



# LSTM-WBLS 模型在日降水量预测中的应用

## 摘要

基于长短时记忆网络 (Long Short-Term Memory, LSTM) 降水量预测模型存在过拟合、时滞现象, 而宽度学习系统 (Broad Learning System, BLS) 无需多次迭代的特点有助于解决 LSTM 的上述缺点. 加权宽度学习系统 (Weighted Broad Learning System, WBLS) 通过在 BLS 中引入加权惩罚因子约束分配样本权重, 降低噪声和异常值对降水量预测精度的影响. 本文提出一种 LSTM-WBLS 日降水量预测模型, 选取湖北省巴东站日降水量进行实证研究, 并考虑气压、气温、湿度、风速和日照等因素对降水量的影响. 实验结果表明, 与现有的预测模型相比, LSTM-BLS 模型在 RMSE、MAE 和  $R^2$  等评价指标上均有显著提升. 不同时间步长下, 本文模型预测精度均优于现有模型, 验证了其稳定性. 与 LSTM 相比, WBLS 直接计算权重的特点使得 LSTM-WBLS 的运算效率并未降低.

## 关键词

降水量预测; 长短时记忆网络; 宽度学习系统; 加权宽度学习系统; 多因素预测

中图分类号 TP183

文献标志码 A

收稿日期 2021-10-18

资助项目 南方海洋科学与工程广东省实验室 (珠海) 基金 (SML2020SP007); 国家自然科学基金 (62076136)

## 作者简介

韩莹, 女, 博士, 副教授, 研究方向为大数据处理方法及其应用. hanyingcs@163.com

罗嘉 (通信作者), 女, 硕士, 研究方向为灾害应急建模与分析、气象数据分析. jeeaaan@qq.com

## 0 引言

短时强降水会造成暴雨洪涝, 继而引发山洪、泥石流等次生灾害, 严重威胁人们生命财产安全. 因此, 熟练掌握降水规律、精准预测日降水量, 对洪涝灾害的研究和控制具有重要指导意义<sup>[1]</sup>.

降水量预测的方法基本分为两类: 基于过程的方法和数据驱动方法. 基于过程的降水量预测方法的优点是对降水物理过程解释清晰, 但物理过程的复杂性增加了建模难度, 需要给出一系列假设才能够对模型求解. 数据驱动的方法是经验型的, 不需要对降水物理过程进行分析, 只根据降水量的历史数据进行预测, 模型简单易操作.

统计方法和机器学习是目前最常见的数据驱动的降水量预测方法. 统计方法方面, 近年来最为流行的是基于差分自回归移动平均 (AutoRegressive Integrated Moving Average, ARIMA) 模型的预测方法<sup>[2-3]</sup>. 研究表明, 当降水量时间序列是线性或接近线性时, 统计模型能产生令人满意的预测结果, 但当时间序列呈现非线性时, 其预测结果往往差强人意. 有鉴于此, 适合复杂非线性过程建模的机器学习方法广泛应用于降水预测中. Hartigan 等<sup>[4]</sup> 使用随机森林 (Random Forest, RF) 和支持向量回归 (Support Vector Regression, SVR) 对悉尼流域内降水和气温进行预测; Xiang 等<sup>[5]</sup> 利用决策树和 FR 的双系统协同影响模型对重庆市 34 个气象观测站的数据进行预测; Peng 等<sup>[6]</sup> 基于极限学习机和基因表达式构建了日降水量预测混合模型; 勾志竟等<sup>[7]</sup> 结合遗传算法和 BP 神经网络的优势研究了天津市日降水等级的预测方法; Rostam 等<sup>[8]</sup> 采用多种优化算法对多层感知器算法进行优化, 以探索伊朗首都大尺度气候指数与降水之间的任何有意义的联系.

然而, 传统机器学习方法无法捕获输入序列的长期记忆<sup>[9]</sup>, 从而影响预测精度. 长短时记忆网络 (Long Short-Term Memory, LSTM) 克服了上述缺点. 王子岳等<sup>[10]</sup> 采用句子状态 LSTM 模型对说话人意图进行识别; 王朋等<sup>[11]</sup> 基于小波长短期记忆网络对风电功率超短期概率进行预测; 罗嘉等<sup>[12]</sup> 等融合 LSTM 与 BLS 对突发气象灾害事件中公众情感倾向分析. 在降水预测方面: Nguyen 等<sup>[13]</sup> 利用 LSTM 改进基于雷达的降雨预报; 沈皓俊等<sup>[14]</sup> 利用 LSTM 研究了中国夏季降水情况; Ni 等<sup>[15]</sup> 给出了两类改进的 LSTM 模型 (WD-LSTM 和 CNN-LSTM), 并分别探讨了其在径流和降雨预测的应用; Kang 等<sup>[16]</sup> 选定多输入变量的

1 南京信息工程大学 自动化学院, 南京, 210044

2 湖北省公众气象服务中心, 武汉, 430074

LSTM 模型对江西景德镇日降水量进行预测.

