



DOI:10.12404/j.issn.1671-1815.2404981

引用格式:魏东,马川,马建民. XGBoost-人工神经网络热力站二次回水温度预测[J]. 科学技术与工程, 2025, 25(17): 7226-7237.

Wei Dong, Ma Chuan, Ma Jianmin. XGBoost-artificial neural networks for secondary return water temperature prediction in thermal stations [J]. Science Technology and Engineering, 2025, 25(17): 7226-7237.

自动化技术、计算机技术

XGBoost-人工神经网络热力站二次回水温度预测

魏东^{1,2}, 马川¹, 马建民³

(1. 北京建筑大学电气与信息工程学院, 北京 100044; 2. 建筑大数据智能处理北京市重点实验室, 北京 100044;
3. 北京起重运输机械设计研究院有限公司, 北京 100010)

摘要 针对热力站二次回水温度预测模型特征多计算量大、模型准确性难以提升的问题,提出一种极限梯度提升-人工神经网络(xtreme gradient boosting-artificial neural network, XGBoost-ANN)二次回水温度预测模型,模型由特征筛选层和预测层组成。特征筛选层利用 XGBoost 算法计算原始数据特征的重要性分数,确定影响二次回水温度的主要特征,从而降低模型复杂度,并提高计算效率;采用贝叶斯正则化算法训练三层前馈 ANN 作为二次回水温度预测层,并通过灰狼优化(grey wolf optimizer, GWO)算法对 ANN 模型的初始权值和阈值进行优化,用灰狼的位置向量表示 ANN 模型的权值和阈值,引入适应度函数来评估每组权值和阈值的性能,帮助模型在训练初期避免陷入局部最优,以提升模型的性能与泛化能力。实验结果表明,所构建的 XGBoost-GWO-ANN 二次回水温度预测模型,相比特征筛选前的模型,均方根误差(root mean squared error, RMSE)性能提升 26.8%, R^2 提升 11.3%, 模型推理时间降低了 46.1%;使用 GWO 算法对 ANN 初始权值和阈值进行寻优,相比于未经优化的 ANN 模型, RMSE 性能提升 20.0%, R^2 提升 3.4%, 预测模型的精度以及泛化能力得到有效提升。

关键词 集中供暖; 热力站系统; 神经网络; XGBoost; 二次回水温度预测

中图分类号 TP183; 文献标志码 A

XGBoost-Artificial Neural Networks for Secondary Return Water Temperature Prediction in Thermal Stations

WEI Dong^{1,2}, MA Chuan¹, MA Jian-min³

(1. School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100010, China;
2. Key Laboratory of Intelligent Processing for Building Big Data, Beijing 100044, China;
3. Beijing Materials Handling Research Institute Co., Ltd., Beijing 100010, China)

[Abstract] To address the challenges of high-dimensional features, large computational demand, and difficulty in improving the accuracy of secondary return water temperature prediction models for heat stations, a secondary return water temperature prediction model based on the xtreme gradient boosting-artificial neural network (XGBoost-ANN) was proposed. The feature screening layer uses XGBoost algorithm to calculate the importance scores of the original data features and determine the main features that affect the secondary backwater temperature, thus reducing the complexity of the model and improving the computational efficiency. Three layers of feedforward ANN were trained by Bayesian regularization algorithm as the secondary backwater temperature prediction layer, and the initial weights and thresholds of the ANN model were optimized by grey wolf optimizer (GWO) algorithm. The weights and thresholds of the ANN model were represented by grey wolf position vector. The fitness function was introduced to evaluate the performance of each set of weights and thresholds to help the model avoid falling into local optimality at the initial stage of training, so as to improve the performance and generalization ability of the model. Experimental results demonstrate that the constructed XGBoost-GWO-ANN secondary return water temperature prediction model achieved significant improvements. Compared to the model before feature filtering, the root mean squared error (RMSE) is reduced by 26.8%, the R^2 is increased by 11.3%, and the model inference time is shortened by 46.1%. Furthermore, the optimization of the initial ANN weights and thresholds using the GWO algorithm improve the RMSE by 20.0% and the R^2 by 3.4% compared to the unoptimized ANN model. These results indicate that the accuracy and generalization abil-

收稿日期: 2024-07-03 修订日期: 2025-03-11

基金项目: 国家自然科学基金面上项目(62371032);北京市自然科学基金面上项目(4232021);住房和城乡建设部科学技术项目(研究开发项目)(2019-K-149);北京建筑大学高级主讲教师培育计划(GJZJ20220803)

第一作者: 魏东(1968—),女,汉族,重庆人,博士,教授。研究方向:预测控制、神经网络优化计算及其在建筑节能领域应用。E-mail: weidong@bucea.edu.cn。

投稿网址: www.stae.com.cn

ity of the proposed prediction model are effectively enhanced.

[**Keywords**] centralized heating; heat station system; neural network; XGBoost; secondary return water temperature prediction

在集中供暖系统中,一次网是指系统中的热源端,其通过二次网向用户输送热能。二次网与一次网在热力站实现换热,热力站后的管道系统,将热能进一步分配给各个用户或建筑物。热力站是承担连接、转换、调节和监控作用的系统,是保障供热系统正常运行和用户获得舒适供暖的关键设备之一^[1]。热力站的二次回水温度是指二次网中的热水在供热后返回热力站的温度,能够表征供热质量,且热力站温度控制策略需要根据二次回水温度变化进行调整。因此,建立热力站二次回水温度预测模型,对确保供热系统的稳定运行至关重要。

然而,集中供暖系统由于分布广、管道长,具有时变性和时滞性,且含有换热器、电磁阀、循环水泵等非线性设备,机理建模困难^[2],因此目前二次回水温度预测多采用数据驱动方式实现,此时模型的训练数据需要通过现场的大量传感设备获取。然而现场采集到的数据规模庞大且维度复杂。大规模高维数据不仅会增加模型的计算成本,还容易导致过拟合,进而可能降低模型的泛化能力和准确性。通过特征筛选技术可以解决这一问题。特征筛选技术是在高维度、已量化的特征向量中选择对指定任务更有效的特征组合^[3]。有效的特征筛选可以改善模型性能,同时降低模型复杂度和过拟合风险。戴娜^[4]利用主成分分析法对热力站二次网供水温度的特征进行筛选,采用长短期记忆(long short-term memory, LSTM)神经网络算法建立二次供水温度预测模型。刘新宇等^[5]利用主成分分析法进行空调系统的特征筛选,采用神经网络(artificial neural network, ANN)模型建立商场冷负荷预测模型。主成分分析法基于线性变换,因此假设数据是线性可分的。如果数据的内在结构是非线性的,主成分分析难以捕捉到数据的真实特征,导致降维效果不佳^[6]。为此,多位学者对不依赖于线性关系的特征筛选方法进行了研究。决策树模型基于特征的阈值判断,而非依赖于输入特征和输出变量之间的线性关系。目前常见的决策树模型有随机森林算法与极端梯度提升(extreme gradient boosting, XGBoost)算法。吴贤国等^[7]利用随机森林算法对特征进行选择,然后建立回归模型,对混凝土抗压强度进行预测,实验表明相比于单一算法,经随机森林特征筛选后的模型预测性能有所提升。然而,随机森林中单棵决策树对数据变化较为敏感,采用带有噪声或者异常值的数据构建模型,

容易出现过拟合。Chen 等^[8]提出 XGBoost 算法,对随机森林算法进行了改进,该方法在模型训练过程中检测验证集性能,当性能不再提升时停止训练,从而避免模型出现过拟合。魏东等^[9]提出了一种基于极限梯度提升(XGBoost)和神经网络的建筑负荷特征筛选及预测方法,旨在解决建筑负荷预测模型中工作量大和泛化能力提升困难的问题。因此,现采用 XGBoost 算法对热力站二次回水温度预测模型的输入特征进行筛选。

