

# 基于 LightGBM 的水厂供水压力预测

耿为民<sup>1</sup> 颜军<sup>2</sup> 张典<sup>3</sup> 马平川<sup>4\*</sup> 阳国华<sup>5</sup>

<sup>1</sup>(上海城建职业学院市政与生态工程学院 上海 200438)

<sup>2</sup>(山东沃特兰德环境科技有限公司 山东 枣庄 277101)

<sup>3</sup>(上海柯林布瑞信息技术有限公司 上海 200233)

<sup>4</sup>(上海出版印刷高等专科学校信息与智能工程系 上海 200093)

<sup>5</sup>(上海市计算技术研究所有限公司 上海 200040)

**摘要** 针对城市供水管网调度问题,提出一种基于 LightGBM(Light Gradient Boosting Machine)的水厂供水压力预测模型。对压力监测点历史数据提取时间特征,并根据特征重要性对测压点排序,以特征权重筛选、特征权重与经验相结合两种方式选取控制点。以南方某城市供水系统为算例,结果表明采用特征权重分析、人工经验相结合选用控制点来预测,具有较高和稳定的预测精度。

**关键词** 供水系统 压力预测 特征权重 LightGBM

中图分类号 TU991

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.04.050

## FORECASTING OF WATER SUPPLY PRESSURE BASED ON LIGHTGBM

Geng Weimin<sup>1</sup> Yan Jun<sup>2</sup> Zhang Dian<sup>3</sup> Ma Pingchuan<sup>4\*</sup> Yang Guohua<sup>5</sup>

<sup>1</sup>(Municipal and Ecological Engineering School, Shanghai Urban Construction Vocational College, Shanghai 200438, China)

<sup>2</sup>(Shandong Water Land Co., Ltd., Zaozhuang 277101, Shandong, China)

<sup>3</sup>(Clinbrain Co., Ltd., Shanghai 200233, China)

<sup>4</sup>(Department of Information and Intelligent Engineering, Shanghai Publishing and Printing College, Shanghai 200093, China)

<sup>5</sup>(Shanghai Institute of Computing Technology Co., Ltd., Shanghai 200040, China)

**Abstract** Aimed at the scheduling problem of urban water distribution system, a water supply pressure prediction model based on LightGBM (Light Gradient Boosting Machine) is proposed. The time characteristics of the historical data on pressure monitoring points were extracted. The monitoring points were sorted according to the feature importance. The control points were selected in two ways: one was according to feature weight, and the other one was combined feature weight and experience. Taking a water supply system in southern China as a research case, the results show it has high and stable prediction accuracy that the control points are selected by combining feature weight analysis and scheduling experience.

**Keywords** Water distribution system Pressure forecast Feature weight LightGBM

## 0 引言

城市供水系统运行调度用以保证用水户的水量、水压、水质,实现安全可靠供水的同时,节能降耗,降低供水费用。供水系统调度涉及水厂、泵站、阀门,庞大

而复杂,现阶段我国供水企业仍以人工经验调度为主。

近年来,研究人员对供水调度运行实施了多方面的研究。基于供水管网水力模型提出了以取水、输水、制水和二级泵房提升成本形成目标函数的多厂协同供水优化调度模式<sup>[1]</sup>、城乡一体化联网的科学调度<sup>[2]</sup>、基于规则控制的水泵运行优化调度<sup>[3]</sup>。供水调度宏观

数学模型则是在历史运行数据的基础上,应用统计分析的方法建立的。对于城市供水系统一级调度,研究人员分别建立了以测压点压力 BP 神经网络模型代替管网水力平衡方程组<sup>[4]</sup>、以宏观模型为基础以最小费用为目标的一级优化模型<sup>[5]</sup>。随着地理信息系统、物联网技术在供水企业的广泛应用以及人工智能的发展,为基于大数据的供水管网工况分析提供了可能性,例如:一种高效的基于动态剪枝的供水管网优化调度算法<sup>[6]</sup>;以出厂压力、增压泵站出站压力、泵站水库液位为输出目标值的深度神经网络模型<sup>[7]</sup>;综合运用朴素贝叶斯等机器学习模型设计的智能供水调度优化算法<sup>[8]</sup>。另外,随着在线模拟系统在大中型城市逐步应用,技术人员建立了基于在线模型的供水管网优化调度系统<sup>[9]</sup>。