虽然基于 LSTM 的降水预测模型已经显示出强大的优势,但现有模型都未解决在预测中存在时滞的问题.这主要是由于 LSTM 训练中需要循环调整权重造成的.注意到新提出的宽度学习系统(Broad Learning System, BLS)具有直接计算权重,运算简单、快捷的优点,可以用来改进 LSTM.但是噪声和异常值对模型会产生不良影响,所以将加权惩罚因子应用于 BLS,提出了加权宽度学习系统(Weighted Broad Learning System, WBLS).通过自动为每个样本分配适当的权重,给高可靠性的样本更高的权重,而可疑的异常值获得较低的权重.因此,减少了异常样本对建模的影响.结合两种算法的优势,本文提出 LSTM-WBLS 日降水量预测模型.

为了有效地验证新模型,本文选取湖北省巴东站进行日降水量预测的实证研究.在预测精度上,与现有降水预测模型相比较,本文模型在均方根误差(RMSE)、平均绝对误差(MAE)和决定系数( $R^2$ )三个评价指标上均表现最佳.在稳定性上,通过分析时间步长分别为 1、3 和 5 d 对各模型预测精度的影响,证明了虽然所有模型的预测精度会随着时间步长增加而降低,但在不同的时间步长下,本文模型在 RMSE、MAE 和  $R^2$  三个评价指标方面仍然表现最佳.在运算效率上,因为 WBLS 计算方便、快捷的特点,加入了 WBLS 的 LSTM-WBLS 模型与 LSTM 模型相比,运算效率并未下降.

## 1 模型原理与结构

本文首先给出 LSTM 和 WBLS 的基本结构和原理,然后给出本文提出的基于多因素的 LSTM-WBLS 预测模型.

### 1.1 LSTM 原理及结构

LSTM 的结构如图 1<sup>[17]</sup>所示.

图 1 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,  $h_t$  和  $C_t$  分别是时间  $t$  中的隐藏状态和单元状态,  $h_{t-1}$  和  $C_{t-1}$  分别是时间  $t-1$  中的隐藏状态和单元状态.以 tanh 和 sigmoid 激活函数  $\sigma$  的形式添加非线性.

LSTM 原理如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (3)$$

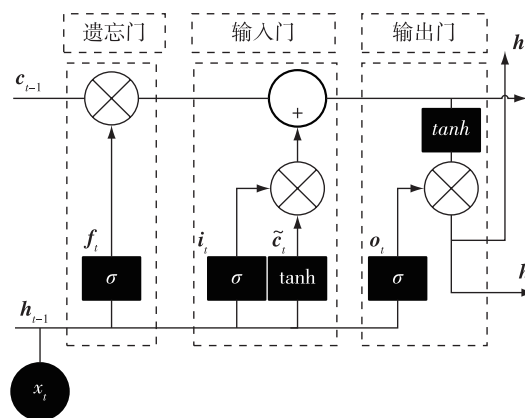


图 1 LSTM 结构原理

Fig. 1 Structure principle of LSTM

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (4)$$

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t * \tanh(C_t), \quad (6)$$

其中  $W_f, W_i, W_c, W_o$  分别代表遗忘门、输入门、记忆单元和输出门的相应的权值向量,  $b_f, b_i, b_c, b_o$  分别代表遗忘门、输入门、记忆单元和输出门的偏差变量,  $*$  是矩阵的 Hadamard 积.

### 1.2 WBLS 原理及结构

Chen 等<sup>[18]</sup>于 2019 年初提出了 BLS,进一步于 2020 年又提出了 WBLS,减少了异常样本对建模的影响<sup>[19]</sup>.

假设  $X$  包含  $N$  个样本,每个样本有  $M$  个维度,  $Y$  是属于  $\mathbf{R}^{N \times Q}$  的输出矩阵,其中  $Q$  是输出的维度. WBLS 的结构如图 2<sup>[19]</sup>所示.