在完成特征筛选之后,可基于所筛选的特征集构建预测模型。目前常用的预测方法有线性回归、时间序列分析和 ANN 等^[10-12]。热力站系统具有较强惯性、非线性以及时变性,而线性回归方法假设特征之间是线性相关的,时间序列分析方法通常假设数据具有一定的平稳性和线性特性,因此这两种方法难以捕捉特征之间的复杂非线性关系。相比于传统线性模型,ANN 具有多层非线性处理单元,其能够更有效地捕获和表达数据中的高度非线性特征。董晨等^[13]通过相关性分析构建预测特征集,并利用 ANN 模型实现对二次回水温度的预测。刘军杰等^[14]根据热力站实测数据和气象数据,通过相关性分析生成预测函数,建立了 ANN 预测模型,对二次回水温度进行预测。在模型训练过程中,如果 ANN 网络结构过于复杂,不利于模型后期的收敛,从而使模型偏离全局最优点,降低模型的泛化能力。通过在贝叶斯框架下引入正则化参数的方法可以使模型自动平衡网络的复杂度。魏东等^[15]使用贝叶斯正则化算法对空调负荷预测模型进行训练,提升了模型的泛化能力。因此现采用贝叶斯正则化算法训练 ANN 模型。

然而,前馈 ANN 对初始权值与阈值等参数敏感,初始参数如果设置过大可能会导致梯度爆炸,而设置过小则可能导致梯度消失。合理的初始参数选择可以帮助模型更快地找到全局最优解或者更接近最优解的位置,避免梯度消失或爆炸等问题,并且可以防止网络在训练初期过于复杂化,帮助模型逐渐学习数据的规律,从而提升模型的泛化能力。王晓辉等^[16]采用粒子群优化算法对 ANN 模型的初始权值和阈值进行寻优,解决了网络的过拟合问题。余峰等^[17]采用 ANN 模型对混凝土剂投加量进行预测,并结合遗传算法对模型初始权值、阈值进行优化。然而粒子群优化算法中,由于粒子的位置和速度更新往往受到局部最优解的影响,有可能

失去探索新区域的能力,特别是对于高维复杂问题,随着维度的增加,搜索空间的复杂性以及局部最优解的数量也会随之增加,粒子可能很快收敛到局部最优解;遗传算法依赖于选择、交叉和变异操作,虽然这些操作能够有效地探索解空间,但它们的随机性使得算法的收敛速度较慢。灰狼优化(grey wolf optimizer, GWO)算法^[18]模仿了灰狼在自然界中的捕猎行为,包括围捕、追踪、攻击猎物等过程,用于解决复杂的优化问题。它通过模拟灰狼群体狩猎的行为对参数进行寻优,引入 α 、 β 、 δ 三种不同等级的灰狼来领导群体,并且其他灰狼跟随这三种领导灰狼进行位置更新。相对于粒子群优化算法以及遗传优化算法,GWO算法具有能够自适应调整的收敛因子以及信息反馈机制,并且参数量较少,易于调节,能够提高模型的收敛速度和全局搜索能力。为此现采用GWO算法对ANN预测模型的初始参数进行寻优。

综上,现针对住宅社区集中供暖系统节能控制需要,提出一种热力站二次回水温度预测方法,首先利用XGBoost对原始数据集进行特征筛选,然后采用三层前馈ANN构建二次回水温度预测模型,使用贝叶斯正则化算法训练该模型,最后利用GWO算法对ANN模型的初始权值阈值进行优化。针对住宅社区集中供暖系统二次回水温度预测的实验结果,验证所提方法的有效性。

1 系统描述

以北京市一个住宅社区集中供暖系统为研究对象。该住宅社区面积约10.4万m²,由20栋高层住宅楼组成,房屋面积在80~120m²,共有1600余户。该住宅社区拥有独立的热力站系统,其结构如图1所示。

在热力站系统中,一次供水通过换热器向二次供水提供热能,之后经过一次回水管道返回热源。二

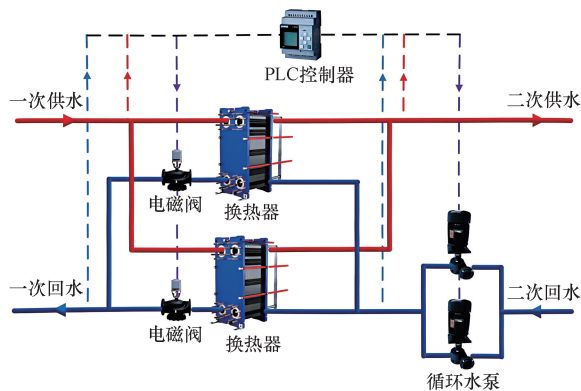


图1 热力站系统

Fig. 1 Thermal power station system

次供水输送到用户端的供暖设备,为用户提供热能,在循环水泵的作用下二次回水从用户端返回热源进行加热循环。通过可编程逻辑控制器(programmable logic controller, PLC)与电磁阀来控制一次供水流量,使二次供水温度达到设定值,以满足用户的供暖需求。

通过预测二次回水温度,系统可以实时了解用户的实际热量需求,当预测到二次回水温度上升,意味着用户热量需求减少,可动态调节一次供水温度和供水流量,以达到按需供热的目的,并实现更精细的调节和控制,减少能源浪费。利用热力站内的水温传感器、电磁式流量传感器来获取二次回水温度与二次循环水流量,并通过压力、湿度、风速传感器来获取室外气象数据,用于建立二次回水温度预测数据集。

2 模型设计

2.1 模型架构

二次回水温度预测模型架构如图2所示,由特征筛选层与模型预测层两部分组成。预测模型构建主要包括以下4个步骤。

步骤1 通过热力站现场的传感器采集二次回水温度预测所需的原始特征数据。

步骤2 利用XGBoost算法对原始特征数据集进行建模,得到每个特征的重要性权重,根据排序筛选出对二次回水温度有显著影响的特征。

步骤3 对筛选后的数据进行预处理操作,对数据中的异常值进行判别,对受噪声干扰的数据进行滤波处理,最后将数据进行归一化。

步骤4 构建XGBoost-GWO-ANN二次回水温度预测模型,使用筛选后的特征集作为模型训练数据,在训练过程中,使用贝叶斯正则化算法训练ANN

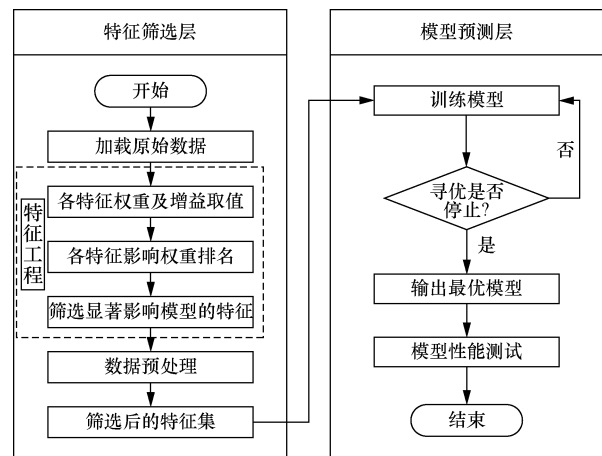


图2 热力站二次回水温度预测模型架构图

Fig. 2 Architecture of secondary return water temperature prediction model for thermal station

模型,同时采用GWO算法对ANN的初始权值和阈值寻找最优解。当训练集和测试集的性能指标达到要求之后,对模型的输出结果进行反归一化处理的操作,从而得出热力站系统二次回水温度预测结果。

2.2 特征筛选层

采用XGBoost算法筛选对二次回水温度具有显著影响的特征集。XGBoost的核心思想是通过构建多个决策树,统计在决策树中每个特征被利用的次數来确定特征的重要性分值。

XGBoost模型公式可表示为

$$\hat{y}_i = \sum_{r=1}^n f_r(x_i), f_r \in G \quad (1)$$

式(1)中: n 为模型中的决策树总数; \hat{y}_i 为第*i*个输入数据的预测值; f_r 为第*r*棵决策树的模型; x_i 为第*i*个输入数据; G 为综合所有决策树的模型。

采用叠加式的训练方式最小化目标函数^[19]。该训练方式在每次迭代过程中添加一棵决策树来拟合之前模型的误差。迭代过程可表示为

$$\begin{cases} \hat{y}_i^{(0)} = 0 \\ \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(r)} = \hat{y}_i^{(r-1)} + f_r(x_i) \end{cases} \quad (2)$$

