doi:10.6041/j.issn.1000-1298.2025.09.060

基于历史场景搜索的灌区渠道泄水调度流量预测

雷国相[1,2]　陈皓锐[1,3]　葛建坤[2]　张宝忠[1,3]　戴　玮[1,3]　陈来宝[4]　白美健[1,3]

(1.中国水利水电科学研究院流域水循环与水安全全国重点实验室, 北京 100048;
2.华北水利水电大学水利学院, 郑州 450046; 3.国家节水灌溉北京工程技术研究中心, 北京 100048;
4.安徽省淠史杭灌区管理总局, 六安 237005)

摘要: 为解决灌区渠道泄水调度决策流程过于复杂的问题,本研究以淠史杭灌区灌口集泄水闸为例,基于历史调度数据,利用 SHAP(Shapley additive explanations)法挑选聚类因子,采用自组织映射(Self-organizing map, SOM)神经网络和 K 均值(K-means)耦合的聚类算法 SOM-K-means 对已有调度数据进行分类,通过寻找历史相似场景预测其调度流量。结果表明:SHAP 法确定的 SOM-K-means 算法聚类因子为过去 6 h 降雨量、过去 9 h 降雨量、未来 6 h 降雨量、灌口集泄水闸闸上水位。SOM-K-means 聚类的 5 种聚类标签(0、1、2、3、4)分别可归类为洪峰期、无调度期、起始期、消退期、衰减期。5 种标签中最大均方根误差(RMSE)为 0.201,最大平均绝对误差(MAE)为 0.146,最大均方误差(MSE)为 0.040,最小决定系数(R^2)为 0.793。研究结果可为灌区水资源优化配置提供重要应用参考。

关键词: 灌区; 调度流量; SHAP; 机器学习; SOM-K-means; 历史场景

中图分类号: TV213.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)09-0697-11　　**OSID**:

Dispatched Flow Prediction of Channel Discharge Scheduling in Irrigation Districts Based on Historical Scenario Searching

LEI Guoxiang[1,2]　CHEN Haorui[1,3]　GE Jiankun[2]　ZHANG Baozhong[1,3]　DAI Wei[1,3]
CHEN Laibao[4]　BAI Meijian[1,3]

(1. *National Key Laboratory of Basin Water Cycle and Water Security*,
China Institute of Water Resources and Hydropower Research, *Beijing* 100048, *China*
2. *College of Water Resources*, *North China University of Water Resources and Hydropower*, *Zhengzhou* 450046, *China*
3. *National Engineering and Technology Research Center for Water Saving Irrigation in Beijing*, *Beijing* 100048, *China*
4. *Anhui Province Pishihang Irrigation District Administration Bureau*, *Lu'an* 237005, *China*)

Abstract: Channel discharge scheduling is an important measure to ensure the safe operation of irrigation districts. Taking Guankouji drainage gate in irrigation area of Pishihang as example, the spillway dispatch flow rate was used as the target variable, and the historical and future rainfall at different time periods, the real-time water level on the spillway gate and its change amount were used as the characterization factors. The Shapley additive explanations (SHAP) method was used to analyze the importance of the feature factors, select the combination of clustering factors, and use SOM-K-means, a clustering algorithm coupled with self-organizing map (SOM) neural network and K-means, to classify the existing scheduling data. Classification was done to find historically similar scheduling scenarios. The results showed that the SHAP method determined the clustering factors for the SOM-K-means algorithm as the past 6 h of rainfall, the past 9 h of rainfall, the next 6 h of rainfall, and the water level on the gates of the irrigation catchment sluice. The five clustering labels (0, 1, 2, 3, 4) of SOM-K-means clustering can be categorized as flood peak period, no scheduling period, onset period, fading period, and decay period, respectively. The root mean square error of the five labels was below 0.3, the average absolute error was below 0.2, the mean square error was below 0.1, and the coefficients of determination were

收稿日期: 2024-06-11　修回日期: 2024-08-29

基金项目: 国家重点研发计划项目(2022YFD1900504)、中国水利水电科学研究院技术创新团队项目(ID145B022021)和河南省高等学校青年骨干教师培养计划项目(2020GGJS100)

作者简介: 雷国相(1998—),男,硕士生,主要从事灌区水资源优化配置研究,E-mail: 1978709470@qq.com

通信作者: 陈皓锐(1984—),男,正高级工程师,主要从事灌区水资源模拟与调控研究,E-mail: chenhr@iwhr.com

above 0.75, and it was feasible to search for historical similar scheduling scenarios by using the SOM - K-means clustering algorithm was feasible for searching historically similar scheduling scenarios. The results of the study had important reference value for flood control and scheduling decisions in irrigation areas.

Key words: irrigation area; dispatched flow; SHAP; machine learning; SOM - K-means; historical scenario

0 引言

在暴雨条件下,丘陵灌区傍山渠道要面对沿程坡面汇流、平交河道洪水汇入的压力[1],决策者需充分了解不同暴雨条件下各渠段的来洪过程、洪量、洪峰及渠道的承洪能力和泄洪效果,以做出合理的流量调度决策,如了解不充分,可能导致渠道水量过度泄放,从而危害下游渠系建筑物安全。

目前,灌区渠道的泄水调度决策主要依赖于传统物理机制模型,其不仅对数据资料和专业知识要求较高,而且各环节的水文水动力方程计算非常复杂。当今人工智能分析技术在水资源优化配置、灾害应急管理等领域日益得到广泛应用[2],机器学习作为其热门分支,可以基于灌区渠道泄水中易获取的主要因素,快速开展调度方案决策。其总体思路一般分为两种:①在识别调度主要影响因素基础上,基于回归算法构建调度方案的知识推理预测模型。如胡乐怡等[3]采用多个机器学习算法构建径流预测方案,为流域防洪防涝提供了重要参考;刘成帅等[4]采用机器学习模型构建 RPV - ML 洪水预报模型,提高了洪水预报精度和效率。②从洪水调度资料中提取降雨、水情等场景特征,并对其进行分类分析,通过构建历史相似场景搜索模型,实现调度方案的快速制订,聚类算法是实现这一想法的可行途径[5]。钱堃等[6]采用 K 均值聚类分析方法对历史洪水进行合理分类,有效提高了洪水预报精度;苑希民等[7]使用自编码器和 K 均值聚类算法实现对原始水文数据的特征提取和洪水分级,该方法能有效提取水文数据特征、提高洪水预报精度;尤烽骅等[8]提出改进的 K 均值聚类算法,为金沙江流域实时防洪预报调度提供参考。

文献[9]基于上述第 1 种思路,以实测水位、不同时段历史和预报降雨量为影响因素,通过线性回归(Linear regression, LR)、K 近邻回归(K-nearest neighbors regressor, KNR)、岭回归(Ridge regression, RDR)、决策树回归(Decision tree regression, DTR)、支持向量回归(Support vector regression, SVR)、自适应提升回归(Adaptive boosting regression, ABR)、极度梯度提升回归(Extreme gradient boosting regression, XGR)、随机森林回归(Random forest regression, RFR)8 种算法进行比较,并结合 SHAP 法对渠道泄水调度流量进行了回归预测。为形成更为全面的渠道泄洪调度方法,本研究采用 SOM 和 K-means 耦合的 SOM - K-means 聚类算法,对历史调度数据进行特征分类,寻找与当前预报场景相似的历史调度场景,作为泄水调度的参考方案,为灌区渠道泄水调度流量预测提供新思路。

1 材料与方法

1.1 研究区概况

淠史杭灌区位于安徽、河南两省,安徽部分由淠河、史河、杭埠河三大灌区组成(图 1)。灌区地貌以丘陵岗地为主,暴雨时降雨会在坡面一侧汇集并流入渠道,导致渠道水位迅速升高,从而引发渠道运行安全问题,该问题在南方丘陵灌区具有典型性[9]。史河灌区汲东干渠来洪主要由灌口集泄水闸排出,其泄水调度工作关乎史河总干渠、汲东干渠和淠河渠首人民的生产生活安全。本文以史河灌区汲东干渠看花楼节制闸-河套汀渡槽为研究对象,其进口为看花楼节制闸,出口为河套汀渡槽,单元内设置灌口集泄水闸用于排除洪水(图 2 中调度单元 4)。

