

DOI:10.19853/j.zgjsps.1000-4602.2022.19.021

基于DM-LSTM的城市降雨径流预测研究

崔忠捷¹, 卿晓霞², 杨森雄³

(1. 重庆大学 环境与生态学院, 重庆 400045; 2. 重庆大学 土木工程学院, 重庆 400045; 3. 中国电建贵阳勘测设计研究院有限公司, 贵州 贵阳 550000)

摘要: 智慧水务背景下,如何基于人工智能理论与技术深化城市降雨径流模型研究,是一项值得探索的课题。由于城市降雨径流时间分辨率高且样本特征分布不具有规律性,直接采用长短期记忆(LSTM)模型进行预测面临着挑战。基于此,提出用数据挖掘(DM)算法及规则对城市降雨径流时序数据集进行聚类 and 重构,并基于深度学习算法对LSTM模型的结构和参数进行优化,构建了DM-LSTM耦合模型,并用于研究区域的降雨径流模拟。结果表明,对于各类降雨事件,与LSTM模型相比,DM-LSTM耦合模型的均方根误差(RMSE)降低了2.1%~41.9%,纳什效率系数(NSE)提高了0.4%~56.4%,决定系数(R^2)提高了0.3%~65.6%。DM-LSTM耦合模型不仅对各类降雨事件均表现出更好的预测性,而且模型运行时间仅为2.044 s,能够很好地满足城市降雨径流预测对实时性、准确性和稳定性的需求。

关键词: 城市降雨径流预测; 数据驱动模型; 长短期记忆(LSTM)模型; 数据挖掘; 深度学习; 智慧水务

中图分类号: TU992 **文献标识码:** A **文章编号:** 1000-4602(2022)19-0132-07

Prediction of Urban Rainfall Runoff Based on DM-LSTM

CUI Zhong-jie¹, QING Xiao-xia², YANG Sen-xiong³

(1. College of Environment and Ecology, Chongqing University, Chongqing 400045, China;
2. School of Civil Engineering, Chongqing University, Chongqing 400045, China; 3. PowerChina
Guiyang Engineering Corporation Limited, Guiyang 550000, China)

Abstract: Under the background of smart water, how to deepen the research on urban rainfall runoff model based on artificial intelligence theory and technology is a topic worthy of exploration. Due to the high temporal resolution of urban rainfall runoff and irregular distribution of sample features, it is challenging to directly use long short term memory (LSTM) model for prediction. Based on this background, data mining (DM) algorithm and rules were proposed to cluster and reconstruct the urban rainfall runoff time series data sets, the structure and parameters of LSTM model were optimized based on deep learning algorithm, and a DM-LSTM coupling model was constructed and applied to simulate rainfall runoff in the study area. For all kinds of rainfall events, compared with the LSTM model, the root mean square error (RMSE) of the DM-LSTM coupling model was decreased by 2.1%~41.9%, the Nash-Sutcliffe efficiency

基金项目: 国家重点研发计划项目(2017YFC0404704); 重庆市教委科学技术研究项目(KJZD-K202100104); 重庆市科委社会民生类重点研发项目(cstc2018jszx-zdyfxmX0010)

通信作者: 卿晓霞 E-mail: 13983707801@126.com

(NSE) coefficient was increased by 0.4%–56.4%, and the coefficient of determination (R^2) was increased by 0.3%–65.6%. The DM-LSTM coupling model showed a better prediction performance for all kinds of rainfall events, and decreased the running time of the model to only 2.044 s, which could well meet the needs of real-time, accuracy and stability for urban rainfall runoff prediction.

