

DOI:10.3876/j.issn.1000-1980.2021.01.002

基于时空特征挖掘的流量过程智能模拟方法

朱跃龙,赵 群,余宇峰,万定生

(河海大学计算机与信息学院,江苏 南京 211100)

摘要:为减轻洪水灾害事件可能带来的严重后果,实现对流量的及时、准确预测,提出一种基于时空特征挖掘的流量过程智能模拟方法。该方法首先从空间角度入手,建立测站之间的拓扑结构关系;再利用图卷积网络进行空间挖掘;最后利用门控循环单元进行时序挖掘。试验结果表明,基于时空特征挖掘的流量过程智能模拟方法比基于单一特征的模拟方法效果更好。

关键词:时空特征;智能模拟方法;水文预报;图卷积网络;门控循环单元

中图分类号:TP39 **文献标志码:**A **文章编号:**1000-1980(2021)01-0007-06

Intelligent simulation method of runoff process based on spatiotemporal feature mining

ZHU Yuelong, ZHAO Qun, YU Yufeng, WAN Dingsheng

(College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: To reduce the serious consequences of flood disaster events, the river discharge should be predicted timely and accurately to provide decision support for the flood forecasting. In this study, an intelligent simulation method of runoff process based on the spatiotemporal feature mining was proposed. Firstly, the topological diagram of hydrological stations was established from the spatial point of view. Then, this study proposed a novel framework that incorporated the graph convolutional networks (GCN) for mining the spatial feature. The gated recurrent unit was adopted to capture temporal features for the hydrological prediction. The experimental results show that the intelligent hydrological forecasting model based on the spatiotemporal feature mining is better than other models with single feature.

Key words: spatiotemporal feature; intelligent simulation method; hydrological forecast; graph convolutional network; gated recurrent unit

近年来,全球气候变暖,极端降水事件频发导致了各种洪涝灾害,给人类带来巨大的经济社会损失。仅 2020 年全国共出现 45 次强降水过程,面降水量 622 mm,较常年同期偏多 13%,引发珠江、长江、黑龙江 3 个流域极端严重的流域性洪涝灾害^[1]。为了减轻洪水灾害带来的不利影响,及时准确的水文预报显得尤为重要。水文预报是根据前期的水文气象资料,对未来一定时期的水文资料进行预报,对防洪救灾的决策、水资源的合理利用等具有重要意义^[2]。

随着大数据时代的到来,基于数据驱动的水文预报模型在水文事件预报中越来越受到重视^[3-4]。数据驱动的水文预报模型从历史水文资料出发,挖掘规律进行预测,为相关水利人员提供了一个新的预测方案。但现有基于数据驱动的水文模型大多侧重于从历史数据中学习时间特征,很少兼顾到数据的空间特征^[5]。水文数据往往包含丰富的时空特征信息,而该特征信息可以为水文时空序列相似性,水文时空数据建模分析等作出贡献^[6],本文提出一种基于时空特征挖掘的流量过程智能模拟方法 GS-Hydro,构建水文结构关系图,利用图卷积网络挖掘空间特征的同时利用门控循环单元捕捉时间特征,以此提高预测精度。

1 水文预报模型相关研究

1.1 基于数据驱动的洪水预报模型

数据驱动模型是从数据着手,通过对时间序列数据的分析,从大量的数据中发现和揭示数据的发展变化

基金项目:国家重点研发计划(2018YFC1508100)

作者简介:朱跃龙(1959—),男,教授,博士,主要从事智能数据处理、水利信息化和智能洪水预报研究。E-mail:ylzhu@hhu.edu.cn

通信作者:赵群,博士研究生。E-mail:qunzhaoqz@outlook.com

引用本文:朱跃龙,赵群,余宇峰,等.基于时空特征挖掘的流量过程智能模拟方法[J].河海大学学报(自然科学版),2021,49(1):7-12.

ZHU Yuelong, ZHAO Qun, YU Yufeng, et al. Intelligent Simulation Method of Runoff Process based on Spatiotemporal Feature Mining [J]. Journal of Hohai University(Natural Sciences), 2021, 49(1): 7-12.