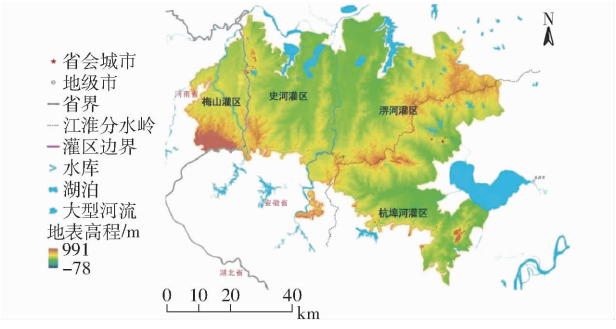


图 1 淠史杭灌区示意图

Fig. 1 Schematic of irrigation area of Pishihang

1.2 场景特征指标

当需要启动渠道防洪调度时,调度单元 4 的进口闸-看花楼节制闸一般处于关闭状态,单元内渠道无上游来水影响,其运行水位主要受渠道现状水位和降雨产生的侧向坡面来洪共同影响。历史降雨较大,集洪区前期土壤含水率相对更大,导致后期降雨也更容易产流;此外,未来降雨量也会对来洪产生影响,因此历史和未来降雨量、渠道水情是渠道泄水调度的主要关注指标[10-11]。以灌口集泄水闸调度流

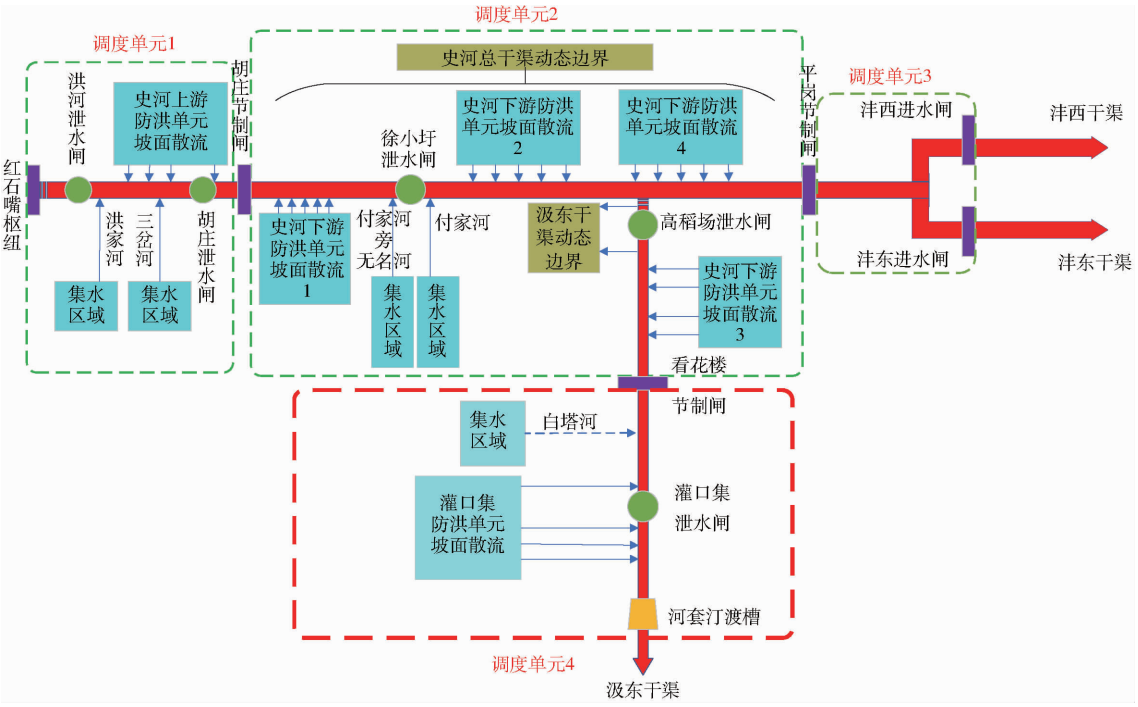


图 2　史河总干渠泄水单元连接关系图

Fig. 2　Connecting relationships of spillway units in Shihe main drainage canal

量 Y 作为目标变量，以历史和未来降雨量、实测闸上水位及水位差（记为 $x_1 \sim x_{10}$）作为场景特征因子，各特征因子及目标变量见表 1。上述各类数据来源于安徽省水文局和淠史杭灌区管理总局，且在历史调度场景中均安全可行。

表 1　特征因子及目标变量说明

Tab. 1　Description of characterization factors and target variables

变量		说明	数据系列长度
特征因子	x_1	过去 1 h 降雨量（mm）	2010—2020 年
	x_2	过去 2 h 降雨量（mm）	
	x_3	过去 3 h 降雨量（mm）	
	x_4	过去 6 h 降雨量（mm）	
	x_5	过去 9 h 降雨量（mm）	
	x_6	未来 1 h 降雨量（mm）	
	x_7	未来 3 h 降雨量（mm）	
	x_8	未来 6 h 降雨量（mm）	
	x_9	灌口集泄水闸闸上水位（m）	
	x_{10}	过去 30 min 闸上水位差（m）	
目标变量	Y	灌口集泄水闸调度流量（m^3/s）	

1.3　聚类因子选取

1.3.1　特征因子重要性分析

机器学习算法中，特征数量选择对算法性能影响较大，选择特征过多可能会导致算法过拟合，过少可能会使聚类结果缺乏代表性，均无法达到最佳聚类效果[12]。因此，在数据聚类前，需在 10 组特征因子中合理选择适量的聚类因子，采用沙普利可加性特征解释（Shapley additive explanation，SHAP）法得到各因子特征重要性系数（SHAP 值）如图 3 所示。

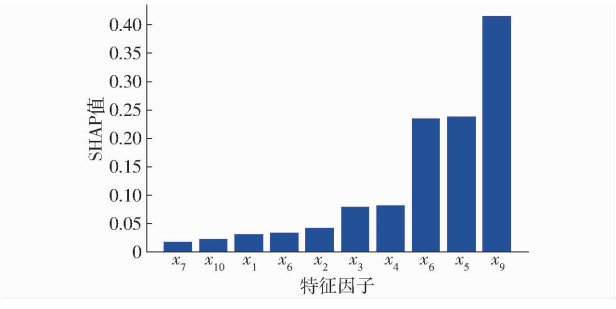


图 3　灌口集泄水闸特征因子 SHAP 值

Fig. 3　SHAP values for characteristic variables of Guankouji drainage gate

1.3.2　特征因子筛选

依据图 3 特征重要性系数按照由大到小的顺序，对特征因子依次累加建立 10 种组合，计算其特征重要性系数累积贡献率（表 2）。由表 2 可得，在组合 4（$x_9 + x_5 + x_8 + x_4$）时特征重要性累积贡献率已达 81% ＞ 80%，参考文献[13]，将 $x_4 + x_5 + x_8 + x_9$ 作为算法聚类因子。

1.4　研究方法

1.4.1　沙普利可加性特征解释法

通过对所有可能的子集进行遍历，计算每个子集对最终预测结果的贡献，从而得出每个特征的 SHAP 值[14]。

1.4.2　SOM－K-means 聚类算法

1.4.2.1　自组织映射神经网络

自组织映射神经网络（Self-organizing map，SOM）

表 2 特征因子组合重要性贡献率

Tab. 2 Importance contribution of combinations of characteristic factors %

组合	特征因子	特征重要性贡献率
1	x_9	34.6
2	x_9+x_5	54.5
3	$x_9+x_5+x_8$	74.1
4	$x_9+x_5+x_8+x_4$	81.0
5	$x_9+x_5+x_8+x_4+x_3$	87.6
6	$x_9+x_5+x_8+x_4+x_3+x_2$	91.1
7	$x_9+x_5+x_8+x_4+x_3+x_2+x_6$	94.0
8	$x_9+x_5+x_8+x_4+x_3+x_2+x_6+x_1$	96.6
9	$x_9+x_5+x_8+x_4+x_3+x_2+x_6+x_1+x_{10}$	98.5
10	$x_9+x_5+x_8+x_4+x_3+x_2+x_6+x_1+x_{10}+x_7$	100