模型的第*k*次目标函数计算过程为

$$X_{\text{obj}} = \sum_{i=1}^n l(y, \hat{y}) + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T \theta_j^2 \quad (4)$$

式中: X_{obj} 为目标函数; $\sum_{i=1}^n l(y, \hat{y})$ 用来度量预测分数和真实分数之间的偏差; $\sum_{k=1}^K \Omega(f_k)$ 为正则化项; T 为叶子节点的个数; θ_j 为第*j*个叶子节点的分数; γ 为控制树的复杂度的参数; λ 为限制叶子节点的分数的参数。

XGBoost对于每一轮迭代,对目标函数进行更新,如式(5)所示,并在 $f_r = 0$ 处,使用二阶泰勒展开式对损失函数进行近似,从而寻求目标函数最小化的解 f_r ,因此,其近似的目标函数为

$$\tau^{(r)} = \sum_{i=1}^n l[y_i, \hat{y}_i^{(r-1)} + f_r(x_i) + \Omega(f_r)] \quad (5)$$

$$\tau^{(r)} \approx \sum_{i=1}^n l[y_i, \hat{y}_i^{(r-1)} + f_r(x_i) + \frac{1}{2} h_i f_r^2(x_i)] + \Omega(f_r) \quad (6)$$

在近似的目标函数求解的过程中,将所有样本的损失函数近似表达式和正则化项相加,得到新的目标函数,其具体计算过程为

$$\begin{aligned} X_{\text{obj}} &\approx \sum_{i=1}^n \left[g_i f_r(x_i) + \frac{1}{2} h_i f_r^2(x_i) \right] + \Omega(f_r) \\ &= \sum_{i=1}^n \left[g_i w_q(x_i) + \frac{1}{2} h_i w_q^2(x_i) \right] + \Omega(f_r) + \\ &\quad \lambda T + \frac{1}{2} \lambda \sum_{j=1}^T \theta_j^2 \\ &= \sum_{j=1}^T \left[\sum_{i \in I_j} (g_i) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \\ &\quad \lambda T \end{aligned} \quad (7)$$

式(7)中: $w_q(x_i)$ 为样本 x_i 对应节点的权重; $g_i = \partial \hat{y}^{r-1} l(y_i, \hat{y}^{r-1})$ 为 $f_r = 0$ 的一阶导数; $h_i = \partial^2 \hat{y}^{r-1} l(y_i, \hat{y}^{r-1})$ 为 $f_r = 0$ 的二阶导数; I_j 为当前节点的实例集。

通过式(7),目标函数最终被简化为一个仅与每个叶子节点的分数相关的一元二次函数形式,该形式的目标函数通过求解每个叶子节点分数的最优值 θ_j^* 来最小化,分别如式(8)、式(9)所示。

$$\theta_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (8)$$

$$X_{\text{obj}} = - \frac{1}{2} \sum_{j=1}^T \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} + \lambda T \quad (9)$$

在XGBoost训练过程中,模型通过计算每个分裂候选点的增益来确定最佳分裂点,通过选择增益最大的分裂点来增强模型的性能。

2.3 模型预测层

选择采用三层前馈ANN来构建热力站二次回水温度模型。首先使用贝叶斯正则化算法训练ANN模型,为进一步提升模型的性能,引入GWO算法,用于优化ANN模型的初始权值和阈值。

2.3.1 模型训练算法

为了优化ANN模型网络结构,避免训练后期出现过拟合现象,导致模型泛化能力下降,引入贝叶斯框架,在模型训练中引入正则化项对ANN模型参数进行显式概率分布假设的分析与推理,以防止模型过拟合,并且通过正则化参数自动调整网络结构复杂度,以提高模型的泛化能力。

首先定义模型损失函数,其中包括数据误差项和正则化项,即

$$E(\mathbf{W}) = E_D(\mathbf{W}) + \varphi E_W(\mathbf{W}) \quad (10)$$

$$E_D(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^N (y_k - \hat{y}_k)^2 \quad (11)$$

$$E_W(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^M w_j^2 \quad (12)$$

式中: \mathbf{W} 为模型的权重矩阵; $E_D(\mathbf{W})$ 为数据误差项;

$E_w(\mathbf{W})$ 为正则化项,用于权重衰减; \hat{y}_k 为模型第 k 个预测值; y_k 为二次回水温度实际值; N 为数据总量; M 为模型权重的数量; w_j 为模型的第 j 个权值; φ 为正则化参数,控制数据误差项与正则化项的平衡。

基于贝叶斯正则化的损失函数需要判断正则化参数是否收敛,若收敛则停止训练;若不收敛则继续更新正则化参数。

正则化参数的更新公式为

$$\begin{cases} \gamma = M - 2\alpha\text{Tr}(\mathbf{H}^{-1}) \\ \varphi = \frac{\gamma}{2E_w(\mathbf{W})} \\ \mathbf{H} = \mathbf{J}^T\mathbf{J} + \varphi\mathbf{I} \end{cases} \quad (13)$$

式(13)中: γ 为有效参数数量,用于衡量模型的复杂度; α 为一个正则化系数,用于控制模型的复杂度; Tr 为求矩阵的迹; \mathbf{H} 为 Hessian 矩阵; \mathbf{J} 为雅各比矩阵; \mathbf{I} 为单位矩阵。

通过不断迭代上述过程,贝叶斯正则化算法可以动态调整正则化参数,使模型在训练过程中自动平衡数据误差与模型复杂度,从而提高模型的泛化能力与鲁棒性。

2.3.2 基于GWO算法的模型初始权(阈)值优化

为解决 ANN 模型初始权值阈值的随机设定易导致预测模型陷入局部最优解从而导致模型性能下降的问题,提出 GWO-ANN 模型,将 GWO 算法中灰狼的位置信息代表 ANN 模型的初始权值阈值,并对该初始参数进行优化。

1) GWO 算法原理

GWO 算法是一种受灰狼群体行为启发而来的优化算法,该算法模拟了灰狼群体中的领导者与追随者之间的协作行为,通过模拟狼群的社会结构和捕猎策略来解决优化问题。灰狼群体内部具有明确的等级结构,分为 4 个层级, α (领导层)、 β (辅助领导)、 δ (底层领导层)、 ω (底层成员),如图 3 所示。

GWO 算法模拟了狼群围捕、追踪直至攻击猎物的一系列协作行为,旨在求解各类复杂的优化问题。为描述灰狼个体如何根据猎物的位置信息调整自己的位置,以逐渐缩小包围圈并最终捕获猎物,将狼群包围猎物的行为抽象为数学模型,如式(14)~式(17)所示。

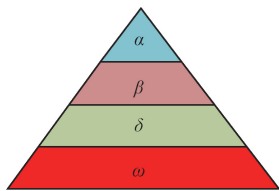


图3 灰狼的等级

Fig.3 Hierarchy of grey wolf

$$\mathbf{A} = 2a\mathbf{v}_1 - \mathbf{a} \quad (14)$$

$$\mathbf{C} = 2\mathbf{v}_2 \quad (15)$$

$$\mathbf{D} = |\mathbf{C}\mathbf{X}_p(t) - \mathbf{X}(t)| \quad (16)$$

$$\mathbf{X}(t+1) = \mathbf{X}_p(t) - \mathbf{A}\mathbf{D} \quad (17)$$

式中: t 为迭代次数; $\mathbf{X}_p(t)$ 为猎物的位置矢量; $\mathbf{X}(t)$ 和 $\mathbf{X}(t+1)$ 均为灰狼不同迭代轮次的位置矢量,这些位置矢量的更新是基于猎物与灰狼之间的相对位置关系,以及灰狼群体内部的协作机制。在更新过程中,引入 3 个随机参数分别是收敛因子 a 、随机向量 \mathbf{v}_1 和 \mathbf{v}_2 ,并在区间 $[0,1]$ 内随机取值。其中,收敛因子 a 更新方式如式(18)所示。