Key words: urban rainfall runoff prediction; data-driven model; LSTM model; data mining (DM); deep learning; smart water

降雨径流预测模型可分为过程驱动模型和数据驱动模型,前者基于水文、水力学产汇流计算和水力计算,需要求解复杂的物理方程;后者直接从数据中学习物理现象的长期规律,建立输入与输出数据之间的映射关系,从而拥有更快的运算速率和良好的模拟性能^[1-2]。近年来,我国正大力推进智慧水务建设,尤其是在城市供排水管网系统的智能监控与调度、城市水系统智慧化运营管理平台构建方面进行了积极的探索和实践^[3]。随着智慧水务系统的持续运行,海量数据的积累为研究水系统信息挖掘、建模和分析创造了很好的条件。如何在深度学习和数据挖掘理论的基础上,进一步深化城市降雨径流数据驱动模型的研究以提高其预测性,是一项值得探索的课题。

深度学习作为机器学习发展的更高级形式,凭借其自动提取数据特征的优势,已在计算机视觉、自然语言处理等领域得到成功应用。由于降雨径流是一种内联相关性极强的时序数据,其某一时刻的数据与之前时刻的数据之间存在着密切的关系,因此要求预测降雨径流的数据驱动模型具有“记忆”的能力。为此,具有循环神经网络架构的深度学习网络——长短期记忆(LSTM)模型被引入水文预测领域,并取得了良好的效果^[4-8]。目前LSTM模型在水文领域主要用于江河流域^[9-12]。江河流域径流数据属于时间分辨率较低(日尺度、月尺度)且特征分布具有明显周期性的中长期时序数据,而城市降雨径流数据具有在单场降雨过程中连续但在整个数据集中离散的特点,不仅时间分辨率高,而且样本特征分布不具有规律性。由于LSTM模型的本质是寻找输入数据与输出数据之间的最优映射关系,对数据集中的样本特征分布十分敏感^[13],用于处理数据特征不显著的城市降雨径流时,模拟精度会受到严重影响。因此,如何提高城市降雨径流数据驱动模型的预测精度是必须面对的问题。数据

挖掘(DM)是从数据中析取有用的、先前未知但最终能变为可理解的知识的过 程^[14]。其基于人工智能、机器学习、统计学等理论及方法,能够在自动分析和处理数据的基础上,对数据集中的属性、模式产生频度、对象簇集等进行描述,重构时序数据和样本特征,有望用于解决上述问题。

为此,笔者在采用DM理论与技术对研究区域降雨数据集进行聚类 and 重构的基础上,提出一种耦合DM与LSTM的数据驱动模型,以提高城市降雨径流的预测精度和稳定性,旨在为智慧水务背景下的城市内涝防控提供技术支撑。

1 研究区域与数据来源

研究区域位于重庆两江新区悦来新城生态城内,如图1所示,地形上东高西低、南高北低,海拔高程为200~350 m,规划总面积为3.53 km²。其西邻嘉陵江,东邻张家溪,下垫面情况复杂,地势起伏多变,具有典型的山地丘陵特征。降雨集中在5月—9月,多年平均降雨量为1 078 mm,容易发生内涝灾害。

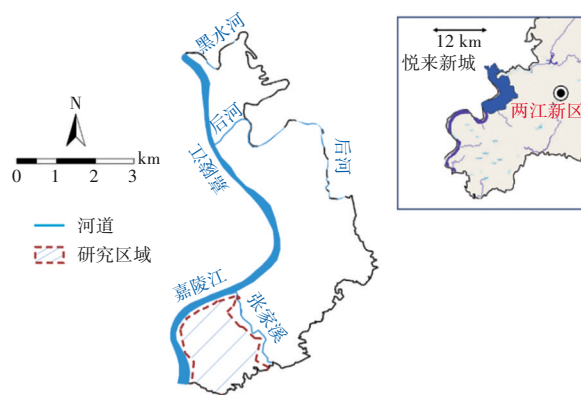


图1 研究区域位置

Fig.1 The case study catchments

以重庆市悦来海绵城市监测与信息平台提供的2018年—2019年降雨及径流数据作为原始数据,采样周期为10 min。考虑到降雨强度小的事件不会

导致城市内涝,因此选择24 h降雨总量 ≥ 10 mm的降雨时序数据,共计9 288条,并以6 h作为最小时间间隔进行划分^[15],得到43场独立降雨事件,编号为R1~R43。

