

基于滚动交叉验证的城市需水预测方法

董增川¹, 王佳晟¹, 崔 璨¹, 韩亚雷¹, 陈荣豪¹, 杨家亮², 王淑云²

(1. 河海大学水文水资源学院, 江苏 南京 210098; 2. 湖南省水利水电勘测设计规划研究总院, 湖南 长沙 410007)

摘要:为提高机器学习算法在城市需水预测中的精度,提出了一种基于滚动交叉验证的系统化预测方法,包括影响因子指标体系构建、需水预测模型构建、结合滚动交叉验证的超参数优化以及模型性能的评估与优选,并以衡阳市为实例进行了方法验证。结果表明:预测的 2025 年衡阳市需水量与规划值具有较高的一致性,验证了该方法的适用性和实际应用价值;该方法具有较强的普适性,可根据不同区域的经济社会发展趋势及用水结构灵活调整指标体系和模型组合,结合滚动交叉验证的超参数优化显著提高了模型的泛化能力和预测精度,更好地满足了真实应用场景的需水预测需求。

关键词:城市需水预测;机器学习算法;超参数优化算法;滚动交叉验证;需水预测模型;衡阳市
中图分类号:TV213.4 **文献标志码:**A **文章编号:**1004-6933(2025)03-0013-07

Method for urban water demand forecast based on rolling cross-validation//DONG Zengchuan¹, WANG Jiasheng¹, CUI Can¹, HAN Yalei¹, CHEN Ronghao¹, YANG Jialiang², WANG Shuyun² (1. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; 2. Hunan Water Resources and Hydropower Survey, Design, Planning and Research Co., Ltd., Changsha 410007, China)

Abstract: To enhance the accuracy of machine learning algorithms in urban water demand forecast, this study proposes a systematic forecasting method based on rolling cross-validation. The method integrates the construction of an influencing factor indicator system, the development of forecast models, hyperparameter optimization guided by rolling cross-validation, and comprehensive model performance evaluation and model selection. The method is validated using Hengyang City as a case study. The results indicate that the forecast urban water demand of Hengyang City for 2025 closely aligns with the planned values, demonstrating the applicability and practical value of the method. Moreover, the method exhibits strong generation ability, allowing flexible adaptation of indicator systems and model structures to accommodate varying regional socio-economic trends and water use characteristics. The incorporation of rolling cross-validation into the hyperparameter optimization process significantly enhances the generalization capability and forecast accuracy of the models, making the method well-suited for real applications in urban water demand management.

Key words: urban water demand forecast; machine learning algorithm; hyperparameter optimization algorithm; rolling cross-validation; water demand forecast model; Hengyang City

在我国经济社会快速发展的时代背景下,城市水资源规划与管理对于保障水资源合理开发利用、促进城市可持续发展起着至关重要的作用^[1]。需水预测是制定合理水资源利用和管理策略的关键,但受气候变化、人类活动、地域发展条件、政策调整等因素影响,水资源需求具有非线性、不确定性的特征,其变化规律难以准确掌握。因此,科学预测城市需水量是保障城市水安全和缓解城市水资源供需矛盾的关键所在^[2]。

传统的统计方法,如时间序列法和相关分析法,虽然在一定程度上能够满足需求,但在应对复杂非

线性关系和多变量因素影响时表现出局限性^[3]。为了应对这些挑战,大数据分析技术、机器学习和深度学习方法逐渐被引入需水预测。基于机器学习算法的城市需水预测通常包括影响因子筛选、需水预测模型构建和超参数优化。影响因子筛选多采用主成分分析^[4]、平均影响值算法^[5]和随机森林(random forest, RF)模型^[6],需水预测模型涵盖灰色预测法^[7]、传统机器学习算法^[8-12]、深度学习算法^[13-14]以及组合预测模型^[5,15]等。部分预测模型需要进行超参数优化,超参数优化算法包括遗传算法^[12,16]、粒子群优化(particle swarm optimization,

基金项目:犬木塘水库工程科技创新项目(QMT-KY/02-2022);湖南省水利科技项目(XSKJ2021000-05)

作者简介:董增川(1963—),男,教授,博士,主要从事水资源规划管理研究。E-mail:zcdong@hhu.edu.cn

通信作者:王佳晟(2001—),男,硕士研究生,主要从事水文水资源研究。E-mail:231301010040@hhu.edu.cn

PSO) 算法^[17-18] 和 贝 叶 斯 优 化 (Bayesian optimization, BO) 算法^[19] 等。当前需水预测研究多聚焦于单一模型和单一优化算法的应用, 缺乏对不同预测模型与超参数优化算法在统一框架下的系统性比较。此外, 需水预测通常表现出显著的时间序列特性, 传统的固定数据集划分或简单交叉验证方法在处理时间动态变化时难以充分挖掘数据特性, 从而导致优化后的预测模型难以满足实际应用场景中的精度要求。

