

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20221908

# 改进 EEMD-WOA-SRU 模型在用水量预测中的应用

刘 扬<sup>a,b</sup>, 杜帅兵<sup>b</sup>

(华北水利水电大学 a. 黄河流域水资源高效利用省部共建协同创新中心; b. 信息工程学院, 河南 郑州 450046)

**摘要:** 针对用水量信号表现出的强随机性和非平稳性状态, 用水量预测存在的精度低、可信度差等问题, 提出了基于改进 EEMD-WOA-SRU 的混合用水量预测模型。首先采用长短期记忆网络(LSTM)预测法抑制集合经验模态分解(EEMD)的端点效应得到改进后的本征模态分量(IMF), 然后使用鲸鱼算法(WOA)优化简单循环单元(SRU)并预测各分量, 最后累加得到最终的预测结果。试验结果表明, EEMD 的分解误差平均降低 0.94%, 相较于 SRU, EEMD-WOA-SRU 模型预测的平均绝对误差降低 45.42%, 均方根误差降低 50.43%, 可信度提高 52.38%。研究结果可为水资源决策提供依据。

**关键词:** 用水量预测; 集合经验模态分解; 鲸鱼优化算法; 端点效应; 简单循环单元

**中图分类号:** TV124      **文献标志码:** A      **文章编号:** 1000-7709(2023)12-0032-04

## 1 引言

水资源是承载社会经济生活发展的关键因素, 准确地预测用水量, 对于合理管理水资源, 实现区域可持续发展具有重要意义<sup>[1]</sup>。近年来, 以 RNN 为基础的循环神经网络在时间序列建模中表现出了优异的性能, 在水文预测中应用广泛<sup>[2]</sup>, 简单循环单元(SRU)是改进的循环神经网络, 具有更高的运算效率。环境变化和人类活动的影响导致了区域用水量信号呈不确定性、非平稳状态<sup>[3]</sup>, 直接使用单一模型进行预测存在精度低和可信度差等问题, 因此引入集合经验模态分解(EEMD)方法对用水量数据进行分解, 得到一系列较为稳定的分量信号, 然而 EEMD 存在端点效应, 导致信号分解过程存在一定的误差<sup>[4]</sup>。鉴此, 本文通过使用 LSTM 预测法抑制 EEMD 的端点效应, 以降低 EEMD 的分解误差, 再使用鲸鱼优化算法(WOA)优化 SRU, 构建了改进的 EEMD-WOA-SRU 混合用水量预测模型, 并将该模型应用于人民胜利渠的日用水量预测研究中, 验证了所构建模型的预测精度和可信度。

## 2 基于改进 EEMD-WOA-SRU 模型的用水量预测方法

### 2.1 改进集合经验模态分解

EEMD 方法于信号分解前在序列中加入了高斯白噪声, 利用噪声的均匀分布特性解决了模态混叠问题<sup>[5]</sup>, 其分解步骤如下。

**步骤 1** 将白噪声信号  $s(t)$  添加至原始用水量信号  $x(t)$  中, 得到新信号  $x'(t)$  并对其进行 EMD 分解, 得到  $n$  个本征模态分量(IMF):

$$x'(t) = \sum_{j=1}^n h_j(t) + r_n(t) \quad (1)$$

式中,  $h_j(t)$  为第  $j$  个分解后得到的 IMF 分量,  $j=1, 2, \dots, n$ ;  $r_n(t)$  为残差分量。

**步骤 2** 重复步骤 1, 每次为待分解的信号加入不同幅值的白噪声:

$$x'(t) = \sum_{j=1}^n h_{i,j}(t) + r_{i,n}(t) \quad (2)$$

式中,  $h_{i,j}(t)$  为第  $i$  次分解后得到的第  $j$  个 IMF 分量,  $i=1, 2, \dots, m$ ;  $r_{i,n}(t)$  为第  $i$  次分解后得到的残差分量。

**步骤 3** 对  $h_{i,j}(t)$  求平均, 得到最后的 IMF 分量  $C_j(t)$ :

$$C_j(t) = \frac{1}{m} \sum_{i=1}^m h_{i,j}(t) \quad (3)$$

**步骤 4** 各 IMF 分量与残差分量重构即可得到原始信号:

$$x(t) = \sum_{j=1}^n C_j(t) + r(t) \quad (4)$$

收稿日期: 2022-09-14, 修回日期: 2022-10-17

作者简介: 刘扬(1979-), 女, 博士, 副教授, 研究方向为深度学习、数据挖掘、水利大数据分析等, E-mail: ly\_research@126.com

