回归模型进行了比较,论证了随机回归模型的精度高于线性模型和非线性模型;受水库调节温度的影响,水温-气温模型在 3 个站点的预测精度并不都是较高的。杨青青等^[18]采用多元线性回归法建立稻虾种养模式的虾沟水温预报模型,以平均气温、平均水温以及迁移 1 d 平均气温为预测因子,构建了日平均水温、日最高水温、日最低水温预报模型,

水温模型预测精度较高,由于在筛选与水温相关要素时主要集中在气温,范围较小,仅通过相关分析无

法准确判断出气温对虾塘水温变化的真实影响。以上研究都是基于气象因子的日数据。



a:8个气象因子组合;b:5个气象因子组合。

图5 8个气象因子组合与5个气象因子组合预测虾塘水温效果

Fig.5 Effects of combination of eight meteorological factors and combination of five meteorological factors on predicting shrimp pond water temperature

本研究借鉴各领域专家经验^[20,27,33-36],选取8个气象因子(小时数据)作为影响虾塘水温的气象因子,充分考虑气象因子与虾塘水温之间的相关性、冗余性和交互性,基于邻域粗糙集和熵理论,引入邻域互信息和条件互信息度量了它们之间的不确定性,并利用构建的评价函数得到影响虾塘水温的气象因子组合。从原来的8个气象因子中得到5个气象因子组合,即最低气温、10 min 风速、气压、降水量、最高气温。通过对测试集均方根误差、均方误差、平均绝对误差的综合分析,一方面论证了气象因子与虾塘水温之间存在相关性、冗余性和交互性;另一方面也表明了基于邻域粗糙集的气象因子选择的有效性。本试验获得了 SFNN 模型、高斯回归模型、岭回归模型预测虾塘水温的效果。它们的预测由好到差排序为:SFNN 模型、岭回归模型和高斯回归模型。本研究从气象因子对虾塘水温影响的角度出发,预测了虾塘水温的变化,但在实际生产中,除了本研究讨论的气象因子,可能还有其他因素对虾塘水温预测产生影响。在之后的工作中,将进一步考虑其他有关因素的影响,以进一步提高虾塘水温预测效果。

4 结论

本研究基于邻域粗糙集构造了虾塘水温预测模型,该模型考虑了气象因子之间的相互作用对虾塘水温变化的影响。结果表明:

(1)将8个气象因子组合作为 SFNN 模型、高斯

回归模型和岭回归模型的输入,进行虾塘水温预测。结果是 SFNN 模型预测效果最好, $RMSE$ 、 MSE 和 MAE 误差最小,分别为1.326 1、1.758 6和1.028 8;决定系数(R^2)最大为0.708 4。

(2)不同气象单因子对虾塘水温预测结果表明:气压对虾塘水温的预测效果较好,预测误差 $RMSE$ 和 MSE 最小,分别为1.868 6和3.491 8。降水量和 10 min 风速分别作为气象单因子预测虾塘水温的效果并不理想,但是在选出的5个气象因子组合中与其他气象因子相互作用后,表现出了良好的预测效果。

(3)将5个气象因子组合(最低气温、最高气温、气压、降水量、10 min 风速)输入 SFNN 模型进行虾塘水温预测,比8个气象因子组合的预测效果要好,预测误差 $RMSE$ 、 MSE 和 MAE 最小分别为1.121 1、1.256 9和0.893 8,决定系数(R^2)最大为0.791 6。因此,在南美白对虾养殖中,基于邻域粗糙集选择的5个气象因子组合,利用 SFNN 模型进行虾塘水温预测的技术具有很好的应用前景。

参考文献:

- [1] 陈美佳,孔杰,谭建,等. 凡纳滨对虾育种基础群体收获体重的未知亲本组效应分析[J]. 中国水产科学, 2021, 28(7): 863-870.
- [2] 张志东,陈爱华,吴杨平,等. 不同饲料投喂模式对脊尾白虾生长、消化酶、体成分及养殖水环境的影响[J]. 中国水产科学, 2020, 27(9): 1075-1084.
- [3] 农成万,张幸,李斌喜,等. 钦州对虾养殖灾害性天气分析和