针对上述问题, 本文提出了一种基于滚动交叉验证的系统化预测方法, 涵盖需水预测的关键环节, 包括影响因子指标体系构建、需水预测模型构建、结合滚动交叉验证的超参数优化以及模型性能的评估与优选。并以衡阳市为实例进行验证, 采用 RF 模型降维筛选城市需水的主要影响因子, 分别构建基于 BP 神经网络、长短期记忆网络 (long short term memory network, LSTM)、支持向量回归模型 (support vector regression, SVR) 和极限梯度提升模型 (extreme gradient boosting, XGBoost) 的需水预测模型, 并通过固定步长滚动交叉验证, 结合 PSO 算法、BO 算法和 Optuna 对模型中关键超参数进行优化。根据模型评估结果比较不同优化算法与预测模型的性能, 优选出的组合模型可进一步应用于衡阳市未来需水预测。

1 基于滚动交叉验证的城市需水预测方法

1.1 影响因子指标体系构建

城市总需水量受经济、社会、气候等多重因素的综合影响。为了科学地刻画这些复杂因素对需水变化的驱动作用, 构建合理的影响因子指标体系成为城市需水预测模型的基础前提。在构建过程中应遵循全面性、代表性、科学性和可获取性原则, 确保所选指标既能覆盖需水预测的多维驱动因子, 又具有良好的数据可得性与区域适应性。在指标体系构建完成后, 为降低冗余信息带来的维度灾难或过拟合风险, 需进一步筛选对城市需水预测具有显著影响的主要影响因子。

本文采用 RF 模型对影响因子进行特征重要性评估。模型通过在多个子样本上构建并训练多棵决策树, 计算每个特征在分裂节点时对不纯度的平均减少量, 以衡量其在预测任务中的相对重要性。根据评估结果对全部特征进行排序, 筛选出对城市需水预测贡献度较高的主要影响因子, 作为后续模型的输入变量^[6,20]。

1.2 需水预测模型构建

以筛选后的主要影响因子为输入, 城市需水量

为输出, 通过训练机器学习模型拟合输入与输出之间的非线性关系。模型训练前应结合模型特征和数据分布合理设定超参数的初始搜索范围, 该范围既要覆盖潜在的最优解空间, 又应避免搜索范围过宽导致优化效率低下, 从而为后续优化算法提供良好的收敛基础。

选取 BP 神经网络、LSTM、SVR、XGBoost 构建需水预测模型。各模型的结构设计与参数范围设定依据其理论特点与历史数据特征。对于 BP 神经网络, 其输入层节点数由已筛选的输入特征维度确定, 隐藏层神经元数量根据经验公式进行初设, 输出层固定为单节点以实现城市需水量的回归预测^[8,21]; LSTM 的初始记忆单元数设置为输入特征维度的 1~4 倍, 训练过程中采用 Adam 优化器进行参数更新, 并在隐藏层后引入 Dropout 机制以缓解过拟合问题^[5,22]; SVR 的正则化参数范围结合数据波动性经验设定, 不敏感损失函数的初始范围通常设为历史需水量标准差的 5%~10%^[11]; XGBoost 的最大树深参数初始范围设定为 6~12, 最小分裂损失初值根据历史需水量数据的离散程度量化设定^[23-24]。

1.3 结合滚动交叉验证的超参数优化

为进一步提升预测模型的泛化能力与鲁棒性, 本文引入滚动交叉验证策略, 结合多种优化算法进行超参数优化。滚动交叉验证用于评估模型在不同数据划分下的性能, 其验证结果被嵌入优化算法的目标函数中, 作为关键依据动态引导参数更新, 从而提升模型整体性能与泛化能力。

滚动交叉验证是一种专为时间序列数据设计的验证方法, 其核心思想是在保持时间顺序的前提下, 持续扩展训练集并滑动更新验证集, 以评估模型在不同阶段的预测能力, 能够有效避免未来信息泄露带来的评估偏差^[25]。本文采用固定步长滚动交叉验证, 按照固定的时间步长依次向训练集追加样本, 同时保持验证集样本量不变, 在每次迭代中使用新增后的训练集训练模型, 并在当前验证集上评估模型性能, 最终汇总各轮验证结果。此外, 本文选取 PSO 算法、BO 算法和 Optuna 进行超参数优化。PSO 算法通过模拟群体协同搜索实现全局优化, 适用于复杂参数空间^[11]; BO 算法基于代理模型和概率采样策略, 在高计算代价场景中具备较高效率^[19]; Optuna 具备搜索空间动态调整能力, 能高效寻优并适应多模型调参需求^[26]。