式中,  $r(t)$  为经 EEMD 分解后的残差分量。

EEMD 在分解过程中通过信号的局部极值来获得包络函数, 由于无法确定序列的端点处是否为极值点, 直接对序列进行分解会出现端点效应。基于此, 通过数据预测法来抑制 EEMD 的端点效应, 具体步骤如下。

**步骤 1** 在进行 EEMD 分解前, 使用 LSTM 作为预测模型, 将用水量数据以“滑动窗口法”输入 LSTM 中, 来预测序列两端的数据。

**步骤 2** 在预测结果中找出极大值和极小值点(左右两端各找出一个极大值和一个极小值), 并与原始序列进行拼接。

**步骤 3** 对拼接后的序列进行 EEMD 分解, 结束分解后截断预测部分对应的分量数据, 获得改进的 EEMD 分解结果。

## 2.2 鲸鱼优化算法

鲸鱼优化算法(WOA)通过模拟鲸鱼在捕食过程中的上升螺旋气泡法进行寻优计算, 具有较快的寻优能力和收敛速度<sup>[6]</sup>。算法过程主要有包围猎物、螺旋气泡网攻击和搜寻猎物。

(1) 包围猎物。鲸鱼确定目标位置, 开始对猎物进行包围捕食。在搜索域内, 鲸鱼通过不断更新位置尽可能地靠近目标猎物, 其位置更新公式为:

$$D = |CX^*(t) - X(t)|$$

$$X(t+1) = X^*(t) - AD \quad (5)$$

式中,  $D$  为鲸鱼当前位置与最优位置的距离;  $C$  为扰动因子,  $C \in [0, 1]$ ;  $X^*(t)$  为鲸鱼当前最优位置;  $X(t)$  为鲸鱼当前位置;  $t$  为当前迭代次数;  $A$  为收敛因子,  $|A| \leq 1$ 。

(2) 螺旋气泡网攻击。鲸鱼在捕食过程中使用螺旋气泡网进行攻击猎物, 对猎物进行收缩环绕包围的同时, 以相同概率  $p$  进行螺旋式的位置更新:

$$X(t+1) = \begin{cases} X(t) - AD & p < 0.5 \\ D' e^{bl} \cos 2\pi l + X(t) & p \geq 0.5 \end{cases} \quad (6)$$

其中  $D' = |X^*(t) - X(t)|$

式中,  $D'$  为包围步长;  $b$  为控制螺旋形状的常量;  $l, p$  均为随机数,  $l \in [-1, 1], p \in [0, 1]$ 。

(3) 搜寻猎物。鲸鱼在捕食过程中, 根据向量  $A$  的变化随机寻找目标。为增强鲸鱼的全局搜索能力, 设定当  $|A| \geq 1$  时, 鲸鱼扩大搜寻范围, 进行随机搜捕, 以保证获得全局最优目标, 位置更新公式为:

$$\begin{cases} D = |CX_{\text{rand}}(t) - X(t)| \\ X(t+1) = X_{\text{rand}}(t) - AD \end{cases} \quad (7)$$

式中,  $X_{\text{rand}}(t)$  为从鲸鱼群体中随机选取的个体位

置向量。

## 2.3 简单循环单元

SRU 是一种轻量级的深度神经网络<sup>[7]</sup>, 相较于其他循环神经网络, SRU 在计算时不依赖于上一步隐藏层的输出, 而仅依赖于当前的输入数据, 因此能并行计算, 具有较高的运算效率。其次, 在 SRU 中融入了残差网络结构, 能有效避免深层网络的梯度消失。SRU 的神经元计算过程为:

$$\tilde{x}_t = Wx_t \quad (8)$$

$$f_t = \sigma(W_f \tilde{x}_t + b_f) \quad (9)$$

$$c_t = f_t \odot c_{t-1} + (1 - f_t) \odot \tilde{x}_t \quad (10)$$

$$r_t = \sigma(W_r \tilde{x}_t + b_r) \quad (11)$$

$$h_t = h_t \odot g(c_t) + (1 - r_t) \odot x_t \quad (12)$$

式中,  $x_t$  为当前的输入值;  $W$  为待训练的神经元状态参数;  $\sigma$  为 sigmoid 激活函数;  $f_t$  为遗忘门, 用来丢弃过去的无效信息;  $W_f$  为待训练的遗忘门状态参数;  $b_f$  为遗忘门偏置;  $c_{t-1}, c_t$  分别为前一个神经元状态的输入、当前神经元状态的输出;  $\odot$  为点乘符号;  $r_t$  为重置门, 用来控制前一个神经元状态输入的信息;  $h_t$  为神经元隐藏层的输出;  $g$  为  $h_t$  的状态激活函数。