2 研究方法

2.1 基于LSTM的深度学习网络模型

LSTM模型设计旨在从时序数据中学习内在关联信息,其“记忆单元”由3个门控机制和1个候选机制构成。图2为LSTM的网络结构。输入门 I_t 用来选择新的信息以更新状态;遗忘门 F_t 对之前的信息进行选择丢弃;输出门 O_t 确定让当前时间步产生的信息有多少能传递到下一时间步;记忆细胞 C_t 通过遗忘门和输入门来控制信息的流动。LSTM神经网络由若干个“记忆单元”组成,其数量一般由输入数据时间跨度决定,即输入数据每一个步长对应一个LSTM“记忆单元”。这些“记忆单元”根据时间先后顺序逐个连接,形成一个完整的LSTM网络。

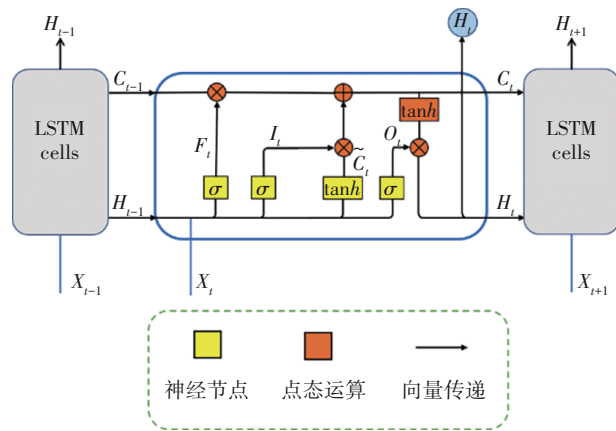


图2 LSTM的网络结构

Fig.2 LSTM network framework

若给定 t 时刻的输入数据 $X_t \in R^{n \times d}$ 和上一时刻的隐藏状态 $H_{t-1} \in R^{n \times h}$,则 t 时刻的输入门 I_t 、遗忘门 F_t 、输出门 O_t 的计算公式如下:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \quad (1)$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \quad (2)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \quad (3)$$

LSTM用候选记忆细胞 \tilde{C}_t 记忆网络当前时刻数据的信息,其计算公式如下:

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \quad (4)$$

当前时刻的记忆细胞 $C_t \in R^{n \times h}$ 由上一时刻的记忆细胞和当前时刻的候选记忆细胞的信息计算得到,即:

$$C'_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (5)$$

式中: W_{xi} 、 W_{xf} 、 W_{xo} 分别为输入门、遗忘门和输出门的输出层权重; W_{xc} 、 W_{hc} 分别为候选记忆细胞的输入层和隐藏状态权重; W_{hi} 、 W_{hf} 、 W_{ho} 分别为输入门、遗忘门和输出门的隐藏状态权重; b_i 、 b_f 、 b_o 、 b_c 分别为输入门、遗忘门、输出门和候选记忆细胞的偏差值; σ 为值域在0~1的Sigmoid函数; $\tanh(\cdot)$ 为值域在-1~1的双曲正切激活函数; \odot 为矩阵元素乘法。

计算隐藏状态变量 $H_t \in R^{n \times h}$:

$$H_t = O_t \odot \tanh(C'_t) \quad (6)$$

计算 t 时刻的输出 y_t :

$$y_t = H_t \odot W_{hy} + b_{hy} \quad (7)$$

式中: W_{hy} 为输出层权重值; b_{hy} 为输出层偏差值。

除了复杂的网络结构外,深度学习网络之所以具有更好的学习能力,还在于其复杂的机器学习算法。由于Adam算法具有收敛快、精度高和稳定性强等优点,因此采用Adam算法对LSTM模型进行优化计算,步骤如下:

① 数据归一化处理。处理后使寻优过程更平缓,能更加准确、快速地收敛到最优点,见式(8):

$$x'_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

式中: x_{\min} 为源数据中的最小值; x_{\max} 为源数据中的最大值; x_t 为源数据中 t 时刻的值; x'_t 为归一化后 t 时刻的值。

② 模型超参数及目标函数的设置。其主要包括初始化超参数神经元的随机舍弃率(Dropout)、一次训练样本数(Batch Size)、优化算法学习率(Learning Rate)和迭代次数(Epochs)。本研究选取均方误差(MSE)作为损失函数,见式(9):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

式中: y 为实测数据样本; \hat{y} 为模型预测数据样本; y_i 和 \hat{y}_i 分别为第 i 时刻的实测数据与预测数据; n 为样本中的数据总量。

