

2-15-2024

Runoff Intelligent Prediction Method Based on Broad-deep Fusion Time-frequency Analysis

Ying Han

School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China; Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China, hanyingcs@163.com

Lehao Wang

School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China

Shumei Wang

Xinjiang Raohe Hydrological and Water Resources Testing Center, Shangrao 334000, China, Eddie3208@163.com

Xiang Zhang

Xinjiang Raohe Hydrological and Water Resources Testing Center, Shangrao 334000, China

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

Runoff Intelligent Prediction Method Based on Broad-deep Fusion Time-frequency Analysis

Abstract

Abstract: Broad learning system(BLS) is introduced to tackle the existed disadvantage that LSTM-based runoff prediction model is easy to fall into local optimization. To reduce the influence of noise on the prediction results, the variational mode decomposition (VMD) is adopted to transform the onedimensional time-domain runoff signal to the two-dimensional time-frequency plane. The runoff prediction model based on VMD-LSTM-BLS is proposed. The simulation results demonstrate that the prediction accuracy of the new model is more significantly improved compared with the baseline model and the existing LSTM-based runoff prediction model.

Keywords

runoff forecast, variational mode decomposition, long and short-term memory network, broad learning system, time-frequency analysis, intelligent prediction

Authors

Ying Han, Lehao Wang, Shumei Wang, Xiang Zhang, and Xingxing Luo

Recommended Citation

Han Ying, Wang Lehao, Wang Shumei, et al. Runoff Intelligent Prediction Method Based on Broaddeep Fusion Time-frequency Analysis[J]. Journal of System Simulation, 2024, 36(2): 363-372.

宽度-深度融合时频分析的径流智能预测方法

韩莹^{1,2}, 王乐豪¹, 王淑梅^{3*}, 张翔³, 罗星星³

(1. 南京信息工程大学 自动化学院, 江苏 南京 210044; 2. 南京信息工程大学 江苏省大气环境与装备技术协同创新中心, 江苏 南京 210044; 3. 信江饶河水文水资源检测中心, 江西 上饶 334000)

摘要: 为解决现有基于 LSTM 的径流预测模型易陷入局部最优的问题, 提出了基于 VMD-LSTM-BLS (variational mode decomposition-LSTM-broad learning system) 的径流预测模型。将宽度学习系统与 LSTM 结合, 针对径流序列多噪音特点, 采用时频分析方法中的变分模态分解, 将径流时间序列的一维时域信号变换到二维时频平面, 减少噪声对预测结果的影响。仿真结果表明: 与基线模型及现有基于 LSTM 的径流预测模型相比, 该模型的预测精度有较为明显的提高。

关键词: 径流预测; 变分模态分解; 长短期记忆网络; 宽度学习系统; 时频分析; 智能预测

中图分类号: TP391.9 文献标志码: A 文章编号: 1004-731X(2024)02-0363-10

DOI: 10.16182/j.issn1004731x.joss.22-1137

引用格式: 韩莹, 王乐豪, 王淑梅, 等. 宽度-深度融合时频分析的径流智能预测方法[J]. 系统仿真学报, 2024, 36(2): 363-372.

Reference format: Han Ying, Wang Lehao, Wang Shumei, et al. Runoff Intelligent Prediction Method Based on Broad-deep Fusion Time-frequency Analysis[J]. Journal of System Simulation, 2024, 36(2): 363-372.

Runoff Intelligent Prediction Method Based on Broad-deep Fusion Time-frequency Analysis

Han Ying^{1,2}, Wang Lehao¹, Wang Shumei^{3*}, Zhang Xiang³, Luo Xingxing³

(1. School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China; 3. Xinjiang Raohe Hydrological and Water Resources Testing Center, Shangrao 334000, China)

Abstract: Broad learning system (BLS) is introduced to tackle the existed disadvantage that LSTM-based runoff prediction model is easy to fall into local optimization. To reduce the influence of noise on the prediction results, the variational mode decomposition (VMD) is adopted to transform the one-dimensional time-domain runoff signal to the two-dimensional time-frequency plane. The runoff prediction model based on VMD-LSTM-BLS is proposed. The simulation results demonstrate that the prediction accuracy of the new model is more significantly improved compared with the baseline model and the existing LSTM-based runoff prediction model.

Keywords: runoff forecast; variational mode decomposition; long and short-term memory network; broad learning system; time-frequency analysis; intelligent prediction

0 引言

径流预测^[1-2]为水资源的合理配置和水量调度提供了科学依据。计算机系统仿真^[3]作为时间序列

预测进行模拟实验分析的方法, 可以通过计算机软件及区域数据信息模拟出相似系统, 对该系统进行性能分析, 为实际情况做出实时反馈^[4]。统计学习预测仿真模型及机器学习预测仿真模型被广

收稿日期: 2022-09-26

修回日期: 2022-11-23

基金项目: 国家自然科学基金(62076136); 教育部新农科研究与改革实践(20200251)