规律,以此估测和控制未来可能出现的状况。基于数据驱动的洪水预报模型从历史水文资料出发,挖掘规律进行预测,相较于其他水文模型具有普适性。

Hwang 等^[7]提出一种基于最小二乘支持向量机(LS-SVM)的非线性水文时间序列预测方法,具有较好的预测精度和预测性能。Xing 等^[8]提出一种新的启发式优化算法 BA 算法,用于优化 SVM 参数和预测 2015 年月平均流量,其预测精度比交叉验证的支持向量机更高。Atiquzzaman 等^[9-10]针对传统基于梯度的慢学习算法在训练和迭代确定网络参数时的不确定性问题,提出一种基于极限学习机(ELM)的水文时间序列快速预测方法。根据洪水形成的基本原理,Chen 等^[11]提出一种基于数据驱动和机器学习的卷积回归水文预报模型,该模型能够反映小时降雨量对未来流量变化的影响,并通过叠加这些影响预测流量变化。

1.2 基于时空特征的预报模型

除了时序特征,自然界大部分的数据还具有空间结构联系。相比于单一的时序预测,基于时空特征结合的预测考虑了数据之间的空间特性,能够挖掘出更多的信息规律,提高预测精度,而空间特征与时间特征的结合预报也被广泛地运用到了各大领域^[12-13]。

Ding 等^[14]利用长短期记忆模型,提出洪水预报模型(STA-LSTM 模型),动态提取关键特征向量,实现中小河流洪水预报。除此之外,空间特征也可用于水文学的许多方面,如水文相似性分析^[15]、流域场景模式库构建^[16]、用水效率分析^[17]等。

目前基于数据驱动的水文预报模型大多考虑时序特征,很少考虑河流之间的地理联系和空间特征。单一测站的水文数据具有时间连续性、周期性,而将多测站的水文数据放在一起,不难看出它们之间的空间联系。本文提出一种基于地理关系的时空智能流量过程模拟模型,通过地理关系建立测站之间的结构联系,辅助预测目标测站的流量值。

2 基于时空挖掘的智能流量过程模拟模型

现实世界中存在大量的图结构或网络结构数据,传统的卷积神经网络(convolutional neural networks, CNN)难以处理。为了解决这一问题,一种处理图结构网络的半监督分类方法——图卷积神经网络(graph convolutional networks, GCN)^[18]被提出,并被广泛应用于社会网络、信息网络等领域的特征学习。与 CNN 不同,GCN 对图像进行像素卷积,并可以挖掘网络结构。目前,GCN 主要对社会网络、信息网络等非结构化数据提取特征进行学习,然而很少有人在水文时间序列数据中使用。

本文构造了一种新的水文站点地理位置关系图,利用 GCN 挖掘空间特征,再利用门控循环单元(GRU)^[19]挖掘时间特征,进一步进行水文预报,方法结构如图 1 所示,图中 X_t 为 t 时刻的输入样本, H_t 为模型输出。