- 气象服务要领[J]. 气象研究与应用, 2013, 34(1): 60-63.
- [4] 王立超, 申子彬, 周溥佳. 南美白对虾养殖气象指数预报方法[J]. 浙江农业科学, 2017, 58(7): 1254-1256.
- [5] 杨 栋, 姚日升, 金志凤, 等. 不同类型海水养殖水体温度日变化谐波分析[J]. 中国农业气象, 2017, 38(9): 558-566.
- [6] 吴迪生, 张 娟, 刘增宏, 等. 赤道西太平洋暖池表层水温与热带气旋的关系[J]. 热带气象学报, 2010, 26(2): 242-249.
- [7] 叶 舟, 龚梦园. 基于下游生境水温需求的水库分层取水调度研究[J]. 人民长江, 2022, 53(6): 105-110.
- [8] 刘杏红, 韩云童, 常晓林, 等. 考虑淤沙影响的高拱坝水库水温数值分析[J]. 水利学报, 2010, 31(2): 6-8.
- [9] 邓爱娟, 刘 敏, 刘志雄, 等. 洪湖地区养殖鱼塘春夏季水温变化及预报研究[J]. 中国农学通报, 2013, 29(29): 61-68.
- [10] 韦耀国, 赵海镜, 杨国华, 等. 神经网络在南水北调中线冬季水温预测中的应用[J]. 水利水电技术(中英文), 2021, 52(S2): 335-338.
- [11] 施 珮, 袁永明, 匡 亮, 等. 基于 EMD-IGA-SELM 的池塘养殖水温预测方法[J]. 农业机械学报, 2018, 49(11): 312-319.
- [12] 陈英义, 程倩倩, 成艳君, 等. 基于 GA-BP 神经网络的池塘养殖水温短期预测系统[J]. 农业机械学报, 2017, 48(8): 172-178.
- [13] 柳海涛, 孙双科, 郑铁刚, 等. 水电站下游鱼类产卵场水温的人工神经网络预报模型[J]. 农业工程学报, 2018, 34(4): 185-191.
- [14] 朱森林, 吴时强. 高斯过程回归模型在河流水温模拟中的应用[J]. 华中科技大学学报(自然科学版), 2018, 46(10): 122-126.
- [15] 程海燕, 脱友才, 许维忠, 等. 气候变暖对稳定分层型供水水库水温影响研究[J]. 人民长江, 2022, 53(10): 22-30.
- [16] 杨开林. 冰期明渠水温模型[J]. 水利学报, 2022, 53(1): 20-30.
- [17] 邹 琳, 刘 伟, 蒋定国. 基于数值方法的大型水库水温结构影响因素分析[J]. 中国农村水利水电, 2022(9): 113-119.
- [18] 杨青青, 曾 月, 邓艳君, 等. 稻虾共作模式虾沟水温预报模型研究[J]. 江苏农业科学, 2021, 49(5): 194-199.
- [19] PAWLAK Z. Vagueness and uncertainty: A rough set perspective[J]. Computational Intelligence, 1995, 11(2): 227-232.
- [20] HU Q H, YU D, LIU J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577-3594.
- [21] SUN B Z, MA W M, XIAO X. Three-way group decision making based on multigranulation fuzzy decision-theoretic rough set over two universes[J]. International Journal of Approximate Reasoning, 2017, 81: 87-102.
- [22] DONG L J, CHEN D G, WANG N L, et al. Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets[J]. Information Sciences, 2020, 532: 61-71.
- [23] ZHANG P F, LI T R, WANG G Q, et al. Multi-source information fusion based on rough set theory: A review[J]. Information Fusion, 2021, 68: 85-117.
- [24] ZHANG X Y, YAO H, LU Z Y, et al. Class-specific information measures and attribute reducts for hierarchy and systematicness[J]. Information Sciences, 2021, 563: 196-225.
- [25] CHEN H M, LI T R, FAN X, et al. Feature selection for imbalanced data based on neighborhood rough sets[J]. Information Sciences, 2019, 483: 1-20.
- [26] DAI J H, WANG W T, XU Q. An uncertainty measure for incomplete decision tables and its applications[J]. IEEE Transactions on Cybernetics, 2012, 43(4): 1277-1289.
- [27] WAN J H, CHEN H M, ZHONG Y, et al. A novel hybrid feature selection method considering feature interaction in neighborhood rough set[J]. Knowledge-Based Systems, 2021, 227: 107167.
- [28] WANG C Z, HUANG Y, SHAO M W, et al. Feature selection based on neighborhood self-information[J]. IEEE Transactions on Cybernetics, 2019, 50(9): 4031-4042.
- [29] SUGANTHAN P N, KATUWAL R. On the origins of randomization-based feedforward neural networks[J]. Applied Soft Computing, 2021, 105: 107239.
- [30] 李亚茹, 张宇来, 王佳晨. 面向超参数估计的贝叶斯优化方法综述[J]. 计算机科学, 2022, 49(S1): 86-92.
- [31] 张慧春, 张 萌, 边黎明, 等. 基于 YOLO v5 的植物叶绿素含量估测与可视化技术[J]. 农业机械学报, 2022, 53(4): 313-321.
- [32] YUAN Z, CHEN H M, ZHANG P F, et al. A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information[J]. IEEE Transactions on Fuzzy Systems, 2021, 39(9): 3395-3409.
- [33] 施 珮, 袁永明, 匡 亮, 等. 基于 GA-SELM 算法的工厂化水产养殖水温预测方法研究[J]. 传感技术学报, 2018, 31(10): 1592-1597, 1612.
- [34] 任 妮, 鲍 彤, 刘 杨, 等. 基于粒子群优化算法和长短时记忆神经网络的蟹塘溶解氧预测[J]. 江苏农业学报, 2021, 37(2): 426-434.
- [35] 李祥铜, 曹 亮, 李湘丽, 等. 基于 WTD-LSTM 的对虾养殖水温组合预测模型[J]. 广东农业科学, 2021, 48(2): 153-160.
- [36] 袁红春, 王 敏, 刘 慧, 等. 基于特征交互与卷积网络的渔场预测模型[J]. 江苏农业学报, 2021, 37(6): 1501-1509.

(责任编辑: 成纾寒)