③ 模型内部参数初始化。利用均值为0、标准差为0.01的正态分布向量初始化输入门、输出门、遗忘门、候选记忆细胞等参数。

④ 模型前向传播计算。将独立降雨事件数据输入模型,通过前向传播计算每一时刻的参数、隐藏状态和输出预测值。第Epochs次迭代得到的

输出和第 t 时刻的参数为:

$$\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_t, \dots, \hat{y}_T] \quad (10)$$

$$\text{para}(t) = [w_{i,t}^T, w_{f,t}^T, w_{o,t}^T, w_{c,t}^T, b_{i,t}^T, b_{f,t}^T, b_{o,t}^T, b_{c,t}^T] \quad (11)$$

式中: $[w_{i,t}^T, w_{f,t}^T, w_{o,t}^T, w_{c,t}^T, b_{i,t}^T, b_{f,t}^T, b_{o,t}^T, b_{c,t}^T]$ 为隐藏层输入门、遗忘门、输出门和候选细胞的权重向量和偏差向量。

⑤ 误差计算与模型反向传播计算。首先计算MSE值,并记录;然后模型反向传播计算神经网络各层的中间变量和参数的梯度,并保存。

⑥ 利用Adam优化算法更新权重。通过⑤保存在网络中各参数的梯度对各参数进行更新。

⑦ 判断是否满足迭代的终止条件,否则令Epochs=Epochs+1,返回④。

2.2 基于DM的降雨时序数据处理

本研究数据集中各独立降雨事件的降雨历时、雨峰位置系数、峰值比例等描述雨型特征的参数差异极大。由于雨型特征能够描述降雨过程的雨量分布情况,不同的雨量分布会导致径流在时间尺度上的变化。因此,平均降雨强度相同的降雨事件会因为雨型特征不同而产生不同的径流结果。为此,基于DM方法对降雨事件进行处理,并根据处理结果重构数据集,使数据驱动模型能更准确地学习信息,从而提高其预测性能。

聚类分析是DM中一种无监督的机器学习方法,其目的是按照某种相似性度量准则将无标签的数据集划分成若干个不同的子集,使得同一个子集内的数据彼此相似,不同子集的数据间具有差异性。其中,层次聚类具有处理多种属性、多种形状数据集的能力,能够反复对数据进行分类和聚合,不需预先设定类的数量,并能直观反映不同尺度上数据集的聚类情况,最后以树状图的形式呈现结果。本研究选择层次聚类算法进行聚类分析。层次聚类的连接规则主要有最小距离规则(Single Linkage)、最大距离规则(Complete Linkage)和组间平均距离规则(Average Linkage)等。根据本数据集特点,采用Average Linkage连接规则,其数学表达式如下:

$$\text{dist}(o_i, o_j) = (1/n_i n_j) \sum_{x \in o_i} [\sum_{y \in o_j} \text{dist}(x, y)] \quad (12)$$

式中: o_i 和 o_j 分别表示样本划分后的第 i 类和第 j 类; x 和 y 表示样本; n_i 和 n_j 分别表示 o_i 和 o_j 中的样

本个数; $\text{dist}(x, y)$ 表示样本 x 与 y 之间的距离; $\text{dist}(o_i, o_j)$ 表示第 i 类(或第 i 个簇)与第 j 类(或第 j 个簇)之间的距离。

相似性度量是聚类算法实现的关键,目的是分析序列之间是否具有相似性特征,如序列的形状、大小等是否相似。针对降雨时序数据,采用动态时间规整算法(DTW)进行相似性度量。

由于降雨时序数据具有高维度属性,不能直接作为输入数据进行聚类计算。首先需将每场降雨事件看作一个独立时序集 X_i , 然后对不同序列集之间的距离进行计算,构造距离矩阵,最后通过层次聚类的连接规则进行分类。计算步骤如下:

① 给定降雨时间序列集 $\{X_1, X_2, \dots, X_{43}\}$ 、目标类的个数 $N(0 < N \leq 43)$ 。

② 把每一个序列归为一类,然后计算两两序列间的DTW距离,采用Average Linkage作为两个类之间的连接规则,构建不同序列之间的距离矩阵 D 。其中序列 X_i 和 X_j 在矩阵中的距离为 D_{ij} 。

