DOI:10.20040/j.cnki.1000-7709.2025.20241615

基于SVM-SARIMA-LSTM模型的城市用水量实时预测

李 轩[1,2],吴永强[1,2],王佳伟[1,2],杨伟超[1,2],张天洋[1,2]

(1.河北建筑工程学院市政与环境工程系,河北 张家口 075000;
2.河北省水质工程与水资源综合利用重点实验室,河北 张家口 075000)

摘要:为提高气象波动下城市用水量预测精度,通过季节性分解的趋势—季节性—残差程序(STL)将城市时用水量分解为趋势分量、季节性分量和残差分量3部分,使用季节性自回归移动平均模型(SARIMA)对季节性部分进行捕捉,利用支持向量机(SVM)提取趋势部分与气温、降水、风速、气压和相对湿度5个气象因素之间的关系,利用长短时记忆网络(LSTM)对波动性明显的残差部分进行关系捕捉,构建了SVM-SARIMA-LSTM用水量实时预测模型,并利用衡水市3个月时用水量数据和气象数据训练SVM-SARIMA-LSTM模型,以随后1周的实测数据作为验证集对模型预测性能进行评估。结果表明,SVM-SARIMA-LSTM模型的平均绝对百分比误差(E_{MAP})比SARIMA模型低4.502%,均方根误差(E_{RMSE})降低了39.084%,确定系数R^2提高了9.965%,最大绝对误差(E_{maxA})减小了55.946%,具有较好的应用价值。所建模型通过整合关键气象因素,准确地捕捉到城市用水量的季节性趋势及非季节性波动,展现了优良的泛化性。

关键词:SARIMA模型;支持向量机;长短时记忆神经网络;SVM-SARIMA-LSTM模型;STL分解程序;气象因素;用水量预测

中图分类号:TU991;TV214 **文献标志码**:A **文章编号**:1000-7709(2025)03-0036-04

1 引言

城市用水量受气象因素影响显著,气象波动时,供水厂的供水量与城市实际用水需求可能产生较大差异,从而导致管网压力波动,增加泄漏风险和水资源浪费。因此,在考虑气象波动等因素准确预测用水量显得尤为关键,可帮助供水厂合理调配资源,并维持管网稳定[1-2]。近年来,在短期用水量预测中广泛采用机器学习、深度学习和时间序列分析法[2]。其中,机器学习擅长捕捉数据的非线性关系,深度学习在处理复杂长序列数据方面表现优异,而传统时间序列法因其结构清晰、参数灵活而被广泛应用[3-5]。然而,单一模型存在局限性,如机器学习和深度学习对计算资源和数据质量要求较高,而传统时间序列法在应对长数据及外部因素变化时表现欠佳。为此,本文通过趋势—季节性—残差分解程序(STL)将城市时用水量分解为趋势、季节性和残差3个分量,利用季节性自回归移动平均模型(SARIMA)预测平稳的季节性分量,采用支持向量机(SVM)捕捉趋势分量与气温、降水、气压、湿度和风速等气象因素的关系,利用长短时记忆网络(LSTM)学习残差分量的波动特征,构建了SVM-SARIMA-LSTM用水量实时预测模型,并以2023年衡水市的时用水数据为例验证了所构建模型在提高城市用水量预测精度和鲁棒性方面的显著效果。

2 数据获取及模型构建

2.1 数据获取

本文使用2023年衡水市核心区小时级给水厂流量监测数据及小时级气象数据(包括气温、气压、风速、相对湿度和降水量)进行分析,选择4月10日~7月9日(连续3个月无异常值)的数据作为训练集,预测7月10~16日1周的时用水量。为避免抹平原始数据中的重要信息,无需额外的滤波操作预处理。衡水市核心区给水厂日设计流

收稿日期:2024-08-29,**修回日期**:2024-11-12

基金项目:河北省教育厅青年拔尖人才项目(BJK202305)

作者简介:李轩(1999-),男,硕士研究生,研究方向为市政管网优化及智慧运维,E-mail:lixuantju@163.com

通讯作者:吴永强(1969-),男,教授、硕导,研究方向为市政管网优化及智慧运维,E-mail:37162172@qq.com

中国知网 https://www.cnki.net

量为 10^5 t,服务 8 个用水分区,覆盖居民生活、工业企业、商业活动、公共设施、消防和环保等多种用水场景。

2.2　SVM-SARIMA-LSTM 模型构建

2.2.1　SARIMA 模型

城市用水量的变化与时间密切相关,通常表现为时间序列形态,并受多种因素影响。为提高对季节性时间序列的预测精度,BOX 和 JENKINS于 1970 年代提出了 ARIMA 模型,并在此基础上发展出引入季节性自回归(SAR)项和季节性移动平均(SMA)项的 SARIMA 模型[6]。SARIMA 模型能够有效地捕捉和预测具有季节性波动的时间序列数据,其数学表达式为:

