

杨瑜玲, 李柱, 王子瑜, 等. 基于特征优选和极端梯度提升树的水厂供水量预测模型[J]. 净水技术, 2026, 45(3): 139-146, 155.

Yang Y L, Li Z, Wang Z Y, et al. Forecasting model of WTP water supply capacity based on feature optimization and extreme gradient boosting tree [J]. Water Purification Technology, 2026, 45(3): 139-146, 155.

基于特征优选和极端梯度提升树的水厂供水量预测模型

杨瑜玲¹, 李柱^{2,*}, 王子瑜², 杨澜², 宋朝阳¹, 崔露苑²

(1. 上海城投水务<集团>有限公司, 上海 200082; 2. 上海城投水务<集团>有限公司制水分公司, 上海 200086)

摘要 【目的】为支撑智慧水务发展, 满足水厂精细化调度需求, 提升供水效率与资源利用率, 需构建时供水量预测模型以优化生产调度。【方法】以华东地区 A 水厂为研究对象, 本文综合时间特征、气象条件、设备检修工况及水力参数构建含 58 个初始特征的数据集, 采用递归特征消除 (RFE) 算法筛选出 16 个关键特征, 同时引入贝叶斯优化 (BO) 对极端梯度提升树 (XGBoost) 模型的超参数进行寻优, 建立时供水量预测模型。【结果】BO 后模型平均绝对百分比误差 (MAPE) 为 4.58%, 平均绝对误差 (MAE) 为 676.02 m³/h。使用 2024 年 4 月实测值评估, 若按模型预测值进行进水量调整, 99.66% 时间段内 1 h 水库水位波动可控制在 ±0.3 m, 无人工干预时日水库液位最大偏差仅为 0.77 m。2024 年 8 月该模型与 A 水厂智能水量调度系统融合并实现在线运行, 通过提供未来 1~4 h 对外供水量趋势预测, 优化调度方案, 使泵启停次数减少 47%。【结论】此模型兼具高精度与实用价值, 可避免设备损耗与能源浪费, 为水厂优化调度提供技术支撑。

关键词 极端梯度提升树 (XGBoost) 递归特征消除 (RFE) 贝叶斯优化 (BO) 特征优选 供水量预测 智慧水务

中图分类号: TU991 **文献标志码:** A **文章编号:** 1009-0177(2026)03-0139-09

DOI: 10.15890/j.cnki.jsjs.2026.03.014

Forecasting Model of WTP Water Supply Capacity Based on Feature Optimization and Extreme Gradient Boosting Tree

Yang Yuling¹, Li Zhu^{2,*}, Wang Ziyu², Yang Lan², Song Zhaoyang¹, Cui Luyuan²

(1. Shanghai Chengtuo Water <Group> Co., Ltd., Shanghai 200082, China;

2. Water Production Branch, Shanghai Chengtuo Water <Group> Co., Ltd., Shanghai 200086, China)

Abstract [Objective] To support the development of smart water, meet the demand for refined scheduling of water treatment plant (WTP), and improve water supply efficiency and resource utilization, it is necessary to construct an accurate hourly water supply forecasting model to optimize production scheduling. [Methods] Taking WTP A as the research object in the east China region, this paper integrated time characteristics, meteorological conditions, equipment maintenance conditions and hydraulic parameters to build a dataset with 58 initial features. The recursive feature elimination (RFE) algorithm was used to screen 16 key features, and Bayesian optimization (BO) was combined to optimize the hyperparameters of the extreme gradient boosting tree (XGBoost) model, thereby establishing an hourly water supply forecasting model. [Results] The model achieved a mean absolute percentage error (MAPE) of 4.58% and a mean absolute error (MAE) of 676.02 m³/h after BO. Verification in April 2024 showed that when scheduled based on predictions, the 1-hour reservoir water level fluctuation could be controlled within ±0.3 m at 99.66% of the time periods, and the maximum daily water level deviation was only 0.77 m without manual intervention. In August 2024, the model was integrated with the intelligent scheduling system and put into online operation. By providing 1 hour to 4 hours water supply forecasting, it reduced the number of pump start-stop cycles by 47%. [Conclusion] This model has both high accuracy and practical value, which can avoid equipment loss and energy waste, and provide technical support for the optimized scheduling of WTPs.

[收稿日期] 2025-09-03

[基金项目] 国家重点研发计划(2022YFC3801000); 上海城投(集团)有限公司科技创新计划项目(启明星专项)(CTKY-PTRC-2023-002-002-005)

[作者简介] 杨瑜玲(1996—), 女, 工程师, 主要从事智慧水务等工作, E-mail: yangyuling@shanghaiwater.com。

[通信作者] 李柱(1980—), 男, 正高级工程师, 主要从事自来水公司运营管理工作, E-mail: lizhu992@126.com。

Keywords extreme gradient boosting tree (XGBoost) recursive feature elimination(RFE) Bayesian optimization (BO) feature selection water supply forecasting smart water

随着上海城市数字化转型的加快推进,智慧水务成为传统水务行业转型升级的核心方向,在供水领域科学准确的水量预测是提升水资源利用效率、优化水量调度的关键支撑^[1]。传统水量预测多聚焦管网侧理论需水量(由用水习惯、气象、社会活动决定),但水厂实际调度还需考虑设备检修、水源水质波动等内部约束,以及外部管网水力条件共同影响。因此,本研究旨在综合各类因素,预测水厂未来小时级预期出厂流量,以贴合调度决策的实际场景,辅助预见性生产控制与资源配置。