1.4 模型性能评估

a. 优化算法评估指标。采用优化性能指数作为优化算法评估指标, 其数值越低表示优化算法性能越优, 且模型具备更强的泛化能力。优化性能指

数计算公式为

$$I=\frac{1}{N}\sum_{i=1}^NR_{\text{MSE}i}\tag{1}$$

式中： I 为优化性能指数； $R_{\text{MSE}i}$ 为第 i 次迭代中验证集的均方误差 (MSE)； N 为迭代总次数。

b. 模型评估指标。采用平均绝对误差 (MAE)、MSE 和平均相对误差 (MAPE) 作为模型评估指标。

2 实例验证

2.1 研究区概况与数据来源

2.1.1 研究区概况

衡阳市位于湖南省中南部,处于 26°07′05″N ~ 27°28′24″N、110°32′16″E ~ 113°16′32″E,总面积 15 310 km²。衡阳市属中亚热带季风气候,多年平均气温为 19. 2℃,多年平均降水量为 1 300 mm。2023 年城市人口为 649. 7 万人,城镇化率达到 57%,万元 GDP 用水量为 72. 25 m³,万元工业增加值用水量为 34. 57 m³,水资源总量为 70. 3 亿 m³,用水总量为 30. 28 亿 m³。衡阳市灌溉面积达 288 610 hm²,农业用水量达 22. 14 亿 m³,占总用水量的 73. 12%,农业用水需求大,导致衡阳市水资源供需矛盾突出,春旱与秋旱时尤为明显。

2.1.2 数据来源

选取 GDP (x_1)、第一产业产值 (x_2)、第二产业产值 (x_3)、第三产业产值 (x_4)、万元 GDP 用水量 (x_5)、万元工业增加值用水量 (x_6)、灌溉面积 (x_7)、城镇化率 (x_8)、人口 (x_9)、年降水量 (x_{10})、年平均气

温 (x_{11})、水资源总量 (x_{12})、一产比例 (x_{13})、二产比例 (x_{14})、三产比例 (x_{15})、农业用水量 (x_{16})、工业用水量 (x_{17})、居民生活用水量 (x_{18})、城镇公共用水量 (x_{19})、生态用水量 (x_{20}) 等社会经济和水文气象指标作为衡量需水量的影响因子^[4, 9-10, 27],其中, $x_1 \sim x_4$ 和 x_{11} 来源于 2005—2023 年《湖南省统计年鉴》, x_5 、 x_6 、 x_{10} 、 x_{12} 和 $x_{16} \sim x_{20}$ 来源于 2005—2023 年《湖南省水资源公报》, $x_7 \sim x_9$ 来源于 2005—2023 年《衡阳市国民经济和社会发展统计公报》, $x_{13} \sim x_{15}$ 由间接计算得到。这些指标能够全面反映经济发展水平、水资源利用效率、社会结构变化、气候条件、水资源基础条件及用水结构对城市需水的多维度影响。此外,城市需水量 (y) 来源于 2005—2023 年《湖南省水资源公报》。

2.2 主要影响因子筛选

将 $x_1 \sim x_{20}$ 作为自变量,需水量作为因变量输入 RF 模型。使用网格搜索进行超参数优化,当决策树数量为 50、最大深度为 5、最小分割样本数为 2、最小叶子节点样本数为 4 时,模型在验证集上的 MSE 达到最小值 0. 932 亿 m³。根据各自变量与因变量的特征重要性排序,选取前 9 个指标作为主要影响因子,即: x_2 、 x_1 、 x_8 、 x_4 、 x_{17} 、 x_3 、 x_6 、 x_5 、 x_9 ,RF 模型筛选结果与皮尔逊相关系数法筛选结果一致。需水量及其主要影响因子如表 1 所示。