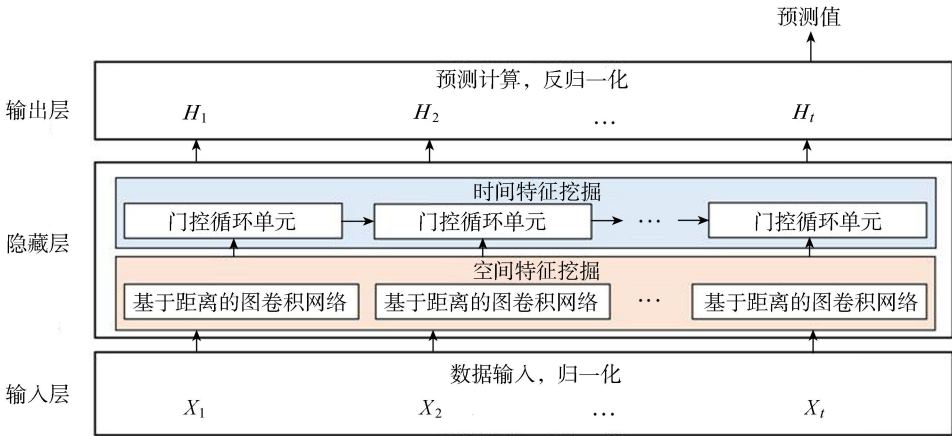


图1 GS-Hydro 模型结构
Fig.1 Structure of GS-Hydro model

输入层,以河流流量信息和当前时段面平均雨量信息为自变量,实现数据规范化、训练集划分、验证集划分、测试集划分等。隐藏层,在空间方面利用图卷积神经网络捕捉河流复杂的网络特征;在时间方面利用门控循环单元将前一时刻的信息传递到下一时刻,其中的更新门和复位门可以更好、更快地确定信息,为模

型留下更多有价值的信息。输出层,对数据进行反规范化处理,并对输出结果进行预测。

2.1 基于距离的水文测站关系结构图

2.1.1 水文测站关系结构图的建立

结构图的生成是训练图卷积网络模型的第一步,也是重要的一步。如果结构图不能很好地反映节点之间的关系,模型可能无法有效地挖掘信息。定义水文测站关系结构图如下:

$$G = (V, E) \quad (1)$$

式中: G ——水文测站地理位置关系图; V ——节点,指各个水文测站; E ——连接点边的集合,指各个水文测站间的地理关系。

同一河流2个水文测站有上下游关系,则其流量相似性将高于其他水文测站。如果2个水文测站没有直接的上下游关系,它们的流量相似性取决于其他水文过程,例如降雨过程。降雨过程驱动的水文过程很大程度上取决于地理距离,本节用水文测站的地理位置结合数字高程模型(DEM)信息确定其上下游关系、河道距离以及地理距离^[20]。如果测站a和测站b具有上下游关系,则河道距离为2个测站间的河道长度 $H_{a,b}$,否则 $H_{a,b}=0$ 。地理距离 $H'_{a,b}$ 为2个测站间的直线距离。

河道距离是目标流域到它的上游流域之间的河流长度,根据DEM估算。当每个网格中的水以最陡的坡度流向其相邻网格时,可以找到水从任何点到海洋或者从研究区域流出的流动路径。如果上游的水流经过下游的水流,则河道距离是它们之间的流路长度。如果它们没有河道联系,则河道距离为零。

邻接矩阵是存放节点间关系(边或弧)数据的数组,定义如下:

$$a_{i,j} = \begin{cases} 1 & i = j \\ \delta \frac{1}{H'_{i,j}} & i \neq j, H'_{i,j} < 100 \text{ km}, H_{i,j} = 0 \\ 0 & i \neq j, H'_{i,j} \geq 100 \text{ km}, H_{i,j} = 0 \\ \alpha \frac{1}{H_{i,j}} + \beta \frac{1}{H'_{i,j}} & i \neq j, H_{i,j} \neq 0 \end{cases} \quad (2)$$

式中: $a_{i,j}$ ——邻接矩阵中第*i*行、第*j*列的元素; α, β, δ ——矩阵中元素的系数。

对模型进行训练拟合出适合的 α, β 和 δ ,然后求出水文测站地理位置关系图的边,以此构造水文测站地理位置关系结构图。值得注意的是,当地理距离较大且没有河道距离时,2个水文测站的流量变化几乎没有关联,本文对水平距离100 km以内的站点计算节点间的关系。

2.1.2 图卷积操作

本文使用图卷积网络对各测站进行空间特征挖掘,如图2所示。

图卷积网络定义如下:

$$\mathbf{X}^{(l+1)} = f(\mathbf{X}^{(l)}, \mathbf{A}) \quad (3)$$

式中: l ——层数; $\mathbf{X}^{(l)}$ —— l 层节点的特征; $\mathbf{A}$ ——邻接矩阵。

首先输入特征矩阵和邻接矩阵,然后对节点进行特征变换,用度矩阵对邻接矩阵进行归一化。加入自循环后,考虑每个节点与相邻节点之间的关系,具体模型如下:

$$x_k^{(l+1)} = \sigma \left(\sum_{m \in N_k} \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} x_m^{(l)} \mathbf{W}^{(l)} + b^{(l)} \right) \quad (4)$$