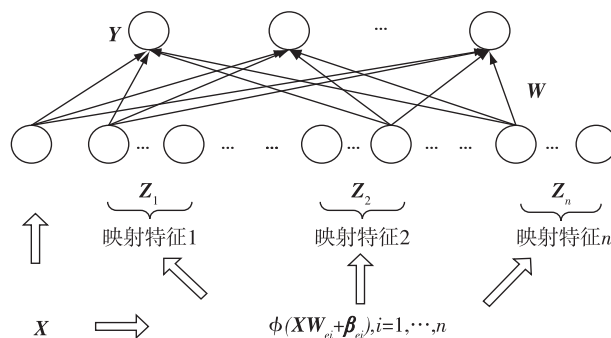


图 2 WBLS 结构原理

Fig. 2 Structure principle of WBLS

第  $n$  个特征由式(7)映射后生成  $p$  个节点:

$$Z_i = \phi(XW_{ei} + \beta_{ei}), i = 1, 2, \dots, n, \quad (7)$$

其中  $W_{ei}$  和  $\beta_{ei}$  是随机生成的权重和偏置.最终, WBLS 模型可以表示为

$$Y = [X | Z_1, \dots, Z_m] W = HW. \quad (8)$$

### 1.3 LSTM-WBLS 日降水预测模型

通过上述分析,融合深度学习和加权宽度学习的优势,本文提出了 LSTM-WBLS 日降水量预测模型.模型整体架构如图 3 所示.

LSTM-WBLS 多因素日降水量预测模型模型的具体运算步骤如下:

#### 1) 降水序列数据的输入

日降水量数据的收集过程中存在人为失误或者机器故障等问题,导致收集到的数据有异常值.为了降低删除异常值对模型预测结果的影响,本文对异常值进行填补 0 值的操作.多因素的日降水量数据存在指标数据尺度不一致问题,每个维度的值范围相差过大,会造成训练过程中不容易收敛,影响算法的学习过程,因此需要对数据进行归一化处理,本文选用最大最小归一化如式(9)所示:

$$x_{scaled} = \frac{x - x_{min}}{x_{max}}, \quad (9)$$

其中  $x_{min}$  表示当前数据的最小值,  $x_{max}$  表示当前数据的最大值.

将归一化的数据进行训练集、验证集和测试集的划分,将训练集作为 LSTM 模型的输入.

#### 2) 基于 LSTM 的训练

LSTM 共有五层,前四层为 LSTM 网络层,目的是能够对输入的序列进行时序特征进行充分提取,第五层为全连接层,目的是将 LSTM 网络层从第一个时间步到最后一个时间步提取到的特征转换为固定的特征向量.为防止 LSTM 在训练过程中出现过拟合,因此每层 LSTM 后都进行 Dropout 处理.

#### 3) 基于 WBLS 的预测

首先,将全连接层的输出作为前  $n$  组映射特征集合  $Z^n = [Z_1, Z_2, \dots, Z_n]$ ,每组包含节点  $p$  个,则映射特征的隐藏层节点数为  $np$ .与输入  $X$  结合可以得到式(8).

其次,通过加权岭回归算法,可以计算出 WBLS 的连接权值,其求解问题见式(10):

$$\min_W f(W) = \min_W \| \theta HW - \theta Y \|_F^2, \quad (10)$$

其中  $\| \cdot \|_F$  是指  $F$  范数,  $\theta$  是样本权重的加权惩罚因子,  $H$  是隐藏层,由式(11)表示.

$$H = [X | E^m] = (h_1, h_2, \dots, h_L) \in \mathbf{R}^{K \times L}, \quad (11)$$

其中:  $h_j \in \mathbf{R}^k, j = 1, \dots, L$  为隐藏层  $H$  的第  $j$  节点,  $L = np + mq$  是隐藏层节点数,  $K$  是时间序列的输入数量.