要实现上述目标,需首先解决现代水厂数据特性带来的挑战及现有研究存在的不足。现代水厂数据具有“高维、异构、强耦合”特征(融合时序、气象、工况、水力等多源数据),现有相关研究多依赖经验进行特征选择,缺乏系统性方法剔除冗余与无关特征,这不仅增加模型计算负担,还可能引入数据噪声,最终损害模型预测性能。基于此,本研究以华东地区 A 水厂为对象,提出基于递归特征消除(RFE)和极端梯度提升树(XGBoost)的水厂供水量预测模型,该模型以 XGBoost 为核心算法,针对水务数据高维、复杂的特性,创新性地引入了 RFE 算法进行系统性的特征筛选,以降低模型复杂度并提升泛化能力。同时,采用贝叶斯优化(BO)对模型关键超参数进行高效自动寻优,以克服传统网格搜索或人工调参的局限性,从而最大化提升模型的预测精度。通过这一系列组合优化,构建出一个兼具高精度与强鲁棒性的预测工具,为水厂的精细化、智能化调度提供可靠的技术支撑。

1 研究区域概况

华东地区 A 水厂设计供水规模为 70 万 m^3/d ,供水服务人口约为 100 万,供水范围内设有 1#、2# 2 个中途增压水库泵站,且与 B 水厂、C 水厂供水范围存在部分交叠,该水厂的供水量不仅受内部生产工况如机泵运行状态、检修计划的影响,还与外部管网水力条件如周边泵站流量、相邻水厂供水分配密切相关,供水量影响因素具有典型的复杂性与关联性特征。

2 时供水量预测模型构造

2.1 模型预测目标

本研究旨在构建一个基于监督学习的回归模型,以解决 A 水厂的小时级供水量预测问题。预测目标是未来 1 h 的出厂瞬时流量,单位为 m^3/h 。输入变量(特征)为当前时刻的一系列数据,涵盖时间、气象、水厂工况及内外部水力情况等多个维度。在实际应用中,模型将根据水厂流量数据采集频率,同步实时预测未来 1 h 的出厂瞬时流量。

2.2 特征构造

城市短期供水量因居民活动与自然系统的深度耦合,存在显著的随机扰动性与不确定性,给小时供水量预测带来挑战。为此,本研究从“全要素覆盖”视角出发,结合特征工程技术,从时间类、气象类、工况类、水力类(内部)、水力类(外部)5 个维度构建特征体系,确保全面捕捉供水量影响因素^[2-5]。

通过原始数据衍生与数值化处理,最终构建包含 58 个特征的初始特征集,如表 1 所示,包括时间类 7 个;气象类 6 个;工况类 1 个;水力类(内部)17 个;水力类(外部)27 个。

2.3 数据收集与处理

A 水厂于 2021 年 10 月完成出水机泵更新改造,对外供水模式发生结构性变化(如机泵效率提升、供水压力稳定区间调整)。为确保数据一致性与模型适用性,本研究选取 2021 年 10 月 1 日—2024 年 3 月 31 日的分钟级数据作为原始数据集,涵盖时间、气象、工况、水力等全维度特征数据。

原始数据收集完成后,需进行数据处理。本文采用拉依达准则^[6]进行异常值筛选与替换, μ 为数据均值, σ 为标准差,分布在 $(\mu-3\sigma, \mu+3\sigma)$ 的数据视为正常值,否则为异常值。异常值处理方法:若 $x_i \leq \mu-3\sigma$,则 $x_i = \mu-3\sigma$;若 $x_i \geq \mu+3\sigma$,则 $x_i = \mu+3\sigma$ (x_i 表示第 i min 的原始数据)。此外,由于实时流量数据毛刺较多,为平滑毛刺,将精度为分钟级的数据计算成滑动小时均值,既保留时序趋势,又消除高频噪声。而针对数据集中的缺失值,采用前向填充的方法处理,即如果某个时间点的数据缺失,则使用其前一个最近的有效值进行填充。对于数据序列开头连续出现的缺失值,则采用后向填充进行

表 1 初始特征集
Tab. 1 Initial Feature Sets

类别	特征	采样间隔	处理方式
时间类	月份、星期几、时刻	—	数值化
	休息日、高低谷(分3段)、高低谷(分4段)、节假日	—	分类编码
气象类	温度、降雨量、湿度	分钟级	直接使用
	最高温、最低温	—	直接使用
	温度划分	—	分类编码
工况类	是否年度大检修日	—	分类编码
水力类(内部)	A 水厂流量(当前)、A 水厂流量(1 h 前)、A 水厂流量(2 h 前)、A 水厂流量(3 h 前)、A 水厂流量(4 h 前)、A 水厂流量(前一天此时)、A 水厂流量(前一周此时)	分钟级	滑动小时均值
	A 水厂流量(当前的一阶导数)、A 水厂流量(当前的二阶导数)、A 水厂流量(1 h 前的一阶导数)、A 水厂流量(1 h 前的二阶导数)、A 水厂流量(2 h 前的一阶导数)、A 水厂流量(2 h 前的二阶导数)、A 水厂流量(3 h 前的一阶导数)、A 水厂流量(3 h 前的二阶导数)、A 水厂流量(4 h 前的一阶导数)、A 水厂流量(4 h 前的二阶导数)	分钟级	由流量数据计算
水力类(外部)	1#泵站水库 1 水位、1#泵站水库 2 水位、1#泵站总进水压力、1#泵站总出水压力、2#泵站水库进水阀开度反馈、2#泵站水库液位 1、2#泵站水库液位 2、2#泵站总进水压力、2#泵站总出水压力、馈水管压力、馈水管累计流量、B 水厂 1#~5#出厂管压力、C 水厂 1#~4#出厂管压力	分钟级	直接使用
	1#泵站出水瞬时流量、2#泵站进水瞬时流量 1、2#泵站进水瞬时流量 2、馈水管瞬时流量、B 水厂总瞬时流量、C 水厂总瞬时流量、中心城区水厂总瞬时流量	分钟级	滑动小时均值