第一作者: 韩莹(1978-), 女, 副教授, 硕士生导师, 博士, 研究方向为大数据处理方法及其应用。E-mail: hanyingcs@163.com

通讯作者: 王淑梅(1976-), 女, 工程师, 学士, 研究方向为水文情报预报。E-mail: Eddie3208@163.com

泛应用于时间序列的预测分析中。

径流时间序列具有非线性^[5]、突变性^[6]、多噪音等特点，因此，基于传统机器学习及统计学的仿真模型预测效果往往差强人意。长短时记忆网络(LSTM)在处理具有较长时间步长及非线性序列数据方面优势明显，在时间预测方面已经取得了优异的成果^[7-8]。单一模型不能完全捕捉径流序列非线性特征，混合模型有助于提高预测精度。Chen等^[9]引入自我注意力机制与LSTM相结合，成功在较短时间步长上获取到径流数据内在相互依赖关系。王军等^[10]引入卷积神经网络(CNN)提取特征改善LSTM模型性能，提高其径流预测能力。时频分析中的集合经验模态分解^[11]、变分模态分解(variational mode decomposition, VMD)^[12-13]被广泛应用于径流时间序列的预处理中，以消除噪音对仿真精度的影响。

受到深层结构的影响，上述基于LSTM的径流预测仿真模型普遍存在易陷入局部最优的缺点，这主要是由于LSTM训练时需要循环调整权重所造成的。已有研究表明宽度学习系统(broad learning system, BLS)^[14]具有求解全局最优、直接计算权重、运算简单、快捷的优点，本文尝试引入BLS改进LSTM用于径流预测来解决LSTM训练数据时出现的低频区陷入局部最优的缺陷。径流数据呈现出多噪音、非平稳等特点，导致预测结果出现偏差^[15-16]。综上，本文提出VMD-LSTM-BLS径流预测模型，模型通过VMD模块对径流序列预处理，消除了噪音对预测精度的影响。利用LSTM有效提取径流长序列中的时间信息^[17]，并承接BLS层有效解决LSTM训练数据时出现的低频区陷入局部最优的缺陷。进一步提高对峰值预测的精度，为水资源的合理配置和水量调度提供了科学依据。

1 基于VMD-LSTM-BLS的径流预测模型

1.1 VMD分解

径流时间序列具有多噪音、非平稳的特点，因

此，需要在输入预测模块之前对原始序列进行降噪处理。与现有降噪方法相比，VMD可以有效分离固有模态分量(intrinsic mode function, IMF)^[18]，搜寻每种模态的最佳中心频率和有限带宽，建立频谱图，对信号频域进行划分，使分解后的模态更平滑。本文选取VMD进行径流时间序列的降噪处理。

步骤1：构造变分问题

$$\mathbf{x}(t) = [(\delta_t + j/\pi t) * \mathbf{u}_k(t)] e^{-j\omega_k t} \quad (1)$$

$$\min_{\{\mathbf{u}_k\}, \{\omega_k\}} \left\{ \sum_k \|\partial_t \mathbf{x}(t)\|_2^2 \right\} \quad (2)$$

$$\text{s.t.} \quad \sum_{k=1}^K \mathbf{u}_k = \mathbf{f} \quad (3)$$

式中： \mathbf{f} 为初始信号； K 为分解模态的个数； $\{\mathbf{u}_k\}$ 、 $\{\omega_k\}$ 为分解后的第 k 个模态分量和中心频率； $\delta(t)$ 为狄拉克函数； $*$ 为卷积运算； $e^{-j\omega_k t}$ 为原始信号的指数信号； ∂_t 为偏导运算； t 为时间变量。

步骤2：求解变分问题

$$L(\{\mathbf{u}_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \|\partial_t \mathbf{x}(t)\|_2^2 + \left\| \mathbf{f}(t) - \sum_k \mathbf{u}_k \right\|_2^2 + \langle \lambda(t), \mathbf{f}(t) - \sum_k \mathbf{u}_k(t) \rangle \quad (4)$$

式中： λ 为拉格朗日乘法算子； α 为二次惩罚因子； $\langle \cdot \rangle$ 为内积运算； $\|\cdot\|_2^2$ 为 L_2 范数的平方。

步骤3：求各模态分量和中心频率

通过搜寻增广拉格朗日函数的鞍点，交替寻找迭代后的 $\{\mathbf{u}_k\}$ 、 $\{\omega_k\}$ 和 λ 的最优结果：

$$\hat{\mathbf{u}}_k^{n+1}(\omega) = \frac{\hat{\mathbf{f}}(\omega) - \sum_{i \neq k} \hat{\mathbf{u}}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{\mathbf{u}}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{\mathbf{u}}_k^{n+1}(\omega)|^2 d\omega} \quad (6)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \gamma(\hat{\mathbf{f}}(\omega) - \sum_k \hat{\mathbf{u}}_k^{n+1}(\omega)) \quad (7)$$

式中： γ 为噪声容忍度； $\hat{\mathbf{u}}_k^{n+1}(\omega)$ 、 $\hat{\mathbf{u}}_i(\omega)$ 、 $\hat{\mathbf{f}}(\omega)$ 、 $\hat{\lambda}(\omega)$ 分别为原信号 $\mathbf{u}_k^{n+1}(t)$ 、 $\mathbf{u}_i(t)$ 、 $\mathbf{f}(t)$ 、 $\lambda(t)$ 对应的傅里叶变换； ω 为频率； ω_k 为相应模态的中心频率。

1.2 LSTM

LSTM 的核心概念在于其用于信息传输的细胞状态以及用于筛选信息的“门”结构。通过 3 个门的函数和细胞状态, LSTM 可以捕获短期和长期时间序列中的复杂相关性。图 1 展示了 LSTM 基本架构。其中, x_t 是输入向量, i_t 是时间步长 t 中的输入状态, f_t 是时间步长 t 中的遗忘状态, o_t 是时间步长 t 中的输出状态, \tilde{c}_t 为添加到新记忆单元状态 c_t 的候选值。以 \tanh 和 sigmoid 激活函数 σ 的形式在 3 个门的顶部添加非线性。