2.3 衡阳市需水预测模型构建

将 9 个主要影响因子作为自变量,需水量作为因变量,分别构建基于 BP 神经网络、LSTM、SVR、XGBoost 的需水预测模型。根据各模型定义与数据

表 1 需水量及其主要影响因子

Table 1 Water demand and its key influencing factors										
年份	x_2 /亿元	x_1 /亿元	x_8 /‰	x_4 /亿元	x_{17} /亿 m ³	x_3 /亿元	x_6 /m ³	x_5 /m ³	x_9 /万人	y /亿 m ³
2005	150. 5	590. 9	33. 8	216. 0	7. 91	224. 3	335. 0	580. 0	722. 6	34. 25
2006	158. 1	672. 1	37. 0	241. 9	8. 00	272. 1	236. 0	519. 0	726. 5	34. 88
2007	192. 4	820. 9	41. 2	299. 7	8. 02	328. 8	262. 0	422. 0	728. 9	34. 33
2008	235. 7	1 000. 1	42. 6	352. 0	7. 50	412. 4	199. 0	340. 0	731. 1	33. 11
2009	240. 5	1 168. 0	43. 2	427. 2	7. 73	500. 4	179. 0	284. 0	739. 8	33. 20
2010	264. 4	1 420. 3	44. 5	510. 2	8. 65	645. 7	154. 0	235. 0	714. 8	33. 34
2011	290. 7	1 734. 3	47. 0	600. 6	8. 73	842. 9	131. 0	206. 0	716. 6	32. 68
2012	322. 9	1 957. 7	47. 9	685. 3	8. 81	949. 5	105. 0	173. 0	719. 8	33. 05
2013	338. 4	2 169. 4	48. 1	791. 6	9. 50	1 039. 4	104. 0	157. 0	725. 0	33. 97
2014	359. 2	2 396. 6	48. 5	917. 4	7. 96	1 119. 9	81. 0	142. 0	730. 3	33. 91
2015	395. 8	2 601. 6	49. 2	1 044. 7	7. 67	1 161. 0	76. 0	125. 0	733. 8	32. 63
2016	430. 6	2 868. 6	51. 1	1 285. 8	7. 46	1 152. 2	69. 6	115. 9	728. 6	32. 53
2017	330. 4	2 921. 5	52. 5	1 480. 0	7. 09	1 111. 1	64. 1	103. 8	720. 5	31. 76
2018	337. 0	3 046. 0	53. 6	1 685. 4	8. 79	1 023. 6	81. 9	111. 3	724. 3	34. 26
2019	380. 1	3 372. 7	54. 9	1 901. 0	8. 00	1 091. 6	80. 0	100. 1	730. 1	32. 71
2020	441. 7	3 508. 5	54. 3	1 907. 5	5. 39	1 159. 3	64. 5	90. 1	664. 2	31. 61
2021	446. 0	3 840. 3	55. 2	2 093. 1	4. 45	1 301. 2	47. 0	83. 2	662. 1	31. 97
2022	474. 1	4 089. 7	55. 8	2 229. 2	3. 52	1 389. 3	34. 9	80. 1	657. 7	32. 77
2023	461. 5	4 190. 9	57. 0	2 343. 6	3. 37	1 385. 8	34. 6	72. 3	649. 7	30. 28

特征,确定各模型关键超参数的范围,如表 2 所示。

表 2 预测模型关键超参数范围设定

Table 2 Key hyperparameter range settings for forecast models			
预测模型	关键超参数	范围	类型
BP 神经网络	隐藏层神经元数	1~19	整数
	激活函数	ReLU/tanh/sigmoid/linear	函数
	学习率	0.0001~0.1	浮点数
LSTM	隐藏单元数	9~36	整数
	丢弃率	0~0.5	浮点数
	学习率	0.0001~0.1	浮点数
SVR	正则化参数	0.1~1000	浮点数
	不敏感损失函数参数	0.01~1	浮点数
	核函数	linear/poly/rbf/sigmoid	函数
	多项式核函数的次数	2~5	整数
XGBoost	树数量	100~1000	整数
	最大深度	6~12	整数
	学习率	0.01~0.3	浮点数
	样本子集比例	0.5~1.0	浮点数
	特征子集比例	0.5~1.0	浮点数
	最小分裂损失	0~5	浮点数
	最小叶子节点权重	1~10	浮点数

2.4 结合固定步长滚动交叉验证的超参数优化

选取 2005—2020 年时间序列数据为训练集, 2021—2023 年时间序列数据为测试集。在训练阶段,采用固定步长滚动交叉验证对模型进行迭代验证,设定初始训练窗口长度为 12 a,固定滑动步长与验证集长度均为 1 a。第一次迭代使用 2005—2016 年数据进行训练,2017 年数据进行验证;第二次迭代使用 2005—2017 年数据进行训练,2018 年数据进行验证,以此类推共进行 4 次迭代。为进一步提升预测模型的性能,采用 PSO 算法、BO 算法和 Optuna 分别对 4 种需水预测模型的关键超参数进行优化调整,优化目标设定为最小化优化性能指数。优化算法的参数配置如表 3 所示。

表 3 优化算法参数设定

Table 3 Parameter setting for optimization algorithms		
优化算法	参数	取值
PSO	粒子数量	30
	迭代次数	50
	c_1	0.5
	c_2	0.3
	w	0.9
BO	随机搜索次数	20
	优化迭代次数	100
Optuna	试验次数	100