处理。选择此方法的原因是,在水力系统中,各项参数(如流量、压力)在短时间内具有较强的惯性,不会发生剧烈跳变,因此前一时刻的值是其最合理的估计。此方法确保了数据集的完整性,为后续模型训练奠定了基础。

2.4 超参数优化

超参数是影响机器学习模型性能的关键因素,不同调参策略会显著改变模型的训练效率与预测精度。传统手动调参方法依赖经验且难以遍历最优参数组合,而贝叶斯高斯过程^[7-8]可通过已知参数点逼近目标函数的方式,高效搜索最优超参数:其核心逻辑是在每次参数搜索时,优先选择已知参数点附近均值大、方差大的区域,通过采集函数平衡均值与方差的关系,对非凸优化问题具有稳健性且耗时短。

2.5 算法选择

供水量预测属于有监督回归问题,常用算法包括决策树类与神经网络类方法。本文选取 XGBoost、反向传播神经网络(BPNN)和长短期记忆(LSTM)神经网络作为代表性模型进行对比。XGBoost 是一种基于 Boosting 的集成树模型,采用前向分布算法与二阶泰勒展开,精度高、灵活性好,并通过正则项防止过拟合^[9]。BPNN 为多层神经网络,

通过正向与反向传播过程更新权重,具有良好的学习效率与泛化能力^[10]。LSTM 通过门控机制改进循环神经网络的长期依赖问题,适用于构建大型深度网络^[11]。

本节选取解释方差分(EVS)、均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)这 4 个指标对 3 类算法的预测结果进行误差分析,各误差指标的计算方法如式(1)~式(4)。

$$E_{VS} = 1 - \frac{\sum_{i=1}^n [y_i - \hat{y}_i - \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)]^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$E_{RMS} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$E_{MA} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$E_{MAP} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (4)$$

其中: \hat{y}_i 、 \bar{y}_i —— y_i 的预测值、均值, m^3/h ;

- E_{VS} ——解释方差分;
- y_i ——第 i 个流量实测数据, m^3/h ;
- E_{RMS} ——预测值和实际值的偏差程度, m^3/h ;
- E_{MA} ——每个样本预测误差绝对值的平均值, m^3/h ;
- n ——供水量预测样本的总数量;
- E_{MAP} ——预测值和实际值相对误差的百分比。

从数值来看, EVS 取值是在 $[0, 1]$, 数值越接近 1, 说明预测效果越好; $RMSE$ 和 MAE 取值是在 $[0, +\infty)$, 数值越小, 说明预测效果越好; $MAPE$ 取值是在 $[0, 100\%]$, 数值越接近 0, 说明预测效果越好。

使用 2021 年 10 月 1 日—2024 年 3 月 31 日, 58 个特征的历史数据分别训练模型, 其中 80% 的数据用于训练模型, 剩余 20% 的数据作为测试集评价模型性能。

不同算法的预测误差指标如表 2 所示, $XGBoost$ 在 4 个误差指标中表现均最优, $BPNN$ 次之, $LSTM$ 效果最差。

表 2 不同算法预测误差对比

Tab. 2 Comparison of Prediction Performance of Different Algorithms

特征保留方式	EVS	$RMSE/$ ($m^3 \cdot h^{-1}$)	$MAE/$ ($m^3 \cdot h^{-1}$)	$MAPE$
$XGBoost$	0.89	936.23	676.02	4.58%
$BPNN$	0.85	1 141.86	896.36	6.08%
$LSTM$	0.80	1 341.45	1 027.98	7.32%

综合来看, $XGBoost$ 算法的计算误差和训练时间成本均最低, 是 3 种所选算法中最优的。主要原因在于: (1) 集成决策树结构简单, 对异常值不敏感, 且可通过并行计算提升训练效率; (2) 前向分布算法与二阶泰勒展开使模型收敛更快, 正则项有效抑制过拟合; (3) 相比 $BPNN$ 、 $LSTM$ 的“黑箱”特性, $XGBoost$ 可输出特征重要性, 模型可解释性更强。综上, 本节选取 $XGBoost$ 算法, 用于构建供水量预测模型。

2.6 特征优选

初始 58 维特征集的选择原则是“全要素覆盖”, 但其中可能存在特征冗余(如压力与流量的信息重叠)、低效特征(如与供水量相关性极低的特征), 不仅增加模型训练成本, 还可能降低预测稳定性。为此, 本研究通过 RFE 算法优选^[12-13], 筛选最

优特征子集。原始 58 维特征集及其重要性如图 1 所示, 采用权重(weight)法计算特征重要性, 即以特征在树中被用作分裂节点的总次数作为该特征的重要性程度。