式中: $x_k^{(l+1)}$ —— $l+1$ 层中节点*k*的特征; $x_m^{(l)}$ ——所有邻居节点的特征层*l*中节点(包括自身)的分布; σ ——非线性变换; $\tilde{\mathbf{A}}$ ——自循环的邻接矩阵; $\tilde{\mathbf{D}}$ ——自循环的度矩阵,对应于 $\tilde{\mathbf{A}}$; $\mathbf{W}^{(l)}$ —— l 层的权重; $b^{(l)}$ —— l 层的截距; N_k ——节点*k*的所有邻居节点(包括节点自身)的个数。

2.2 基于时空特征的智能流量过程模拟模型

进行空间特征提取后可以对时间特征进行处理,本文使用门控循环单元(GRU)进行时间特征的提取。GRU可以改进长短时记忆网络^[21](LSTM)的门设计,克服梯度消失问题。同时,它具有更快的收敛速度,改

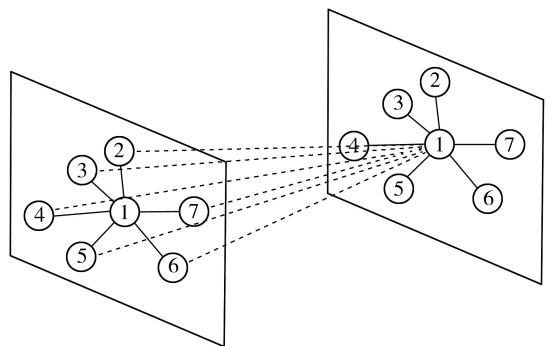


图2 空间特征提取

Fig. 2 Spatial feature extraction

善了 LSTM 训练耗时的问题。
设定模型损失函数为均方误差函数 e , 计算公式如下。

$$e = \frac{1}{n} \sum_{t=1}^n (y_t - Y_t)^2$$

(5)

式中: y_t —— t 时刻的预测值; Y_t —— t 时刻的真实值; n ——模型输入的样本数目。
模型具体计算过程如下。

a. 空间特征提取(以两层卷积为例):

$$f(X_t, A) = \sigma[\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} \sigma(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X_t W^{(0)}) W^{(1)}]$$

(6)

b. 更新门构建:

$$z_t = \sigma_z(W_z[h_{t-1}, f(X_t, A)])$$

(7)

c. 重置门构建:

$$r_t = \sigma_r(W_r[h_{t-1}, f(X_t, A)])$$

(8)

d. 神经元中待定输出:

$$\widetilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, f(X_t, A)])$$

(9)

e. 神经元输出:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h}_t$$

(10)

式中: X_t ——特征矩阵; σ_z 、 σ_r ——更新门、重置门的非线性变换; W_z 、 W_r ——更新门、重置门的权重; W_h ——神经元中待定输出的权重; h_{t-1} ——前一时刻神经元的输出; $\odot$ ——Hadamard 乘积运算, 计算重置门 r_t 与 h_{t-1} 的 Hadamard 乘积, 即计算 r_t 与 h_{t-1} 的对应元素乘积; z_t ——更新门; r_t ——重置门; $\widetilde{h}_t$ ——神经元中的待定输出; h_t ——神经元输出。

通过全连接层得到最终输出: $y_t = O(W_c h_t + b)$, 其中 $O(\cdot)$ 为线性函数; W_c 为权重; b 为偏置。

3 试验分析

3.1 数据集

试验选取江西省鄱阳湖流域外洲站作为模型流量过程测试站点, 如图 3 所示。试验选取外洲站 1998—2010 年夏季小时时段洪水流量 21 060 条数据以及降雨资料作为数据, 对区域内雨量站的降雨量进行算术平均从而获得面降雨量, 选取前 6 h 内的小时段降雨作为降雨信息对流量预测提供辅助, 区域内选取的雨量站如图 4 所示。由于部分降雨不完全, 采用水文常用的线性插值方法对缺失的雨量进行补充。试验以各个水文站为节点, 站点之间的地理关系为邻接矩阵, 各站点的历史小时流量数据和小时降雨数据为特征矩阵。试验使用 60% 的数据集作为训练数据集, 20% 作为验证数据集, 20% 作为测试数据集。