$$a = 2 - l \frac{2}{\text{Max_iteration}} \quad (18)$$

式(18)中: l 为当前迭代轮次;Max_iteration 为更新过程中最大的迭代次数。

在 GWO 算法中,每当灰狼的位置根据先定的搜索策略更新后,会计算更新后每只灰狼的适应度值。适应度值反映了灰狼个体在搜索空间中的优劣程度。GWO 算法会参考这些适应度值,选取适应度值最小的 3 个解 $X_\alpha(t)$ (最佳解)、 $X_\beta(t)$ (次优解)、 $X_\delta(t)$ (第三优解),并将它们作为当前灰狼群体的领导位置。在后续的迭代中,GWO 算法能够逐步引导灰狼群体向搜索空间中的更优区域移动,从而找到问题的近似最优解或最优解,其过程如式(19)~式(22)所示。

$$\mathbf{X}_1 = \mathbf{X}_\alpha(t) - \mathbf{A}_1\mathbf{D}_\alpha \quad (19)$$

$$\mathbf{X}_2 = \mathbf{X}_\beta(t) - \mathbf{A}_2\mathbf{D}_\beta \quad (20)$$

$$\mathbf{X}_3 = \mathbf{X}_\delta(t) - \mathbf{A}_3\mathbf{D}_\delta \quad (21)$$

$$\mathbf{X}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3} \quad (22)$$

式中: \mathbf{D}_α 、 \mathbf{D}_β 、 \mathbf{D}_δ 分别为 α 狼、 β 狼和 δ 狼与其他个体间的距离; \mathbf{A}_1 、 \mathbf{A}_2 、 \mathbf{A}_3 为区间 $[-a, a]$ 中的随机向量,模拟灰狼逼近猎物,当 $|\mathbf{A}| > 1$ 时,灰狼会远离当前猎物,体现出算法在该阶段的全局搜索能力,从而增强解空间的探索效果,当 $|\mathbf{A}| < 1$ 时,灰狼会靠近猎物进行精确追捕,此时局部搜索能力表现良好; \mathbf{X}_1 、 \mathbf{X}_2 、 \mathbf{X}_3 分别为受 α 狼、 β 狼和 δ 狼影响; ω 为狼调整后的位置。在这一过程中,灰狼的行为模拟了其在自然界中的动态捕猎机制,逐步接近并包围猎物。该设计使得算法在早期阶段具有较强的全局探索能力,能够使模型避免陷入局部最优,而在后期阶段则逐步过渡到局部搜索,以提升解的精度。灰狼捕猎过程的位置更新示意图如图 4、图 5 所示,可以直观地展示灰狼根据算法的设计逐步接近并包围猎物,以及在不同阶段平衡全局搜索和局部搜索的能力。

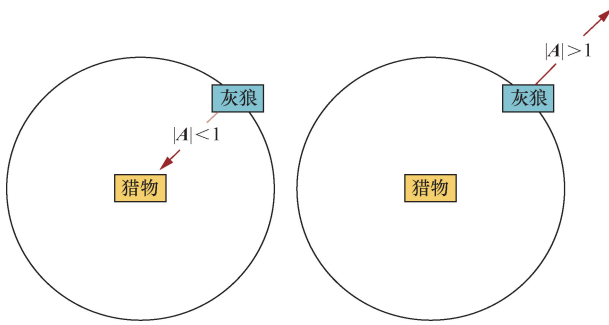
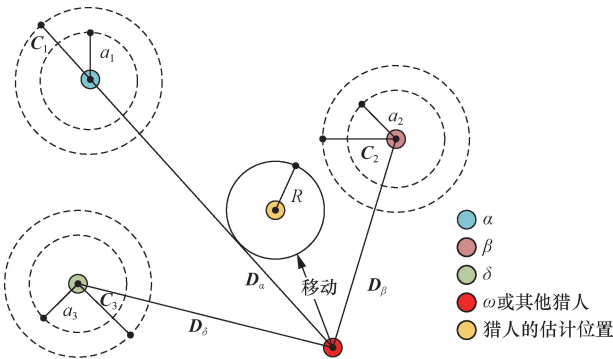


图4 攻击猎物与寻找猎物

Fig. 4 Attacking prey versus searching for prey



a_1, a_2, a_3 为狼群的位置; C_1, C_2, C_3 为随机协同向量, 模拟狼群在搜索空间中的随机移动行为; R 为猎人的估计位置到当前狼位置的半径距离

图5 灰狼个体的位置更新图

Fig. 5 Location update plot of the individual gray wolf

2) 初始权(阈)值优化

初始权(阈)值优化算法中灰狼种群规模 N 的维度被划分为 4 个主要部分。

(1) 输入层到中间层的权值矩阵 W_1 , 其维度取决于输入层节点数和中间层节点数。

(2) 输入层到中间层的阈值矩阵 B_1 , 其维度与中间层节点数相同, 代表了中间层每个神经元的阈值。

(3) 中间层到输出层的权值矩阵 W_2 , 其维度取决于中间层节点数和输出层节点数。

(4) 中间层到输出层的阈值矩阵 B_2 , 其维度与输出层节点数相同, 代表了输出层每个神经元的阈值。

GWO 算法中适应度函数用来评估每个解的质量, 并指导算法向优质解的方向搜索, 其构建表达式为

$$F = \frac{1}{M} \sum_{s=1}^M \sqrt{\sum_{k=1}^b (\gamma_k^s - o_k^s)^2} \quad (23)$$

式(23)中: M 为训练样本数; b 为输出层节点数; γ_k^s 为第 k 个输出节点在样本 s 作用下的输出; o_k^s 为样本 s 的第 k 个输出层节点的期望输出。

基于 GWO 算法的热力站系统二次回水温度预

测模型步骤如下, 其流程图如图 6 所示。

步骤 1 对输入数据进行归一化处理, 加快算法收敛速度。

步骤 2 在配置 ANN 模型时, 设置模型的最大迭代次数为 Max_epochs、学习率为 η 、目标误差为 Target_error。

步骤 3 随机初始化 ANN 模型的所有参数, 包括权值和阈值, 其初始范围设定为 $[-1, 1]$ 。然后, 将初始化后的参数代入 ANN 模型进行初步训练。随后, 根据预定的适应度函数, 计算每只灰狼的适应度值, 并按照这些值进行排序, 选择适应度最小的 3 只灰狼, 分别记为 α 、 β 、 δ 。

步骤 4 根据 GWO 算法的更新机制, 更新每只灰狼的位置, 并利用新的位置参数重新构建 ANN 模型并再次进行训练, 最后计算每只灰狼的适应度值, 并选出适应度最佳的灰狼, 更新其位置信息。

步骤 5 检查 GWO 算法是否已经达到设定的最大迭代次数 Max_iteration。若未达到, 则返回重复步骤 3, 继续迭代优化过程; 若已达到, 则记录当前最优灰狼个体 α 对应的 ANN 模型初始权值和阈值。

步骤 6 利用 GWO 算法优化得到的权值和阈值, 使用贝叶斯正则化算法对 ANN 模型进行下一步训练, 在模型训练不断迭代的过程中判断正则化参数 λ 是否收敛, 若不收敛, 继续更新 λ ; 若收敛, 则停止训练, 输出最优模型。

步骤 7 最后, 将测试数据输入经过训练的 ANN 模型中, 模型将根据输入数据输出热力站系统二次回水温度的预测结果。

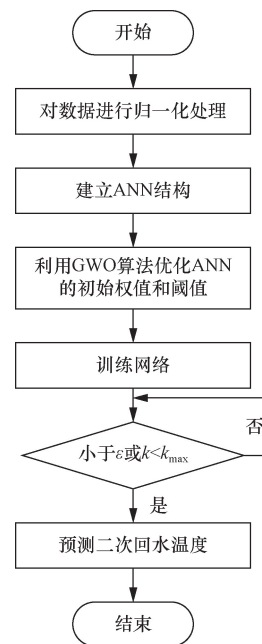


图6 GWO-ANN 模型计算流程

Fig. 6 The GWO-ANN model calculation flow

3 实验研究

3.1 实验数据集

采集了研究对象 2021 年 11 月 15 日—2022 年 3 月 15 日供暖季的运行数据和室外气象参数,所采集的特征变量共 8 个,如表 1 所示。鉴于集中供暖系统固有的惯性和滞后性,及其系统参数变化相对缓慢的特性,采取每小时一次的采样频率来确保数据的准确性和代表性,部分收集到的数据信息已整理,如表 2 所示。