为了探索适用于大中型城市、切实可行的智能化供水调度模式,避免优化求解的时间效率问题,实现供水系统运行的实时决策,本研究以某大型城市为例,构建了基于机器学习 LightGBM 与实时监测数据的水厂供水压力预测模型。该模型未来与管网水力模拟、需水量预测相结合,输出各水厂的出厂流量和压力,水厂基于此边界条件进行出水泵房的开停和频率组合,使供水调度运行逐步向智能化转变。

## 1 水厂供水压力预测模型

供水企业通常在水厂出厂管、管网中关键和最不利压力点安装有压力监测设备,测压点压力值与出厂压力值之间存在一定的变化规律,因此通过提取时间特征并选用一定量的控制点来预测出厂压力具有可实施性。预测模型采用树模型中的 LightGBM。树模型属于机器学习算法,它是一种非参数的监督学习方法。通过树形结构表示决策过程,树的每个节点表示一个特征或属性,每个叶节点表示一个类别或数值,通过树状结构呈现出数据的决策规则。树模型算法的主要目标是通过学习输入特征和目标变量之间的关系,构建一棵最优决策树,使错误率或损失函数最小化<sup>[10]</sup>。LightGBM 是一种结合了梯度增强决策树 (Gradient Boosting Decision Tree) 和随机森林 (Random Forest) 的算法,它通过构建多个弱学习器 (决策树) 来提高预测性能。LightGBM 使用更少的特征片段,可以更快地训练模型,并且具有更好的泛化能力<sup>[11]</sup>。此外,LightGBM 能处理海量数据并且可以直接使用类别型特征,其高效性主要体现在处理多样本和多特征方面,常用两种方法实现:基于梯度的单边采样算法和互斥特征

捆绑算法<sup>[12]</sup>。

基于 LightGBM 与压力监测数据的水厂供水压力预测模型建设主要包括以下 4 个步骤:

**步骤 1 数据前处理。**分别将水厂出厂以及监测点压力数据根据类型按行进行整合,选取不同类型数据时间轴中的最小、最大值作为时间区间起始,时间间隔为 5 分钟,构造标准时间轴。再分别按照水厂出厂和监测点压力对每个类型通过标准时间轴为基准按列进行整合,整合后得到两份数据文件。对这两份数据文件进行数据清洗,按时间轴从小到大进行排序,压力数据小于 0.1 或者大于 1 的替换为空 (压力单位为 MPa),并按列进行向上的线性插值。

**步骤 2 特征提取。**除压力数据作为特征以外,供水运行与时间特征也存在紧密联系。根据时间数据提取时间特征:年、月、日、时、分、星期、非工作日、节日、干支纪日、时辰和节气,作为补充的时间特征维度。

**步骤 3 模型建立与评估。**分别构建管网监测点压力预测出厂压力、出厂压力预测管网监测点压力的 LightGBM 模型。两个模型均将步骤 2 中提取的时间特征加入到特征数据中作为模型特征,通过 5 折交叉验证的方式,每一折选取 80% 数据作为训练集,20% 数据作为测试集<sup>[10]</sup>。使用平均绝对误差 (MAE/MPa)、数据与模型的拟合程度 ( $R^2$ , 值越接近 1, 说明模型训练得越好)、均方根误差 (RMSE/MPa)、平均相对误差 (MRE) 4 种方式进行模型效果的评估,确保 5 次数据划分后的模型效果均衡,这 5 次的模型平均分数作为最终模型评估结果。误差计算公式如下:

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (3)$$

$$E_{MR} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (4)$$

式中: $y_i$ 是实测值; $\hat{y}_i$ 是预测值; $\bar{y}_i$ 是实测均值; $n$ 为样本个数。

**步骤 4 特征权重分析。**在机器学习建模过程中,特征的选择和组合对模型训练效果有较大影响,衡量特征的重要性成为建模的关键环节。在树模型中,特征重要性是指某个特征在分类或回归过程中对模型预测结果的影响程度。本研究将特征重要性作为特征权重分析的指标,筛选出与预测目标相关的特征。基于步骤 3 中训练好的模型,输出其特征重要性数据,根据平均特征重要性权重对特征进行排序。设置不同的特

