

智能水文预报模型的研究与应用

何 健¹, 余宇峰², 冯胜男¹, 邓劲柏², 李 凯¹

(1. 江苏省水文水资源勘测局, 江苏 南京 210029; 2. 河海大学 计算机与信息学院, 江苏 南京 211110)

摘要:洪水过程具有高度非线性、复杂性和非平稳性特征。将自适应步长的布谷鸟搜索(ASCS)算法应用于神经网络水文模型参数优化中, 构建ASCS-LSTM洪水预报模型, 并采用注意力机制进一步提高输入输出的相关性, 实现高精度的智能洪水预测。在秦淮河流域的水位预测实验表明, ASCS-LSTM预报模型的预报结果要优于传统机器学习模型, 稳定性和精确度得到提升, 可为水文预报提供新思路。

关键词:水文预报; 记忆网络; 搜索算法; LSTM; ASCS

中图分类号: [TV124] **文献标识码:** B **文章编号:** 1007-7839(2023)10-0001-0005

Research and application of intelligent hydrological forecasting model

HE Jian¹, YU Yufeng², FENG Shengnan¹, DENG Jinbai², LI Kai¹

(1. Jiangsu Province Hydrology and Water Resources Investigation Bureau, Nanjing 210029, China;
2. College of Computer and Information, Hohai University, Nanjing 211110, China)

Abstract: The flood process is highly nonlinear, complex and non-stationary. The ASCS-LSTM flood forecasting model is built, which adopts adaptive step size cuckoo search (ASCS) algorithm to optimize parameter of LSTM neural network hydrological model, and which applies the attention mechanism to further improve the relevance of input and output to achieve high-precision intelligent flood prediction. The water level prediction experiments in Qinhuai River basin show that ASCS-LSTM forecasting model can achieve more stable and accurate forecasting results than those of traditional machine learning model, and thus providing a new idea for hydrological forecasting.

Key words: hydrological prediction; memory network; search algorithm; LSTM; ASCS

1 概 述

准确的洪水预报可有效支撑流域洪水预警、调度与防洪减灾, 同时对区域规划、灌溉取水、泥沙输移以及其他水文应用等具有重要意义^[1-3]。随着信息技术的快速发展, 水利行业已经形成了空、天、地一体的水雨情观测资料, 如何将在自动控制、图像识别等领域广泛应用的机器学习算法应用于水文

预报领域, 构建具有物理含义的数据驱动预报模型, 提升洪水预报的智能化业务能力和预报精度, 成为当前解决水文学难题、推动水文学进步的驱动力^[4-8]。

数据驱动洪水预报模型采用决策树和人工神经网络等机器学习算法, 对历史水位流量样本进行训练, 构建学习机以实现未来某个时刻的水位流量预测, 在水文预报研究和应用领域得到广泛

收稿日期: 2023-02-23

基金项目: 江苏省水利科技项目(2021065); 国家重点研发计划(2021YFB3900605)

作者简介: 何健(1984—), 男, 高级工程师, 博士, 主要从事水文预报工作。E-mail: longlivehj@hotmail.com

应用^[9-12]。机器学习与人工智能技术的发展,给智能洪水预报技术的进步提供了新的机遇。长短时记忆网络(long short-term memory, LSTM)模型^[13]通过独特的门结构和单元状态记忆长时间信息,能有效提高洪水预报网络的稳定性与可靠性^[12-14]。Cui等^[14]提出融合新安江XAJ和LSTM的混合预报模型(XAJ-LSTM),将XAJ模型的洪水预报结果作为LSTM模型的输入变量,以增强水文建模的物理机制,实现精确的洪水多步提前预报。周研来等^[15]提出了长短期记忆神经网络与批量学习、正则化、筛选神经元技术相结合的深度学习网络,以解决变化环境下雨洪过程统计特征的非线性、随机性和时变性问题。Ding等^[16]提出可解释的时空注意力长短期记忆模型,利用注意力机制动态提取洪水过程中的关键特征向量,实现洪水过程的精准预测。

使用机器学习技术建立预报模型在中小河流水文预报研究中取得了一定的进展,SVM、BP等机器学习模型虽然具有较强的时间序列问题求解能力,但也存在局部最优解且难以从根本上提升预测结果的准确度和可靠性等问题。本文将自适应步长的布谷鸟搜索(ASCS)算法^[17]应用于LSTM神经网络水文模型参数优化中,构建基于ASCS-LSTM的智能预报模型,并采用注意力机制进一步提高输入输出的相关性,实现高精度的智能洪水预测。