表 1 采集到的特征变量
Table 1 Collected feature variables

序号	名称
1	二次供水温度设定值
2	二次供水温度
3	二次回水温度
4	二次循环水流量
5	室外温度
6	相对湿度
7	瞬时风速
8	地面气压

3.2 数据处理

由于热力站系统中的数据采集设备容易受到电磁辐射、环境噪声等干扰,运行数据中存在异常数据,为此引入了三西格玛(3-Sigma)异常值判别方法以识别和修正这些异常情况,如式(24)所示。

$$(\mu - 3\sigma, \mu + 3\sigma) \quad (24)$$

式(24)中: μ 为特征序列均值; σ 为标准差,反映序列的离散程度。对于不符合 3-Sigma 准则的特征异常值,采用相邻正常值的平均值来进行替代。对于因噪声干扰和传感器异常导致的部分数据缺失的现象,使用均值填充法作为处理手段,通过计算特征序列的均值,并将其作为缺失值的填充替代,从而保证了数据的完整性和一致性。如图 7 所示,本文方法在图中 4 处二次供水温度缺失部分进行了有效的填充。

同时采用滑动平均滤波方法对采集数据进行处理,滑动平均滤波方法的计算公式为

$$y[i] = \frac{1}{N} \sum_{j=0}^{N-1} y[i+j] \quad (25)$$

式(25)中: x 为需要滑动平均滤波处理的变量; y 为滤波后的结果; N 为要平均的数据数量, N 取值范围一般为[3,14],这里取 $N = 4$ 。滤波处理后的效果如图 8 所示,可以看出尖峰和随机噪声得到了有效的抑制。

为提高数据的稳定性以及模型的收敛速度,采用 Mapminmax 函数对数据进行归一化处理^[20],公式为

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min} + y_{\min}} \quad (26)$$

式(26)中: x 为原始输入值; x_{\min} 为输入样本中的最小值; x_{\max} 为输入样本的最大值; y 为归一化处理后的数值; y_{\max} 为归一化操作后的最小值; y_{\min} 为归一化后的最大值。

对特征数据集进行数据预处理后再进行切分,将样本数据按 7:3 的比例划分为训练集与验证集。其数据处理流程图如图 9 所示。

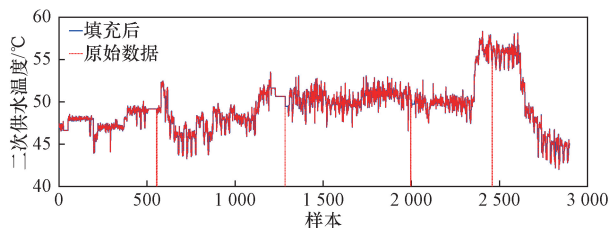


图 7 二次供水温度缺失值填充

Fig. 7 Missing value filling of the secondary water supply temperature

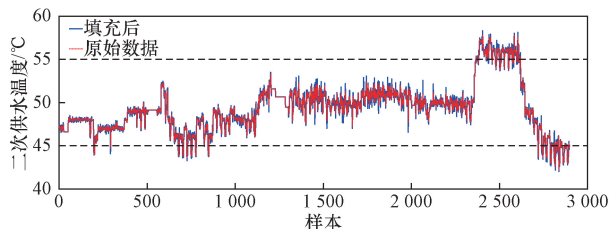


图 8 二次供水温度采样值滤波前后

Fig. 8 The filtering of the temperature sampling value before and after the secondary water supply

表 2 部分数据信息

Table 2 Selected data information

二次供水温度设定值/°C	二次供水温度/°C	二次回水温度/°C	二次循环水流量/(m ³ ·h ⁻¹)	室外温度/°C	瞬时风速/(m·s ⁻¹)	相对湿度/%	地面气压/hPa
52	51.43	38.46	143.75	8.5	3.1	28	1 008.8
52	51.99	39.48	141.37	6.3	0.7	34	1 009.6
52	51.80	39.75	141	3.8	1.6	41	1 010.2
52	52.20	39.78	144.48	1.5	0.9	51	1 010.4
52	51.75	39.73	143.5	0.4	1.4	54	1 011.0

3.3 输入特征筛选

基于热力站系统的历史运行数据和室外气象数据特征筛选模型。通过 XGBoost 算法计算各特征的权重和增益,并对特征数据进行排序。权重和增益数值的高低直接反映了特征变量在预测二次回水温度时的重要性,数值越高,表示该特征对预测结果的影响越大。各特征变量优先级如图 10 所示。根据 XGBoost 算法对原始数据特征得分排序的结果,选取二次循环水流量 [$F_2(k)$]、二次供水温度 [$F_2(k)$]、二次回水温度 [$T_{2,r}(k)$]、室外温度 [$T_{out}(k)$]、瞬时风速 [$WS_{out}(k)$]、相对湿度 [$RH_{out}(k)$]为二次回水温度预测模型的输入变量。

3.4 输入特征筛选

选取均方误差根 (RMSE) 和决定系数 (R^2) 作为二次回水温度预测模型的评价指标。RMSE 和 R^2 的计算公式分别为

$$RMSE = \sqrt{\frac{1}{d} \sum_{i=1}^d (y_i - \hat{y}_i)^2} \quad (27)$$

$$R^2 = 1 - \frac{\sum_{i=1}^d (y_i - \hat{y}_i)^2}{\sum_{i=1}^d (y_i - \bar{y}_i)^2} \quad (28)$$

式中: y_i 为二次回水温度数据的实际值; \hat{y}_i 为二次回水温度数据的预测值; \bar{y}_i 为二次回水温度数据的平均值; d 为二次回水温度数据的数量。

3.5 神经网络模型结构与参数实验

以筛选后的特征为基础构建三层前馈 ANN 模型。ANN 模型包括输入层、隐藏层、输出层,其中隐藏层神经元节点的数量对模型性能有着重要的影响,通过经验公式 [式 (29)] 获取隐藏层节点数^[21] 范围为 [5, 11]。

$$m = \sqrt{bl} \quad (29)$$

式 (29) 中: b 、 m 和 l 分别为 ANN 模型输出层、隐藏层和输入层的节点数。

图 11 展示了 RMSE 和 R^2 随隐藏层节点数变化的趋势。以 RMSE 最低、 R^2 最高为目标,对不同数量的隐藏层节点数目进行实验,具有 9 个节点的隐藏层的 ANN 模型 RMSE 为 0.30, R^2 为 0.88。因此,将隐藏层神经元的数量设置为 9。

在运用 XGBoost 算法进行特征筛选后,预测模型的输入层设计为 6 个神经元,分别对应不同的输入特征。其中前 3 个神经元 (1、2、3) 负责接收热力站的历史运行数据,具体为二次供水温度、二