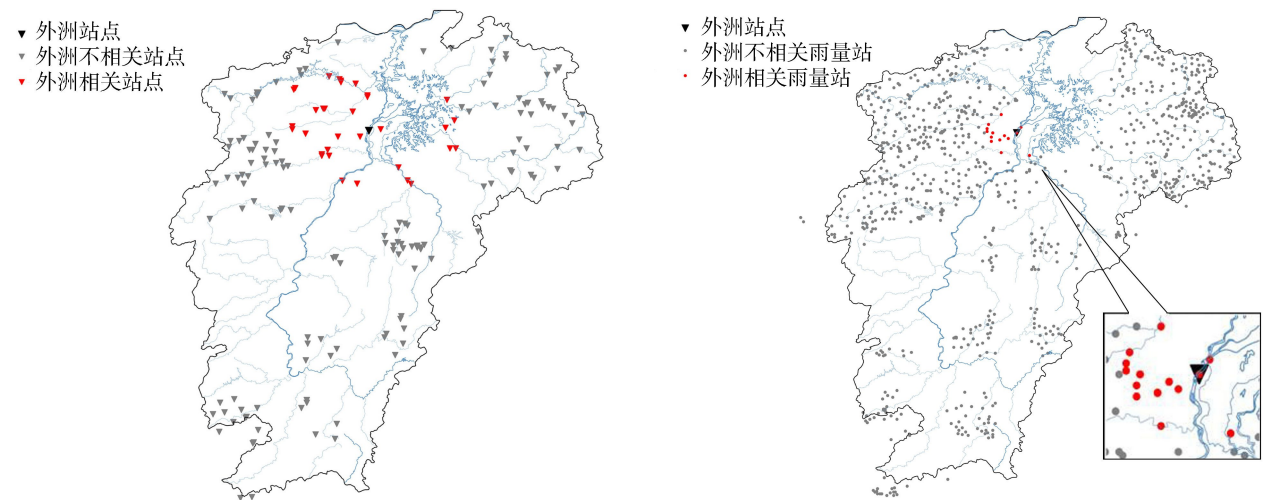


图3 江西省试验模拟测试站点

图4 外洲站周围雨量站示意图

Fig.3 Experimental simulation test stations in Jiangxi Province

Fig.4 Schematic diagram of rainfall stations around Waizhou Station

3.2 评价指标及参数选取

评价指标采用均方根误差和确定性系数。采用支持向量机模型(SVM)^[22]、长短时记忆网络(LSTM)、门

控循环单元 (GRU) 和提出的 GS-Hydro 模型进行训练。训练参数根据试验调试,最优超参配置如下:SVM 的惩罚系数为 0.3,核函数为径向基核函数,区间为 0.05。LSTM 的隐节点数为 100,学习因子的区间为 0.005,批处理大小为 128,迭代次数为 40。GUR 中隐藏节点数为 50,学习速率为 0.005,批处理大小为 32,迭代次数为 50。GS-Hydro 模型中的隐藏节点数为 32,学习速率为 0.005,批处理大小为 64,迭代次数为 100。