注: c_1 、 c_2 分别为认知和社会学习因子, w 为惯性权重。

2.5 预测模型性能评估

经过超参数优化得到各优化算法下需水预测模型的最优超参数率定值,如表 4 所示。采用优化性能指数来评估优化算法的性能和模型的泛化能力,

利用训练集和测试集中需水量预测值的 MAE、MSE 和 MAPE 来评估不同模型的预测性能,结果如表 5 所示。

表 4 模型最优超参数率定结果

Table 4 Optimal hyperparameter calibration results of models				
预测模型	关键超参数	PSO 算法	BO 算法	Optuna
BP 神经网络	隐藏层神经元数	8	9	12
	激活函数	tanh	tanh	tanh
	学习率	0.0116	0.1000	0.0200
LSTM	隐藏单元数	27	20	29
	丢弃率	0.0012	0.2490	0.0019
	学习率	0.0091	0.0092	0.0072
SVR	正则化参数	733.7	0.100	0.624
	不敏感损失函数参数	0.493	0.010	0.533
	核函数类型	poly	linear	sigmoid
XGBoost	树数量	131	355	738
	最大深度	12	6	6
	学习率	0.224	0.300	0.255
	样本子集比例	0.640	1.00	0.502
	特征子集比例	0.985	1.00	0.901
	最小分裂损失	0.317	0.782	0.453
	最小叶子节点权重	3.98	2.48	2.35

对同一需水预测模型在不同优化算法下的预测效果进行评估,结果显示,PSO-BP 神经网络、Optuna-LSTM、PSO-SVR 和 PSO-XGBoost 模型表现最佳,其对应的优化性能指数分别为 0.408、0.701、0.688 和 0.428。PSO 算法和 Optuna 表现出更为稳健的全局搜索能力。相比之下,BO 算法的最优超参数搜索能力相对较弱,尤其在优化 XGBoost 时陷入了局部最优解,样本子集比例和特征子集比例均为 1,模型每棵树都使用全部的训练数据和特征进行训练,从而导致过拟合。PSO 算法通过模拟粒子群的集体行为,在全局搜索和局部开发之间保持良好的平衡,其粒子群体共同协作,能够快速找到较优解。Optuna 能够直接处理离散的整数超参数,使其相比于传统的贝叶斯优化算法更加高效和灵活。

优选的 PSO-BP 神经网络、Optuna-LSTM、PSO-SVR 和 PSO-XGBoost 模型在训练集中表现较好,而测试集中模型的表现更能反映其在实际应用中的预测能力。通过仿真预测得到最优超参数下 4 种预测模型对衡阳市 2005—2023 年需水量的预测结果,如图 1 所示。从图 1 和表 5 中可以看出,测试集中 PSO-BP 神经网络模型各项指标均优于其他模型,综合来看,PSO-BP 神经网络模型表现最佳。

2.6 滚动交叉验证效果评估

为评估滚动交叉验证的效果,对比其与 4 折交叉验证在结合超参数优化算法时的测试集评估指标。选取 2005—2020 年时间序列数据为训练集,

表 5 预测模型性能评估结果

Table 5 Evaluation results of performance of forecast models

预测模型	超参数优化算法	优化性能指数	训练集 (2005—2020 年)			测试集 (2021—2023 年)		
			MAE/亿 m ³	MSE/亿 m ³	MAPE/%	MAE/亿 m ³	MSE/亿 m ³	MAPE/%
BP 神经网络	PSO	0.408	0.207	0.068	0.617	0.440	0.214	1.394
	BO	0.485	0.312	0.074	0.825	0.820	0.731	2.594
	Optuna	0.419	0.124	0.031	0.372	0.653	0.484	2.061
LSTM	PSO	0.753	0.151	0.038	0.455	1.110	1.463	3.471
	BO	0.812	0.230	0.075	0.687	1.113	1.424	3.547
	Optuna	0.701	0.176	0.051	0.529	1.000	1.136	3.166
SVR	PSO	0.688	0.438	0.223	1.316	1.027	1.166	3.219
	BO	0.883	0.392	0.331	1.162	1.247	1.611	3.928
	Optuna	0.712	0.589	0.500	1.756	1.213	1.531	3.822
XGBoost	PSO	0.428	0.196	0.059	0.59	0.840	0.736	2.639
	BO	0.447	0.227	0.100	0.682	1.925	1.874	4.898
	Optuna	0.445	0.209	0.068	0.628	1.097	1.238	3.453

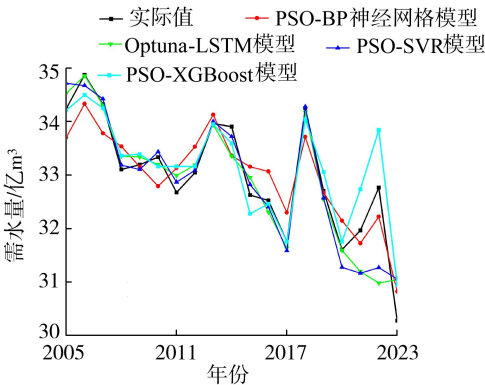


图 1 需水量预测结果

Fig. 1 Results of water demand forecast

2021—2023 年时间序列数据为测试集。训练阶段采用 4 折交叉验证策略,将 2005—2020 年数据按时间顺序等分为 4 个不重叠、连续的子集。在每一轮交叉验证中,依次选取其中 1 个子集作为验证集,其余 3 个子集作为训练集,整个过程重复 4 次,确保每个子集均作为一次验证集参与性能评估。采用 PSO 算法、BO 算法和 Optuna 分别对 4 种模型的关键超参数进行优化调整,优化目标设定为最小化优化性能指数,模型超参数范围与优化算法参数不变。表 6 为 4 折交叉验证效果评估结果。