2 预测模型

2.1 模型结构

洪水过程成因与传播机理复杂,如何高效识别并提取洪水时空特征信息是理解洪水模式并进行成功预报预警的关键。本文将自适应步长的布谷鸟搜索(ASCS)算法应用于LSTM神经网络水文模型参数优化中,同时在LSTM时间序列预测中,采用注

意力机制对模型输入进行选择性的学习以提升模型输入输出之间的关联,构建基于ASCS-LSTM的智能水文预测模型如图1所示。

2.2 模型构建

基于ASCS-LSTM的神经网络水文预测模型包括输入层、隐含层和输出层。输入层以预报断面水位(流量)及对该断面有影响的雨量站雨量信息为自变量,通过数据归一化、训练测试集划分等操作构建模型输入集;输出层对模型预报结果采用反归一化等操作,得到最终输出结果;隐含层包含一个二维结果,其中网络方向包含多个LSTM网络,能够学习预报断面水文时间序列数据的本质特征,时序方向可以实现前一时刻的信息向下一时刻传递。LSTM网络的时间记忆单元可以记忆更多有用的水文时空信息,而遗忘门可以将无用的信息进行过滤。

ASCS-LSTM采用自适应步长布谷鸟搜索算法确定隐含层节点数和学习率参数,模型构建步骤如下:

(1)选择水文时间序列样本数据,对数据进行划分和归一化处理。

(2)采用AM为模型输入分配不同权重,提高输入输出之间的相关性,建立输入输出序列的长距离依赖关系。

(3)初始化LSTM神经网络的隐含层节点数和学习率参数取值范围。

(4)将水文时间序列训练样本数据输入到LSTM神经网络模型中进行训练,采用ASCS算法进行参数寻优。随机产生 n 个孵化鸟巢位置,每个孵化鸟巢对应一组LSTM神经网络的参数,将水文流量数据预测的均方根误差作为各孵化鸟巢位置适应度值,确定当前最优孵化鸟巢位置及其

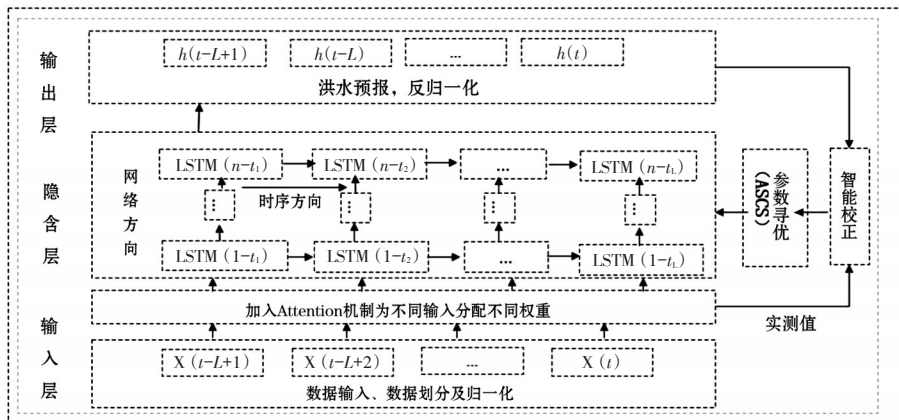


图1 基于ASCS-LSTM的智能水文预测模型

适应度值。

(5)保留该组最优孵化鸟巢位置,利用新的莱维飞行机制更新其他孵化鸟巢位置,得到1组新的孵化鸟巢位置,并计算其位置对应适应度值,依据适应度值采用优胜劣汰策略进行2组位置更新,得到1组较优孵化鸟巢位置。

(6)按照概率丢弃差的鸟巢位置,用偏好随机游动生成新的鸟巢位置替代已丢弃的鸟巢位置。

(7)比较得到1组较优鸟巢位置,判断当前迭代次数是否符合终止条件,如果能够满足,则使用该最优解作为 ASCS-LSTM 模型的结构进行水文流量时间序列的预测,否则返回上述步骤继续对孵化鸟巢位置进行优化。

2.3 模型评估

智能预报模型通过训练历史水文资料去学习构建最优模型,再以该模型预测未来某时刻的水位流量值。因此,本文采用均方根误差、决定系数、平均绝对误差和 Nash-Sutcliffe 系数等预测常用的评估标准对模型结果进行评估。均方根误差表示整个数据集的总体性能,决定系数越接近1,说明模型预测性能越好,平均绝对误差代表预测值与实际值之间的偏差,Nash-Sutcliffe 系数与1的差距越小,模型模拟效果越好,说明预测结果具有较高的可信度,反之则表明模型的预测结果不可信。