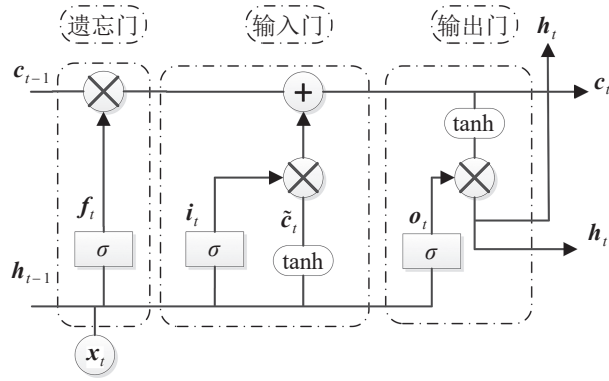


图 1 LSTM 的基本架构
Fig. 1 Basic architecture of LSTM

LSTM 的数学原理:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (8)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (9)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (10)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (11)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \quad (12)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (13)$$

式中: W_{xf} 、 W_{hf} 、 W_{xi} 、 W_{hi} 、 W_{xo} 、 W_{ho} 、 W_{xc} 、 W_{hc} 为输入向量与遗忘门、输入门、记忆单元、输出门相对应的权向量; b_f 、 b_i 、 b_c 、 b_o 为偏置变量; \otimes 为矩阵的 Hadamard 积。

1.3 BLS

LSTM 存在易陷入局部最优的缺点, 本文通过引入 BLS 以解决此问题, 其网络结构如图 2 所示。

假设输入数据集 X 包含 N 个样本, 每个样本有 M 个维度, Y 是属于 $\mathbb{R}^{N \times C}$ 的输出矩阵。对于 n 个特征通过式(14)映射后生成 n 个节点:

$$Z_i = \phi(XW_{ei} + \beta_{ei}), i = 1, 2, \dots, n \quad (14)$$

将所有特征节点表示为 $Z_i = [Z_1, Z_2, \dots, Z_n]$, 第 m 组增强节点为

$$H_m = \zeta(Z^n W_{hj} + \beta_{hj}), j = 1, 2, \dots, m \quad (15)$$

式中: W_{ei} 、 W_{hj} 和 β_{ei} 、 β_{hj} 分别为随机生成的连接权重和偏置值。因此, BLS 模型为

$$Y = [Z_1, Z_2, \dots, Z_n | \zeta(Z^n W_{h1} + \beta_{h1}), \dots, \zeta(Z^n W_{hm} + \beta_{hm})] W^m = [Z_1, Z_2, \dots, Z_n | H_1, H_2, \dots, H_m] W^m = [Z^n | H^m] W^m \quad (16)$$