3.3 试验结果分析

外洲站 SVM、LSTM、GRU 和 GS-Hydro 的模型流量 3 h 预见期预测模拟计算结果如图 5 所示,不同模型在模拟预测中的评价指标见表 1。

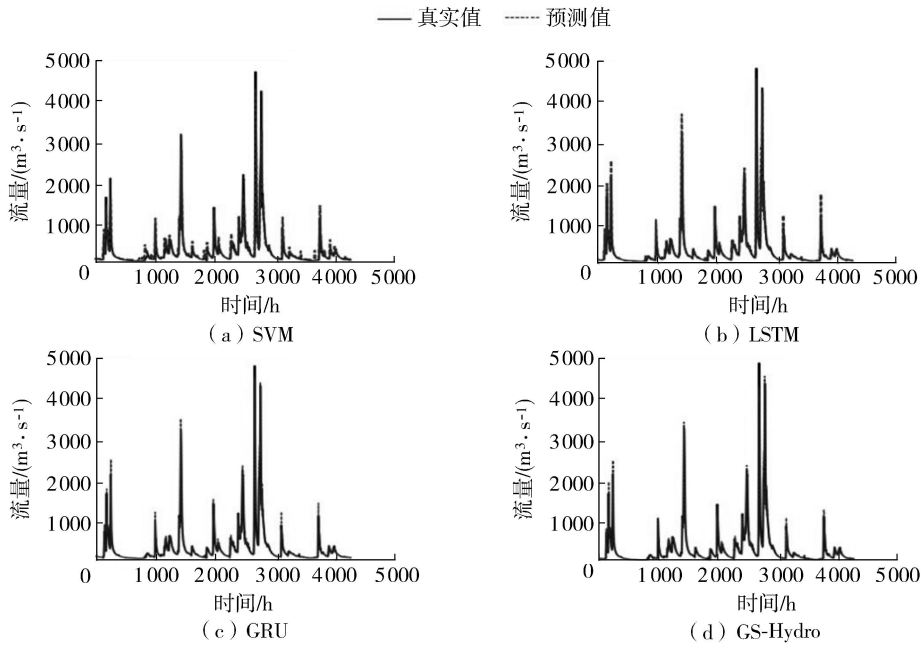


图5 江西省外洲站 3 h 预见期下各模型流量预测结果

Fig.5 Runoff prediction results of Waizhou Station in 3-hour forecast period with different models

表 1 各模型预测结果评价指标对比

预见期/h	确定性系数				均方根误差			
	SVM	LSTM	GRU	GS-Hydro	SVM	LSTM	GRU	GS-Hydro
1	0.86	0.92	0.93	0.96	128.45	92.66	92.73	72.34
2	0.83	0.92	0.91	0.95	135.49	101.23	98.54	81.67
3	0.82	0.90	0.92	0.95	161.04	104.76	103.71	87.77
4	0.80	0.89	0.90	0.93	176.47	112.23	105.85	103.72
5	0.78	0.88	0.88	0.91	192.31	128.72	126.84	117.13
6	0.76	0.82	0.84	0.89	199.83	132.66	130.98	121.32

从图 5 中可以看出,SVM 的预测结果比较抖动,得到的预测曲线不如 LSTM、GRU 和 GS-Hydro 稳定。LSTM 和 GRU 的预测结果相似,GS-Hydro 比 GRU 好,比 LSTM 更稳定。

从表 1 可以看出,1h 流量预测相对最准确,预见期越长,精度越低。GS-Hydro 是所有不同预测期的最佳模型,LSTM 和 GRU 的性能略好于 GRU,支持向量机表现最差。GS-Hydro 的确定性系数最高,LSTM 与 GRU 的差异不显著,SVM 稍差。因此,GS-Hydro 模型在整体预测结果中是最好的,是因为 GCN 捕捉了流量的空间特征,在预测过程中加入了空间内多个相关站的雨量,避免了梯度消失和均值漂移的问题。

4 结 语

本文提出了一种新的流量过程智能模拟方法 GS-Hydro,该方法将图卷积网络 (GCN) 与门控循环单元 (GRU) 相结合,先建立基于河道关系的水文站点地理位置关系图用以进行特征挖掘,再将 GCN 用于获取样本的空间特征,GRU 用于捕捉样本的时间特征。试验表明 GS-Hydro 模型的性能优于其他神经网络模型。

参考文献:

[1] 庞博. 应急管理部发布 2020 年前三季度全国自然灾害情况[EB/OL]. [2020-11-16]. http://www.gov.cn/xinwen/2020-10/08/content_5549725.htm.

[2] 包为民. 水文预报[M]. 北京:中国水利水电出版社, 2009.

[3] WANG J, SHI P, JIANG P, et al. Application of BP neural network algorithm in traditional hydrological model for flood forecasting[J]. Water, 2017, 9 (1) :48.