3 实验结果与分析

3.1 实验结果分析

3.1.1 研究区域概况

选择长江流域支流秦淮河流域为例进行模型实验和分析。秦淮河位于江苏省西南部,流域面积 2 631 km²,具有典型的平原地形。降水主要集中在4—9月,汛期短,滞洪能力弱。本研究分别以秦淮河流域14个雨量站和1个出口断面水文站(东山)2014年5月28日至2019年1月23日的数据验证本模型。

3.1.2 预报因子确定

秦淮河流域属于平原区的中小流域,采用相关系数衡量预报因子与预报目标断面间的相关性。相关系数计算式为

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中: r 为相关系数; n 为时段总长度; x_i 为 i 时刻的监测数据; y_i 为 i 时刻对应的预测值; \bar{x} 为序列的平均值; \bar{y} 为序列对应预测值的平均值。

以东山站水位和流域内14个关联雨量站降水量为预报因子,计算分析得到各测站的相关系数如表1所示。

由表1可知,天生桥最优相关系数在 $t-5$ 时刻,

表1 秦淮河流域各测站相关系数

测站	t	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$	$t-6$	距离/km
东山(水位)	0.893	0.864	0.834	0.803	0.773	0.743	0.713	0
方便	0.382	0.400	0.407	0.415	0.413	0.408	0.397	48.5
土桥	0.355	0.380	0.379	0.377	0.371	0.361	0.357	24.1
卧龙山	0.574	0.555	0.579	0.632	0.621	0.605	0.589	45.1
赵村	0.400	0.425	0.434	0.434	0.433	0.427	0.416	46.4
钟山	0.371	0.362	0.369	0.370	0.386	0.390	0.397	55.5
东山	0.720	0.757	0.749	0.745	0.742	0.687	0.642	0
天生桥	0.371	0.400	0.406	0.410	0.411	0.412	0.401	49.5
前埠村	0.710	0.743	0.743	0.685	0.657	0.603	0.579	6.9
句容站	0.332	0.310	0.341	0.346	0.351	0.358	0.358	54.4
句容水库	0.343	0.331	0.354	0.359	0.361	0.364	0.371	58.0
赤山闸	0.397	0.402	0.413	0.415	0.417	0.415	0.399	45.5
二圣桥	0.360	0.344	0.349	0.354	0.357	0.361	0.363	66.2
天王寺	0.370	0.351	0.359	0.365	0.366	0.369	0.372	58.7
北山水库	0.364	0.321	0.336	0.347	0.359	0.365	0.369	62.5

表示东山站($t+1$)时刻预报水位与天生桥站 $t-5\sim t$ 时刻时段雨量相关,因此,选择该站($t-5$)~ t 时刻的时段降水量数据作为模型的输入。同理可得,东山站 $t+1$ 时刻的水位预测模型以东山站 t 时刻水位和(东山站、前埠村站)($t-1$)~ t 时刻、(土桥站)($t-2$)~ t 时刻、(赵村、卧龙山、方便)($t-3$)~ t 时刻、(天生桥、中山)($t-3$)~ t 时刻、(赤山)($t-5$)~ t 时刻、(天王寺、二圣桥、句容水库、句容、北山水库)($t-6$)~ t 时刻雨量数据、未来1 h的雨量预报数据平均值作为模型的输入。

3.1.3 模型结果分析

(1)不同参数寻优算法对比分析

将 PSO、GA、CS 和 ASCS 算法分别用于优化 LSTM 模型参数,以均方根误差 RMSE 作为适应度, LSTM 模型的学习因子范围为 $[0.001, 0.01]$,隐含层节点数的范围为 $[150, 250]$ 。

PSO 算法的参数配置:种群最大迭代次数为 20,粒子数量为 20,解的维度为 2,惯性参数为 0.8,学习因子数值为 2,随机数分别为 0.6 和 0.3。

GA 算法的参数配置:最大迭代次数为 20,变异概率为 0.01,交叉概率为 0.8,群体大小为 20。

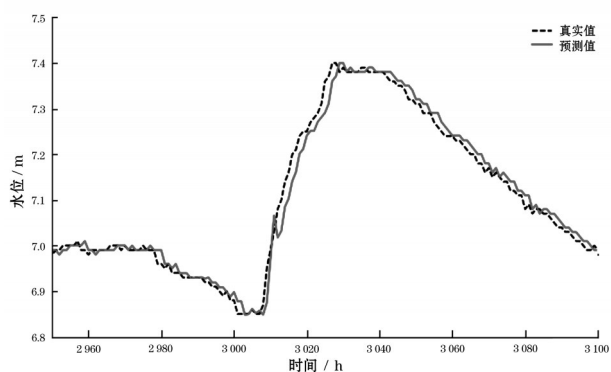
CS、ASCS 算法的参数配置:巢穴数量为 20,需要寻优的参数个数为 2,发现概率为 0.25,最大迭代次数为 20。