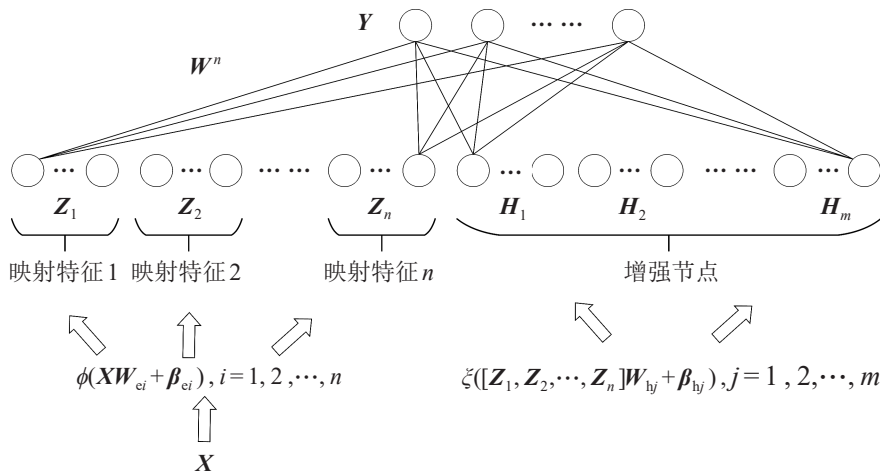


图 2 BLS 网络结构
Fig. 2 BLS network structure

1.4 VMD-LSTM-BLS 预测仿真模型

为提高水文时间序列的预测精度，提出了一种基于 VMD-LSTM-BLS 的预测方法，如图 3 所示。

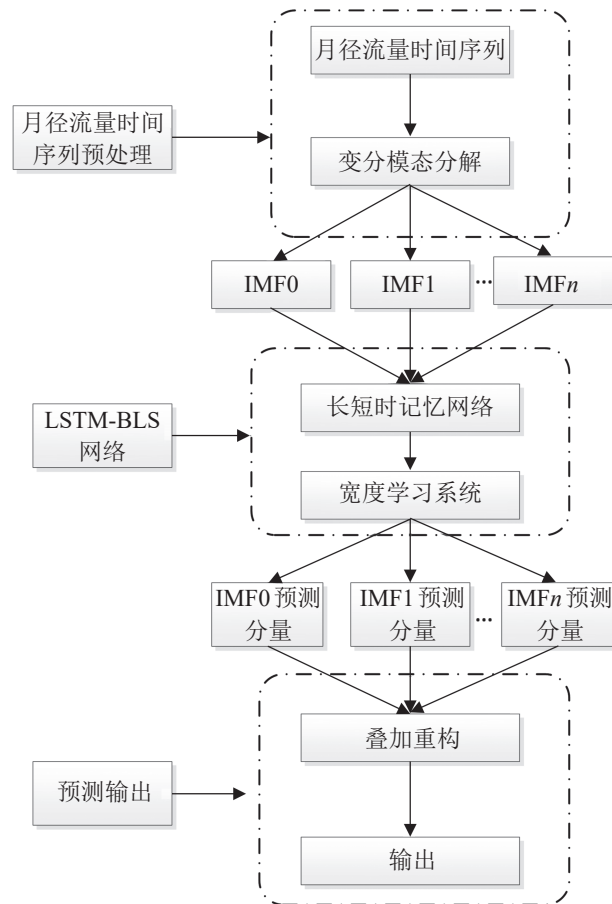


图3 VMD-LSTM-BLS 径流预测模型整体结构图
Fig. 3 Overall structure of VMD-LSTM-BLS runoff prediction model

VMD用于分解和重构原始水文时间序列，根据重构后的水文时间序列建立新的特征集合；LSTM用于提取新特征集合的特征数据；BLS对提取到的特征数据进行线性变换生成映射特征，进一步提取深层特征信息并计算最终权重。LSTM-BLS 预测模块如图 4 所示。其中，LSTM 层与 BLS 网络层中嵌入 2 层全连接层用于降低数据维度。

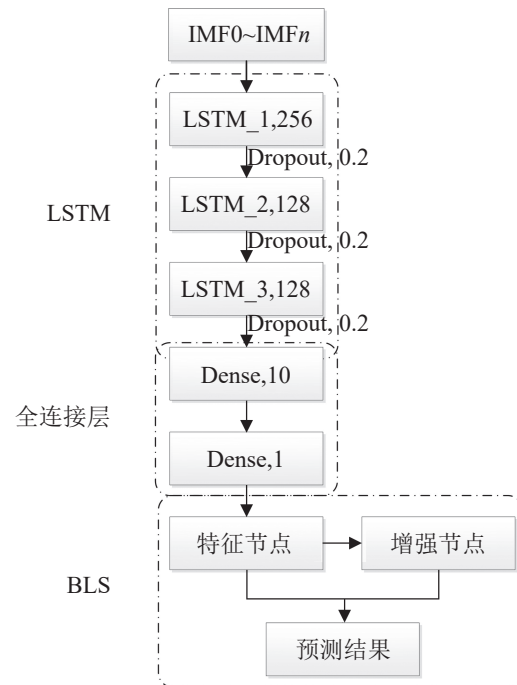


图4 LSTM-BLS 预测模块
Fig. 4 LSTM-BLS prediction module

基于 VMD-LSTM-BLS 的径流预测模型运算步骤如下。

步骤 1: 数据准备。数据查验，确定训练集和测试集。

步骤 2: VMD 处理训练集的原始水文时间序列，其中，通过中心频率法确定分解模态数 K 。

步骤 3: 数据重构。将分解后的一维 IMF 数据，重构为三维数据，作为 LSTM 模型的输入数据。

步骤 4: LSTM 特征提取。在每一层 LSTM 后增加 Dropout 层，以防止过拟合。

步骤 4.1: 前向计算每个神经元的输出值。

步骤 4.2: 反向计算每个神经元的误差项值，包括 2 个方向：①沿时间的反向传播，计算每个时刻的误差项；②将误差项向上一层传播。

步骤 4.3: 根据相应的误差项，计算每个权重和偏置项的梯度。

步骤 4.4: 将各个时刻梯度加在一起得到最终

梯度,并计算得到最终的权重值和偏置值。

步骤4.5:最终输出信号与最终权重值相乘再加上最终偏置,经过激活函数,得到LSTM网络输出结果,即提取到特征数据。

步骤5: Dense层降维。将经过LSTM提取后获得的三位特征数据经过两层Dense层进行降维,重构为一维数据。

步骤6: BLS深层特征提取。

步骤6.1:将降维后的二位数据作为模型输入数据,进行线性变换生成映射特征。

步骤6.2:通过映射特征生成多个映射节点。

步骤6.3:将映射节点经过式(15)生成 m 组增强节点。

步骤6.4:将映射节点和增强节点作为数据输入,通过矩阵求逆计算最终权重值。

步骤7:将最终权重值应用于测试集中,证明方法可行性和提出模型的性能。

2 实例分析

2.1 数据来源

本文的数据来源于鄱阳湖水系西河石门街水文站。测验河段大致顺直长约1 000 m,河槽为V型,中高水主槽宽度约48~120 m,当水位达30 m时,右岸漫滩宽度数公里。石门街水文站实测最高水位30.58 m(假定基面,2020-07-08),实测最大流量为 $2\,520\text{ m}^3\cdot\text{s}^{-1}$ (2020-07-08)。

本文选用1972—2020年共576个月径流量进行实证研究,并按照9:1的比例划分训练集和测试集。

2.2 参数设置与评价指标

分解模态数 K 的选取:由于分解模态数 K 对模型训练结果及降噪存在较大影响。因此,采用中心频率法对 K 值进行定位。

步骤1:对原始数据进行 K 次分解, $K=2,3,\dots,n$ 。导出每次分解时各个模态的中心频率。

步骤2:分析中心频率,如表1所示。确定最

终 K 值:当模态选取值较小时,VMD相当于自适应滤波器,导致信号中重要信息被过滤,影响后续模型预测精度;当模态选取较大时,相邻模态分量中心频率会相距较近,导致通频带重合或产生额外噪声从而造成过分解。