③ 若矩阵 D 中的最小元素为 D_{ij} , 则将序列 X_i 和 X_j 合并为一类。

④ 重新计算类间距离。假设类 o_i 和类 o_j 是已经存在的类,则类 o_i 和类 o_j 之间的类间距离为:

$$\text{avgdist}(o_i, o_j) = \frac{1}{|o_i| \cdot |o_j|} \sum_{X_i \in o_i} \sum_{X_j \in o_j} \text{DTW}(X_i, X_j) \quad (13)$$

⑤ 重复步骤③和④,直至满足计算终止条件。

2.3 基于DM-LSTM的降雨径流模型

由于深度学习神经网络模型结构在很大程度上决定了模型的预测效果。本研究采用网络层数分别为1、2、3、4,对应的神经元数分别为8、16、32、64、128、256、1 024的组合方案,同时以43场降雨事件的预测值与实测值的均方根误差(RMSE)作为评价指标,采用Adma算法对LSTM模型的网络结构进行优化,迭代次数为200,得到LSTM最佳网络结构为(2, 1 024),模型超参数Learning Rate为0.001, Batch Size为5。

DM-LSTM耦合模型。首先基于DM对数据集进行层次聚类并重构 N 个数据子集,然后将重构后的数据子集分别划分训练集和测试集,并对各LSTM模型进行训练,最后集成为DM-LSTM耦合模

型用于预测。

2.4 性能评估指标

选择 RMSE、纳什效率系数(NSE)、决定系数(R^2)、峰值相对误差(PE)4个量化指标对模型进行验证,各指标的计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Q_{\text{obs},t} - Q_{\text{fore},t})^2} \quad \text{RMSE} > 0 \quad (14)$$

$$\text{NSE} = 1 - \frac{\sum_{t=1}^n (Q_{\text{obs},t} - Q_{\text{fore},t})^2}{\sum_{t=1}^n (Q_{\text{obs},t} - Q_{\text{obs,mean}})^2} \quad (15)$$

$$R^2 = \frac{\sum_{t=1}^n [(Q_{\text{fore},t} - Q_{\text{fore,mean}})(Q_{\text{obs},t} - Q_{\text{obs,mean}})]}{\sum_{t=1}^n (Q_{\text{obs},t} - Q_{\text{fore,mean}})^2 \sum_{t=1}^n (Q_{\text{obs},t} - Q_{\text{obs,mean}})^2} \quad (16)$$

$$\text{PE} = \frac{Q_{\text{fore}}^{\text{peak}} - Q_{\text{obs}}^{\text{peak}}}{Q_{\text{obs}}^{\text{peak}}} \quad (17)$$

式中: $Q_{\text{obs},t}$ 和 $Q_{\text{fore},t}$ 分别表示第 t 个时刻的实测流量和预测流量; $Q_{\text{obs,mean}}$ 和 $Q_{\text{fore,mean}}$ 分别表示实测流量和预测流量的均值; $Q_{\text{obs}}^{\text{peak}}$ 和 $Q_{\text{fore}}^{\text{peak}}$ 分别表示实测流量和预测流量的峰值。

3 结果与讨论

3.1 基于DM处理的降雨时间序列数据

图3为基于DM处理的43场降雨事件结果。图3中数值表示叶子节点簇之间的无量纲距离,根据确定的距离值即可划分簇的数量。以无量纲距离0.965划分树状图,共得到7个簇。其中1、2、7号簇的降雨事件数均为1场;3、4号簇的降雨事件数均为2场;5号簇的降雨事件数为9场,6号簇为27场。进一步地,将6号簇的降雨事件重构为A类数据子集,占总降雨事件场次数的62.8%;5号簇的降雨事件重构为B类数据子集,占总降雨事件场次数的20.9%。考虑到1、2、3、4、7号簇中7场降雨事件的降雨历时普遍偏长,且与其他样本的相似度较小,为了平衡重构数据集的数量和数据类型,将其组成为C类数据子集,占比为16.3%。由图3可以看出,A类子集多为雨峰位置靠前的单峰雨型,并且峰值较高、降雨历时较短;B类子集多为双峰或多峰雨型,并且雨峰之间有一定的时间间隔,降雨历时适中,雨量呈衰减趋势;C类子集的雨型具有较多的雨