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L^s)^D(1-L)^d Y_t = c+\varepsilon_t+\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}+\sum_{j=1}^{Q}\Theta_j L^{j\times s}\varepsilon_{t-j} \tag{1}$$

式中,p 为滞后阶数,代表了模型会考虑的之前观测值的数量;φ_i 为自回归系数;L 为滞后算子;s 为季节的长度;D 为季节性差分的阶数;d 为差分的阶数;Y_t 为第 t 个时间点的用水量;c 为常数项,反映用水量的平均值;q 为滞后阶数;θ_j 为移动平均系数;ε_{t-j} 为 j 个时间单位之前的误差;Q 为季节性移动平均项的滞后阶数;Θ_j 为季节性移动平均系数。

2.2.2　支持向量机(SVM)

SVM 是一种有效的监督学习方法,特别适合处理高维数据和非线性问题。在本文研究中,SVM 被用来捕捉城市用水量中趋势分量与关键气象因素之间的复杂关系。其模型表达式为:

$$y_t = \boldsymbol{w}[\boldsymbol{T}_t,\boldsymbol{P}_t,\boldsymbol{W}_{S_t},\boldsymbol{P}_{R_t},\boldsymbol{R}_{H_t}]+b+e_t \tag{2}$$

目标函数为:

$$\min_{w,b}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{N}(\delta_i+\delta_i^*) \tag{3}$$

约束条件为:

$$\begin{cases} y_i-(\boldsymbol{w}x_i+b)\leqslant\varepsilon+\delta_i \\ (\boldsymbol{w}x_i+b)-y_i\leqslant\varepsilon+\delta_i^* \\ \delta_i,\delta_i^*\geqslant 0 \end{cases} \tag{4}$$

式中,y_t 为在时间 t 的趋势分量;$\boldsymbol{w}$ 为待估计的权重向量;$\boldsymbol{T}_t$、$\boldsymbol{P}_t$、$\boldsymbol{W}_{S_t}$、$\boldsymbol{P}_{R_t}$、$\boldsymbol{R}_{H_t}$ 均为特征向量,其中 $\boldsymbol{T}_t$ 为气温,$\boldsymbol{P}_t$ 为降水,$\boldsymbol{W}_{S_t}$ 为风速,$\boldsymbol{P}_{R_t}$ 为气压,$\boldsymbol{R}_{H_t}$ 为相对湿度;b 为偏置项;e_t 为误差项;ε 为误差允许区间;C 为正则化参数;δ_i、δ_i^* 均为松弛变量。

2.2.3　长短时记忆神经网络(LSTM)

LSTM 是一种特殊类型的递归神经网络(RNN),专门用来解决传统 RNN 在处理长序列数据时容易出现的梯度消失和梯度爆炸问题。LSTM 的核心特征在于它的记忆单元和输入门、遗忘门和输出门 3 个控制信息流动的门。这些组件共同工作,使 LSTM 能够在长时间序列中有效地保存和利用过去的信息。其中,输入门(Input Gate)为:

$$i_t=\sigma\{\boldsymbol{W}_i[h_{t-1},x_t]+b_i\} \tag{5}$$

遗忘门(Forget Gate)为:

$$f_t=\sigma\{\boldsymbol{W}_f[h_{t-1},x_t]+b_f\} \tag{6}$$

记忆单元状态更新(Cell State Update):

$$C_t=f_tC_{t-1}+i_t\widetilde{C}_t \tag{7}$$

$$\widetilde{C}_t=\tanh\{\boldsymbol{W}_c[h_{t-1},x_t]+b_c\} \tag{8}$$

输出门(Output Gate)为:

$$o_t=\sigma\{\boldsymbol{W}_o[h_{t-1},x_t]+b_o\} \tag{9}$$

$$h_t=o_t\tanh(C_t) \tag{10}$$

式中,σ 为 sigmoid 激活函数,用于将值压缩到 0～1之间,表示信息的接收程度;$\boldsymbol{W}_i$ 为输入门的权重矩阵;h_{t-1} 为前一个时间步的隐藏状态通;x_t 为当前时间步的输入;b_i 为输入门的偏置项,输入门的作用是决定当前输入和过去的记忆中哪些信息是重要的,应该被保留;$\boldsymbol{W}_f$ 为遗忘门的权重矩阵;b_f 为遗忘门的偏置项,遗忘门的功能是决定哪部分旧信息需要被遗忘;C_t 为记忆单元状态;$\widetilde{C}_t$ 为当前时间步的候选记忆单元;$\boldsymbol{W}_c$ 为候选单元状态的权重矩阵;$\boldsymbol{W}_o$ 为输出门的权重矩阵;b_c 为偏置项;o_t 为输出门的输出;h_t 为前时间步的隐藏状态。

2.2.4　STL 程序

STL 是一种高效的时间序列分解程序,通过数值运算将时间序列拆分为趋势、季节性和残差 3 部分,不依赖于传统的数学模型。这种基于数值的处理方法使得 STL 在应对异常数据时表现出卓越的鲁棒性。分解后的各部分具有明确的物理意义,相较于将时间序列分解为本征模态函数(IMF),STL 提供了更高的可解释性[7]。这种分解策略增强了数据分析的灵活性和精确性,并允许独立研究趋势分量与气象因素之间的关系,有效避免了周期性波动的影响,从而提升用水量预测的精度和可靠性。

2.2.5　综合预测模型的实现

SVM-SARIMA-LSTM 模型构建步骤如下。

步骤 1　使用 STL(季节性和趋势分解)方法将城市用水量数据分解为趋势分量、季节性分量和残差分量。

步骤 2　设置 SVM 参数,以捕捉趋势分量与

中国知网　https://www.cnki.net

气温、降水量、风速、气压和相对湿度 5 个气象因素之间的关系。

步骤 3 通过观察季节性分量的季节性差分前后的自相关图和偏自相关图，进行 ADF（Augmented Dickey-Fuller）检测，并结合 AIC（Akaike 信息准则）和 BIC（贝叶斯信息准则）设置SARIMA模型的参数$(p,d,q)(P,D,Q,s)$。

步骤 4 使用 LSTM 捕捉残差分量中具有时间序列特性的非平稳波动。

步骤 5 综合各分量的预测结果，整合 SVM、SARIMA 和 LSTM 模型的预测结果，构建最终的 SVM-SARIMA-LSTM 模型。

2.3 模型评价指标

为全面评估 SVM-SARIMA-LSTM 模型在城市用水量预测中的效能，本文采用平均绝对百分比误差（E_{MAP}）、最大绝对误差（E_{maxA}）、均方根误差（E_{RMSE}）、确定系数（R^2）4 种评价指标，其中 E_{MAP} 衡量预测值与实际值之间的平均偏差，E_{MAP} 越低，预测效果越好；E_{maxA} 反映模型在极端情况下的最大偏差，E_{maxA} 越小，模型在不利情况下越稳健；E_{RMSE} 量化整体偏差，E_{RMSE} 越低，预测准确性越高；R^2 表示模型对实际值变动的解释力，R^2 越高，预测的可靠性越强。

3 数据分析与结果讨论

3.1 气象因素分析

在分析 2023 年衡水市核心区的日用水量时，发现用水量除了呈现夏季增加、冬季减少的季节性趋势外，还存在显著波动，特别在 4～8 月期间更为频繁。用水量与气象因素（如气温和降水）显示出用水量波动与天气变化的紧密关联。具体表现为大部分时段用水量随气温升高而增加，而降水的出现通常使用水量和气温同步达到极小值，增强了用水量的波动性。然而，7～8 月期间由于降水频繁且气温较高，导致气温与用水量的相关性有所减弱。这表明气象因素对城市用水量有一定的影响。

进一步，采用斯皮尔曼秩相关系数分析用水量与气象因素的关系，见图 1。由图 1 可知，用水量与气压、湿度和降水量的相关系数分别为－0.38、－0.58、－0.44，呈中等负相关，与气温的相关系数为 0.45，呈中等正相关，而与风速的相关性较弱。虽然这些相关性在统计学上未达到高度显著的水平，但考虑到城市人口规模的相对固定性和日常用水需求的基本稳定性，气象波动对日常用水量的变化仍有重要影响。