SRU 的神经网络结构见图 1。

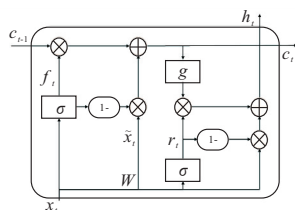


图 1 SRU 的神经网络结构

Fig. 1 Neuronal structure of SRU

## 2.4 EEMD-WOA-SRU 用水量预测模型流程

首先对用水量数据进行预处理, 并使用 LSTM 预测法抑制 EEMD 的端点效应, 得到一系列的信号分量并输入到 WOA-SRU 中, 累加各分量的预测值, 从而得到最终的用水量预测结果, 模型构建流程见图 2。

## 2.5 评估指标

选用分解误差  $e$  来作为抑制 EEMD 端点效应的评价指标, 其公式为:

$$e_i = \frac{1}{n} \text{sqrt} \left( \sum_{t=1}^n [x(t) - \text{imf}_i(t)]^2 \right) \quad (13)$$

式中, sqrt 为平方根;  $x(t)$  为原始用水量信号;  $\text{imf}_i$  为经 EEMD 分解后的第  $i$  个本征模态分量。  $e$  越小, 表示端点效应抑制的效果越好。

选用平均绝对误差 ( $M_{\text{MAE}}$ )、均方根误差 ( $R_{\text{RMSE}}$ )、相对误差 ( $R_{\text{RE}}$ )、纳什效率系数 ( $N_{\text{NSE}}$ ) 来作为评判模型预测效果的评价指标。其中,

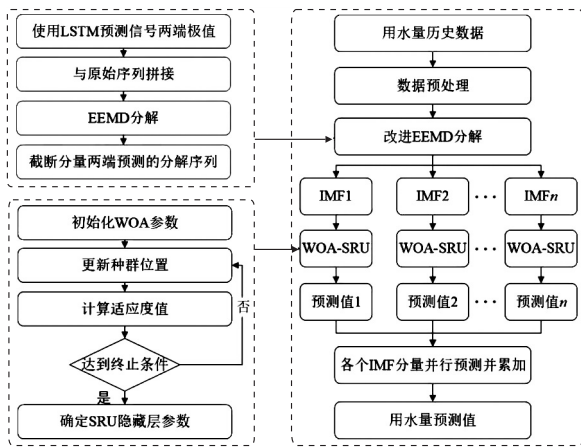


图 2 EEMD-WOA-SRU 模型构建流程图

Fig. 2 Construction flow chart of EEMD-WOA-SRU model

$M_{MAE}$ 、 $R_{RMSE}$ 、 $R_{RE}$  用以衡量预测值与观测值之间的偏差； $N_{NSE}$  用以评估水文模型模拟的优劣。

### 3 实例应用

人民胜利渠是第一个兴建在黄河下游的水利工程,使黄河转害为利。本文使用 2010~2019 年以天为单位的引黄用水量观测数据为试验对象,以 3 : 1 的比例划分训练集和测试集,分别用过去 7、14 d 预测未来 1 d 的用水量,用过去 14、21 d 预测未来 7 d 的用水量。

表 1 为采用 LSTM 预测法、镜像极值延拓法与未进行极值延拓的分量误差对比结果。由表 1 可知,LSTM 预测法比镜像极值延拓法分解误差小 0.41%,比未经过极值延拓的分量误差小 0.94%。说明使用 LSTM 预测法较好地抑制了 EEMD 的端点效应。