表1 分解模态中心频率表

Table 1 Decomposition mode center frequency

K	IMF0	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
3	0.359	0.185	0.025				
4	0.467	0.347	0.183	0.025			
5	0.467	0.348	0.182	0.026	0.006		
6	0.467	0.348	0.271	0.182	0.026	0.006	
7	0.486	0.467	0.348	0.271	0.182	0.026	0.006

由表1可知,分解模态数 $K=4$ 时,模态取值过小,IMF1与IMF3信号分解不完全,未达到最佳分解频率,易导致原始信号中一些重要信息被过滤,影响后续预测;当 $K=6$ 时,IMF1与IMF2中心频率差值为0.078,IMF4与IMF5的中心频率差值为0.02,较为接近,易产生通频带、模态重合现象,造成过分解。故本文选取 $K=5$ 作为研究对象。

本文模型均采用单输入单输出的形式,将分解后的模态作为输入,输入到3层单通道LSTM中进行初次特征提取,分别设置层数为256、128、128,为了防止训练出现过拟合的状况,在每一层末尾都加入了dropout层。再将初次特征提取后的数据作为输入经过全连接层,生成映射特征输入到BLS层中,进一步提取高层次特征,并计算出最终的权重值。具体参数见表2。

为验证本文模型的有效性,使用均方根误差 E_{RMS} 、平均绝对误差 E_{MA} 、平均绝对百分比误差 E_{MAP} 和纳什系数 S_{N} 为评价指标。

$$E_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$E_{\text{MA}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$E_{\text{MAP}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (19)$$

表2 本文模型主要参数表
Table 2 Main parameters of this model

层	参数	值
Input	时间步长	2
ARIMA	$\{p, d, q\}$	$\{10, 1, 0\}$
SVM	C	100 000
	缓冲大小	32
VMD	γ	0.000 1
	α	10
CNN	K	5
	卷积核个数	512
	卷积核大小	5
	填充	same
Maxpooling	激活函数	ReLU
	池化核大小	2
LSTM	填充	valid
	层数	3
	神经元个数	$\{256, 128, 128\}$
	优化器	Adam
	批大小	20
	训练次数	80
Dropout	损失函数	MAE
	P	0.2
Dense	层数	2
	神经元个数	$\{10, 1\}$
	激活函数	ReLU
BLS	s	0.8
	C	0.000 1
	N_1	10
	N_2	30
	N_3	280

注： N_1 为每个映射节点的个数； N_2 为映射特征个数； N_3 为增强结点个数； C 为 L_2 正则化参数； s 为缩放尺度。

$$S_N = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (20)$$

式中： y_i 为实际值； \hat{y}_i 为预测值； \bar{y}_i 为真实值的平均值； n 为样本数量。

E_{RMS} 、 E_{MA} 、 E_{MAP} 常用于评估时间序列的预测误差，其值越小，则预测精度越高。根据《水文情报预报规范》(GB/T 22482-2008)相关规定：当 $0.9 \leq S_N$ 时，预测精度等级为甲级； $0.7 \leq S_N < 0.9$

时，预测精度等级为乙级； $0.5 \leq S_N < 0.7$ 时，预测精度等级为丙级； $S_N < 0.5$ 时，预测结果不可信。

2.3 结果对比分析

为验证本文模型各模块的有效性和适用性，在相同数据集的基础上，以 E_{RMS} 、 E_{MA} 、 E_{MAP} 和 S_N 为参考指标，进一步测试VMD分解的必要性，并对比基线模型和现有模型的预测效果，实验结果如表3所示。

表3 现有模型的评价指标值
Table 3 Evaluation index values of existing models

模型	E_{RMS}	E_{MA}	E_{MAP}	S_N
ARIMA	0.921	0.532	1.659	0.133
SVR	0.914	0.513	1.619	0.145
LSTM	0.858	0.494	1.422	0.247
CNN-LSTM	0.823	0.482	1.302	0.308
LSTM-BLS	0.795	0.453	1.025	0.353
VMD-LSTM	0.381	0.286	0.688	0.515
VMD-CNN-LSTM	0.282	0.266	0.174	0.877
本文	0.004	0.003	0.025	0.985

表3所示，基线模型、单一的深度学习以及CNN-LSTM、LSTM-BLS组合模型的 S_N 指标方面均 <0.5 ，预测结果不可信；VMD-LSTM模型 S_N 指标达到0.515，可信度达到丙级；VMD-CNN-LSTM组合模型的 S_N 指标达到0.877，可信度达到乙级；本文模型 S_N 指标高达0.985，可信度达到甲级标准。可见，经VMD分解处理的模型，预测效果比单一模型更好，本文模型进一步提高了预测精度。