在预见期为 6 h 的预测模型中,GA、PSO、CS 和 ASCS 分别需要 18 次、14 次、13 次和 12 次迭代可以获得最优 LSTM 模型参数。ASCS 算法步长因子随迭代次数增加不断减小,寻优前期下降速度较快,寻优后期下降速度变慢,使得算法前期收敛时间短,后期步长变化幅度减小,提升了搜索的准确率,可以更快地搜索到最佳解。从结果可知,ASCS 算法的搜索速度和全局搜索能力均要优于 PSO 算法、GA 算法和 CS 算法,用 ASCS 算法对 LSTM 网络进行参数优化可以获得较优的 LSTM 参数。

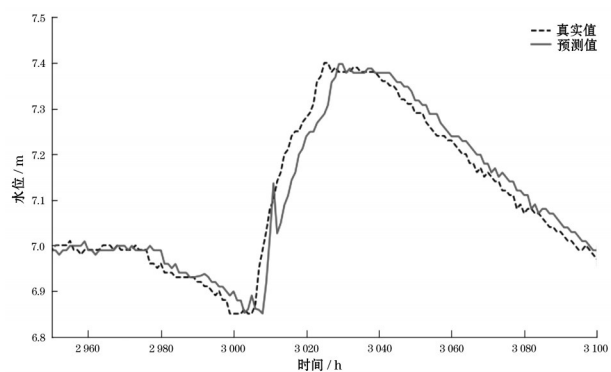
(2)ASCS-LSTM 模型预报结果分析

根据上述实验中获得的最优解,将 ASCS 算法中参数解的数量设为 20,解的维度设为 2,发现概率设为 0.25,最大迭代次数设为 20;LSTM 神经网络水文模型的隐含层节点数的范围设为 150~250,学习因子的范围设为 0.001~0.01。并将依据算法确定的预报因子作为输入,ASCS-LSTM 模型在预见期分别为 2 h、4 h 和 6 h 的预测结果如图 2 所示(2018 年 5 月 28 日 9:00 至 2018 年 6 月 3 日 15:00 的数据)。

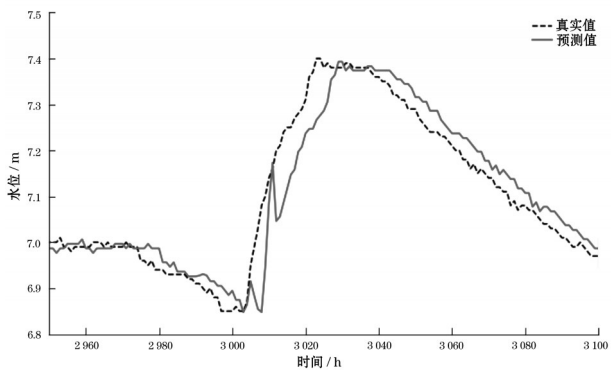
由图 2 可知,ASCS-LSTM 模型可以很好地模拟



(a) 预见期为 2 h 的预测结果



(b) 预见期为 4 h 的预测结果



(c) 预见期为 6 h 的预测结果

图 2 不同预见期的模型预测结果

水位涨落趋势,虽然随着预见期的延长,模型预测能力出现一定程度的削弱,但预测过程线整体较平滑,与实测曲线较为接近。此外,ASCS-LSTM 模型在 2 个流域内的 NSE 和 R^2 值均在 85% 以上, RMSE、MAE 均取得良好的拟合效果,说明了 ASCS-LSTM 模型在智能洪水预报应用中的可靠性和适用性。

3.2 不同模型对比分析

为验证 ASCS-LSTM 模型的适用性与可靠性,将其与 SVM、BP 模型在实验流域内的预测结果进行比较,其中 3 个模型的输入与上述实验中确定的预

报因子保持一致。3种模型预见期6 h的预测结果对比如图3所示。

根据RMSE、MAE、NSE等评价指标,对3种预测模型的不同预见期进行误差比较,结果表明,ASCS-LSTM模型的评价指标均优于其他模型,表明该算法具有较高的精度。