从表 5 和表 6 可以看出,滚动交叉验证结合超参数优化表现出显著优势,模型预测结果在各项评估指标上均优于采用 4 折交叉验证的模型。其中 PSO-BP 神经网络模型表现最优,预测值的 MAE、MSE、MAPE 分别优化了 42.5%、63.4%、42.4%。滚动交叉验证通过多次迭代动态扩展训练数据,能够更充分地挖掘数据集的时间特征信息,特别是在样本量有限的情况下,该方法优势更为明显。此外,该方法能够显著增强模型的泛化能力,使其在多样化数据环境中保持较高的稳定性和预测精度。结合滚动交叉验证的超参数优化在需水预测实际应用

场景中表现出较强的适应性。

表 6 4 折交叉验证效果评估结果

Table 6 Evaluation results of performance of 4-fold cross-validation

预测模型	优化算法	MAE/亿 m ³	MSE/亿 m ³	MAPE/%
BP 神经网络	PSO	0.765	0.585	2.420
	BO	0.966	0.931	3.050
	Optuna	0.887	0.787	2.800
LSTM	PSO	1.360	1.850	4.290
	BO	1.220	1.490	3.850
	Optuna	1.235	1.525	3.900
SVR	PSO	1.129	1.276	3.570
	BO	1.376	1.895	4.350
	Optuna	1.246	1.553	3.940
XGBoost	PSO	0.890	0.792	2.810
	BO	1.361	1.852	4.300
	Optuna	1.176	1.381	3.710

2.7 模型预测结果验证

需水预测模型的核心是通过训练某区域现状年的数据优选一个具有良好泛化能力和预测精度的预测模型。后续实践中将主要影响因子的预测值输入具有最优超参数的模型中,以得到规划年的需水量预测值,而主要影响因子的预测值根据政策制定的要求或者不同的预测方法得到。在衡阳市的实例中,通过自回归积分滑动平均模型预测经济类指标和人口数,采用复杂函数回归趋势法拟合用水指标和城镇化率^[28-29],得到 2025 年的主要影响因子预测值。将这些预测值输入最优参数配置的 BP 神经网络模型,得到 2025 年需水量的预测值为 33.27 亿 m³,与《衡阳市节约用水“十四五”规划》中的用水总量控制指标 35.5 亿 m³ 具有较高的一致性,验证了该方法在城市需水预测中的适用性和实际应用价值。

3 普适性分析

需水预测方法的普适性主要体现在以下几个方

面:①能够根据不同区域的经济社会发展趋势及用水结构灵活构建影响因子指标体系,并通过筛选主要影响因子确保输入变量的相关性和代表性,减少冗余信息对模型性能的干扰,为模型提供更契合实际需求的输入数据;②支持多种机器学习模型的构建与优化,可根据数据特性灵活选择模型与超参数优化算法的组合方案,同时设定模型的关键参数范围,以确保优化过程在合理的空间内高效进行,这种设计有效增强了模型对不同数据集的适应能力;③结合滚动交叉验证的超参数优化算法充分捕捉了时间序列数据的动态特性,可以有效应对样本量有限的情况,通过多次迭代滚动更新训练集与验证集的划分方案,提升了模型对数据变化的适应性,显著提高了模型的泛化能力和预测精度。

4 结 论

a. 结合滚动交叉验证的超参数优化算法显著提高了需水预测模型的泛化能力和预测精度,实例验证中 PSO-BP 神经网络模型表现最佳,与结合 4 折交叉验证的超参数优化算法相比,模型预测值的 MAE、MSE、MAPE 分别优化了 42.5%、63.4%、42.4%,能够更好地满足真实应用场景中的需水预测需求。

b. 衡阳市 2025 年需水量的预测值为 33.27 亿 m³,与《衡阳市节约用水“十四五”规划》中的用水总量控制指标 35.5 亿 m³ 具有较高的一致性,验证了该方法在城市需水预测中的适用性和实际应用价值。

c. 该方法具有较强的普适性,可根据不同区域的经济社会发展趋势及用水结构灵活调整影响因子指标体系和模型组合,并且通过结合滚动交叉验证的超参数优化算法保证预测精度,适用于不同区域或行业多场景的需水预测。

参考文献:

[1] 吴迪,孙莺,夏管军,等. 生态文明思想下的盐城市水环境治理对策[J]. 水利经济,2023,41(2):63-67. (WU Di,SUN Ying,XIA Guanjun,et al. Countermeasures of water environment management in Yancheng City under the concept of ecological civilization [J]. Journal of Economics of Water Resources,2023,41(2):63-67. (in Chinese))

[2] 杨少康. 水资源需求预测的研究进展[J]. 农业与技术,2020,40(1):39-41. (YANG Shaokang. Research progress on water resources demand forecasting [J]. Agriculture and Technology,2020,40(1):39-41. (in Chinese))

[3] 郭晓茜,闫强,王高尚. 水资源需求预测研究方法概述 [J]. 中国矿业,2019,28(9):160-165. (GUO Xiaoqian,

YAN Qiang, WANG Gaoshang. Review of prediction methods on water resources demand [J]. China Mining Magazine,2019,28(9):160-165. (in Chinese))

[4] 杨登元,鞠茂森,唐德善. 基于改进 PCA-BP 神经网络模型的海宁市需水预测 [J]. 水电能源科学,2024,42(5):68-71. (YANG Dengyuan, JU Maosen, TANG Deshan. Water demand prediction in Haining City based on improved PCA-BP neural network model [J]. Water Resources and Power,2024,42(5):68-71. (in Chinese))

[5] 陈星,沈紫菡,许钦,等. 基于 CNN-LSTM-Attention 的月生活需水预测研究 [J]. 三峡大学学报(自然科学版), 2024,46(5):1-6. (CHEN Xing, SHEN Zihan, XU Qin, et al. Research on monthly water demand prediction based on CNN-LSTM-Attention [J]. Journal of China Three Gorges University (Natural Sciences), 2024, 46 (5): 1-6. (in Chinese))

[6] GUO Jun,SUN Hui,DU Baigang. Multivariable time series forecasting for urban water demand based on temporal convolutional network combining random forest feature selection and discrete wavelet transform [J]. Water Resources Management,2022,36(9):3385-3400.

[7] 李嘉欣,彭少明. 黄河流域用水现状及需水预测研究 [J]. 人民黄河,2024,46(7):66-71. (LI Jiaxin, PENG Shaoming. Study on the current situation of water use and water demand forecast in the Yellow River Basin [J]. Yellow River,2024,46(7):66-71. (in Chinese))

[8] 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 Jiaming, GUO Chunqing, LI Xinjian, 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. (in Chinese))

[9] 崔东文. 加权平均集成神经网络模型在城市需水预测中的应用 [J]. 水资源保护,2014,30(2):27-32. (CUI Dongwen. Application of weighted average integrated neural network model to urban water demand forecasting [J]. Water Resources Protection,2014,30(2):27-32. (in Chinese))

[10] 王盼,陆宝宏,张瀚文,等. 基于随机森林模型的需水预测模型及其应用 [J]. 水资源保护,2014,30(1):34-37. (WANG Pan, LU Baohong, ZHANG Hanwen, et al. Water demand prediction model based on random forests model and its application [J]. Water Resources Protection,2014, 30(1):34-37. (in Chinese))

[11] 单义明,杨侃. 基于灰色关联度分析的山西省 PSO-SVR 需水量预测模型 [J]. 水电能源科学,2021,39(2):18-21. (SHAN Yiming, YANG Kan. Forecasting model of PSO-SVR water demand in Shanxi Province based on grey correlation analysis [J]. Water Resources and Power, 2021,39(2):18-21. (in Chinese))