是一种无监督学习的神经网络模型,它在无监督情况下对输入数据进行自动分类,并对输入模式进行自组织学习,反复调整连接权值,最终在竞争层将分类结果表示出来[15]。

(1)初始化网络参数设置:设定网络连接权值 w_{ij}、初始学习率 $\eta(0)$、不通邻节点半径 sigma、迭代次数 C_n、特征数量 n_{features} 以及神经元行列数[16-17]。网络连接权值使用较小随机值进行初始化;n_{features} 为 4(SHAP 法确定的 4 个特征因子(x_4、x_5、x_8、x_9));一般来说,输出层中神经元总数为[18-19]

$$N=5\sqrt{n} \tag{1}$$

式中 n——样本数

输出层网络由行列数一致的神经元组成,则神经元行列数为

$$h=\sqrt{N} \tag{2}$$

本研究中灌口集泄水闸样本数 $n=230$,代入式(1)、(2)得到 SOM 神经元尺寸为 9×9。

(2)寻找输入数据匹配的节点,计算式为

$$d_j=\sum_{i=1}^{N}(x_i-w_{ij})^2 \tag{3}$$

(3)调整输出节点连接权值,计算式为

$$w_{ij}(t+1)=w_{ij}(t)+\eta(t)T_{j,i(x)}(x(t)-w_{ij}(t)) \tag{4}$$

其中

$$T_{j,i(x)}=\mathrm{e}^{-\frac{S_{j,i(x)}^2}{2\sigma^2}} \tag{5}$$

$$\eta(t)=\eta(0)(1-t/C_n) \tag{6}$$

式中 $T_{j,i(x)}$——获胜神经元周围的邻域函数

$S_{j,i(x)}$——节点 j 和获胜神经元 $i(x)$ 间距离

σ——邻域半径,均采用欧氏距离

其中学习效率 $\eta(t)$ 随时间 t 单调递减以保证学习过程的收敛性。

(4)增加学习步数。令 $t=t+1$ 重复步骤(2)、(3)直到学习效率 $\eta(t)$ 为 0,或者达到迭代次数 C_n 为止,最后得到输出的聚类数与聚类中心。

1.4.2.2 K 均值聚类算法

K-means 通过计算样本点与簇中心的距离来将相似的样本划分为同一个簇,从而使相似数据划分为同一组,差异大的数据划分为不同组。K-means 处理较大数据时复杂度低、效率高,但 K-means 算法对于簇数量 k 需要提前指定,且对初始聚类中心的选择比较敏感,容易达到局部最优值,影响聚类效果,对数据具有一定的适用性[20]。

1.4.2.3 DBI 指数

利用 DBI 指数(Davies-bouldin index)作为衡量聚类结果质量的指标[21]。它通过比较聚类簇之间的聚集度和分散度来评估聚类结果的紧密度和分离度。DBI 取值范围为[0, +∞),DBI 越小表示聚类结果越好。

1.4.2.4 SOM-K-means 算法

SOM-K-means 算法具体流程如图 4 所示,步骤如下:

(1)依据 SOM 算法得到初始聚类数 K_m 及聚类中心,将初始聚类数输入 K-means 算法得到处理后聚类结果。

(2)取 $k=1$,并将 k 依次加 1,重复输入到 K-means 算法中,直到 $k=K_m$,计算所有聚类数 DBI 指数。

(3)比较 DBI 指数,选择最小 DBI 指数对应的聚类数 k 作为最佳聚类结果。

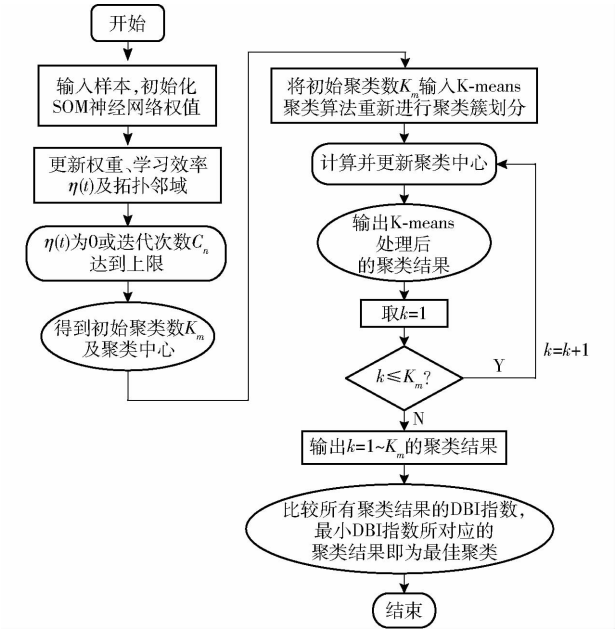


图 4 SOM-K-means 算法流程图

Fig. 4 Flowchart of SOM-K-means algorithm

将灌口集泄水闸调度数据(其中 50 组数据泄水

闸调度流量为 0）进行标准化处理并进行编号，编号形式为：年 + 月 + 日 + 小时（24 h 计）+ 分钟。如 2020 年 7 月 18 日 00:30 数据编号为 202007180030，选取 SHAP 法确定的特征因子组合作为聚类因子，采用 SOM - K-means 聚类算法中得到的每个样本聚类标签，确定每个样本所属的聚类标签后，计算各聚类标签中每个样本与其他样本间的欧氏距离，挑选距离最小的样本作为最佳相似样本，以最佳相似样本的调度流量作为预测值，并与该样本的实际值进行比较，计算精度评价指标评判其可行性。技术路线图如图 5 所示。

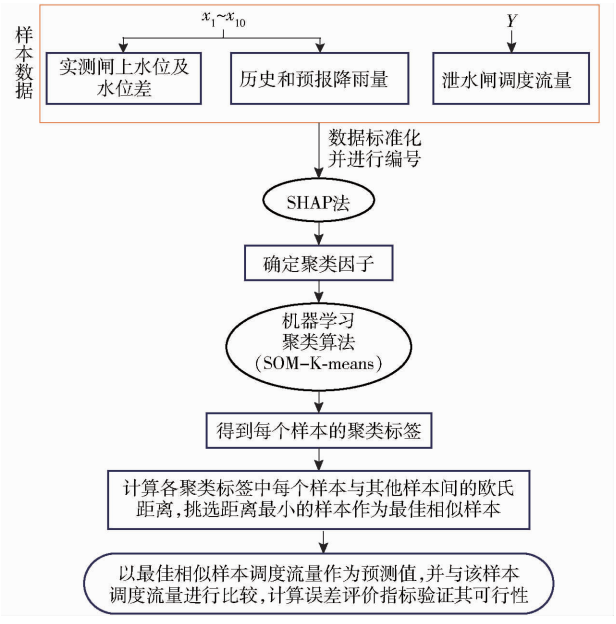


图 5　历史场景搜索模型技术路线图

Fig. 5　Historical scene search modeling technology roadmap

1.5　数据标准化与算法评价指标

将搜集到的灌口集泄水闸数据调用 Python 3.9 进行算法分类与预测。为避免数据量纲对研究结果造成影响，采用 Z-score 对模型数据进行标准化；利用灌口集泄水闸调度流量预测值与实际值间的均方根误差（RMSE）、平均绝对误差（MAE）、均方误差（MSE）和决定系数（R^2）作为精度指标评估算法预测精度[22-23]，具体计算公式见文献[9]。

2　结果与分析

2.1　聚类数确定

确定 SOM 神经网络初始参数后，依据图 4 得到 SOM 初始聚类数为 $k=8$，再将初始聚类数和聚类中心输入 K-means 算法得到处理后的聚类结果，计算不同聚类数的 DBI 指数，得到聚类数 k 为 2 ~ 8 的 DBI 指数如图 6 所示（$k=1$ 时不存在 DBI 指数）。由图 6 得到，$k=5$ 时 DBI 指数最小，即 $k=5$ 为最佳聚类数。

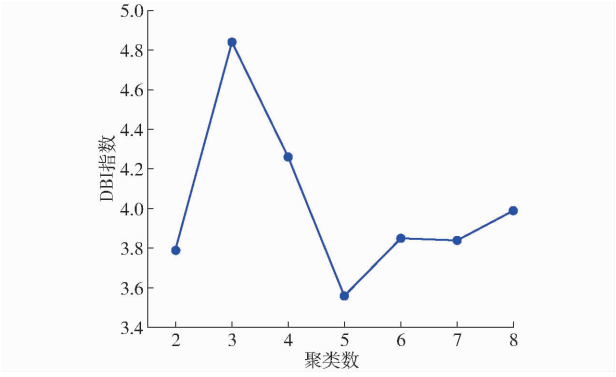


图 6　不同聚类数下 DBI 指数变化曲线

Fig. 6　Change curve of DBI index under different numbers of clusters

2.2　聚类标签特征分析

以 x_4、x_5、x_8、x_9 作为聚类因子，采用 SOM - K-means 聚类算法对灌口集泄水闸数据进行聚类并进行标签划分（5 种聚类标签分别记为 0（洪峰期）、1（无调度期）、2（起始期）、3（消退期）、4（衰减期）），图 7 ~ 11 为灌口集泄水闸 5 类聚类标签中聚类因子（x_4、x_5、x_8、x_9）及调度流量（Y）原数据的分布频数。