征权重阈值来筛选监测点,分别使用大于等于该特征权重阈值的压力监测点作为特征构建模型,并比较不同特征权重阈值下模型的精度。

基于步骤 3 建立的模型,压力控制点的选择采用了 2 种方法进行对比:基于特征权重;特征权重与人工经验调度相结合,即将人工经验调度中重点关注的压力监测点与基于特征权重法筛选的控制点相融合(两者之间存在部分重合)。经验控制点的选择主要包括供水最不利点、供水边界点(水厂间的控制测压点有重复,共同控制点主要位于水厂供水边界,边界上的管道在水压、水质方面易出现问题,因此需要重点关注)。

## 2 算例分析

以南方某城市供水系统为例,该城市有水厂出厂压力监测点 10 个,管网压力监测点 124 个,均有连续 30 个月的历史监测数据,对异常数据进行了数据清洗。算例基于 LightGBM 算法,构建了监测点与出厂压力之间的关系模型。基于监测点压力预测所需的出厂供水压力,并进一步优化压力控制点的选取,使预测更加准确。所有预测模型均考虑了全部时间特征。

### 2.1 水厂供水压力预测模型结果

基于建模区域 30 个月的历史监测数据,建立了水厂供水压力预测模型,预测值与实测数据对比见图 1。误差区间分别为:MAE(0.001 1 ~ 0.003 5)、 $R^2$ (0.874 ~ 0.989 4)、RMSE(0.002 1 ~ 0.005)、MRE(0.003 5 ~ 0.012)。水厂出厂压力的预测是基于 124 个压力监测点,从预测结果中可看出,基于监测点压力可以较为准确地预测出对应的出厂压力,因此只需确定控制点压力值,即可基于模型输出水厂供水压力。在实际工作中,使用全部 124 个压力测点作为控制点不具备可操作性。因此在构建的模型基础上,使用各个监测点的特征权重进行筛选,最终仅保留部分作为控制点。

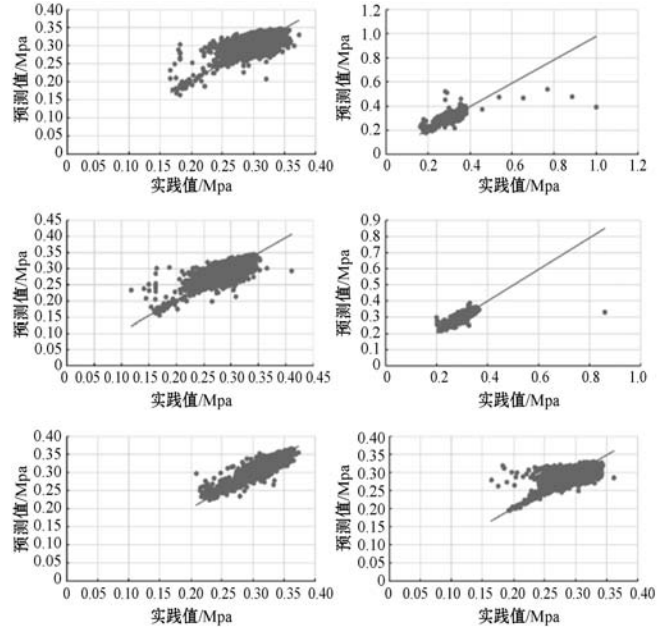
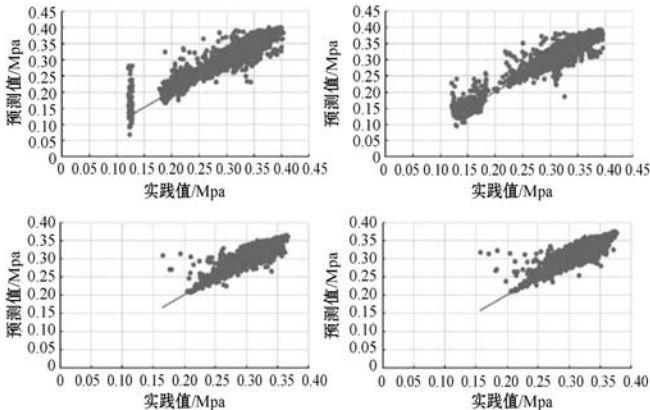
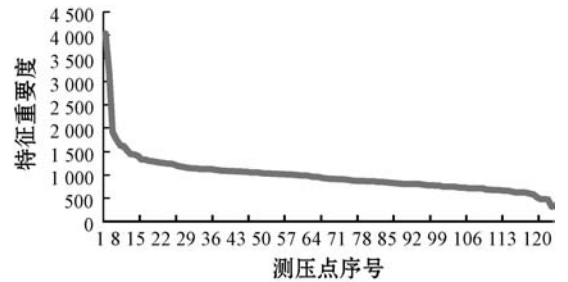