式(10)是一个最小二乘问题,是关于  $W$  的凸优化估计,旨在求出使训练误差最小时的输出权重  $W$ .对其进行求解,结果见式(12):

$$W = H^+ Y, \quad (12)$$

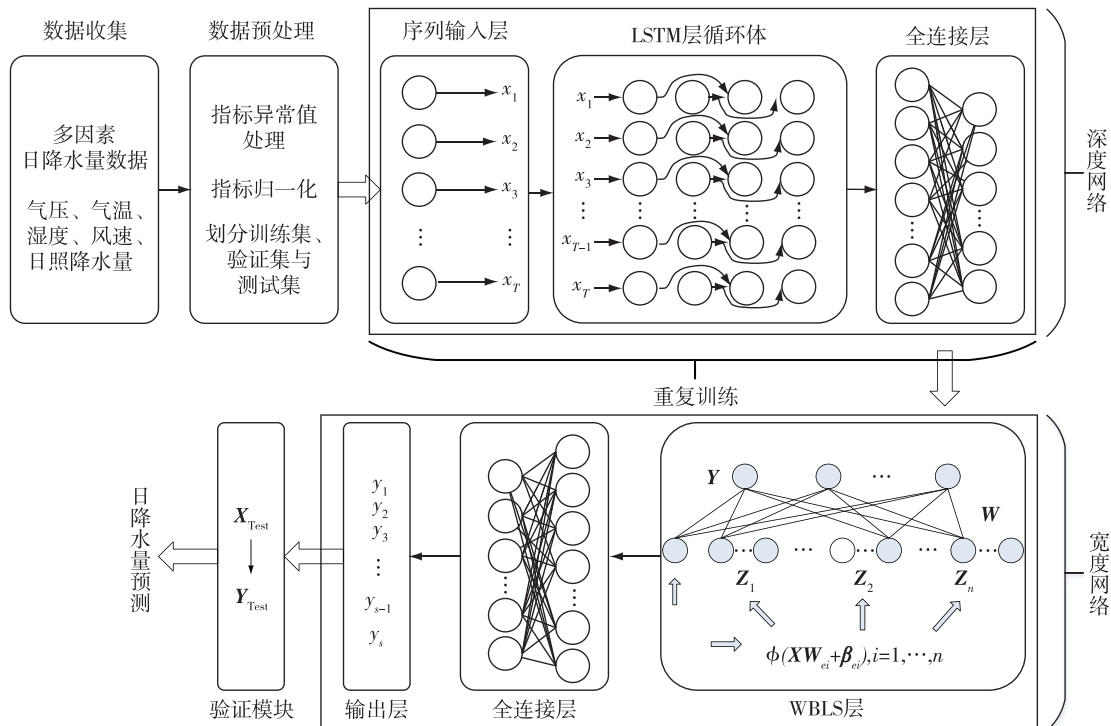


图 3 LSTM-WBLS 模型整体框架

Fig. 3 Overall framework of the LSTM-WBLS model

其中  $H^+$  是  $H$  的伪逆.但通常情况下,上述解的泛化误差可能会很大,特别是对于一些病态问题.为了提升网络的泛化能力,在原式引入  $F$  范数正则项以防止网络过拟合,得到式(13):

$$\min_W f(W) = \min_W \|\theta HW - \theta Y\|_F^2 + C \|W\|_F^2. \quad (13)$$

式(13)为一个岭回归问题.  $C$  表示对权重平方和的进一步约束,可以由式(14)最终求解得到权重  $W$ :

$$W = (CI + H^T \theta^2 H)^{-1} H^T \theta^2 Y. \quad (14)$$

4) 输出预测结果

将式(14)中通过运算得到的权重  $W$  与隐藏层  $H$  结合得到最终的预测结果.

## 2 实例分析

### 2.1 研究地区以及数据集描述

巴东县,隶属湖北省恩施土家族苗族自治州,位于湖北省西南部,属于亚热带季风气候,温暖多雨,湿热多雾,四季分明.最热月平均气温一般高于  $22\text{ }^\circ\text{C}$ ,最冷月气温在  $0\sim 15\text{ }^\circ\text{C}$  之间;年降水量多在  $800\sim 1\ 600\text{ mm}$ .巴东天气的非周期性变化和降水季节变化都很显著,所以对其日降水预测比较困难.

本文数据在国家气象中心网站获取.数据的范围为 2000—2020 年巴东地区气象观测站实测降水量观测数据.将共 7 671 d 的数据以 7:2:1 的比例设为训练集、验证集和测试集,测试集为最近几年的降水量数据.

### 2.2 参数设置与评价指标

将日降水量映射为  $S \times \tau \times D$  个张量数据作为模型的输入.其中,  $S$  为样本数量(samples),  $\tau$  为时间步长(time steps),  $D$  为特征个数(features),本文模型为气压、气温、湿度、风速、日照以及降水量六个维度的输入和降水量一个维度的输出.所以  $D$  为 6.