由图 7 可看出，x_9 大部分位于 54.5 ~ 56.0 m 之间，水位不高；x_4 在 20 ~ 90 mm 波动，x_5 在 40 ~ 90 mm 间分布较多，x_8 在 40 ~ 60 mm 间分布较多，部分达到 80 mm 以上，降雨强度较大。Y 总体位于 0 ~ 90 m^3/s，在 30 ~ 80 m^3/s 间分布较多，属于中高水平。结合数据编号可知，聚类标签 2 中包含的数据均为有调度过程且为洪水发生的起始阶段，其中有 2016 年的前 9 组，2015 年的前 10 组，2020 年的前 6 组，2010 年的前 14 组。综合来看，聚类标签 2 所在时段呈现历史及未来降雨较强但水位不高的特点，可归类为灌口集泄水闸调度的起始期。

由图 8 可看出，聚类标签 0 的 x_4 绝大部分位于 40 mm 以上，部分位于 100 mm 以上，强度较大，与聚类标签 2 相比，部分 x_4 已超过 90 mm，最高可达 160 mm；x_5 位于 60 mm 以上，部分 x_5 更是在 160 mm 以上，最高可达 200 mm，降雨强度极大。x_9 位于 55.0 ~ 58.5 m 之间，其中绝大部分水位高于 56.0 m，最高时接近 58.5 m，参考灌口集泄水闸的闸上设计水位为 57.32 m，该聚类标签时段内是泄水闸防洪的高风险期，大部分调度流量位于 100 m^3/s 以上，最大可达 220 m^3/s，可见此时段内洪水强度极大。聚类标签 0 所在时段呈现历史降雨强度和水位极高且调度流量极大的特点，可归类为灌口集泄水闸调度的洪峰期。

由图 9 可看出，x_4 在 20 ~ 45 mm 分布较多，降雨

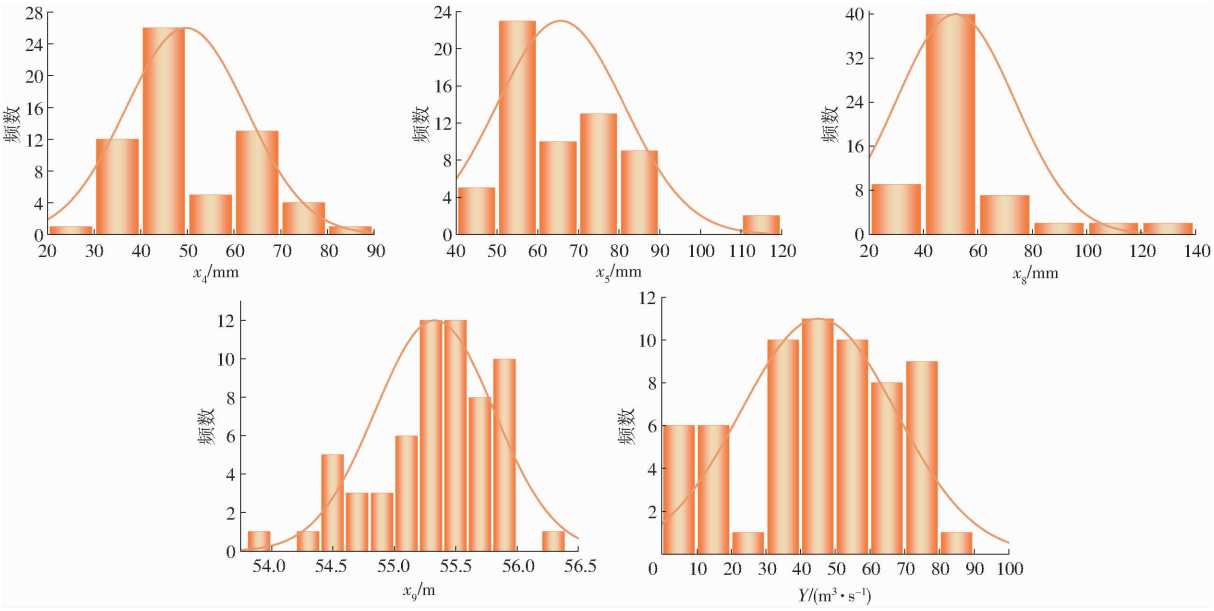


图 7 聚类标签 2 分布频数

Fig. 7 Frequency plot of distribution of clustered labels 2

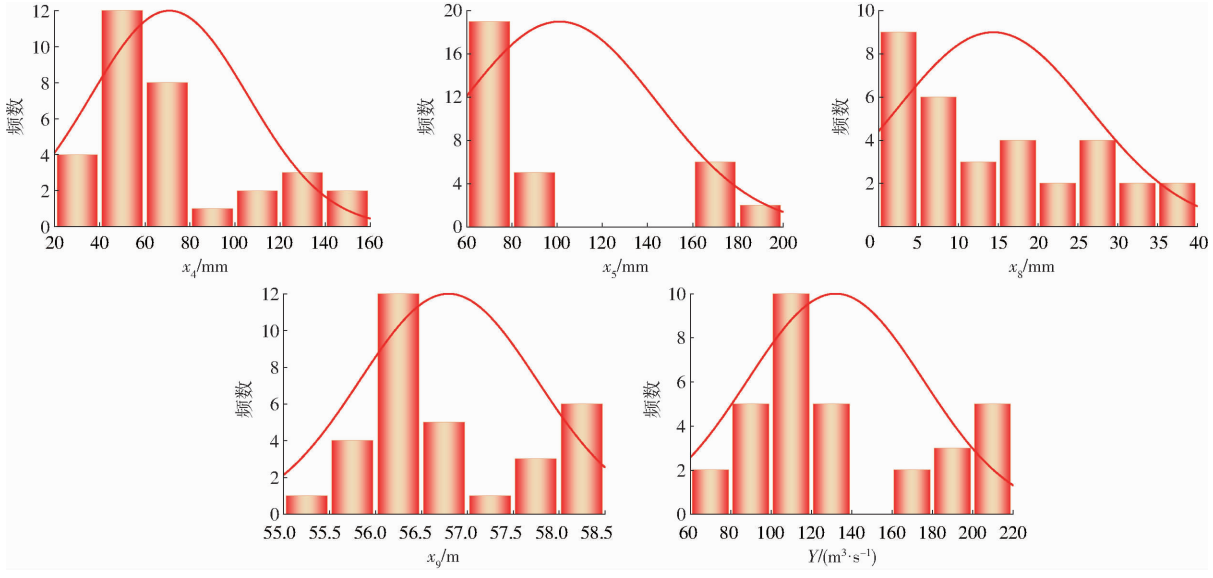


图 8 聚类标签 0 分布频数

Fig. 8 Frequency plot of distribution of clustered labels 0

强度不大;x_5在 25 ~ 65 mm 波动,大部分位于 40 mm 以上,降雨强度较大;相比聚类标签 2,历史降雨 x_4、x_5较小。x_9位于 55.0 ~ 58.5 m 之间,与聚类标签 0 分布类似,但水位高于 56.0 m 的频数减少;聚类标签 4 调度流量在 50 ~ 100 m^3/s 间分布较多,在 100 m^3/s 以上也有一定比例,对比 5 类聚类标签,聚类标签 4 中 Y 在 100 m^3/s 以上的分布频数仅次于聚类标签 0,可见该时段内的洪水强度相对于聚类标签 0 正在减弱,但仍处于较高状态,聚类标签 4 时段呈现历史降雨强度较大、水位极高且调度流量较大的特点,可归类为灌口集泄水闸调度的衰减期。