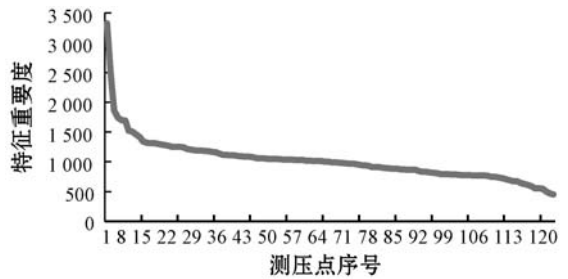
图 1 10 个出厂压力预测结果

### 2.2 基于特征权重分析的控制点筛选

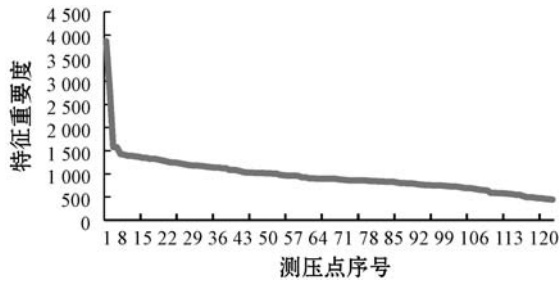
在建立 LightGBM 模型时,模型会为每个输入数据计算出一个特征重要度,此重要度作为特征重要性指标。图 2 列出了 4 个典型出厂压力点特征重要度值曲线,图中每个测压点序号对应的具体测点均不相同。可以看出,每个水厂均有 10 ~ 15 个较为相关、重要度值较高的测压点。因此,设置了不同的特征权重阈值来筛选监测点,分别使用大于等于该特征权重阈值的监测点作为特征构建模型,并对不同特征权重阈值下模型的精度进行比较(当阈值为 0 时,表示使用全部监测点)。以出厂压力 B 为例,水厂特征权重模型精度比较见表 1,随着筛选阈值的增加预测精度逐渐降低。



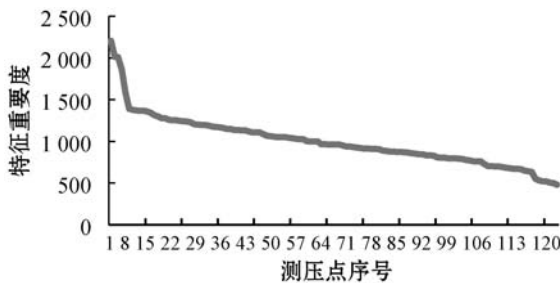
(a) 出厂压力 A



(b) 出厂压力 B



(c) 出厂压力 C



(d) 出厂压力 D

图 2 测压点重要度特征排序

表 1 出厂压力 B 特征权重模型精度比较

阈值	MAE/MPa	R <sup>2</sup>	RMSE/MPa	MRE
0	0.001 4	0.959 7	0.002 8	0.004 7
1 000	0.001 4	0.959 6	0.002 8	0.004 7
1 500	0.001 5	0.957 9	0.002 9	0.004 8
2 000	0.001 6	0.952 8	0.003 0	0.005 3
2 500	0.001 8	0.946 0	0.003 3	0.006 0
3 000	0.001 8	0.946 0	0.003 3	0.006 0
3 500	0.003 5	0.874 0	0.005 0	0.011 4
4 000	0.003 5	0.874 0	0.005 0	0.011 4

供水企业调度技术人员根据日常运行形成了一定经验,其中包含了基于压力控制点对水厂出厂压力的调控。因此在使用 LightGBM 算法预测出厂压力时,可以结合经验调度,将优化算法与经验控制相融合。对算例中区域进行了调研,选取了总计 20 个经验压力控制点,每个水厂有 5~10 个,水厂之间有相互重叠,重叠的监测点主要在供水范围边界(研究区域中已经建立了管网水力模型,经验点的选取参考了水力模拟结果)。将经验控制点与基于特征权重法筛选的控制点进行结合(两者存在一定程度的重合),融合构建新的模型。图 3(以出厂压力 B 为例)是仅使用基于特征权重法筛选的控制点构建的模型与使用经验、特征权重法筛选控制点融合构建的模型的效果对比。可以看出,仅使用测压点特征权重的模型随着筛选阈值的增加精度有所降低,这是由于随着筛选阈值的增加,选择的特征控制点逐渐减少,模型效果就会随之降低。而使用融合控制点构建的模型效果相对稳定,主要由于