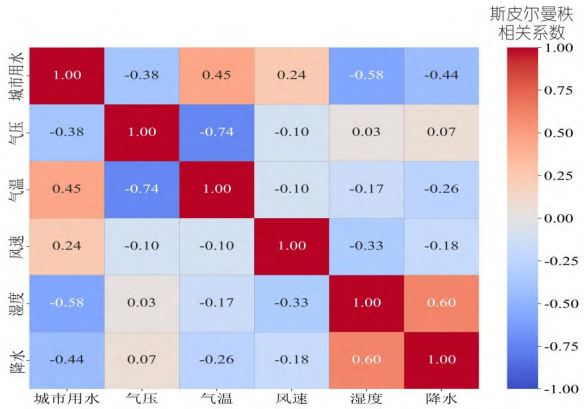

图 1 斯皮尔曼秩相关系数图

Fig.1 Spearman rank correlation coefficient chart

3.2 时用水量分析

现有研究表明[8]，日用水量受日平均气温和降水量的影响已得到广泛认可。然而，在研究时用水量与气温、气压之间的关系时面临一个挑战，即时用水量与气温、气压均存在周期性变化，这会干扰 SVM 模型对变量间真实关系的捕捉。图 2 为气温、气压和用水量差异较大的 4 d 变化情况。由图 2 可看出，不同日期的用水量模式相对稳定，气温和气压也呈现出周期性变化。在不同的时段，用水量与气温、气压可能表现出同向或异向变化，而 SVM 会对这些周期性趋势进行平均处理，导致在预测用水量时引入干扰。

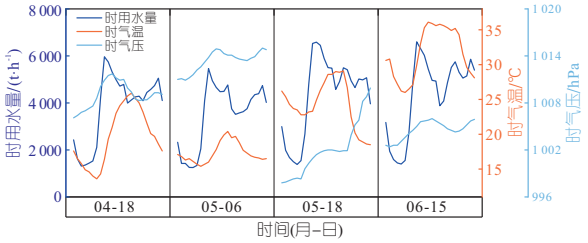

图 2 时用水量与气温气压图

Fig.2 Chart of hourly water consumption, temperature, and air pressure

为解决这一问题，本文采用 STL 分解方法对 4 月 10 日～7 月 9 日的连续数据进行了分析，将时用水量分解为趋势分量、季节性分量和残差分量，并绘制了相应的折线图，见图 3。由图 3 可知，趋势分量与用水量的整体变化趋势基本一致，说明气温、气压等气象因素对用水量的影响主要

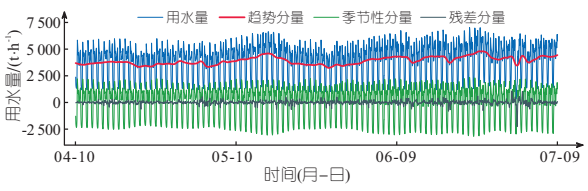

图 3 STL 用水量分解图

Fig.3 STL decomposition of water consumption

中国知网 https://www.cnki.net

体现在趋势分量上;季节性分量表现为稳定的周期性变化,反映了日常用水的规律性;而残差分量数值较小,呈现出不规则波动,主要反映短期随机变化。通过专注于趋势分量,可更清晰地评估气象因素对时用水量的长期变化影响[9]。

3.3　模型性能比较

根据水务系统的监测频率,本文设置 1 h 的预测步长,利用 3 个月的小时级用水量与气象数据构建了训练集(共 2 184 条数据)。随后使用 1 周的气象数据进行用水量预测,评估 SARIMA 和 SVM-SARIMA-LSTM 模型的性能,并与单独使用的 SVM 和 LSTM 模型进行对比,以验证预测结果的准确性。根据对季节性分量的分析,SARIMA 模型参数设置为$(p,d,q)(P,D,Q,s)=(1,0,1)(1,1,1,24)$。表 1 为各模型的评估结果。

表 1　预测误差评估结果

Tab.1　Evaluation results of prediction errors

模型	E_{MAP}/%	E_{RMSE}	R^2	E_{maxA}
SVM	34.915	1 176.361	0.397	2 457.787
LSTM	10.742	495.044	0.890	1 350.342
SRAIMA	11.281	552.040	0.863	2 555.376
SVM-SARIMA-LSTM	6.779	336.281	0.949	1 125.742