峰,并且峰值较小,呈锯齿状,雨量分布较为均匀,降雨历时偏长。

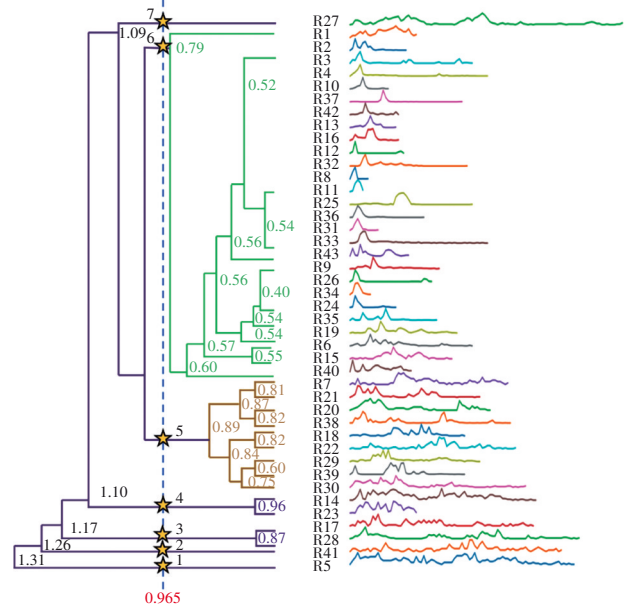


图3 DM得到的降雨时序数据结果

Fig.3 Results of rainfall pattern clustering for rainfall events by DM

3.2 模型预测性能评估

用37场训练集数据对LSTM和DM-LSTM模型进行训练后,分别从A、B、C三类数据集中各选取两场降雨事件构成测试集数据。其中,A类选取R11和R31,B类选取R21和R39,C类选取R14和R41。用LSTM模型和DM-LSTM耦合模型进行径流预测,表1和图4分别展示了两模型的径流预测性能指标和径流预测过程趋势(间隔10 min采样1次)。

表1 LSTM与DM-LSTM的模拟性能指标

Tab.1 Evaluation index of LSTM and DM-LSTM

类别	降雨事件编号	模型	RMSE/ ($\text{m}^3 \cdot \text{s}^{-1}$)	R^2	NSE	PE
A类	R11	LSTM	0.009 3	0.893	0.860	-0.07
		DM-LSTM	0.007 9	0.912	0.900	-0.09
	R31	LSTM	0.002 2	0.926	0.908	-0.13
		DM-LSTM	0.002 1	0.929	0.912	-0.14
B类	R21	LSTM	0.005 7	0.824	0.823	-0.08
		DM-LSTM	0.005 3	0.856	0.845	-0.35
	R39	LSTM	0.026 8	0.798	0.772	-0.14
		DM-LSTM	0.020 6	0.878	0.866	-0.20
C类	R14	LSTM	0.026 0	0.640	0.607	-0.14
		DM-LSTM	0.015 1	0.905	0.867	-0.28
	R41	LSTM	0.009 4	0.512	0.500	0.02
		DM-LSTM	0.006 2	0.848	0.782	-0.29