经验控制点和基于特征权重法筛选的控制点存在一定程度的重合,随着筛选阈值的增加,基于特征权重法筛选的特征控制点逐渐变少,但是经验控制点不变,维持了模型精度的稳定。因此,将两种方法融合,即理论与实践相结合,既保证了精度,又使优化结果更好地具备实操性。技术人员基于构建模型,根据需要输入控制点的压力需求,即可输出相应的 10 个出厂供水压力值,未来控制点还可以根据实际运行工况进行增补和调换。

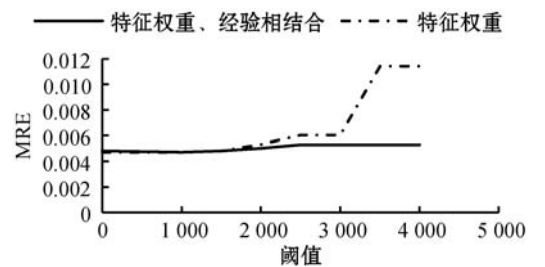
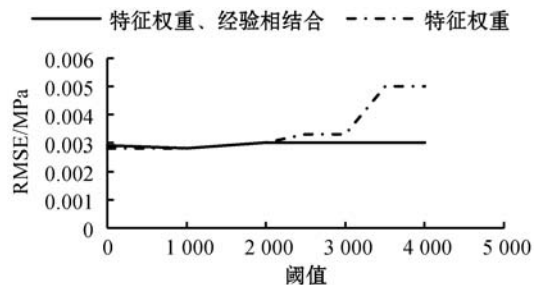
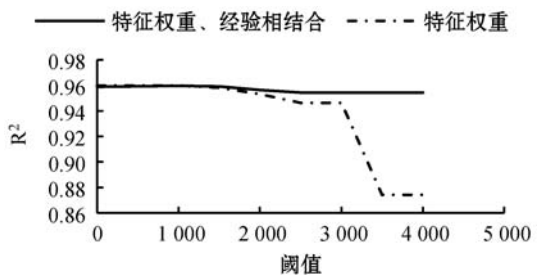
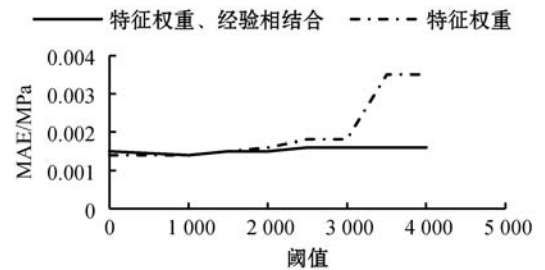


图 3 不同阈值情况下“特征权重、经验相结合”与“特征权重”筛选控制点预测出厂供水压力 B 精度对比

### 3 结 语

随着物联网技术在供水企业的逐步应用以及运行数据的不断积累,为实施供水智能调度提供了数据条件。本研究基于某供水系统压力监测点历史数据与机器学习中的 LightGBM 算法,构建了水厂供水压力预测