2.6.1 特征优选方法原理

RFE 是一种基于模型的特征选择方法, 通过反复训练模型和剔除非必要特征的方式来选择最优的特征子集, 实现了全部特征的遍历, 是一种寻找最优特征子集的贪心算法。结合以 $XGBoost$ 算法为基础建立的预测模型, RFE 算法进行特征选择的步骤如下。

(1) 初始化: 选择 $XGBoost$ 算法建立基础模型, 并设定目标特征数量。

(2) 特征重要性评估: 用原始特征集的数据训练基础模型, 并采用 weight 法计算特征重要性。

(3) 特征排序: 根据重要性大小对特征进行排序, 选择重要性最大的一个特征作为待剔除的特征。

(4) 特征剔除: 从特征集中剔除待剔除的特征, 得到新的特征子集。

(5) 判断停止条件: 如果特征数量达到预设的目标或者所有特征都被剔除完毕, 算法自动终止; 否则回到步骤(3), 在新的特征子集上重新训练 $XGBoost$ 模型, 继续执行步骤。

(6) 结束算法: 停止算法时选择剩余的特征作为最终的特征子集。

2.6.2 特征优选实践

本研究使用上述方法, 从“无预筛选”与“有预筛选”2 个维度开展 RFE 试验, 以测试集的 $MAPE$ 为性能评价标准。

(1) 无预筛选特征的 RFE 算法特征优选

由于难以获取模型最优的目标特征数量先验知识, 在测试中预设目标特征数量为 1~58, 逐一运行 RFE 算法并计算模型性能。结果显示(图 2), 当特征数量为 40 时, 模型 $MAPE$ 为 4.57%, 性能最优(略优于全量特征模型)。但该方案未剔除冗余特征, 仍存在数据传输与存储成本较高的问题。

(2) 有预筛选特征的 RFE 算法特征优选

通过特征数据相关性分析和业务经验对全量特征进行预筛选, 之后由 RFE 算法对特征进一步优化。

首先基于数据相关性分析, 与预测目标的数据相关性大的特征通常对模型性能有较大作用, 从全量特征中剔除与 y 的相关性绝对值小于 0.1 的 18 个特征, 剩余 40 个特征构成初步特征子集。