由图 10 可看出,聚类标签 3 中 x_4中大部分位于 0 ~ 1 mm 之间;x_5在 0 ~ 10 mm 间分布较广,x_8大部分为 0 mm,由 x_4、x_5、x_8分布情况可知,该聚类标签时段内降雨强度极小;对比前 3 类聚类标签,该聚类标签内 x_9基本位于 55.0 ~ 56.5 m,普遍高于聚类标签 2 但低于聚类标签 0 和 4;Y 分布范围与聚类标签 2 较为类似,但该聚类标签中调度流量在 80 m^3/s 以上的频数较聚类标签 2 多,聚类标签 3 降雨较少但水位及流量偏大的原因可能为:该时段前降雨强度较大导致水位较高,该聚类标签时段虽然降雨较小,但前期强降雨导致的高水位依然需要泄水闸宣泄较大流量以保证安全,结合降雨、水位及调度流量分布情况,聚类标签 3 可归类为灌口集泄水闸调度的消退期。

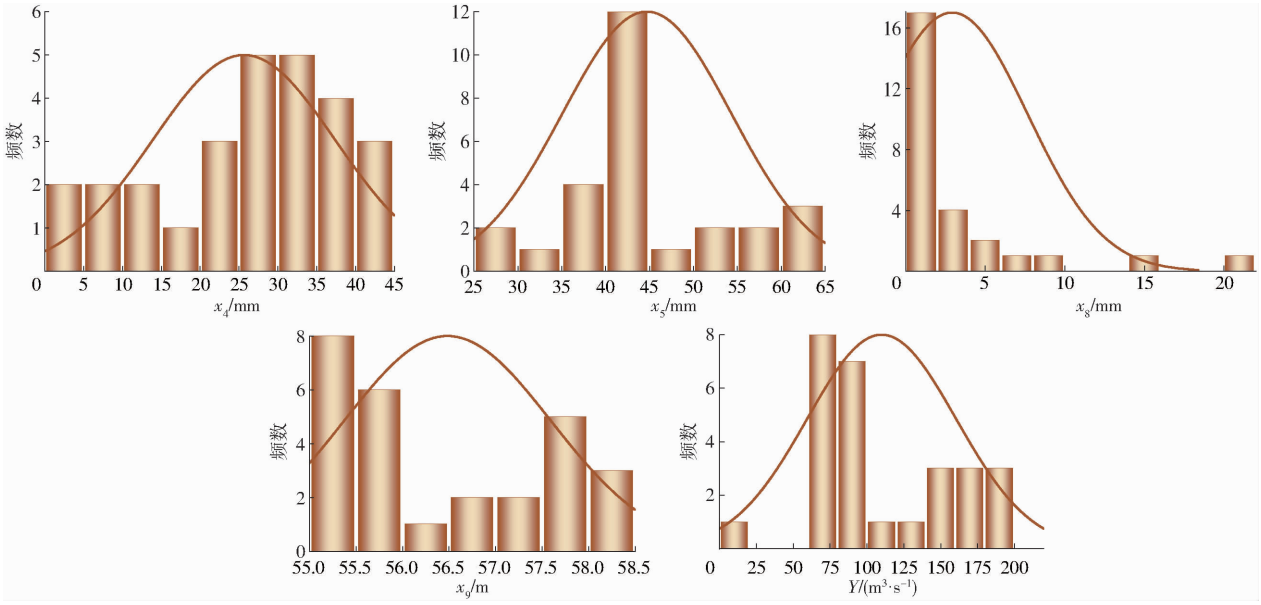


图 9　聚类标签 4 分布频数

Fig. 9　Frequency plot of distribution of clustered labels 4

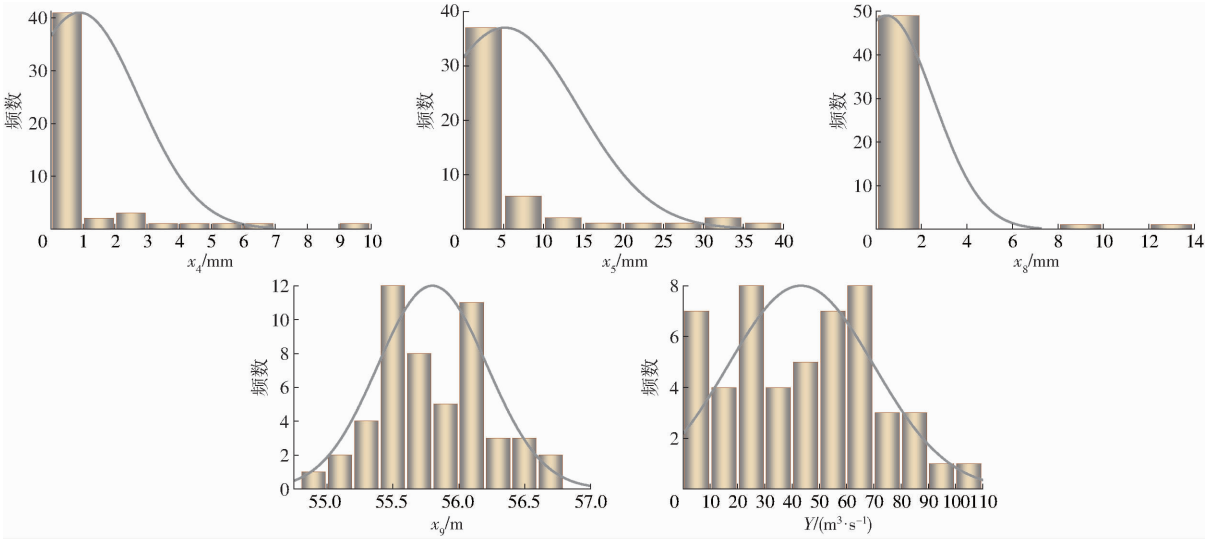


图 10　聚类标签 3 分布频数

Fig. 10　Frequency plot of distribution of clustered labels 3

由图 11 可看出，各时段降雨量（x_4、x_5、x_8）分布范围较小；x_4、x_5、x_8 分别集中于 0 ~ 10 mm、0 ~ 24 mm、0 ~ 2 mm 区间内，降雨强度较小；水位集中分布在 53.5 ~ 55.0 m 之间，普遍低于其余 4 类标签水位。由 Y 分布频数可以得到，绝大部分调度流量均为 0 m^3/s，由该聚类标签的原数据可得，50 组无调度数据中，有 45 组位于该聚类标签中，且分布于该聚类标签中有调度数据的 Y 均处于极低水平，最大不超过 5 m^3/s，聚类标签 1 所在调度时段呈现历史降雨强度较小、水位较低且调度流量极小的特点，可归类为灌口集泄水闸的无调度期。

2.3　各聚类标签预测精度评价

将每个样本均作为原样本，以同标签中最佳相似样本的调度流量作为预测值，灌口集泄水闸调度流量标准化实际值与预测值线性拟合结果如图 12 所示。

表 3 为灌口集泄水闸各聚类标签及其归类时期的 RMSE、MAE、MSE 和 R^2。由表 3 可看出，SOM - K-means 聚类算法可将泄水闸的调度阶段进行聚类标签划分，并可较好归类为不同时期。对比算法内标签，无调度期 R^2 较其他聚类标签虽不高，但 RMSE、MSE、MAE 均为算法中最小，这是因为无调度数据的调度流量均为 0，而聚类标签中无调度期内预测得到的调度流量大部分也均为 0，因而无调度期内多数调度流量实际值与预测值相同。表 3 聚类得到的 5 个标签中，RMSE 均在 0.3 以下，MAE 均在 0.2 以下，MSE 均在 0.1 以下，R^2 均在 0.75 以上，可见采用 SOM - K-means 聚类算法对泄水闸调度流