采用 Dropout 退出部分神经元来防止过拟合,确定随机丢弃比例  $P$  值.再通过全连接层,将其输出作为 WBLS 层的映射特征,与输入  $X$  一起构成隐藏层  $H$ ,最后算出输出权重  $W$ .  $N_1$  为每个映射特征节点个数,  $N_2$  为映射特征个数,  $C$  为  $L_2$  正则化参数.本文利用验证集对本文模型的超参数进行实验,取值为多次实验后选取的最优值.本文所用的参数如表 1 所示.

选取 RMSE、MAE 和  $R^2$  对算法的精确度进行评

估. RMSE 对预测值误差十分敏感,能够体现预测的精确度. MAE 可以避免误差相互抵消的问题,可以准确反映实际预测误差.  $R^2$  常用于判断回归方程的拟合程度,数值在 0 到 1 之间,越大表示模型的预测性能越好.

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{y}(i) - y(i))^2}, \quad (15)$$

$$E_{\text{MAE}} = \sum_{i=1}^n |(\tilde{y}(i) - y(i))|, \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\tilde{y}(i) - y(i))^2}{\sum_{i=1}^n (y(i) - \bar{y})^2}, \quad (17)$$

其中,  $y_i$  表示真实月降水量,  $\tilde{y}$  表示预测月降水量,  $\bar{y}$  表示平均月降水量.

表 1 LSTM-WBLS 模型主要参数

Table 1 Main parameters of LSTM-WBLS model

层	参数	值
Input	Timesteps	1,3,6
LSTM	Layers	4
	Number of neurons	{256,128,64,32}
	Loss	mse
	Optimizer	Adam
	Batch_size	64
Dropout	epochs	80
	$P$	0.2
Dense	Layers	1
	Number of neurons	20
	Activation	relu
WBLS	$N_1$	20
	$N_2$	30
	$C$	1e-5

### 2.3 与现有模型对比分析

将现有模型与本文模型进行对比分析,以预测长度 1 d 为例,对比结果如表 2 所示.本文模型与现有的 SVM<sup>[4]</sup>、EEMD-ARIMA<sup>[3]</sup>、LSTM<sup>[13]</sup>、CNN-LSTM<sup>[15]</sup> 和 LSTM-BLS 模型相比:RMSE 值分别减少了 50.20%、47.58%、37.00%、34.80% 和 17.54%; MAE 值分别减少了 55.29%、53.19%、49.20%、48.00% 和 22.72%;  $R^2$  值分别增加了 0.209、0.189、0.078、0.058 和 0.015.显然,本文模型表现在三个指标上都是最优的,证明了本文模型的有效性和准确性.



表 2 各个模型评价指标对比

Table 2 Comparison of evaluation indicators of each model

模型	RMSE/mm	MAE/mm	$R^2$
SVM	4.191	2.427	0.711
EEMD-ARIMA	3.981	2.318	0.731
LSTM	3.313	2.136	0.842
CNN-LSTM	3.201	2.087	0.861
LSTM-BLS	2.531	1.404	0.905
LSTM-WBLS	2.087	1.085	0.920

为进一步验证本文模型有效性,对 LSTM 系列相关模型预测进行可视化.将测试集的降水序列和各个模型的预测值进行拟合,对比可视化如图 4 所示.为了方便作图,其中第 1 天对应 2018 年 11 月 26 日的降水真实值与预测值,一直到 2020 年 12 月 31 日共 767 d.

从图 4 可以看出,在降水量突变的日期本文模型的预测结果要明显优于现有的所有模型.注意到,现有的基于 LSTM 模型(图 4a、4b)在预测上都不可避免地存在滞后性,因此无法精准预测.图 4c 因加入 BLS 基本解决了滞后性问题,但是噪声和异常值对预测的不良影响依然存在.本文模型在图 4c 基础

上加入了加权惩罚因子,预测结果最优(图 4d).

### 2.4 与单因素模型对比

为进一步验证本文模型的有效性,与单因素降水量输入的模型进行对比,结果如表 3 所示.可以看出多因素输入的预测要远远高于单因素输入模型.原因是数据中零值过多,单输入模型无法准确预测.部分数据集如表 4 所示.综合考虑各种气象因素的影响,本文模型可以准确地对降水量进行预测.