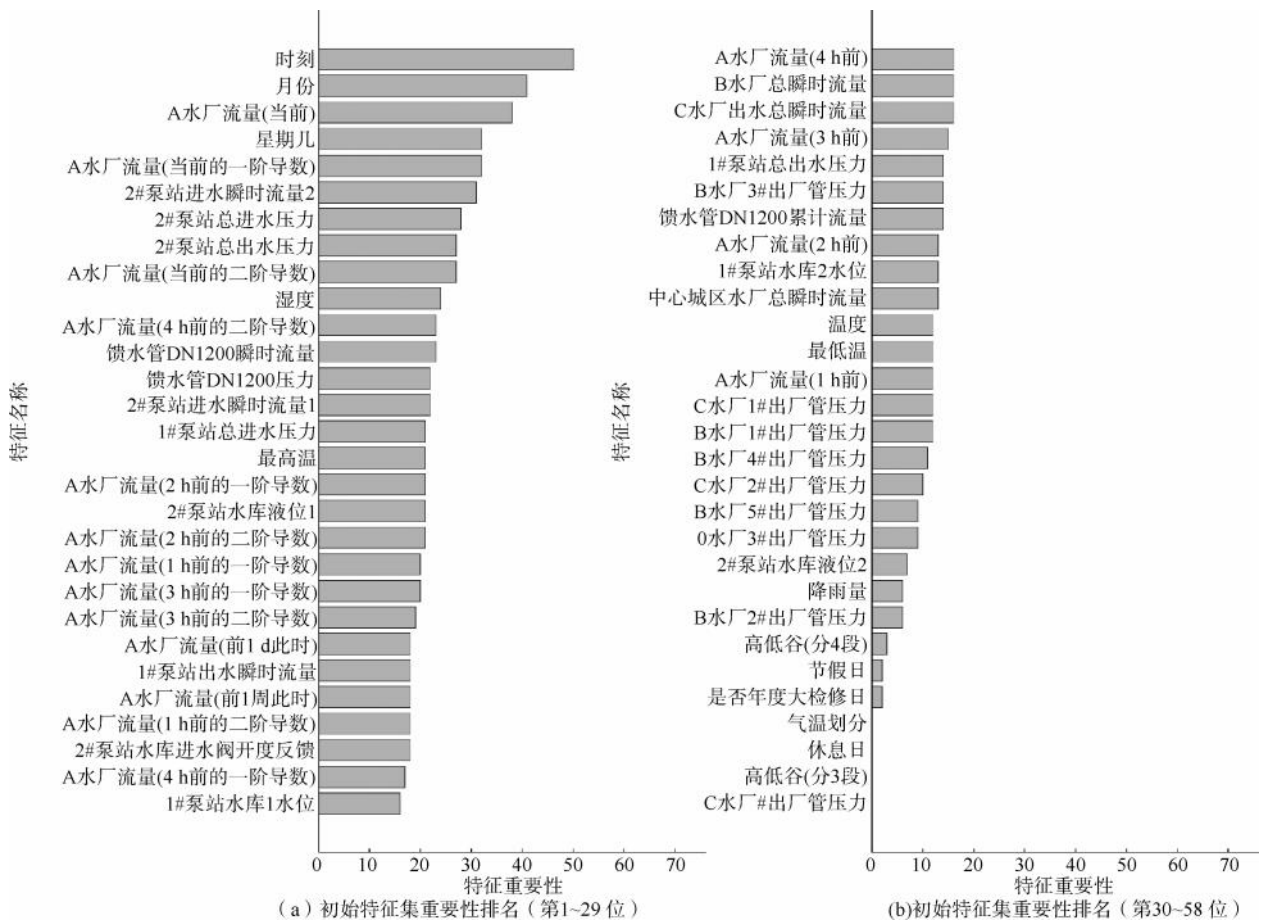


图1 初始特征集及重要性

Fig. 1 Initial Feature Sets and the Importance

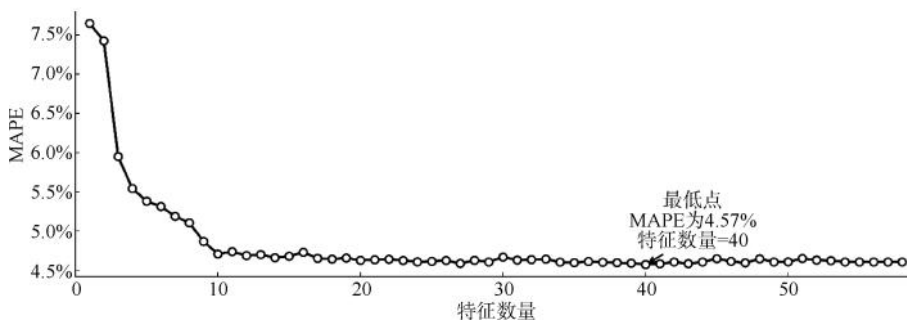


图2 无预筛选特征的RFE算法特征优选结果

Fig. 2 Results of Feature Selection by RFE Algorithm without Pre-Screening

在此基础上,根据业务经验简化水力类(外部)特征。水厂的压力和流量特征以不同的数据形式反映相似的水厂供水量变化规律,数据信息存在相似和冗余。A水厂、B水厂的压力、流量宜保留其中一类。从已选出的40个特征中删除两厂全部压力或流量特征重新训练模型,对比测试集上的模型性能表现,删除水厂压力特征时模型精度更高,此时剩余特征为32个。

预设目标特征数量为1~32,运行RFE算法,结果显示(图3):当特征数量为16时,模型MAPE为4.64%,虽精度略低于无预筛选方案,但特征数量减少56.25%,大幅降低模型复杂度与实时计算成本。

2.6.3 特征优选结果分析

按上述特征优选方式所得特征子集分别训练模型,在测试集上评估模型性能(表3)。

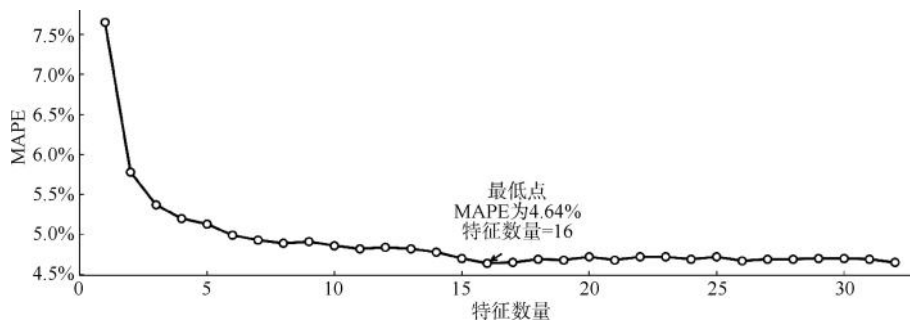


图3 有预筛选特征的 RFE 算法特征优选结果

Fig. 3 Results of Feature Selection by RFE Algorithm with Pre-Screening

表3 特征优选结果分析

Tab. 3 Analysis of Feature Selection Results

特征保留方式	特征数量	MAE/ ($m^3 \cdot h^{-1}$)	MAPE	
无预筛	特征优选前	58	676.02	4.58%
	特征优选后	40	675.54	4.57%
有预筛	特征优选前	32	686.12	4.65%
	特征优选后	16	677.58	4.64%

对比发现无特征预筛选的特征优选模型测试性

能最优,且略优于采用全量特征的结果。在有特征预筛选的情况下,尽管各步骤所得模型测试性能均不是最优,但大幅精简了特征集合,而模型性能与最优情况差异微小。

综合权衡特征子集数量和模型预测精度,本节认为结合特征预筛选和 RFE 算法的特征优选结果最佳,特征为 16 个,最优特征子集的特征及重要性排序结果如图 4 所示。