由表 1 可知,SVM 模型的 E_{MAP} 和 E_{RMSE} 较高,R^2 值仅为 0.397,表明其对数据的拟合能力较差,且 E_{maxA} 值较大,显示出在极端情况下误差显著。相比之下,LSTM 模型的 E_{MAP} 和 E_{RMSE} 较低,R^2 值达 0.890,E_{maxA} 值较小,显示出其在捕捉时间序列中长期依赖关系方面的优势。SARIMA模型的 E_{MAP} 和 E_{RMSE} 略高于 LSTM,R^2 值为 0.863,但 E_{maxA} 值较高,可能因其在处理随机波动时不及 LSTM。城市用水量具有周期性、趋势性和随机波动特性,单一模型难以全面捕捉。SVM-SARIMA-LSTM 组合模型融合了各模型的优势,显著提升了预测精确度和鲁棒性。与 SARIMA 相比,组合模型的 E_{MAP} 降低了 4.502%,E_{RMSE} 减少了 39.084%,R^2 增加了 9.965%,E_{maxA} 下降了 55.946%,显著减少了极端情况下的误差。这表明组合模型能够更全面地捕捉用水量的复杂特性,从而提高预测的准确性和可靠性。

图 4 为 SARIMA 和 SVM-SARIMA-LSTM 模型的预测值与实际用水量的对比。由图 4 可知,SVM-SARIMA-LSTM 模型在预测最高和最低用水量时表现更为贴合,特别是在 7 月 12、13 日衡水市经历持续降水事件,准确反映了气温骤降对用水量的影响,而 SARIMA 模型则高估了用水量。这进一步验证了组合模型在极端天气情

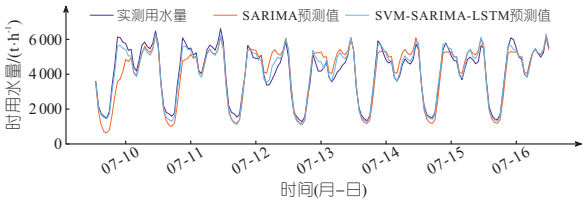

图 4　预测值与实测用水量对比图

Fig.4　Comparison of predicted and measured water consumption

况下的适应性和预测精度。分析其原因在于 SARIMA模型主要依赖于历史数据进行预测,在应对非周期性或突发性因素引起的波动时表现较弱。而 SVM-SARIMA-LSTM 通过子模型的协同作用,提升了对复杂时间序列的捕捉能力。SARIMA 分析季节性波动,SVM 揭示趋势分量与气象因素的复杂关系,而 LSTM 的门控机制则强化了对不稳定因素的响应。各技术的综合应用大幅提高了组合模型的预测精度和可靠性,使其更能适应城市用水量的多变性和复杂性。

4　结论

a. 本文结合关键气象因素构建了 SVM-SARIMA-LSTM 用水量预测模型,相较于单一模型,其展现出更高的预测精度和适应性,显著提升了预测性能,为城市供水系统的高效管理提供了有效工具。

b. 通过斯皮尔曼秩相关性分析和 STL 分解,证明了气象因素对城市用水量波动的显著影响,特别是趋势分量主要受气象因素的驱动。

参考文献:

[1]　陆维佳,朱建文,叶圣炯,等. 基于多因素长短时神经网络的日用水量预测方法研究[J]. 给水排水,2020,56(1):125-129.

[2]　KAVYA M, MATHEW A, SHEKAR P R, et al. Short term water demand forecast modelling using artificial intelligence for smart water management [J]. Sustainable cities and society, 2023, 95: 104610.

[3]　高学平,陈玲玲,刘殷竹,等. 基于 PCA-RBF 神经网络模型的城市用水量预测[J]. 水利水电技术,2017,48(7):1-6.

[4]　曹敬椿,卢敏. 基于混合策略改进 SSA-SVM 的工业用水量预测[J]. 水电能源科学,2023,41(9):28-31.

[5]　王健,刘丽,查淳膺,等. 基于多层长短期记忆神经网络的用水量预测[J]. 水电能源科学,2023,41(12):24-27.