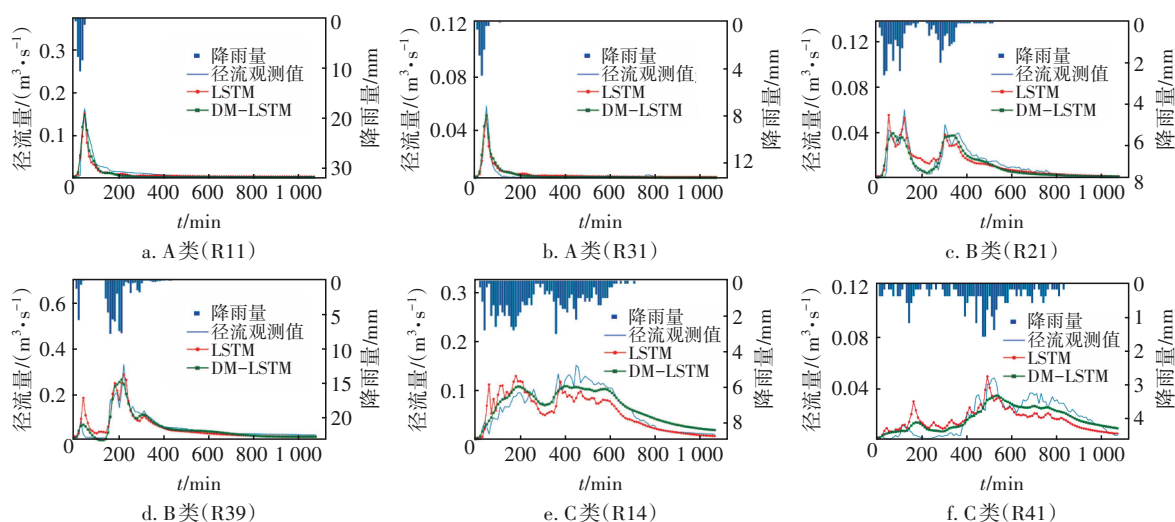


图4 LSTM与DM-LSTM的模拟结果

Fig.4 Simulation results of LSTM and DM-LSTM

从表1和图4可以发现,在A、B、C三类降雨事件中,DM-LSTM耦合模型均显示出更好的预测效果。对于A类降雨事件,DM-LSTM耦合模型相较于LSTM模型,其RMSE最大降低了15.1%, R^2 最大提高了2.1%,NSE系数最大提高了4.7%。在B类降雨事件中,DM-LSTM耦合模型比LSTM模型的RMSE值最大降低了23.1%, R^2 最大提高了10.0%,NSE系数最大提高了12.2%。对于C类降雨事件,DM-LSTM耦合模型比LSTM模型的RMSE值最大降低了41.9%, R^2 最大提高了65.6%,NSE系数最大提高了56.4%。表明DM-LSTM耦合模型对各类降雨事件在预测精度、可信度、相关性等方面都有显著提升,表现出良好的稳定性,能整体提高对各类降雨事件的预测性能。且由图4可见,DM-LSTM耦合模型的径流预测过程曲线更加平滑,说明其对径流过程变化趋势的捕捉更为准确。

此外,DM-LSTM耦合模型的建模效率和运行效率的实验结果也表现出非常突出的优势,其建模时间为377 s,单次运行时间为2.044 s(软件开发环境为Windows10 64位操作系统支持的Python,硬件配置为AMD Ryzen5 2600六核十二线程3.85 Hz的CPU以及8 GB的运行内存),完全能够满足城市降雨径流实时预测需要。

观察DM-LSTM耦合模型的峰值流量预测指标,发现其PE值介于-0.35~0.09之间,相比于LSTM模型仅有微小的提升甚至出现下降现象。分析可能是由以下原因导致:①虽然聚类处理后,具

有相似特征的数据归在一类数据集中能提高模型的整体预测性能,但由于重构后数据集的规模减小,使得模型可学习到的知识减少,导致对某些降雨事件的峰值预测效果出现回落的现象;②本模型尚缺乏局部关注能力,在处理时序数据的局部突变(如雨峰的出现使径流量突增)与学习历史信息时,不能自适应加大突变信息的权值以提升对降雨径流过程突变特征的提取能力。

4 结论

① 提出了一种DM-LSTM耦合模型,该模型基于DM对数据集进行聚类与重构,并采用Adam算法对LSTM模型的结构和参数进行优化,以提高LSTM模型对时间分辨率高且样本特征分布不具有规律性的城市降雨径流预测性能。

② 基于DTW算法和层次聚类的Average Linkage连接规则,将43场独立降雨事件划分为3类数据子集。各子集内表征降雨事件雨型特征的雨峰数量、位置、大小和降雨历时等彼此相似,而各子集间上述雨型特征差异显著,使得LSTM模型能针对不同雨型特征的各类降雨事件进行更好的学习和预测。