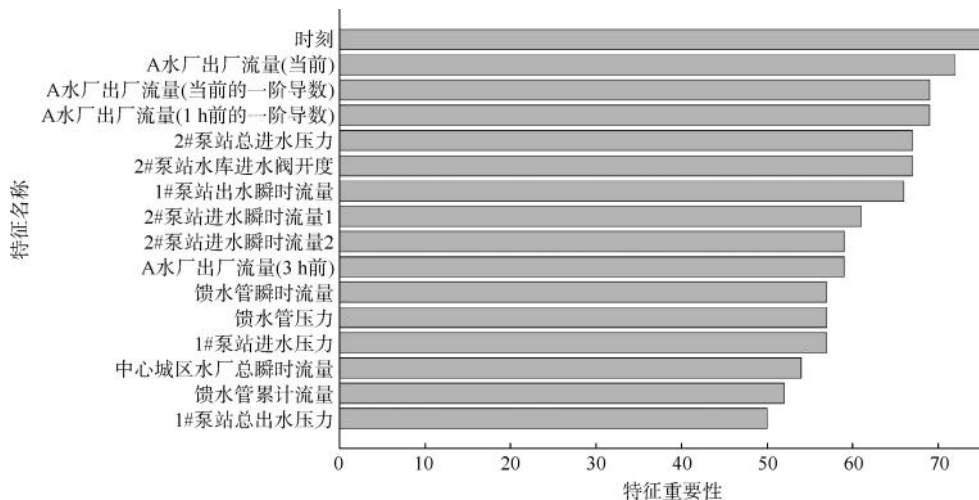


图4 最优特征子集及重要性

Fig. 4 Optimal Feature Sets and the Importance

3 结果和讨论

本文构建的模型是为了向 A 水厂的智能水量调度系统提供未来小时级供水量趋势预测。该智能水量调度系统基于水量平衡原理,建立了进水量与出水量之间的关系式,可以实现自动计算水厂所需总进水量,实现供水-制水之间的平衡。水厂进水量、出水量和水库液位的关系简化表达式如式(5)。

$$\Delta h = \frac{Q_{进} - Q_{出}}{S} \quad (5)$$

其中: Δh ——水库液位变化高度, m;

$Q_{进}$ ——水厂进水流量, m^3/h ;

$Q_{出}$ ——水厂出水流量, m^3/h ;

S ——全水厂水库底面积,取值为 $14\,300\,m^2$, m^2 。

模型上线前,本文从水厂实际应用角度出发,以水库液位变化情况为重点,使用 2024 年 4 月 1 日—4 月 16 日数据,评估不同特征构造的模型预测误差造成的小时级水库液位高度变化情况,以及对水库

液位的日累计影响,确定最终上线运行的模型版本。参与分析的模型包括4个,分别是全量特征构造的模型(特征数量为58个)、全量特征进行特征优选后构造的模型(特征数量为40个)、预筛选后构造的模型(特征数量为32个)、预筛选且特征优选后特征优选的模型(特征数量为16个)。

3.1 绝对误差分析

A水厂水库底面积总和约为14 300 m²,若智能水量调度系统直接按照模型预测的供水量来调整水厂进水量,根据式(5),MAE的绝对值大于

1 430 m³/h时,在1 h内会对水库液位造成±0.1 m的影响。本文分别统计了4个模型的MAE的绝对值大于1 430 m³/h(±0.1 m)、2 860 m³/h(±0.2 m)、4 290 m³/h(±0.3 m)及5 720 m³/h(±0.4 m)的比例(表4)。4个模型表现相近,全量特征进行特征优选(特征保留数量N=40)的MAE的绝对值大于1 430 m³/h的比例最小,为12%。此外,仅有约0.34%的MAE的绝对值大于4 290 m³/h,即约99.66%的预测值,在1 h内对水库液位造成的影响小于±0.3 m,风险可控。

表4 不同特征保留方式预测误差对比

Tab. 4 Comparison of Prediction Errors under Different Feature Retention Methods

特征保留方式	全量特征 (N=58)	全量特征进行特征优选 (N=40)	预筛选后特征 (N=32)	预筛选后特征优选 (N=16)
对水库液位造成±0.1 m影响的比例	13.00%	12.00%	14.00%	13.00%
对水库液位造成±0.2 m影响的比例	1.00%	1.37%	0.68%	0.68%
对水库液位造成±0.3 m影响的比例	0.34%	0.34%	0.34%	0.34%
对水库液位造成±0.4 m影响的比例	0	0	0	0

进一步以天为单位计算模型预测误差对水库液位的累计影响,若液位为正值,代表按照预测值来调度,会使水库液位偏高;若液位为负值,代表按照预测值来调度,会使水库液位偏低。4个模型中,预筛选后特征优选(N=16)的模型对应的最大累计液位值为0.77 m,是所有模型中偏移最小的,代表若无人工干预,当日水库液位将最终偏高0.77 m。

3.2 模型预测结果对智能水量调度系统运行效果的影响分析

为满足实际生产需求,本研究基于上文所述方法,使用预筛选结合特征优选(N=16)的数据集,构建XGBoost模型,提供A水厂未来1~4 h的对外瞬时供水流量预测,并与厂内智能水量调度系统深度融合。

2024年8月,模型于A水厂在线运行,每3 min计算1次,连续60 d的运行结果显示MAPE为7.4%(表5)。为了清楚显示预测值与真实值,选取连续30 d的数据曲线进行曲线对比,结果表明拟合效果较好(图5)。

与此同时,本研究将预测模型与A水厂智能水量调度系统融合(V1版本),同时后台运行不含预测模型的原系统(V0版本),通过46 d的对比运行,分析模型对调度效果的提升作用。

表5 连续60 d运行的供水量预测模型性能

Tab. 5 Performance of Water Supply Capacity Forecasting Model during 60 d Consecutive Operation