(下转第 6 页)

中国知网　https://www.cnki.net

Adaptive Cross-fusion Model Based on Variational Modal Decomposition and Its Application to Monthly Runoff Prediction

SUN Yu-hui[1], WANG Qing-jie[2], YUE Chun-fang[3]

(1. Urumqi Water Industry Investment and Development Group Co., Ltd., Urumqi 830002, China;
2. Sichuan Water Development Investigation, Design & Research Co., Ltd., Chengdu 610065, China;
3. College of Water Conservancy and Civil Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: "Decomposition-integration" runoff prediction model is one of the mainstream approaches to improve prediction accuracy in existing studies. There are three main approaches of decomposition-integration modeling: hindcast experiment (HE), forecast experiment (FE) and adaptive forecast experiment (AFE). Previous studies have mainly focused on the improvement of HE, with little research on the practicality of various modeling approaches. This paper analyzes the characteristics of various modeling approaches. Various decomposition-integrated prediction models are constructed using three benchmark prediction models (back propagation neural network(BP), support vector machine(SVM), and random forest(RF) and two decomposition algorithms (variational modal decomposition(VMD) and complete ensemble empirical modal decomposition(CEEMD)), respectively. Finally, two typical inland rivers in the Tianshan mountain system are used as examples to verify the prediction performance of various models. In addition, the prediction capability of the cross-fusion of AFE and the benchmark model is also explored. The simulation results show that the HE model is built with test data information in advance, which is not consistent with the prediction reality; However, the prediction accuracy of the FE model is extremely low due to the severe influence of the decomposition endpoint effect; The AEF is consistent with the actual time-by-time observation-rolling decomposition-real-time modeling prediction, and the AFE model based on VMD and CEEMD has a high prediction accuracy for runoff maxima. Among the AFE-type models, VMD is more adaptable and can achieve high accuracy of monthly runoff forecasting in flood season. The adaptive cross-fusion model based on VMD can achieve a prediction accuracy comparable to or even higher than that of HE model, which has practical significance for the improvement of runoff prediction accuracy.

Key words: streamflow forecasting; endpoint effects; variational modal decomposition; posterior experiment; forecast experiment; adaptive forecast experiment

(上接第 39 页)

[6] GEORGE E P BOX, GWILYM M JENKINS, GREGORY C REINSEL. Time series analysis forecasting and control[M]. Hoboken: Wiley, 2016.

[7] 王晨, 李又轩, 吴其琦, 等. 基于 SVM-STL-LSTM 的区域短期电力负荷预测研究[J]. 水电能源科学, 2024, 42(4): 215-218.

[8] 曹文静, 孙傅, 刘益宏, 等. 极端高温事件对城市用水量和供水管网系统的影响[J]. 气候变化研究进展, 2018, 14(5): 485-494.

[9] 李元梦, 李登华, 丁勇. 基于 FFT-STL 的大坝监测数据时效分量提取与分析算法[J]. 水电能源科学, 2024, 42 (7): 129-132.

Real-time Urban Water Demand Forecasting Based on SVM-SARIMA-LSTM Model

LI Xuan[1,2], WU Yong-qiang[1,2], WANG Jia-wei [1,2], YANG Wei-chao[1,2], ZHANG Tian-yang[1,2]

(1. Department of Municipal and Environmental Engineering, Hebei University of Architecture, Zhangjiakou 075000, China; 2. Key Laboratory of Water Quality Engineering and Comprehensive Utilization of Water Resources, Zhangjiakou 075000, China)

Abstract: In order to improve the forecasting accuracy of urban water consumption under meteorological fluctuation, seasonal-trend decomposition based on loess (STL) was used to decompose urban hourly water consumption into three components: trend, seasonality, and residuals. The SARIMA was used to capture the seasonal component. The support vector machines (SVM) was used to extract the relationship between the trend component and five meteorological factors—temperature, precipitation, wind speed, atmospheric pressure, and relative humidity. Additionally, long short-term memory networks (LSTM) is employed to capture relationship of residuals with significant volatility. The real-time urban water demand forecasting model based on SVM-SARIMA-LSTM was established. The model was trained using three months of water consumption and meteorological data in Hengshui City, and assessed the model's forecasting performance using the actual data of the subsequent week as the validation set. The results indicate that the mean absolute percentage error (E_{MAP}) of the SVM-SARIMA-LSTM model is 4.502% lower than that of the SARIMA model, and the root mean square error (E_{RRMS}) is reduced by 39.084%; The coefficient of determination R^2 is raised by 9.965%, and the maximum absolute error (E_{maxA}) is reduced by 55.946%, which has better application value. By integrating the key meteorological factors, the model accurately captures both seasonal trends and non-seasonal fluctuations in urban water demand, demonstrating good generalization.

Key words: SARIMA model; support vector machine; long short-term memory neural network; SVM-SARIMA-LSTM model; STL decomposition; meteorological factors; water demand forecasting

中国知网 https://www.cnki.net