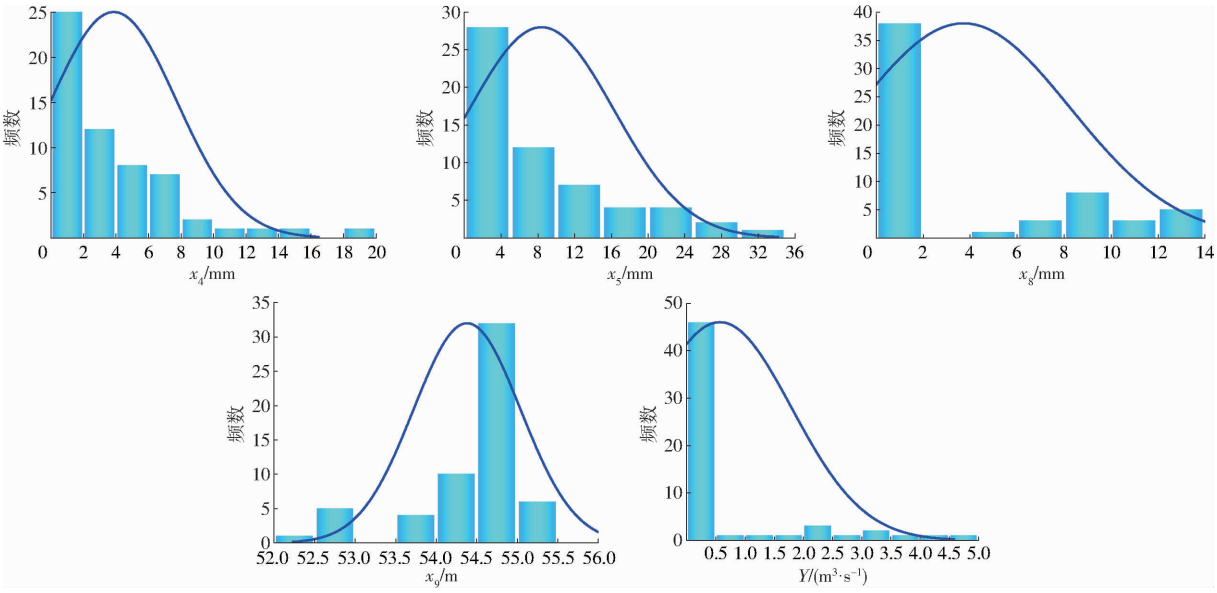


图 11 聚类标签 1 分布频数

Fig. 11 Frequency plot of distribution of clustered labels 1

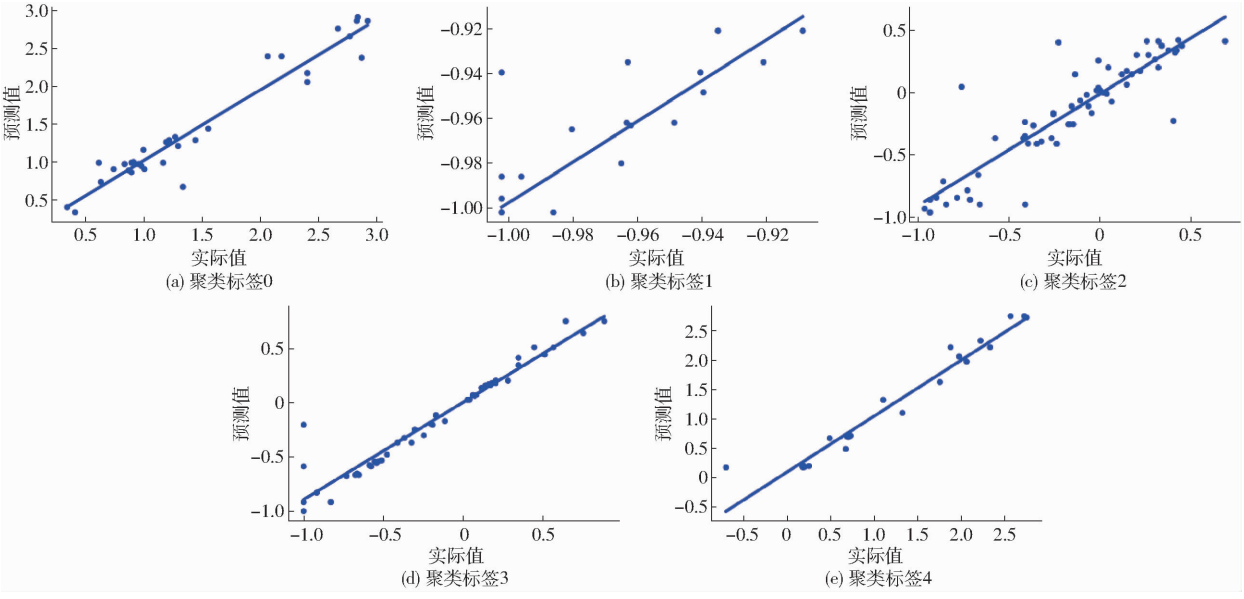


图 12 灌口集泄水闸调度流量预测结果线性拟合图

Fig. 12 Linear fit plots of dispatch flow prediction results for Guankouji drainage gate

表 3 预测评价指标

Tab. 3 Projected evaluation indicators

聚类标签	归类时期	RMSE	MAE	MSE	R^2
0	洪峰期	0.201	0.146	0.040	0.934
1	无调度期	0.010	0.005	0.001	0.802
2	起始期	0.191	0.123	0.037	0.793
3	消退期	0.122	0.064	0.015	0.931
4	衰减期	0.190	0.138	0.039	0.953

量进行预测可行。需要注意的是,聚类算法实际输出运行结果时,其标签对应的数字在不同时间输出结果可能不同,但同一聚类标签内所含数据特征是相同的,本文在 2.2 节的重点是对聚类标签内数据特征进行分析,进而分为所属时期,在实际应用中,可为新样本所处洪水调度发生阶段提供参考。

3 讨论

3.1 调度流量特征因子分析

本研究各调度单元内,泄水闸调度流量取决于渠道侧面集水区域的产汇流以及洪水入渠后的水位推进这两个物理过程,因此将泄水闸调度流量的影响因子归类为水位及降雨量两种因素。张宇等[24]发现,在不同形状的流域内,防洪控制站的洪水流量变化与降雨强度、降雨时间、降雨过程等因素有显著相关性;梁慧真等[25]指出率定精确的水位流量关系曲线对于灌区计划用水、渠道流量运行调度具有重要意义。以上研究均表明,降雨和水位是影响洪水

流量的重要因素。图 3 计算特征因子重要性系数总和为 1.198≠1 的原因是，SHAP 值表示的特征重要性大小不仅受目标变量的影响，还受各特征因子间的影响，SHAP 法通过检测特征间的交互作用，从而提供更全面的特征重要性系数[26]。

3.2　特征因子筛选对场景分类的影响

由图 3 可得，初始选取的特征因子 SHAP 值大于 1，这表明 10 组特征因子间存在冗余问题，选择合理数量的因子组合不仅易于样本训练解释，而且能提高分类准确性[27]。周俊伟等[27]采用随机森林算法（Random forest，RF）结合斯皮尔曼相关性对特征因子进行分层聚类，显著提高了冬小麦的分类精度；廖娟等[28]利用最小冗余最大相关算法（Minimum redundancy maximum relevance，mRMR）确定最优植被指数组合，实现棉花黄萎病危害等级的精准分类；赵玉清等[29]使用较优特征组合的支持向量机（SVM）检测模型能更准确地对缺陷生咖啡豆进行分类。本研究采用 SHAP 法分析特征因子重要性并挑选出聚类因子组合（$x_4 + x_5 + x_8 + x_9$），不仅降低了前期数据收集工作的难度，而且能较好地对调度数据进行分类，进而为调度流量预测提供指导。

3.3　算法耦合与洪水场景分类标准

采用 SOM 和 K-means 聚类算法进行耦合的原因为：①SOM 算法无需预先指定聚类数，能自动对输入模式进行聚类。但无法针对每个样本输出明确的聚类标签，收敛时间长，且易受不平衡数据集（本研究中降雨因素类别显著多于水位类别）的影响，造成分类不准确[30]。②K-means 收敛速度快、分类结果明确。但聚类结果依赖于初始簇中心的选择，而聚类中心在经典 K-means 聚类算法中是随机确定的，本研究中降雨、水位数据维度相差较大，初始聚类中心和 k 值的设定较为困难，易造成聚类效果不佳[31]。将 2 种聚类算法进行结合，既克服了 SOM 算法中收敛速度较慢和分类结果不准确的问题，又解决了 K-means 中初始聚类中心难以设定的问题，因而达到较为准确的分类结果。表 3 聚类标签中洪峰期的 RMSE、MAE、MSE 在归类时期中均最大的原因为洪峰期流量波动较大，增加了预测难度[32]；5 类聚类标签的 R^2 中，除起始期在 0.8 以下，其余均在 0.8 以上，其原因为：灌口集泄水闸位于调度单元下游，起始洪水阶段降雨的产汇流过程对泄水流量的影响较大[33]，因而起始期调度流量的预测难度较大。