预测对象	MAPE
未来1 h瞬时供水流量	7.4%
未来2 h瞬时供水流量	9.7%
未来3 h瞬时供水流量	12.5%
未来4 h瞬时供水流量	12.5%

核心评价指标为机泵启停次数——频繁启停会导致设备损耗加剧、能耗上升。结果显示,V0版本的机泵启停次数为70次;V1版本的机泵启停次数为37次,较V0版本减少47%。减少的启停主要集中在出水流量临近15 000 m³/h的短时波动场景,由于模型提前预测流量变化趋势,避免了流量微小波动即调整泵组的冗余操作,显著提升机泵运行稳定性与能源利用效率。

3.3 总结

本研究综合特征数量(16维,精简高效)、绝对误差(99.66%时段水库液位≤±0.3 m、日最大水库液位偏差为0.77 m)、调度效果(实际运行时机泵启停减少47%)三大因素,最终确定预筛选结合特征优选(N=16)的XGBoost模型为上线版本,并与水厂智能水量调度系统深度融合,实现供水量预测与生产调度的一体化。

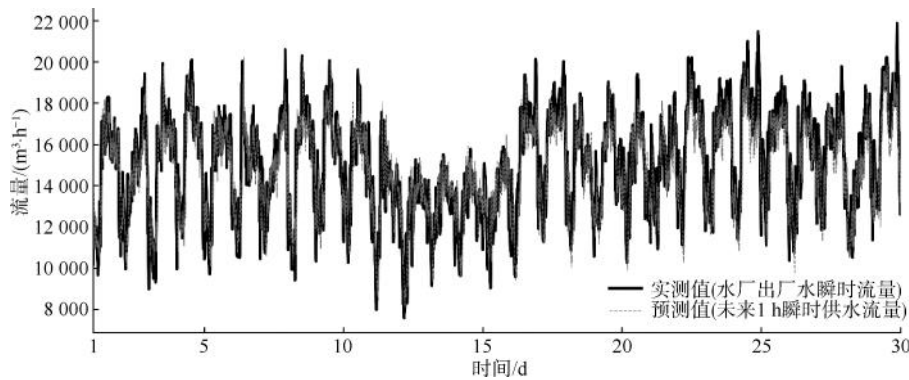


图5 连续30 d运行的供水量预测模型预测曲线

Fig. 5 Prediction Curve of the Water Supply Forecasting Model during 30 d Consecutive Days of Operation

4 结论

本文结论如下。

(1) 本文构建了覆盖时间、气象、工况、水力(内/外部)的58维初始特征集,通过结合皮尔逊相关系数以及RFE算法优选出16个关键特征,结合BO的XGBoost模型,实现了A水厂时供水量的高精度预测,为智慧水务调度提供了系统化特征体系与模型框架。

(2) 模型性能验证表明,基于XGBoost构建的模型MAPE为4.6%、MAE为649.86 m³/h,显著优于BPNN(MAPE = 6.08%)与LSTM(MAPE = 7.32%);模拟智能水量调度系统直接按照模型预测的供水量来调度,99.66%的时间段内1 h水库液位 $\leq \pm 0.3$ m,无人工干预日水库液位最大偏差仅为0.77 m,兼具高精度与稳定性。

(3) 实际应用效果显示,模型与智能水量调度系统融合后,通过提供未来1~4 h供水量变化趋势,使机泵启停次数减少47%,有效降低设备损耗与能源浪费,验证了模型的实际应用价值,为智慧水务实现“无人运行、少人管理”的智慧管理模式提供了技术支撑。

参考文献

[1] 李原园,曹建廷,黄火键,等. 国际上水资源综合管理进展[J]. 水科学进展, 2018, 29(1): 127-137.
Li Y Y, Cao J T, Huang H J, et al. International progresses in integrated water resources management[J]. Advances in Water Science, 2018, 29(1): 127-137.
[2] 孙晓峰. 基于社会大数据的典型区域供水量时空变化规律研究[J]. 给水排水, 2023, 49(s1): 908-915.
Sun X F. Study on the laws of time and space change of water

supply based on social big data [J]. Water & Wastewater Engineering, 2023, 49(s1): 908-915.
[3] 常远勇,谈建国,彭杰,等. 上海市日供水量与气象要素的相关性分析及预测模型的建立[J]. 水资源与水工程学报, 2015, 26(1): 32-36.
Chang Y Y, Tan J G, Peng J, et al. Relativity analysis of daily water supply and meteorology factor and establishment of forecast model in Shanghai [J]. Journal of Water Resources and Water Engineering, 2015, 26(1): 32-36.
[4] 徐东,张曾,周迅,等. 基于机器学习的水厂短期供水量预测模型构建[J]. 四川水利, 2022, 43(4): 37-40, 57.
Xu D, Zhang Z, Zhou X, et al. Construction of short-term water supply prediction model of waterworks based on machine learning [J]. Sichuan Water Resources, 2022, 43(4): 37-40, 57.
[5] 杨存满,鞠佳伟,袁芳,等. 基于BP神经网络的城市自来水厂供水量预测研究[J]. 灌溉排水学报, 2021, 40(s2): 35-37.
Yang C M, Ju J W, Yuan F, et al. Study on water supply prediction of urban waterworks based on BP neural network [J]. Journal of Irrigation and Drainage, 2021, 40(s2): 35-37.
[6] 周欣磊,顾海挺,刘晶,等. 基于集成学习与深度学习的日供水量预测方法[J]. 浙江大学学报:工学版, 2023, 57(6): 1120-1127.
Zhou X L, Gu H T, Liu J, et al. Daily water supply prediction method based on integrated learning and deep learning [J]. Journal of Zhejiang University: Engineering Science, 2023, 57(6): 1120-1127.
[7] 项佳宇,王晓冰,李冰,等. 改进贝叶斯优化与集成学习短期负荷预测模型[J]. 电力系统及其自动化学报, 2025, 37(9): 34-44.
Xu J Y, Wang X B, Li B, et al. Short-term load forecasting model based on improved Bayesian optimization and ensemble learning [J]. Proceedings of the CSU-EPSCA, 2025, 37(9): 34-44.