表 1 EEMD 分量的分解误差

Tab. 1 Decomposition error of EEMD components

延拓方式	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
无延拓	2.895 26	2.500 51	2.345 81	2.261 09	2.190 17	2.249 42	1.097 74
镜像延拓	2.912 02	2.488 23	2.357 33	2.206 25	2.200 22	2.198 54	1.095 33
LSTM 预测法	2.889 51	2.468 30	2.322 34	2.227 53	2.179 62	2.215 67	1.091 29

选用 LSTM、SRU、EMD-RNN 作为基准模型进行对比,预测曲线结果见图 3。由图 3(a)~(d)可知,相较于其他预测模型,EEMD-WOA-SRU 模型的拟合度更好,预测结果与实际观测值更接近。

图 3(e)~(h)中,虚线为误差平均值,实线为误差的中位数。由图 3(e)~(h)可看出,相较于其他基准模型,EEMD-WOA-SRU 模型的误差区间、中位数和平均值更小。

由不同模型在不同预测期下预测的  $M_{MAE}$ 、 $R_{RMSE}$ 、 $N_{NSE}$  的指标取值可知,EEMD-WOA-SRU 模型相较于其他基准模型有更低的  $M_{MAE}$ 、 $R_{RMSE}$

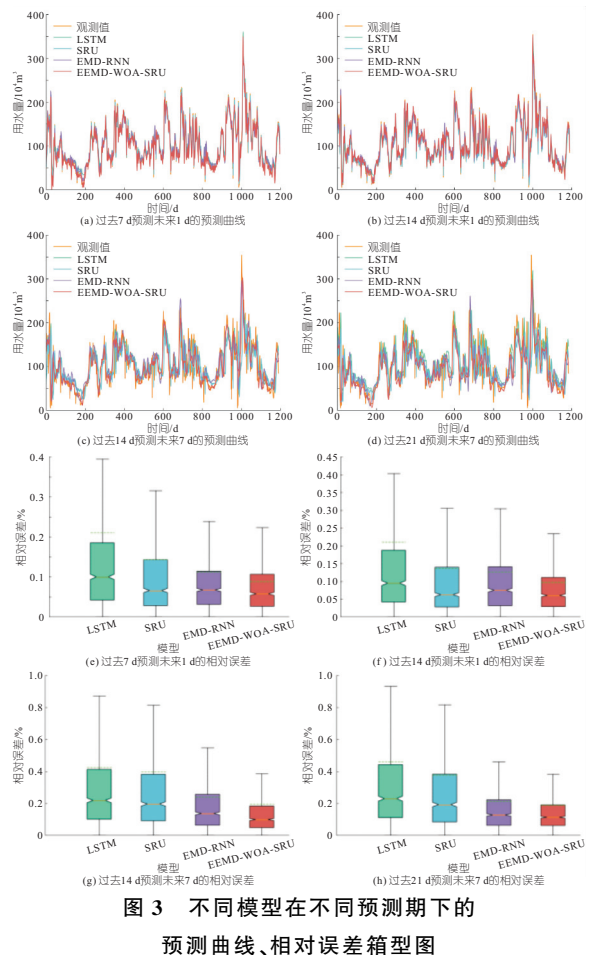


图 3 不同模型在不同预测期下的预测曲线、相对误差箱型图

Fig. 3 Forecast curves and relative error box plots of different models in different prediction periods

取值和更高的  $N_{NSE}$  取值。与 SRU 相比,在引入改进的 EEMD 和 WOA 方法后,EEMD-WOA-SRU 的  $M_{MAE}$  平均降低 45.42%, $R_{RMSE}$  平均降低 50.43%, $N_{NSE}$  平均提升 52.38%。与 EMD-RNN 相比, $M_{MAE}$  平均降低 14.87%, $R_{RMSE}$  平均降低 13.64%, $N_{NSE}$  平均提升 3.23%。

综上所述,相较于其他基准模型,EEMD-WOA-SRU 模型在不同预测期的预测结果均表现出了更好的精准度、可信度,具有更强的抗干扰性和稳定性。说明改进的 EEMD 方法和 WOA 算法有效地提高了 SRU 模型的抗干扰能力、预测精度和可信度。

### 4 结论

a. 使用 LSTM 预测法抑制了 EEMD 的端点效应,使用 WOA 算法优化了 SRU,进而提高了模型整体的预测精度和可信度。

b. 虽然使用“分解—合成—预测”的方式提高了模型的预测精度,然而引入信号处理过程和优化算法降低了模型的运行效率,未来有待进一步提升。