在描述洪水场景分类时，前人多以洪水自身的形状特征作为标准，如以“大/中/小”等强度指标，“高/矮”和“胖/瘦/尖瘦”、“多洪峰长历时”等形态指标，“急/缓”等变化速率指标[34-35]等。本文按照渠道洪水从产生到结束的时间过程演进特点，将其分为 5 类（起始期、洪峰期、衰减期、消退期、无调度期）。本研究中洪水过程的划分标准与前人不同的原因为：①前人研究多为水库和流域洪水过程，其洪水特征因子包含洪峰流量、涨退速率、高脉冲历时等诸多指标，本研究领域仅为调度单元内泄水闸的调度流量，其影响因素主要取决于实时水位及降雨量两种指标，与前人进行特征选取上有较大差别。②洪水流量与其发生阶段有着密切关系[36]，本文通过划分为起始期、洪峰期、衰减期、消退期、无调度期，也可更方便了解目前洪水调度发生的紧急程度，进而做好调度部署工作。本文与上述研究的分类标准虽有差异，但本质基本相同，即通过归纳分析泄水调度时的降雨、水情特征，将渠道的洪水过程线分成几个特征时段，每个特征时段的流量、水位涨退趋势、脉冲历时都有相对明显的区分，从而方便进行场景的精准定位搜索。

3.4　聚类算法与回归算法预测方法对比

参考文献[9]，选取过去 6 h 降雨量、过去 9 h 降雨量、未来 6 h 降雨量、灌口集泄水闸闸上水位 4 组特征变量，通过多种机器学习算法比选后，推荐采用随机森林回归算法进行灌口集泄水闸调度流量预测，该模型在训练阶段精度指标为：RMSE 为 0.126、MAE 为 0.080、MSE 为 0.016、R^2 为 0.982，在模型测试阶段的精度指标为：RMSE 为 0.263、MAE 为 0.164、MSE 为 0.069、R^2 为 0.950；本研究采用 SOM－K-means算法在预测灌口集泄水闸调度流量上只有聚类标签 4 的 RMSE、MAE、MSE 和 R^2 优于随机森林回归算法的测试集，由此可得，在预测灌区渠道泄水闸调度流量上，回归算法精度高于聚类算法。虽然聚类算法预测精度略低于回归算法，但根据其最佳相似样本的编号，可迅速找到历史相似调度场景，进而为当前调度情况提供更加丰富的信息。在实际调度中，如需快速获得该时刻泄水闸的调度流量，可利用特定因子直接采用机器学习回归算法进行预测；如需借鉴历史经验进行调度，可利用本文确定的特征因子采用聚类算法寻找历史相似场景，两类方法可相互验证，为调度管理者提供多样化选择。

4　结论

（1）采用 SHAP 法确定的算法聚类因子组合为 $x_4 + x_5 + x_8 + x_9$（过去 6 h 降雨量、过去 9 h 降雨量、未来 6 h 降雨量、灌口集泄水闸闸上水位）。

（2）按照洪水发生阶段，将 SOM－K-means 算法

聚类标签划分为起始期、洪峰期、衰减期、消退期、无调度期。5 个标签中,RMSE 均在 0.3 以下,MAE 均在 0.2 以下,MSE 均在 0.1 以下,R^2均在 0.75 以上,可见采用 SOM - K-means 聚类算法对泄水闸调度流量进行预测可行。采用机器学习聚类算法预测泄水闸调度流量精度上虽略低于回归算法,但根据其最佳相似样本编号,可迅速找到历史相似调度场景,进而为当前调度情况提供更加丰富的信息。

参 考 文 献

[1] 胡笑妍,崔振华,李荣容. 封丘倒灌区贯孟堤扩建工程对黄河下游防洪影响及消减措施研究[J]. 水科学与工程技术, 2023(4): 18 - 20.
HU Xiaoyan, CUI Zhenhua, LI Rongrong. The influence of flood control and eliminating measures in the lower reaches of the Yellow River of the Guanmengdi expansion project of Fengqiu inverted irrigation area[J]. Water Sciences and Engineering Technology, 2023(4): 18 - 20. (in Chinese)

[2] 张潮. 数字孪生水利建设中筑牢数字安全屏障的思考[J]. 中国水利, 2024(7): 62 - 68.
ZHANG Chao. Reflections on establishing a solid digital security barrier in digital twin water conservancy construction[J]. China Water Resources, 2024(7): 62 - 68. (in Chinese)

[3] 胡乐怡,付晓雷,蒋晓蕾,等. 基于 LSTM、RF、SVR 三种机器学习方法的径流预测研究[J]. 水文, 2024, 44(5): 17 - 24.
HU Leyi, FU Xiaolei, JIANG Xiaolei, et al. Research on runoff prediction based on machine learning: LSTM, RF and SVR[J]. Journal of China Hydrology, 2024, 44(5): 17 - 24. (in Chinese)

[4] 刘成帅,解添宁,李文忠,等. 考虑径流过程矢量化的机器学习洪水预报模型[J]. 水科学进展, 2024, 35(3): 420 - 429.
LIU Chengshuai, XIE Tianning, LI Wenzhong, et al. Machine learning-based flood forecasting models considering runoff process vectorization[J]. Advances in Water Science, 2024, 35(3): 420 - 429. (in Chinese)

[5] 倪鹏,包为民,张乾,等. 基于主成分分析的系统聚类分析方法在洪水预报中的应用[J]. 三峡大学学报(自然科学版), 2018, 40(1): 1 - 4.
NI Peng, BAO Weimin, ZHANG Qian, et al. Application of hierarchical cluster analysis to flood forecasting based on principal component analysis[J]. Journal of China Three Gorges University(Natural Sciences), 2018, 40(1): 1 - 4. (in Chinese)

[6] 钱堃,包为民,李偲松,等. K 均值聚类分析方法在洪水预报中的应用[J]. 水电能源科学, 2012, 30(5): 41 - 44.
QIAN Kun, BAO Weimin, LI Yinsong, et al. Application of K-means cluster analysis in flood forecasting[J]. Water Resources and Power, 2012, 30(5): 41 - 44. (in Chinese)

[7] 苑希民,李达,田福昌,等. 基于 AE - RCNN 的洪水分级智能预报方法研究[J]. 水利学报, 2023, 54(9): 1070 - 1079.
YUAN Ximin, LI Da, TIAN Fuchang, et al. Research on flood classified intelligent forecasting method based on AE - RCNN[J]. Journal of Hydraulic Engineering, 2023, 54(9): 1070 - 1079. (in Chinese)

[8] 尤烽骅,余玉聪,刘招,等. 基于联合改进人工蜂群及 K 均值聚类算法的洪水分类研究[J]. 水文, 2021, 41(4): 40 - 47.
YOU Fenghua, YU Yucong, LIU Zhao, et al. Research on flood classification based on joint improved artificial bee colony and K means clustering algorithms[J]. Journal of China Hydrology, 2021, 41(4): 40 - 47. (in Chinese)

[9] 葛建坤,雷国相,陈皓锐,等. 基于 SHAP 重要性排序和机器学习算法的灌区渠道调度流量预测[J]. 农业工程学报, 2023, 39(13): 113 - 122.
GE Jiankun, LEI Guoxiang, CHEN Haorui, et al. Irrigation district channel dispatch flow prediction based on SHAP importance ranking and machine learning algorithm[J]. Transactions of the CSAE, 2023, 39(13): 113 - 122. (in Chinese)

[10] 王顺利. 丰乐水库灌区南干渠防洪问题与对策[J]. 江淮水利科技, 2012, 37(1): 22 - 24.
WANG Shunli. Flood control problems and countermeasures in the south trunk canal of Fengle reservoir irrigation area[J]. Jianghuai Water Resources Science and Technology, 2012, 37(1): 22 - 24. (in Chinese)

[11] 范营营. MIKE11 在泄洪闸闸宽比选中的应用——以江西省某灌区为例[J]. 海河水利, 2022(6): 102 - 105, 116.
FAN Yingying. Application of MIKE11 hydrodynamic model in the width comparison of spillway gate—taking an irrigation district in Jiangxi Province as the case[J]. Haihe Water Resources, 2022(6): 102 - 105, 116. (in Chinese)

[12] 罗博炜,罗万红,谭家驹. 融合 SOM 神经网络与 K-means 聚类算法的用户信用画像研究[J]. 铁路计算机应用, 2024, 33(7): 14 - 19.
LUO Bowei, LUO Wanhong, TAN Jiaju. User credit profile integrating SOM neural network and K-means clustering algorithm[J]. Railway Computer Application, 2024, 33(7): 14 - 19. (in Chinese)