及时获得反馈,因此,接下来的研究重点是提升溯源用时。

## 参 考 文 献

- [1] Alkadri M F, Turrin M, Sariyildiz S. A computational workflow to analyse material properties and solar radiation of existing contexts from attribute information of point cloud data [J]. *Building and Environment*, 2019, 155(1):268-282.
- [2] 李维刚,杨威,赵云涛,等.融合大数据与冶金机理的热轧带钢力学性能预报模型[J].*钢铁研究学报*,2018,30(4):302-308.
- [3] 王剑波,李腾飞.基于大数据挖掘的高速公路货车能耗统计方法研究与应用[J].*公路*,2019,64(4):228-232.
- [4] Ai Y, Lu J, Zhang L. Location-allocation model of maritime emergency supplies repertory under joint of government and enterprise[J]. *Fifth International Conference on Transportation*, 2015,21(1):883-892.
- [5] 宋焕,王瑞梅,胡好.食品供应链中溯源信息共享的演化博弈分析[J].*哈尔滨工业大学学报(社会科学版)*,2017,19(2):111-118.
- [6] 任斌,朱昌锋,钟校.基于前景理论的应急物资需求动态调整模型[J].*中国安全科学学报*,2018,28(3):179-184.
- [7] 肖文,胡娟,周晓峰.基于 MapReduce 计算模型的并行关联规则挖掘算法研究综述[J].*计算机应用研究*,2018,35(1):13-23.
- [8] Wang X, Zhai Y, Lin Y, et al. Mining layered technological information in scientific papers: A semi-supervised method [J]. *Journal of Information Science*,2019,45(6):779-793.
- [9] 盛虎宜,刘长石,鲁若愚.震后初期应急物资短缺情景下的定位-路径问题研究[J].*运筹与管理*,2019,28(6):41-47.
- [10] 黎茵.有机农产品的食源性风险与可追溯体系[J].*北京交通大学学报(社会科学版)*,2017,16(4):46-53.
- [11] 孟燕萍,申慢慢.考虑灾后道路恢复情况下动态应急物资选址问题[J].*重庆交通大学学报(自然科学版)*,2019,38(1):89-96.
- [12] Hong S H, Lee S K, Yu J H. Automated management of green building material information using web crawling and ontology[J]. *Automation in Construction*, 2019, 102(7):230-244.
- [13] Christensen T H, Rommes E. Don't blame the youth: The social-institutional and material embeddedness of young people's energy-intensive use of information and communication technology [J]. *Energy Research & Social Science*, 2019, 49(1):82-90.
- [14] 杨建亮,侯汉平.基于自然灾害的大众应急物资快速投送问题研究[J].*北京交通大学学报(社会科学版)*,2017,16(4):72-79.
- [15] 宋英华,葛艳,杜丽敬,等.考虑车辆等待的应急物资调配

方案优化研究[J].*控制与决策*,2019,34(10):2229-2236.

## (上接第343页)

模型。该模型通过对压力监测点历史数据的学习,发现运行中管网监测点与出厂压力的变化规律,并融合时间特征、特征权重筛选的控制点、经验调度控制点来预测各水厂出厂压力。对比结果表明,采用特征权重筛选与经验调度相结合选择压力控制点的方法具有较高且稳定的预测精度。

水厂压力预测模型作为供水智能调度系统的重要组成部分,未来与管网水力模拟系统相组合,根据对未来工况的预判生成水厂调度指令,在保障水量、水压、水质的前提下,实现智能化供水调度。

## 参 考 文 献

- [1] 李树平,赵子威,周艳春,等.多厂协同供水优化调度分析[J].*给水排水*,2021,47(6):148-154.
- [2] 金晓静,王金辉,毛丽萍.基于管网模型的多水源供水科学调度系统实践与应用[J].*给水排水*,2020,46(S1):942-944.
- [3] Marchi A, Simpson A R, Lambert M F. Pump operation optimization using rule-based controls[J]. *Procedia Engineering*,2017,186:210-217.
- [4] 黄廷林,戴雪峰,张卉,等.改进 PSO 算法在多水源供水系统优化调度中的应用[J].*中国给水排水*,2013,29(23):64-68.
- [5] 王彤,张浩祥,徐杰,等.城市供水系统两级优化调度研究[J].*人民黄河*,2019,41(7):81-86.
- [6] 李俊禹,刘书明,吴雪,等.基于动态剪枝的城市供水管网优化调度算法[J].*环境工程*,2022,40(6):226-232,153.
- [7] 蒲政衡,赵平伟,冯偲懋,等.基于深度学习的供水管网实时智能调度研究[J].*给水排水*,2022,48(11):166-172.
- [8] 林峰,李旭,曾翰,等.某大型城市智能供水调度优化算法研究与模拟实践[J].*中国给水排水*,2023,39(9):109-115.
- [9] 赵美玲,张巧珍,朱俊,等.基于在线模型的供水管网优化调度系统设计[J].*中国给水排水*,2022,38(16):35-39.
- [10] 周志华.机器学习[M].北京:清华大学出版社,2016:73-91.
- [11] Meng Q, Ke G L, Wang T F, et al. A communication-efficient parallel algorithm for decision tree [C]//30th Conference on Neural Information Processing Systems,2016:1279-1287.
- [12] Ke G L, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree [C]//31st Conference on Neural Information Processing Systems,2017:3149-3157.