(1) VMD分解分析

实测月径流序列如图5所示。由图5可知，月径流的原始序列数据包含大量噪音，直接影响预测精度，需对原始数据进行平稳性处理。

本文采用VMD对原始数据进行降噪分解。其中，Alpha为中等带宽限制，IMF0为趋势项分量。

由原始径流序列(图5)和模态分解图(图6)和可以看出：

1) 原始径流时间序列中，由于季节影响，个别时刻呈现大幅波动，整体呈季节性规律，交替

出现高峰与低谷。峰值的剧烈增长提升了模型的预测难度。

2) 原始径流序列为非平稳时间序列, 且完全随机, 噪声影响显著。

3) 通过 VMD 分解, 可以使月径流数据趋于平稳, 使预测模型能够更好地识别其变化规律, 提升预测精度。

4) IMF 随分量的增加振幅逐渐减小, 频率逐渐降低, 反映了径流数据在此时间段内的整体变化趋势。

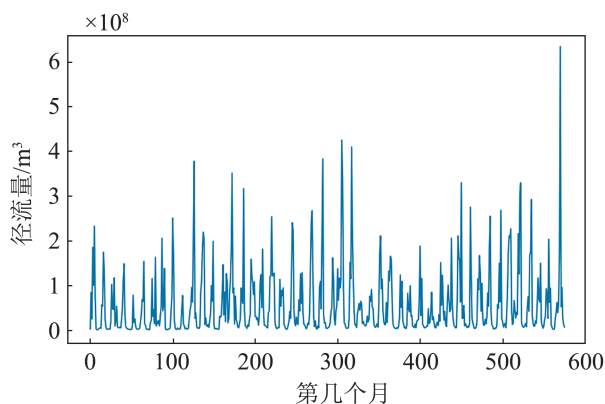


图5 石门街站月径流序列
Fig. 5 Monthly runoff sequence of shimenjie station

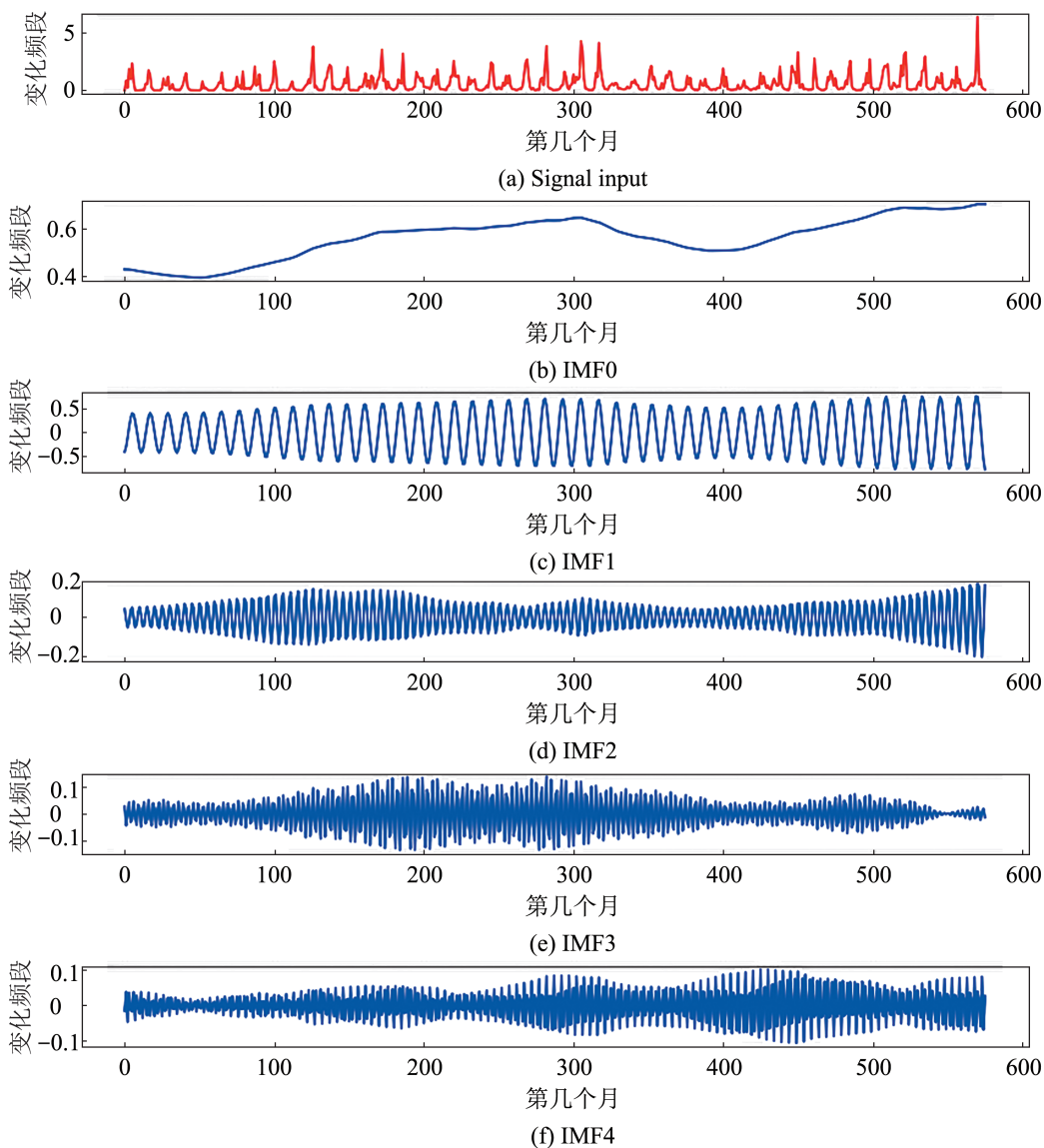


图6 模态分解图
Fig. 6 Modal decomposition

<http://www.china-simulation.com>

为了进一步验证VMD的必要性,将不加入VMD分解模块的LSTM-BLS模型与VMD-LSTM-BLS模型进行预测精度对比分析。本文模型相较于LSTM-BLS在 E_{RMS} 指标方面降低了99.49%; E_{MA} 值减少了99.31%; E_{MAP} 值减少了97.61%。由此可见,加入VMD噪音分解模块在径流预测中对精度的提高是十分必要的。