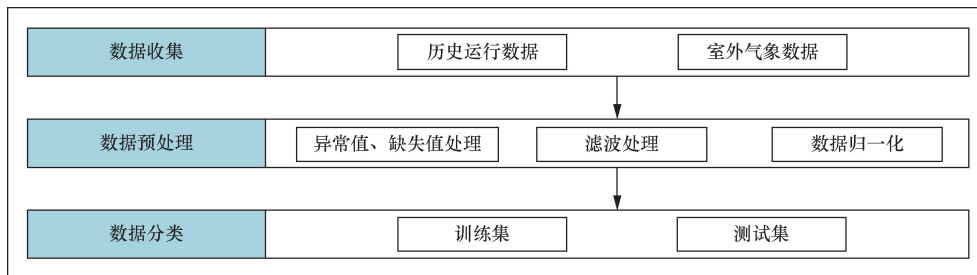


图 9 数据预处理流程图

Fig. 9 Flow chart of data preprocessing

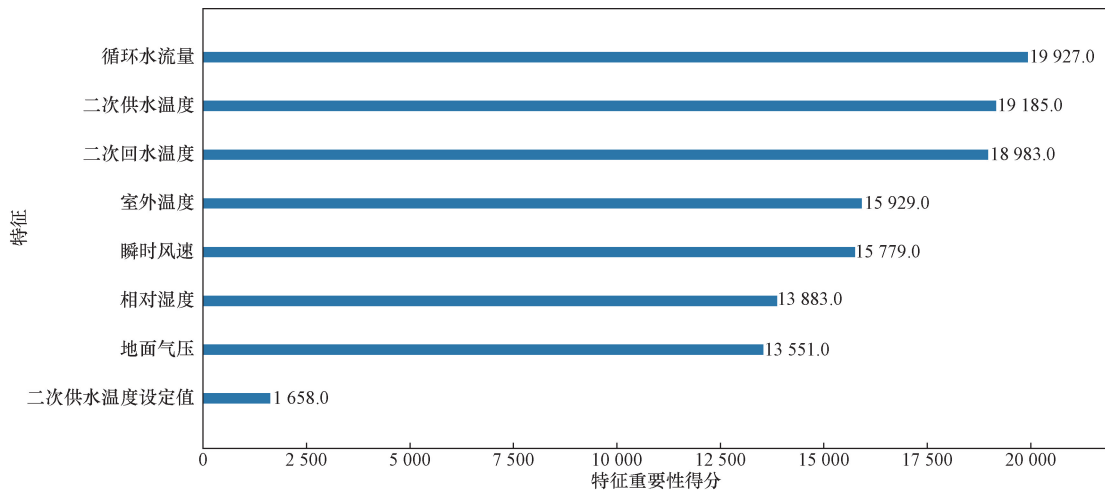


图 10 特征重要性分析结果

Fig. 10 Results of the feature importance analysis

次回水温度和二次循环水流量;后3个神经元(4、5、6)则负责接收室外气象数据,包括室外温度、相对湿度和瞬时风速。为了有效地处理这些输入特征并预测下一时刻的二次回水温度,在模型中设置了包含9个神经元的隐藏层,使其在多元的输入中学习复杂的模式和关联信息。最后,模型的输出层包含1个神经元,用于预测输出下一时刻二次回水温度数据。

建立的二次回水温度预测模型的网络结构可以概括为6-9-1,即输入层有6个神经元,隐藏层有9个神经元,输出层有1个神经元。模型的具体拓扑结构如图12所示,清晰地展示了各层神经元之间的连接关系和数据流向。

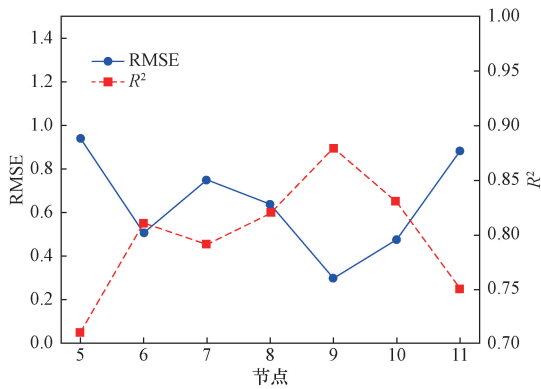


图 11 隐藏层节点数量选择实验结果图
Fig. 11 Experimental results of choosing the number of nodes in the hidden layer

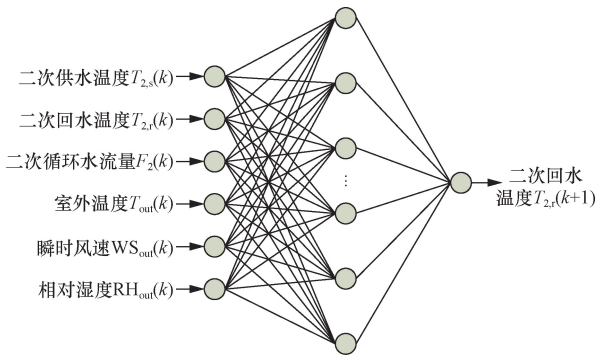


图 12 二次回水温度预测模型拓扑图
Fig. 12 Topological map of the prediction model for quadratic backwater temperature

3.6 特征筛选前后模型预测实验

为了验证特征筛选对模型预测性能的影响,对比了特征筛选前后 ANN 模型的 RMSE 和 R^2 以及模型推理时间。由图 13 可以看出,特征筛选后的 ANN 模型相比未筛选前的模型具有更好的拟合效果。图 14 显示,特征筛选后的 ANN 模型在第 25 次迭代时便趋于收敛,最终 RMSE 为 0.30,而未经过特征筛选的 ANN 模型在第 56 次迭代时才开始收

敛,最终 RMSE 为 0.41。

经过特征筛选后的二次回水温度预测模型,在预测性能方面 R^2 提升了 11.3%,RMSE 降低了 26.8%。在时间性能方面,进行了 5 次推理时间测试,模型推理时间较特征筛选前缩短了 46%。特征筛选前后的模型预测实验结果如表 3 所示。

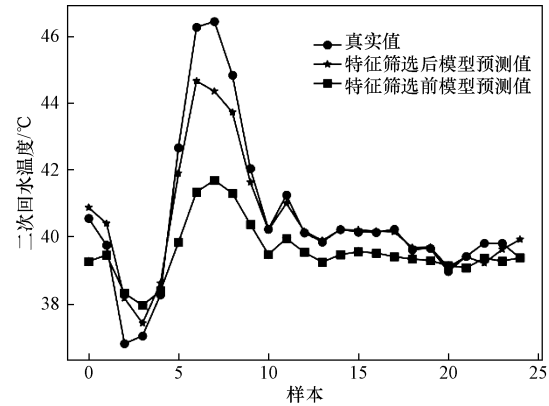


图 13 特征筛选前后 ANN 模型预测结果对比图
Fig. 13 Comparison of ANN model prediction results before and after feature screening

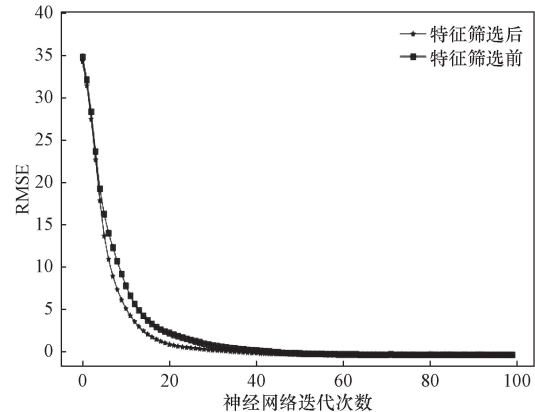


图 14 特征筛选前后 RMSE 下降曲线图
Fig. 14 RMSE decrease before and after feature screening

表 3 特征筛选前后对比

模型	RMSE	R^2	推理时间/s
特征筛选前	0.41	0.79	0.13 ± 0.03
特征筛选后	0.30	0.88	0.07 ± 0.02

3.7 优化模型初始权值(阈值)实验

为验证所提方法的有效性,对于训练数据,分别利用未经过优化的 XGBoost-ANN 模型,以及经过 GWO 算法优化后的 XGBoost-GWO-ANN 模型,在相同条件下进行训练,并在相同测试集上进行测试。XGBoost-ANN 模型的预测结果与相对误差如图 15 和图 16 所示,XGBoost-GWO-ANN 模型的

预测结果与相对误差如图 17 和图 18 所示,两种预测方法预测值对比如图 19 所示,相对误差对比如图 20 所示。由图 19 可知,XGBoost-GWO-ANN 模型在使用XGBoost算法进行特征筛选后,通过GWO算法优化ANN模型的初始权值与阈值,其预测结果和真实值高度吻合。相比之下,尽管XGBoost-ANN模型也使用了XBoost进行特征筛选,但是由于未经过灰狼算法的优化,在测试第500个样本后,预测值与真实值之间出现了较大的偏差,导致模型准确性下降。由图20可知,XGBoost-GWO-ANN模型预测值与真实值的相对误差显著低于XGBoost-ANN模型的误差。

通过多次实验取得实验结果,并对实验结果求平均值,XGBoost-ANN模型和XGBoost-GWO-ANN两种模型的RMSE和 R^2 对比如表4所示。采用灰狼算法优化后的ANN模型相较于未经过优化的模型RMSE降低了20%, R^2 提升了3.4%,这充分证明了采用GWO算法优化后的ANN模型XGBoost-GWO-ANN具有更强的泛化能力和更高的精度。