③ 与LSTM模型相比,DM-LSTM耦合模型对各类降雨事件表现出了更好的预测性能,且稳定性好,不仅预测精度、可信度和相关性等各项指标都有显著提升,而且径流预测过程曲线更平滑,对径流过程变化趋势的捕捉更准确。此外,DM-LSTM耦合模型的建模效率和运行效率极高,完全能够满

足城市降雨径流实时预测的需要。

④ 尽管耦合数据挖掘方法提高了LSTM模型的预测性能,但其对径流峰值的预测能力尚有不足,可通过扩大数据集规模和采用深度学习的注意力机制进一步改善模型性能。

参考文献:

- [1] BARTOLETTI N, CASAGLI F, MARSILI-LIBELLI S, *et al.* Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system [J]. *Environmental Modelling & Software*, 2018, 106:35-47.
- [2] CHANG W, CHEN X. Monthly rainfall-runoff modeling at watershed scale: a comparative study of data-driven and theory-driven approaches[J]. *Water*, 2018, 10(9): 1116.
- [3] 张金松, 李旭, 张炜博, 等. 智慧水务视角下水务数字化转型的挑战与实践 [J]. *给水排水*, 2021, 47(6):1-8.
ZHANG Jinsong, LI Xu, ZHANG Weibo, *et al.* Challenges and practice of water digital transformation from the view of smart water [J]. *Water & Wastewater Engineering*, 2021, 47(6):1-8 (in Chinese).
- [4] WU C C, ZHANG X Q, WANG W J, *et al.* Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model [J]. *Science of the Total Environment*, 2021, 783:146948.
- [5] 胡庆芳, 曹士圯, 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.
HU Qingfang, CAO Shiyi, YANG Huibin, *et al.* Daily runoff predication using LSTM at the Ankang Station, Hanjiang River [J]. *Progress in Geography*, 2020, 39(4):636-642 (in Chinese).
- [6] FENG D P, FANG K, SHEN C P. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales [J]. *Water Resources Research*, 2020. DOI:10. 1029/2019WR026793.
- [7] TENNANT C, LARSEN L, BELLUGI D, *et al.* The utility of information flow in formulating discharge forecast models: a case study from an arid snow-dominated catchment [J]. *Water Resources Research*, 2020. DOI: 10. 1029/2019WR024908.
- [8] CUI Z J, QING X X, CHAI H X, *et al.* Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis [J]. *Journal of Hydrology*, 2021, 603:127124.
- [9] AN L X, HAO Y H, YE H T, *et al.* Simulation of karst spring discharge using a combination of time-frequency analysis methods and long short-term memory neural networks [J]. *Journal of Hydrology*, 2020, 589: 125320.
- [10] ZHANG J W, CHEN X H, KHAN A, *et al.* Daily runoff forecasting by deep recursive neural network [J]. *Journal of Hydrology*, 2021, 596:126067.
- [11] ZUO G G, LUO J G, WANG N, *et al.* Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting [J]. *Journal of Hydrology*, 2020, 585:124776.
- [12] GAO S, HUANG Y F, ZHANG S, *et al.* Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation [J]. *Journal of Hydrology*, 2020, 589:125188.
- [13] GUO X, DU Z L, LIU C, *et al.* A new uncertainty analysis-based framework for data-driven computational mechanics [J]. *Journal of Applied Mechanics*, 2021, 88(11):111003.
- [14] 尹刚, 王涛, 刘冰珣, 等. 面向开源生态的软件数据挖掘技术研究综述 [J]. *软件学报*, 2018, 29(8): 2258-2271.
YIN Gang, WANG Tao, LIU Bingxun, *et al.* Survey of software data mining for open source ecosystem [J]. *Journal of Software*, 2018, 29(8): 2258-2271 (in Chinese).
- [15] WANG W T, YIN S Q, XIE Y, *et al.* Effects of four storm patterns on soil loss from five soils under natural rainfall [J]. *Catena*, 2016, 141: 56-65.

作者简介:崔忠捷(1995-),男,安徽宿州人,硕士研究生,主要研究方向为数值模拟、海绵城市技术。

E-mail:17815382014@163.com

收稿日期:2022-02-20

修回日期:2022-05-09

(编辑:任莹莹)