表 3 与单因素模型对比

Table 3 Comparison with single factor model mm

模型	RMSE	MAE
LSTM(单因素)	7.610	4.008
LSTM-WBLS(单因素)	6.610	3.890
LSTM-WBLS	2.087	1.085

### 2.5 稳定性分析

不改变模型中的参数,将预测长度分别设置为 3 d 和 5 d,对日降水量进行预测,结果如表 5 所示.结合预测长度为 1 d 的预测结果,可以看出随着预测长度的增加,所有预测模型的预测精度都有所下降.但是,LSTM-WBLS 模型在不同预测长度下,预测精

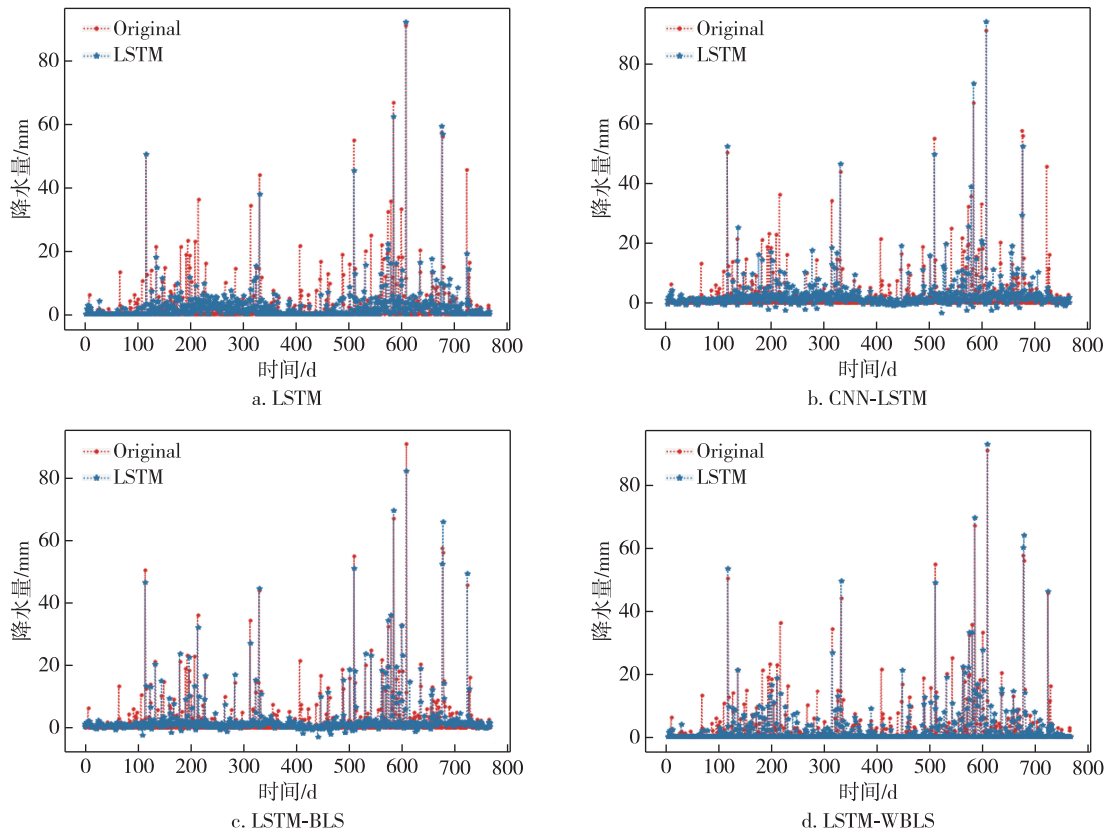


图 4 LSTM 系列相关模型预测可视化对比

Fig. 4 Visualized comparison of prediction results by LSTM series related models

表 4 2001 年 1 月部分数据集

Table 4 Part of the dataset in January 2001

气压/ kPa	气温/ ℃	降水量/ mm	相对湿度/ %	风速/ (m/s)	日照/ h
97.3	9.2	0.0	82	0.3	1.2
97.4	8.2	0.0	84	1.8	1.5
98.3	7.9	0.9	69	3.0	0.0
98.1	4.7	0.1	83	1.3	0.0
97.7	5.9	0.0	68	2.0	1.0
97.8	7.4	0.0	75	1.5	0.6
97.7	7.5	1.1	76	3.8	0.1
97.8	6.6	3.3	82	3.0	0.0
98.1	6.3	0.0	77	2.8	0.0
97.9	6.6	0.0	67	2.0	0.5