(下转第155页)

- Qi L Q, Liu S W, Sun X F. Study on risk assessment of urban water supply network based on dynamic simulation [J]. *Municipal Engineering Technology*, 2022, 40(3): 140-144.
- [5] Zhang W J, Lai T T, Li Y. Risk assessment of water supply network operation based on ANP-fuzzy comprehensive evaluation method [J]. *Journal of Pipeline Systems Engineering and Practice*, 2022, 13(1): 0000602.
- [6] Akbarinia G R, Gilani N, Mohhammadzadeh S, et al. Application of Kent Muhlbauer method to risk assessment of urban gas pipelines [J]. *Journal of Pipeline Systems Engineering and Practice*, 2022, 13(1): 0000601.
- [7] 熊凡, 李沛鸿, 袁逸敏, 等. 基于 GIS 和组合赋权的城市洪涝灾害风险评估[J]. *人民长江*, 2023, 54(8): 60-66.
- Xiong F, Li P H, Yuan Y M, et al. Urban flood disaster risk assessment based on GIS and combined weights [J]. *Yangtze River*, 2023, 54(8): 60-66.
- [8] W. Kent M. 管道风险管理手册[M]. 杨嘉瑜, 张德彦, 吴俊, 等, 译. 2 版. 北京: 中国石化出版社, 2005.
- W. Kent M. Pipeline risk management manual Second Edition [M]. trans. by Yang J Y, Zhang D Y, Wu J, et al. 2th ed. Beijing: China Petrochemical Press, 2005.
- [9] Zhao J R, Tian J D, Meng F X, et al. Safety assessment method for storage tank farm based on the combination of structure entropy weight method and cloud model [J]. *Journal of Loss Prevention in the Process Industries*, 2022, 75: 104709. DOI: 10.1016/j.jlp.2021.104709.
- [10] Aljanabi A R A, Ghafour K M. Fuzzy AHP and fuzzy TOPSIS methods of analysing online impulsive buying of organic food: A cognitive-affective decision-making perspective [J]. *Journal of Intelligent & Fuzzy Systems*, 2024, 46(4): 7823-7838.
- [11] Sun P, Zhang R Z, Qiu X W. A Survey on Cloud Model [J]. *Journal of Internet Technology*, 2023, 24(5): 1159-1167.

(上接第 146 页)

- [8] 余维, 王欣, 陈斌, 等. 基于贝叶斯优化极端梯度提升树的电缆状态分类研究 [J]. *郑州大学学报: 理学版*, 2025, 57(6): 1-7.
- She W, Wang X, Chen B, et al. Research on cable condition classification based on Bayesian optimization of extreme gradient boosting tree [J]. *Journal of Zhengzhou University: Natural Science Edition*, 2025, 57(6): 1-7.
- [9] 贺波, 马静, 高赫余. 基于多粒度特征和 XGBoost 模型的城市日供水量预测 [J]. *长江科学院院报*, 2020, 37(5): 43-49.
- He B, Ma J, Gao H Y. Predicting urban daily water supply based on multi-granularity feature and XGBoost integrated model [J]. *Journal of Yangtze River Scientific Research Institute*, 2020, 37(5): 43-49.
- [10] 温家鸣, 郭纯青, 李新建, 等. 基于 BP 神经网络的桂林生态城市建设需水量预测 [J]. *水资源保护*, 2012, 28(3): 47-50.
- Wen J M, Guo C Q, Li X J, et al. Water demand forecast of eco-city construction in Guilin City based on BP neural network [J]. *Water Resources Protection*, 2012, 28(3): 47-50.
- [11] Narayanan L K, Sankaranarayanan S, Rodrigues J, et al. Water demand forecasting using deep learning in IoT enabled water distribution network [J]. *International Journal of Computers Communications & Control*, 2020, 15(6): 141-147.
- [12] 马飞虎, 雷皓安, 孙翠羽, 等. 基于 RFE-OPTUNA-XGBoost 模型的高速公路逃费模式识别 [J]. *应用科学学报*, 2024, 42(5): 857-870.
- Ma F H, Lei H A, Sun C Y, et al. Highway toll evasion patterns identification based on RFE-OPTUNA-XGBoost model [J]. *Journal of Applied Sciences*, 2024, 42(5): 857-870.
- [13] 崔晓彤, 靳亚红, 兰凤, 等. 基于多光谱影像的冬小麦土壤含盐量反演研究 [J]. *河北农业大学学报*, 2025, 48(1): 107-115, 126.
- Cui X T, Jin Y H, Lan F, et al. Inversion of salt content in winter wheat soil based on multi-spectral images [J]. *Journal of Agricultural University of Hebei*, 2025, 48(1): 107-115, 126.