(2) 与基线模型的对比分析

本文模型相较于支持向量回归(SVR)、差分自回归移动平均模型(ARIMA)基线模型在 E_{RMS} 指标方面降低了99.56%、99.57%,在 E_{MA} 指标方面分别降低了99.42%、99.44%;在 E_{MAP} 指标方面分别降低了98.46%、98.49%。SVR模型在训练速度方面远超本文模型,但其预测结果误差大、精确度低,不符合我们的预期。ARIMA模型在训练平稳数据方面优势更强,但本文训练的月径流数据受自然环境及人类活动等多方面因素影响,呈现非平稳性特点,使得其预测精度不及本文模型。

(3) 与现有模型的对比分析

本文模型相较于LSTM、CNN-LSTM、VMD-LSTM和VMD-CNN-LSTM模型的 E_{RMS} 值分别降低了99.53%、99.51%、98.95%、98.58%; E_{MA} 值分别减少了99.39%、99.38%、98.95%、98.87%; E_{MAP} 指标方面分别降低了98.24%、98.08%、96.37%、85.63%。本文模型在径流预测精度方面提升明显。

与CNN-LSTM和VMD-CNN-LSTM预测模型相比,本文模型在预测精度方面提升也十分明显。分析原因,CNN和LSTM两者结合虽然可以同时提取径流序列的时空特征,相对单独的LSTM预测模型预测精度有所提高。但是,CNN和LSTM同属深度学习模型,因此,都不可避免地易陷入局部最优,两者结合并不能解决该缺陷。而作为平行结构,BLS具有求解全局最优,直接计算权重的特点,因此,将LSTM与BLS结合,可以有效避免LSTM陷入局部最优,与LSTM和CNN-LSTM相比,预测精度提升明显。

(4) 预测效果图分析

为了进一步证明本文模型在径流预测方面的优势,将现有径流预测模型和本文模型的预测效果进行可视化,如图7所示。分析结果如下:

单一LSTM模型在处理月径流数据时容易在低频区陷入局部最优,引入CNN能够更好地学习到径流时间序列数据的特征,使模型训练速度更加迅速,可以更好地拟合峰值的数据,使得精度有所提升,但不能解决其低频区局部最优的缺陷。

由于月径流时间序列的多噪音、非平稳性,去噪模块VMD可以对原始的径流序列进行分解得到平稳序列的IMF及残差,从不同的频率去拟合原始数据,从而降低原始数据中噪声的影响,提高预测精度。再次印证了解析模块在径流预测中的必要性。

VMD-CNN-LSTM模型在流量变化较大、出现极值的预测点的预测拟合仍然较差,极大影响了预测精度。这是由于深度组合模型仍存在局部最优的缺陷没有解决。本文模型利用BLS解决了LSTM模型的上述缺点,即使在极值点的拟合效果仍然十分理想。

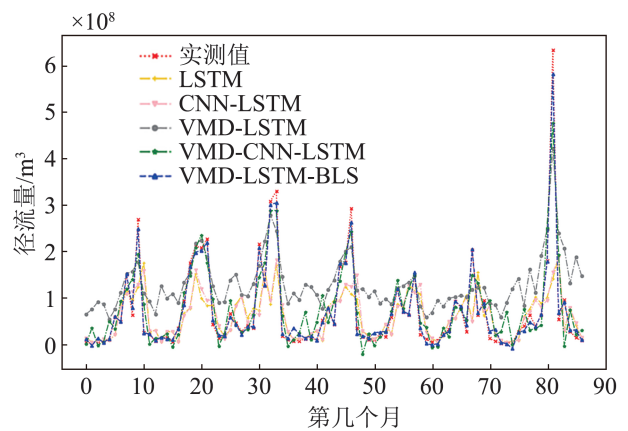


图7 本文模型与现有模型预测效果对比图
Fig. 7 Comparison of prediction effect between this model and existing models

3 结论

为解决由非线性和不稳定性特征对月径流预

测精度的影响, 以及深度学习模型易陷入局部最优的问题, 本文提出了VMD-LSTM-BLS组合模型, 并以石门街站576个月径流量为数据集展开了验证。对比分析后表明: ①针对月径流量存在的非线性和不稳定性特征, 需要对原始数据采用VMD分解, 分解后的多个分量可以有效地减少中心频率混叠的问题, 使模型能够更好地学习到数据特征, 提高模型的精度。②LSTM模型能够捕捉长期依赖关系, 借助BLS可以直接计算权重, 能够从训练数据中进一步学习到深层次的重要特征的特点。本文有效地将宽度学习模型与深度学习模型相结合, 解决了训练数据时深度学习网络易陷入局部最优的缺陷, 提高了对极端峰值的预测拟合精度, 证明了本文模型的适用性和有效性。

VMD-LSTM-BLS模型仅对历史时间序列数据进行数据挖掘及分析预测, 并未考虑到天气、经纬度等外部因素对预测效果的影响。未来可结合多因素输入进行分析, 改进现有模型, 更好地应用于径流时间序列的预测研究中。

参考文献:

- [1] 金菊良, 魏一鸣, 丁晶. 预测日径流过程的最近邻仿真模型[J]. 系统仿真学报, 2002, 14(11): 1494-1496.
Jin Juliang, Wei Yiming, Ding Jing. Nearest Neighbor Bootstrap Model for Predicting Daily Flow Process[J]. Journal of System Simulation, 2002, 14(11): 1494-1496.
- [2] Huseyin Cagan Kilinc, Adem Yurtsever. Short-term Streamflow Forecasting Using Hybrid Deep Learning Model Based on Grey Wolf Algorithm for Hydrological Time Series[J]. Sustainability, 2022, 14(6): 3352.
- [3] 李伟, 周玉臣, 林圣琳, 等. 仿真模型验证方法综述[J]. 系统仿真学报, 2019, 31(7): 1249-1256.
Li Wei, Zhou Yuchen, Lin Shenglin, et al. Review of Simulation Model Validation Methods[J]. Journal of System Simulation, 2019, 31(7): 1249-1256.
- [4] Qin Jinbo, Liang Ji, Chen Tao, et al. Simulating and Predicting of Hydrological Time Series Based on TensorFlow Deep Learning[J]. Polish Journal of Environmental Studies, 2019, 28(2): 795-802.
- [5] Peng Tian, Zhou Jianzhong, Zhang Chu, et al. Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks[J]. Water, 2017, 9(6): 406.
- [6] 曹爽, 秦天玲, 石晓晴, 等. 亚马逊流域降雨径流时空变化分析[J]. 水文, 2018, 38(1): 90-96.
Cao Shuang, Qin Tianling, Shi Xiaoping, et al. Time-spatial Distribution of Rainfall and Runoff in Amazon Basin[J]. Journal of China Hydrology, 2018, 38(1): 90-96.
- [7] 黄克威, 王根绪, 宋春林, 等. 基于LSTM的青藏高原冻土区典型小流域径流模拟及预测[J]. 冰川冻土, 2021, 43(4): 1144-1156.
Huang Kewei, Wang Genxu, Song Chunlin, et al. Runoff Simulation and Prediction of a Typical Small Watershed in Permafrost Region of the Qinghai-Tibet Plateau Based on LSTM[J]. Journal of Glaciology and Geocryology, 2021, 43(4): 1144-1156.
- [8] 夏容, 江官星. 基于LSTM的软件时间序列延迟预测仿真[J]. 计算机仿真, 2021, 38(12): 435-439.
Xia Rong, Jiang Guanxing. Software Time Series Delay Prediction Simulation Based on LSTM[J]. Computer Simulation, 2021, 38(12): 435-439.
- [9] Chen Xi, Huang Jiayu, Han Zhen, et al. The Importance of Short Lag-time in the Runoff Forecasting Model Based on Long Short-term Memory[J]. Journal of Hydrology, 2020, 589: 125359.
- [10] 王军, 高梓勋, 朱永明. 基于CNN-LSTM模型的黄河水质预测研究[J]. 人民黄河, 2021, 43(5): 96-99, 109.
Wang Jun, Gao Zixun, Zhu Yongming. Research on Yellow River Water Quality Prediction Based on CNN-LSTM Model[J]. Yellow River, 2021, 43(5): 96-99, 109.
- [11] Tan Qiaofeng, Lei Xiaohui, Wang Xu, et al. An Adaptive Middle and Long-term Runoff Forecast Model Using EEMD-ANN Hybrid Approach[J]. Journal of Hydrology, 2018, 567: 767-780.
- [12] Dragomiretskiy K, Zosso D. Variational Mode Decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [13] 包苑村, 解建仓, 罗军刚. 基于VMD-CNN-LSTM模型的渭河流域月径流预测[J]. 西安理工大学学报, 2021, 37(1): 1-8.
Bao Yuancun, Xie Jiancang, Luo Jungang. Monthly Runoff Prediction of Weihe River Basin Based on VMD-CNN-LSTM Model[J]. Journal of Xi'an University of Technology, 2021, 37(1): 1-8.
- [14] 任长娥, 袁超, 孙彦丽, 等. 宽度学习系统研究进展[J]. 计算机应用研究, 2021, 38(8): 2258-2267.
Ren Chang'e, Yuan Chao, Sun Yanli, et al. Research of Broad Learning System[J]. Application Research of Computers, 2021, 38(8): 2258-2267.
- [15] Zhou Jianzhong, Peng Tian, Zhang Chu, et al. Data Pre-

- analysis and Ensemble of Various Artificial Neural Networks for Monthly Streamflow Forecasting[J]. *Water*, 2018, 10(5): 628.
- [16] 李彤, 胡国华, 顾庆福, 等. 近55年来降水及人类活动对资水流域径流的影响[J]. *水文*, 2018, 38(6): 54-58, 88.
Li Tong, Hu Guohua, Gu Qingfu, et al. Impact of Precipitation and Human Activities on Runoff in Zishui River in Recent 55 Years[J]. *Journal of China Hydrology*, 2018, 38(6): 54-58, 88.
- [17] 胡庆芳, 曹士坦, 杨辉斌, 等. 汉江流域安康站日径流预测的LSTM模型初步研究[J]. *地理科学进展*, 2020, 39(4): 636-642.
- [18] Youngmin Seo, Sungwon Kim, Singh V P. Machine Learning Models Coupled with Variational Mode Decomposition: A New Approach for Modeling Daily Rainfall-runoff[J]. *Atmosphere*, 2018, 9(7): 251.