[4] 张弛, 王本德, 李伟. 数据挖掘技术在水文预报中的应用及水文预报发展趋势研究[J]. 水文, 2007, 27 (2) :74-77. (ZAHNG Chi, WANG Bende, LI Wei. Application of data mining technology in hydrological forecasting and research on development trend of hydrological forecasting[J]. Journal of China Hydrology, 2007, 27 (2) :74-77. (in Chinese))

[5] 程扬, 王伟, 王晓青. 水文时间序列预测模型研究进展[J]. 人民珠江, 2019, 40 (7) :18-23. (CHENG Yang, WANG Wei, WANG Xiaoqing. Research progress on hydrological time series prediction model[J]. Pearl River, 2019, 40 (7) :18-23. (in Chinese))

[6] 袁定波, 艾萍, 熊传圣. 水文时空数据挖掘方法及其应用评述[J]. 水利信息化, 2018 (1) :14-17, 22. (YUAN Dingbo, AI Ping, XIONG Chuansheng. A review of hydrological spatio-temporal data mining methods and its application [J]. Water Resources Informatization, 2018 (1) :14-17, 22. (in Chinese))

[7] HWANG S H, HAM D H, KIM J H. Forecasting performance of LS-SVM for nonlinear hydrological time series[J]. KSCE Journal of Civil Engineering, 2012, 16 (5) :870-882.

[8] XING Bing, GAN Rong, LIU Guodong, et al. Monthly mean streamflow prediction based on bat algorithm-support vector machine[J]. Journal of Hydrologic Engineering, 2015, 21 (2) :04015057.

[9] ATIQUZZAMAN M, KANDASAMY J. Prediction of hydrological time-series using extreme learning machine[J]. Journal of Hydroinformatics, 2016, 18 (2) :345-353.

[10] ATIQUZZAMAN M, KANDASAMY J. Robustness of extreme learning machine in the prediction of hydrological flow series [J]. Computers and Geosciences, 2018, 120:105-114.

[11] CHEN Chen, HUI Qiang, PEI Qingqi, et al. CRML: a convolution regression model with machine learning for hydrology forecasting[J]. IEEE Access, 2019, 7:133839 -133849.

[12] YAN Sijie, XIONG Yuanjun, LIN Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32 (1) :7444-7452.

[13] SAN K T, KYUNG L W, YOUNG S S. Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects[J]. PloS One, 2019, 14 (9) :e0220782.

[14] DING Yukai, ZHU Yuelong, FENG Jun, et al. Interpretable spatio-temporal attention LSTM model for flood forecasting[J]. Neurocomputing, 2020, 403:348-359.

[15] ZHAO Qun, ZHU Yuelong, WAN Dingsheng, et al. Similarity analysis of small- and medium-sized watersheds based on clustering ensemble model[J]. Water, 2020, 12 (1) :69.

[16] 巫义锐, 汪浩航, 魏大宝, 等. 时空多特征流域场景模式库构建方法[J]. 河海大学学报(自然科学版), 2020, 48 (6) :514-520. (WU Yirui, WANG Haohang, WEI Dabao, et al. Construction method of watershed scene pattern library via spatio-temporal multiple features[J]. Journal of Hohai University (Natural Sciences), 2020, 48 (6) :514-520. (in Chinese))

[17] 刘波, 汪紫薇, 王文鹏, 等. 我国城市用水效率关键指标时空格局分析[J]. 河海大学学报(自然科学版), 2020, 48 (6) :534-541. (LIU Bo, WANG Ziwei, WANG Wenpeng, et al. Spatiotemporal characteristics analysis of major indicators of urban water use efficiencies over mainland China[J]. Journal of Hohai University (Natural Sciences), 2020, 48 (6) :534-541. (in Chinese))

[18] ZHANG Si, TONG Hanghang, XU Jiejun, et al. Graph convolutional networks: a comprehensive review[J]. Computational Social Networks, 2019, 6 (1) :1-23.

[19] PATIL S, STIEGLITZ M. Hydrologic similarity among catchments under variable flow conditions[J]. Hydrology and Earth System Sciences, 2011, 15 (135) :989-997.

[20] HUI Tao, MARCO G, BAI Yungang, et al. Trends of streamflow in the Tarim River Basin during the past 50 years: human impact or climate change? [J]. Journal of Hydrology, 2011, 400 (1/2) :1-9.

[21] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9 (8) :1735-1780.

[22] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9 (3) :293-300.