[12] 苏志伟,吴元梅,张丽娟,等. 基于 MEA-GA-BP 耦合模型的城市需水量预测 [J]. 水电能源科学,2022,40

- (11):31-34. (SU Zhiwei, WU Yuanmei, ZHANG Lijuan, et al. Urban water demand prediction based on MEA-GA-BP coupling model [J]. Water Resources and Power, 2022, 40(11):31-34. (in Chinese))
- [13] ZANFEI A, BRENTAN B M, MENAPACE A, et al. Graph convolutional recurrent neural networks for water demand forecasting[J]. Water Resources Research, 2022, 58(7): e2022WR032299.
- [14] 许月萍, 曾田力, 周欣磊, 等. 基于主要驱动因子筛选法和深度学习算法的浙江省动态需水量预测[J]. 水利水电科技进展, 2024, 44(2): 47-53. (XU Yueping, ZENG Tianli, ZHOU Xinlei, et al. Forecast of dynamic water demand in Zhejiang Province based on main driving factor screening method and deep learning algorithm [J]. Advances in Science and Technology of Water Resources, 2024, 44(2): 47-53. (in Chinese))
- [15] HU Piao, TONG Jun, WANG Jingcheng, et al. A hybrid model based on CNN and Bi-LSTM for urban water demand prediction [C]//2019 IEEE Congress on Evolutionary Computation (CEC). Wellington: IEEE, 2019:1088-1094.
- [16] 马于航, 索梅芹. 基于 GA-BP 神经网络与正态区间估计的需水预测: 以邯郸市为例[J]. 中国农村水利水电, 2022(3): 19-24. (MA Yuhang, SUO Meiqin. Research on water demand forecasting in Handan City based on GA-BP neural network and normal interval estimation[J]. China Rural Water and Hydropower, 2022(3): 19-24. (in Chinese))
- [17] DU Baigang, HUANG Shuo, GUO Jun, et al. Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks[J]. Applied Soft Computing, 2022, 122: 108875.
- [18] YAN Kun, YANG Minzhi. Water demand forecast model of least squares support vector machine based on particle swarm optimization [J]. MATEC Web of Conferences, 2018, 246: 01029.
- [19] WANG Xudong, XU Haojie, PAN Yanxia, et al. Forecasting ecological water demand of an arid oasis under a drying climate scenario based on deep learning methods [J]. Ecological Informatics, 2024, 82: 102721.
- [20] 郭敏丽, 刘天航, 毕二平, 等. 地下水位机器学习模型中的特征筛选及应用效果分析[J]. 水资源保护, 2025, 41(3): 179-186. (GUO Minli, LIU Tianhang, BI Erping, et al. Feature selection in machine learning models of groundwater level and its application effect analysis[J]. Water Resources Protection, 2025, 41(3): 179-186. (in Chinese))
- [21] 刘娣, 孙佳倩, 余钟波. 基于机器学习模型的多层土壤湿度反演[J]. 河海大学学报(自然科学版), 2024, 52(3): 7-14. (LIU Di, SUN Jiaqian, YU Zhongbo. Multi-layer soil moisture inversion based on machine learning models [J]. Journal of Hohai University (Natural Sciences), 2024, 52(3): 7-14. (in Chinese))
- [22] 石卓, 史东华, 姚成, 等. 基于径流模数的 LSTM 模型在无资料嵌套流域的应用[J]. 河海大学学报(自然科学版), 2024, 52(3): 51-57. (SHI Zhuo, SHI Donghua, YAO Cheng, et al. Application of runoff modulus-based LSTM in ungauged nested watersheds [J]. Journal of Hohai University (Natural Sciences), 2024, 52(3): 51-57. (in Chinese))
- [23] 佟大威, 杨传会, 余佳, 等. 基于 XGBoost-PSO 的混凝土重力坝体型多目标优化设计[J]. 河海大学学报(自然科学版), 2023, 51(3): 91-98. (TONG Dawei, YANG Chuanhui, YU Jia, et al. Multi-objective shape optimization of concrete gravity dam based on XGBoost-PSO [J]. Journal of Hohai University (Natural Sciences), 2023, 51(3): 91-98. (in Chinese))
- [24] CHEN Tianqi, GUESTRIN C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Association for Computing Machinery, 2016: 785-794.
- [25] WANG Xiaoyu, LUO Dongkun, ZHAO Xu, et al. Estimates of energy consumption in China using a self-adaptive multi-verse optimizer-based support vector machine with rolling cross-validation[J]. Energy, 2018, 152: 539-548.
- [26] AKIBA T, SANO S, YANASE T, et al. Optuna: a next-generation hyperparameter optimization framework [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: Association for Computing Machinery, 2019: 2623-2631.
- [27] 孔祥仟, 陈园, 刘博懿, 等. 基于主成分分析的神经网络在需水预测中的应用[J]. 水电能源科学, 2018, 36(4): 26-28. (KONG Xiangqian, CHEN Yuan, LIU Boyi, et al. Application of neural network based principal component analysis in prediction of water demand [J]. Water Resources and Power, 2018, 36(4): 26-28. (in Chinese))
- [28] 魏豪杉, 王红瑞, 郑鹏鑫, 等. 基于水资源空间均衡的“四水四定”调控模型构建[J]. 水资源保护, 2024, 40(3): 71-77. (WEI Haoshan, WANG Hongrui, JIA Pengxin, et al. Construction of “defining city, land, population, and industry based on water” regulation and control model based on spatial equilibrium of water resources [J]. Water Resources Protection, 2024, 40(3): 71-77. (in Chinese))
- [29] 魏豪杉, 王红瑞, 郑鹏鑫, 等. 基于水资源空间均衡的“四水四定”调控模型的应用[J]. 水资源保护, 2024, 40(5): 78-85. (WEI Haoshan, WANG Hongrui, JIA Pengxin, et al. Application of water resources spatial equilibrium-based “defining city, land, population, and industry based on water” regulation and control model [J]. Water Resources Protection, 2024, 40(5): 78-85. (in Chinese))

(收稿日期: 2024-09-30 编辑: 施业)