表 5 不同预测长度下各个模型评价指标对比

Table 5 Comparison of evaluation indicators for each model under different prediction lengths

模型	RMSE/mm		MAE/mm		R <sup>2</sup>	
	3 d	5 d	3 d	5 d	3 d	5 d
SVM	4.325	4.457	2.601	2.725	0.701	0.682
EEMD-ARIMA	4.056	4.216	2.407	2.421	0.721	0.705
LSTM	3.408	3.421	2.234	2.351	0.831	0.822
CNN-LSTM	3.307	3.392	2.179	2.306	0.855	0.843
LSTM-BLS	2.641	2.728	1.503	1.581	0.889	0.873
LSTM-WBLS	2.145	2.216	1.202	1.272	0.909	0.886

度依然优于其他模型.这一结果验证了本文模型的稳定性.

### 2.6 运算效率分析

运算效率也是算法的主要评价指标.在保证 LSTM-WBLS 与 LSTM 训练都达到最优结果的情况下,运算效率对比如表 6 所示.由表 6 可以看出,LSTM-WBLS 训练时间只比 LSTM 长 2 s 左右,效率未明显下降.其原因是 WBLS 不需大量运算、直接计算权重的特点使得 LSTM-WBLS 相比 LSTM,在运算效率上不会有太大的下降.

表 6 LSTM-WBLS 和 LSTM 运算效率对比

Table 6 Comparison of operation efficiency between LSTM-WBLS and LSTM

模型	训练时间/s		
	时间步长=1 d	时间步长=3 d	时间步长=5 d
LSTM	655.925	945.031	1 194.134
LSTM-WBLS	657.231	947.132	1 196.891

## 3 结论

鉴于现有日降水预测模型的缺点,本文提出一

种 LSTM-WBLS 日降水预测模型.通过实证研究,本文模型借助 WBLS 不用大量训练、直接通过伪逆计算权重的特点解决了 LSTM 预测中存在的滞后问题,且运算效率没有下降.通过自动为每个样本分配适当的权重,给高可靠性的样本更高的权重,而可疑的异常值获得较低的权重,减少了异常样本的影响,提高了预测精度与稳定性.本文探讨了在降水量预测中,同时融合深度学习与宽度学习优势的可能性,为降水量预测研究提供了新的思路.本文模型仅考虑历史气象数据和具体日降水数据,以后将加入地理、地貌等特征,进一步提高日降水预测精度.

## 参考文献

### References

- [1] 王海鹏,张斌,刘祖涵,等.基于混沌理论的武汉、宜昌近 60 年来月降水特征的对比研究[J].自然灾害学报,2012,21(6):111-118  
WANG Haipeng, ZHANG Bin, LIU Zuhan, et al. Chaos theory-based comparative study on monthly rainfall characteristics in Wuhan and Yichang during recent 60 years [J]. Journal of Natural Disasters, 2012, 21(6): 111-118
- [2] 张改红.基于 ARIMA 模型的渭南市降水量趋势分析与预测[J].价值工程,2019,38(34):197-199  
ZHANG Gaihong. Analysis and prediction of precipitation trend in Weinan city based on ARIMA model [J]. Value Engineering, 2019, 38(34): 197-199
- [3] 胡盈,吴静.基于 ARIMA 模型的降水空间特征分析及预测[J].江西科学,2021,39(1):99-104  
HU Ying, WU Jing. Analysis and prediction of precipitation spatial characteristics based on ARIMA model [J]. Jiangxi Science, 2021, 39(1): 99-104
- [4] Hartigan J, MacNamara S, Leslie L M, et al. Attribution and prediction of precipitation and temperature trends within the Sydney catchment using machine learning [J]. Climate, 2020, 8(10): 120
- [5] Xiang B, Zeng C F, Dong X N, et al. The application of a decision tree and stochastic forest model in summer precipitation prediction in Chongqing [J]. Atmosphere, 2020, 11(5): 508
- [6] Peng Y Z, Zhao H S, Zhang H, et al. An extreme learning machine and gene expression programming-based hybrid model for daily precipitation prediction [J]. International Journal of Computational Intelligence Systems, 2019, 12(2): 1512-1525
- [7] 勾志竟,任建玲,徐梅,等.基于 Hadoop 的 GA-BP 算法在降水预测中的应用[J].计算机系统应用,2019,28(9):140-146  
GOU Zhijing, REN Jianling, XU Mei, et al. Application of GA-BP algorithm based on Hadoop in precipitation forecast [J]. Computer Systems & Applications, 2019, 28(9): 140-146
- [8] Rostam M G, Sadatinejad S J, Malekian A. Precipitation forecasting by large-scale climate indices and machine