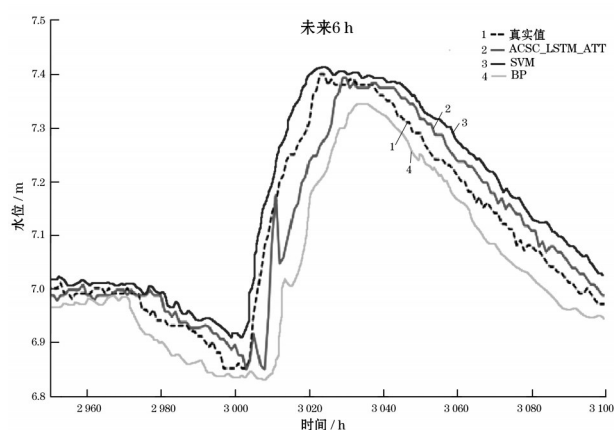


图3 不同模型的预测结果对比

4 结 语

针对传统机器学习模型存在局部最优解且难以从根本上提升预测结果的准确度和可靠性等问题,本文将自适应步长的布谷鸟搜索算法应用于LSTM神经网络水文模型参数优化中,并采用注意力机制进一步提高输入、输出的相关性,构建ASCS-LSTM模型,实现高精度的智能洪水预测。将上述模型应用于秦淮河流域的水位预测并进行比较分析。实验结果表明,本文所构建的模型在一定程度取得比SVM和BP等模型更好的预报效果,具有良好的应用价值。今后可着手将气象、蒸发量等特征因素作为预报因子,进一步提高预报模型的准确性,并探索将洪水过程的物理含义融入到深度学习模型的构建过程中,提升模型的可解释性。

参考文献:

- [1] 张建云. 中国水文预报技术发展的回顾与思考[J]. 水科学进展, 2010, 21(4): 435-443.
- [2] 刘志雨, 刘玉环, 孔祥意. 中小河流洪水预报预警问题与对策及关键技术应用[J]. 河海大学学报(自然科学版), 2021, 49(1): 1-6.
- [3] 张旭旻, 瞿思敏, 李倩, 等. 基于协整理论的淮河流域上

- 游洪水预报实时校正方法[J]. 水资源保护, 2022, 38(6): 88-95, 145.
- [4] MOSAVI A, OZTURK P, CHAU K W. Flood prediction using machine learning models: literature review[J]. Water 2018(10), 1536: 1-40.
- [5] SUN A Y, SCANLON B R. How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions[J]. Environmental Research Letters, 2019(1): 48.
- [6] SCHMIDT L, HEE F, ATTINGER S, et al. Challenges in applying machine learning models for hydrological inference: a case study for flooding events across Germany [J]. Water Resources Research, 2020, 56(5): 1-10.
- [7] 芮孝芳. 水文学与“大数据”[J]. 水利水电科技进展, 2016, 36(3): 1-4.
- [8] 张珂, 牛杰帆, 李曦, 等. 洪水预报智能模型在中国半干旱半湿润区的应用对比[J]. 水资源保护, 2021, 37(1): 28-35.
- [9] GHORBANI M A, ZADEH H A, ISAZADEH M, et al. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction [J]. Environmental Earth Sciences, 2016, 75(6): 476.
- [10] 冯钧, 潘飞. 一种LSTM-BP多模型组合水文预报方法[J]. 计算机与现代化, 2018(7): 82-85.
- [11] ZHANG D, LIN J, PENG Q, et al. Modelling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm [J]. Journal of Hydrology, 2018, 56(5): 720-736.
- [12] 马凯凯, 李士进, 王继民, 等. 数据驱动的中小河流智能洪水预报方法对比研究[J]. 中国科学技术大学学报, 2016, 46(9): 774-779.
- [13] Hochreiter, Sepp, Schmidhuber, et al. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [14] CUI Z, ZHOU Y, GUO S, et al. A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting [J]. Hydrology Research, 2021(3): 1436-1454.
- [15] 周研来, 郭生练, 张斐章, 等. 人工智能在水文预报中的应用研究[J]. 水资源研究, 2019, 8(1): 1-12.
- [16] DING Y, ZHU Y, FENG J, et al. Interpretable spatio-temporal attention LSTM model for flood forecasting [J]. Neurocomputing, 2020, 40(3): 348-359.
- [17] LI K, WAN D, ZHU Y, et al. The applicability of ASCS-LSTM-ATT model for water level prediction in small- and medium-sized basins in China [J]. Journal of Hydroinformatics, 2020, 22(6): 1693-1717.