表4 预测结果的误差对比

Table 4 The error contrast of the prediction results

模型	RMSE	R^2
XGBoost-ANN	0.30	0.88
XGBoost-GWO-ANN	0.24	0.91

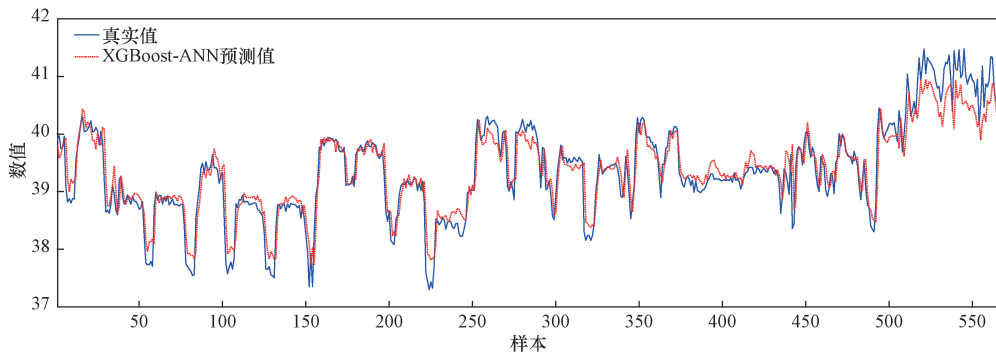


图15 XGBoost-ANN模型测试集的预测表现

Fig. 15 Predictive performance of the XGBoost-ANN model test set

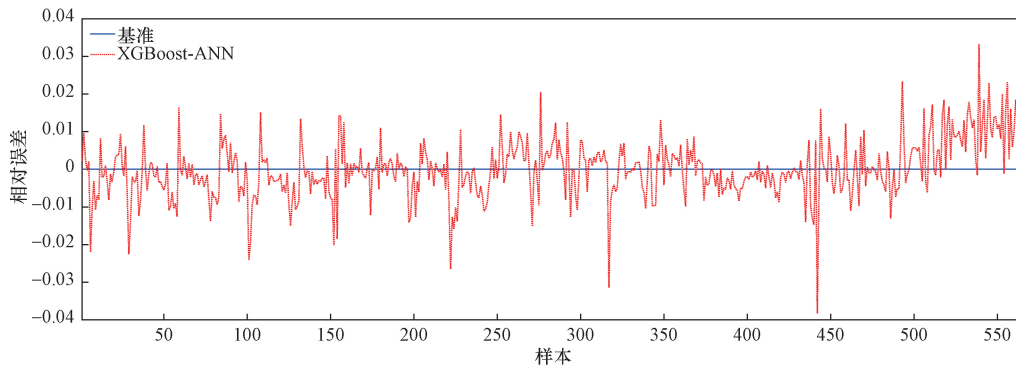


图16 XGBoost-ANN模型测试集的相对误差

Fig. 16 Relative errors in the XGBoost-ANN model test set

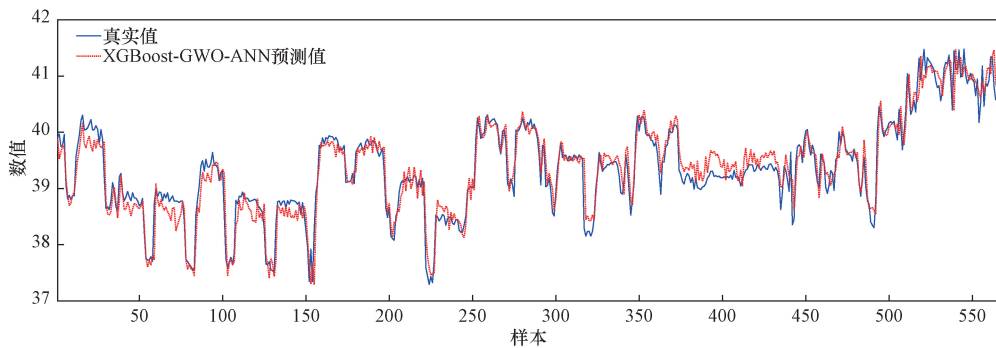


图17 XGBoost-GWO-ANN模型测试集的预测表现

Fig. 17 Predictive performance of the XGBoost-GWO-ANN model test set

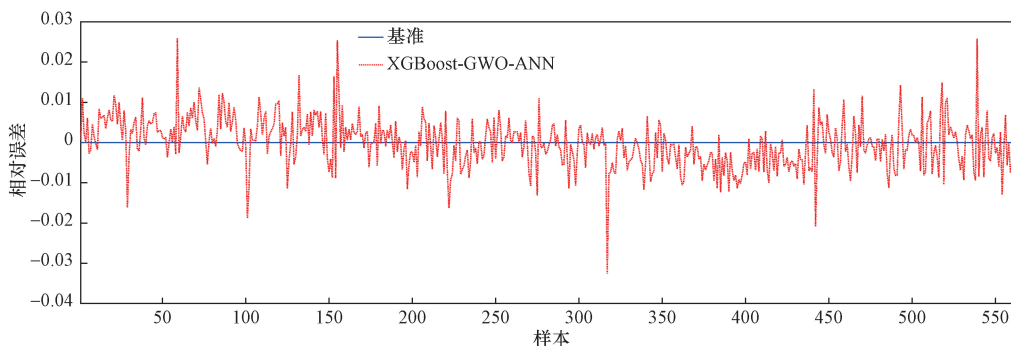


图 18 XGBoost-GWO-ANN 模型测试集相对误差

Fig. 18 The relative error of the XGBoost-GWO-ANN model test set

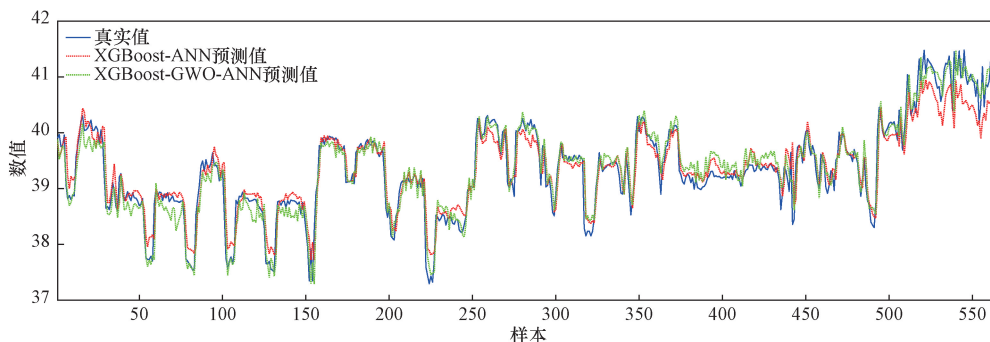


图 19 两种预测模型预测值对比

Fig. 19 Comparison of the predicted values of the two prediction models

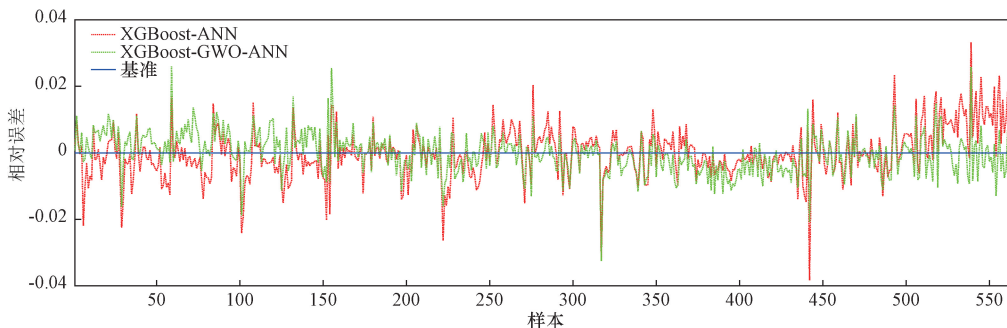


图 20 两种预测模型相对误差对比

Fig. 20 Comparison of the relative error between the two prediction model

4 结论

为解决热力站二次回水温度预测模型特征多、计算量大、模型准确性难以提升的问题,提出了一种 XGBoost-神经网络二次回水温度预测模型。首先,利用 XGBoost 算法进行特征筛选,识别出影响二次回水温度的主要特征,降低模型的复杂度和过拟合风险,从而提升模型预测性能以及计算效率;然后,将筛选后的特征输入到 ANN 模型,为优化网络结构,避免过拟合风险,使用贝叶斯正则化算法对其进行训练,并利用 GWO 算法对 ANN 模型的初始权值和阈值进行优化,通过引入适应度函数来评估每组权值和阈值的效果,并不断判断适应度值以满足收敛条件;最