- learning techniques [J]. *Journal of Arid Land*, 2020, 12(5):854-864
- [9] Shen C P. A trans-disciplinary review of deep learning research for water resources scientists [J]. *Water Resources Research*, 2018, 54(11):8558-8593
- [10] 王子岳, 邵曦. 基于 S-LSTM 模型利用‘槽值门’机制的说话人意图识别[J]. *南京信息工程大学学报(自然科学版)*, 2019, 11(6):751-756  
WANG Ziyue, SHAO Xi. Speaker intention recognition based on S-LSTM model and slot-gate [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2019, 11(6):751-756
- [11] 王朋, 孙永辉, 翟苏巍, 等. 基于小波长短期记忆网络的风电功率超短期概率预测[J]. *南京信息工程大学学报(自然科学版)*, 2019, 11(4):460-466  
WANG Peng, SUN Yonghui, ZHAI Suwei, et al. Ultra-short-term probability prediction of wind power based on wavelet decomposition and long short-term memory network [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2019, 11(4):460-466
- [12] 罗嘉, 王乐豪, 涂姗姗, 等. 基于 LSTM-BLS 的突发气象灾害事件中公众情感倾向分析[J]. *南京信息工程大学学报(自然科学版)*, 2021, 13(4):477-483  
LUO Jia, WANG Lehao, TU Shanshan, et al. Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2021, 13(4):477-483
- [13] Nguyen D H, Kim J B, Bae D H. Improving radar-based rainfall forecasts by long short-term memory network in urban basins [J]. *Water*, 2021, 13(6):776
- [14] 沈皓俊, 罗勇, 赵宗慈, 等. 基于 LSTM 网络的中国夏季降水预测研究[J]. *气候变化研究进展*, 2020, 16(3):263-275  
SHEN Haojun, LUO Yong, ZHAO Zongci, et al. Prediction of summer precipitation in China based on LSTM network [J]. *Climate Change Research*, 2020, 16(3):263-275
- [15] Ni L L, Wang D, Singh V P, et al. Streamflow and rainfall forecasting by two long short-term memory-based models [J]. *Journal of Hydrology*, 2020, 583:124296
- [16] Kang J L, Wang H M, Yuan F F, et al. Prediction of precipitation based on recurrent neural networks in Jingdezhen, Jiangxi province, China [J]. *Atmosphere*, 2020, 11(3):246
- [17] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8):1735-1780
- [18] Chen C L P, Liu Z L, Feng S. Universal approximation capability of broad learning system and its structural variations [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(4):1191-1204
- [19] Chu F, Liang T, Chen C L P, et al. Weighted broad learning system and its application in nonlinear industrial process modeling [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(8):3017-3031

## Application of improved LSTM-WBLS model in daily precipitation forecast

HAN Ying<sup>1</sup> GUAN Jian<sup>1</sup> CAO Yunzhong<sup>1</sup> LUO Jia<sup>2</sup>

<sup>1</sup> School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044

<sup>2</sup> Hubei Public Meteorological Service Center, Wuhan 430074

**Abstract** The popular Long Short-Term Memory (LSTM) based precipitation prediction models suffer from overfitting and time lag. Broad Learning System (BLS), which does not require multiple iterations, helps to solve the above disadvantages of LSTM. Weighted Broad Learning System (WBLS) reduces the impact of noise and outliers on precipitation prediction accuracy by introducing a weighted penalty factor constraint to assign sample weights in the BLS. Thus a LSTM-WBLS daily precipitation prediction model is proposed in this paper. The daily precipitation at Badong station in Hubei province is selected for empirical study. And the influence of air pressure, temperature, humidity, wind speed and sunshine on precipitation is considered. The experimental results demonstrate that the LSTM-BLS model has significantly improved the prediction accuracy in the evaluation indexes of RMSE, MAE and  $R^2$  compared with existing prediction models. The prediction accuracy of the new model outperforms existing models at different time steps, proving its stability. In particular, the direct calculation of weights by WBLS does not make any reduction in operational efficiency of LSTM-WBLS.

**Key words** precipitation forecast; long short-term memory (LSTM) network; broad learning system (BLS); weighted broad learning system (WBLS); multi-factor prediction