后确定模型的最佳权值和阈值,避免模型在训练初期陷入局部最优,从而提升模型的性能与泛化能力。利用北京市某住宅社区集中供暖系统传感器采集的数据进行了建模实验,实验结果如下。

(1) 经过特征筛选后的 ANN 预测模型相比未特征筛选前, R^2 提升了 11.3%, RMSE 降低了 26.8%;模型的训练时间减少了 3.4%,模型推理时间缩短了 46%。

(2) 采用 GWO 算法优化后的 ANN 模型相比未经优化的 ANN 模型, RMSE 降低了 20%, R^2 提升了 3.4%。

在接下来的研究中,将所提二次回水温度预测模型用于集中供暖控制系统设计,以降低集中供暖系统的综合能耗。

参 考 文 献

- [1] Yuan J, Zhou Z, Wang C, et al. A novel model for improving the prediction accuracy of the new heating station [J]. *Energy and Buildings*, 2020, 229: 110521.
- [2] 周浩杰. 集中供热系统换热站负荷预测与控制算法研究[D]. 天津: 天津理工大学, 2019.
Zhou Haojie. Research on load forecasting and control algorithm of heat exchange station in central heating system[D]. Tianjin: Tianjin University, 2019.
- [3] 牛勇, 李华鹏, 刘阳惠, 等. 超高维数据特征筛选方法综述[J]. *应用概率统计*, 2021, 37(1): 69-110.
Niu Yong, Li Huapeng, Liu Yanghui, et al. Overview of feature screening methods for ultra-high dimensional data [J]. *Chinese Journal of Applied Probability and Statistics*, 2021, 37(1): 69-110.
- [4] 戴娜. 基于多种机器学习算法的超短期二次网供水温度预测研究[D]. 天津: 天津大学, 2021.
Dai Na. Research on ultra-short-term secondary supply temperature prediction based on multiple machine learning algorithms[D]. Tianjin: Tianjin University, 2021.
- [5] 刘新宇, 赵安军, 程仁印, 等. 基于改进灰色神经网络的商场冷负荷预测模型研究[J]. *工业控制计算机*, 2021, 34(8): 63-66.
Liu Xinyu, Zhao Anjun, Cheng Renyin, et al. Research on cold load forecasting model of shopping malls based on improved grey neural network[J]. *Industrial Control Computer*, 2021, 34(8): 63-66.
- [6] 赵蕾. 主成分分析方法综述[J]. *软件工程*, 2016, 19(6): 1-3.
Zhao Qiang. Overview of principal component analysis methods [J]. *Software Engineering*, 2016, 19(6): 1-3.
- [7] 吴贤国, 刘鹏程, 陈虹宇, 等. 基于随机森林的高性能混凝土抗压强度预测[J]. *混凝土*, 2022(1): 17-20, 24.
Wu Xianguo, Liu Pengcheng, Chen Hongyu, et al. Characteristic screening and prediction of high-performance concrete compressive strength based on random forest method[J]. *Concrete*, 2022(1): 17-20, 24.
- [8] Chen T, Guestrin C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2016: 785-794.
- [9] 魏东, 杨洁婷, 韩少然, 等. 基于 XGBoost-神经网络的建筑负荷预测模型构建[J]. *科学技术与工程*, 2023, 23(29): 12604-12611.
Wei Dong, Yang Jieting, Han Shaoran, et al. Construction of building load prediction model based on XGBoost-neural network algorithm[J]. *Science Technology and Engineering*, 2023, 23(29): 12604-12611.
- [10] 王佳明, 杨海滨, 赵天怡, 等. 基于温度多元线性回归模型的空调制冷站在线预测控制方法研究[J]. *暖通空调*, 2023, 53(2): 140-147.
Wang Jiaming, Yang Haibin, Zhao Tianyi, et al. Online predictive control method of air conditioning chiller plants based on temperature multiple linear regression model[J]. *Heating Ventilating*, 2023, 53(2): 140-147.
- [11] 陈卫雄. 基于时间序列模型的青藏铁路路基变形预测[J]. *科学技术与工程*, 2021, 21(35): 15203-15208.
Chen Weixiong. Prediction of roadbed deformation of Qinghai-Tibet railway based on time series model [J]. *Science Technology and Engineering*, 2021, 21(35): 15203-15208.
- [12] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(6): 1229-1251.
Zhou Feiyan, Jin Linpeng, Dong Jun. Review of convolutional neural network [J]. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251.
- [13] 董晨, 周浩杰, 宫铭举, 等. 基于神经网络的换热站二次回水温度预测方法[J]. *计算机应用与软件*, 2020, 37(6): 83-87, 125.
Dong Chen, Zhou Haojie, Gong Mingju, et al. Prediction method of secondary backwater temperature of heat exchange station based on neural network [J]. *Computer Applications and Software*, 2020, 37(6): 83-87, 125.
- [14] 刘军杰, 王书铭, 姜威. 基于改进 BP 神经网络的供热系统二网回水温度预测方法[J]. *中国新技术新产品*, 2023(17): 73-76.
Liu Junjie, Wang Shuming, Jiang Wei. Prediction method of backwater temperature of heating system based on improved BP neural network [J]. *New Technology & New Products of China*, 2023(17): 73-76.
- [15] 魏东, 焦焕炎, 冯浩东. 基于负荷预测的冷冻站系统非线性预测控制[J]. *控制理论与应用*, 2021, 38(10): 1619-1630.
Wei Dong, Jiao Huanyan, Feng Haodong. Nonlinear predictive control of refrigeration system based on load forecasting [J]. *Control Theory & Applications*, 2021, 38(10): 1619-1630.
- [16] 王晓辉, 李兆巍, 杨亚龙. 考虑学习效率的教室内热环境调控系统研究 [J]. *科学技术与工程*, 2023, 23(36): 15588-15596.
Wang Xiaohui, Li Zhaowei, Yang Yalong. Thermal climate control systems in classrooms considering learning efficiency [J]. *Science Technology and Engineering*, 2023, 23(36): 15588-15596.
- [17] 余峰, 王珂佳, 张文龙, 等. 基于遗传算法优化 BP 神经网络的水生态修复原位控浊混凝投药预测[J]. *环境工程*, 2023, 41(4): 154-163.
Yu Fei, Wang Kejia, Zhang Wenlong, et al. Prediction of coagulant dosage for insitu turbidity control in water ecological restoration basei on BP neural tworkoptmz by gene algorithm [J]. *Environmental Engineering*, 2023, 41(4): 154-163.
- [18] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer [J]. *Advances in Engineering Software*, 2014, 69: 46-61.
- [19] Gómez-Ríos A, Luengo J, Herrera F. A study on the noise label influence in boosting algorithms: AdaBoost, GBM and XGBoost [C]//International Conference on Hybrid Artificial Intelligence Systems. Cham: Springer, 2017: 268-280.
- [20] 陈科, 张保明, 王宇灵, 等. 基于统计分析的数据挖掘中数据处理综述[J]. *测绘科学*, 2008(S1): 76-77, 81.
Chen Ke, Zhang Baoming, Wang Yuling, et al. A survey of data processing in data mining based on statistical analysis [J]. *Science of Surveying and Mapping*, 2008(S1): 76-77, 81.
- [21] 丁思超. 随机配置网络模型参数及其激活函数选择方法研究 [D]. 沈阳: 沈阳工业大学, 2023.
Ding Sichao. Selection of parameters and activation function for stochastic configuration network [D]. Shenyang: Shenyang University of Technology, 2023.