[13] 周婷,莫小明,查志华,等. 基于力声信号锯齿化多特征融合的香梨脆度评价[J]. 农业工程学报, 2022, 38(13): 305 - 312.
ZHOU Ting, MO Xiaoming, ZHA Zhihua, et al. Evaluation of crispness fusing features extracted from jagged curves of mechanical-acoustic signals of Korla pears[J]. Transactions of the CSAE, 2022, 38(13): 305 - 312. (in Chinese)

[14] 周亚男,陈绘,刘洪斌. 基于多源数据和 Stacking - SHAP 方法的山地丘陵区土地覆被分类[J]. 农业工程学报, 2022, 38(23): 213 - 222.
ZHOU Ya'nan, CHEN Hui, LIU Hongbin. Land cover classification in hilly and mountainous areas using multi-source data and Stacking - SHAP technique[J]. Transactions of the CSAE, 2022, 38(23): 213 - 222. (in Chinese)

[15] 张可,柴毅,匡金骏. 基于征兆邻搜索优化聚类和自组织映射神经网络的多病害诊断[J]. 农业工程学报, 2011,

27(1)：215－222.
ZHANG Ke, CHAI Yi, KUANG Jinjun. Method for multi-disease diagnosis based on optimized symptom adjacent-searching clustering and SOM NN[J]. Transactions of the CSAE, 2011, 27(1): 215－222. (in Chinese)
[16] KOHONEN T, KASKI S, LAGUS K, et al. Self organization of a massive document collection[J]. IEEE Transactions on Neural Networks, 2000, 11(3): 574－585.
[17] 李寒,陶涵虓,崔立昊,等. 基于 SOM－K-means 算法的番茄果实识别与定位方法[J]. 农业机械学报, 2021, 52(1): 23－29.
LI Han, TAO Hanxiao, CUI Lihao, et al. Recognition and localization method of tomato based on SOM－K-means algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 23－29. (in Chinese)
[18] ASTEL A, TSAKOVSKI S, BARBIERI P, et al. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets[J]. Water Research, 2007, 41(19): 4566－4578.
[19] 郅瑞卿,关侠,鄢旭久,等. 基于自组织神经网络的耕地自然质量评价方法及其应用[J]. 农业工程学报, 2014, 30(23): 298－305.
QIE Ruiqing, GUAN Xia, YAN Xujiu, et al. Method and its application of natural quality evaluation of arable land based on self-organizing feature map neural network[J]. Transactions of the CSAE, 2014, 30(23): 298－305. (in Chinese)
[20] 张志勇,路小娟,孔令刚,等. K-means 结合 RBF 神经网络预测线性菲涅尔集热回路出口熔盐温度[J]. 农业工程学报, 2021, 37(3): 213－222.
ZHANG Zhiyong, LU Xiaojuan, KONG Linggang, et al. Predicting molten salt temperature at the circuit outlet of linear Fresnel heat collector using K-means combined with RBF neural network[J]. Transactions of the CSAE, 2021, 37(3): 213－222. (in Chinese)
[21] 陈凯,雷琪,李豆萌. 基于改进 K-means 聚类的轨道交通基础设施分布式光伏发电典型场景生成及出力特性分析[J]. 电气工程学报, 2024, 19(2): 364－372.
CHEN Kai, LEI Qi, LI Doumeng. Typical scenario generation and output characteristics analysis of distributed photovoltaic power generation for railway transportation infrastructure based on improved K-means clustering[J]. Journal of Electrical Engineering, 2024, 19(2): 364－372. (in Chinese)
[22] 周修理,秦娜,王开宇,等. 基于机器学习的机械压实对大豆产量的影响预测研究[J]. 农业机械学报, 2023, 54(11): 139－147.
ZHOU Xiuli, QIN Na, WANG Kaiyu, et al. Effect of mechanical compaction on soybean yield based on machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(11): 139－147. (in Chinese)
[23] 曹守启,葛照瑞,张铮. 基于物联网的浮标水质监测系统与溶解氧浓度预测模型[J]. 农业机械学报, 2021, 52(11): 210－218
CAO Shouqi, GE Zhaorui, ZHANG Zheng. Buoy water quality monitoring system and prediction model based on internet of things[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(11): 210－218. (in Chinese)
[24] 张宇,杨令强,梁延涛,等. 不同暴雨条件下扇形和羽形流域的洪水规律[J]. 济南大学学报(自然科学版),2024, 38(2): 149－155.
ZHANG Yu, YANG Lingqiang, LIANG Yantao, et al. Flood laws of sector and pinnate watersheds with different rainstorm conditions[J]. Journal of University of Jinan(Science and Technology), 2024, 38(2): 149－155. (in Chinese)
[25] 梁慧真,宗全利,刘贞姬,等. 明渠非标准断面法测流水位流量关系的率定及影响因素分析[J]. 排灌机械工程学报, 2024, 42(2): 147－153,160.
LIANG Huizhen, ZONG Quanli, LIU Zhenji, et al. Calibration and influencing factors analysis of stage-discharge relation of non-standard section method in open channels[J]. Journal of Drainage and Irrigation Machinery Engineering, 2024, 42(2): 147－153,160. (in Chinese)
[26] 潘梦瑶,任瑛,王思源,等. 基于梯度提升算法和 SHAP 的石家庄 PM2.5 和臭氧浓度预测及影响因素分析[J]. 环境科学学报, 2024, 44(7): 402－409.
PAN Mengyao, REN Ying, WANG Siyuan, et al. Prediction of PM2.5 and ozone concentration in Shijiazhuang and analysis of influencing factors based on gradient boosting algorithm and SHAP[J]. Acta Scientiae Circumstantiae, 2024, 44(7): 402－409. (in Chinese)
[27] 周俊伟,冯浩,董勤各. 基于不同时间尺度与特征优选的黄淮海平原冬小麦识别[J]. 农业机械学报, 2024, 55(9): 262－274.
ZHOU Junwei, FENG Hao, DONG Qin'ge. Identification of winter wheat in Huang－Huai－Hai plain based on different time scales and feature preference[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(9): 262－274. (in Chinese)
[28] 廖娟,梁业雄,姜锐,等. 基于 Transformer－FNN 和无人机高光谱遥感技术的棉花黄萎病危害等级分类研究[J]. 农业机械学报, 2025, 56(2): 240－251.
LIAO Juan, LIANG Yexiong, JIANG Rui, et al. Classification of cotton *Verticillium wilt* severity levels based on Transformer－FNN and UAV hyperspectral remote sensing technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 240－251. (in Chinese)

(下转第 717 页)

SHI Xiang. Research on the anti-evaporation water-saving rate and benefit analysis of floating ball covering water surface in floating box enclosure[D]. Urumqi: Xinjiang Agricultural University, 2020. (in Chinese)

[23] 李勋,严新军,侍克斌. 覆盖面积和网围结构对水面蒸发抑制率的影响[J]. 人民长江, 2019, 50(3): 117 - 123.
LI Xun, YAN Xinjun, SHI Kebin. Influence of coverage area and enclosure structure on suppression rate of water-surface evaporation[J]. Yangtze River, 2019, 50(3): 117 - 123. (in Chinese)

[24] LI C L, SHI K B, YAN X J, et al. Experimental analysis of water evaporation inhibition of plain reservoirs in inland arid area with light floating balls and floating plates in Xinjiang, China[J]. Journal of Hydrologic Engineering, 2021, 26(2): 04020060.

[25] HAO G, HAN K, SHI K. Effect of floating balls on evaporation inhibition, surface energy balance and biological water quality parameters at different coverage fractions[J]. Agricultural Water Management, 2023, 287: 108460.

[26] AMINZADEH M, LEHMANN P, OR D. Evaporation suppression and energy balance of water reservoirs covered with self-assembling floating elements[J]. Hydrology and Earth System Sciences, 2018, 22(7): 4015 - 4032.

[27] BENJAORAN V, RAKSAKSRI L, SETTAKHUMPOO J. Experimentation of evaporation reduction by a use of plastic bottles covering[J]. Water Conservation & Management, 2023, 7(2): 114 - 121.

[28] LI C, LI C, GU J, et al. Anti-evaporation effect and water-saving benefit analysis of water surface evaporation by different physical materials and coverage rate[J]. Physics and Chemistry of the Earth, 2022, 128: 103281.

[29] HAN K W, SSHI K B, YAN X J. Evaporation loss and energy balance of agricultural reservoirs covered with counterweighted spheres in arid region[J]. Agricultural Water Management, 2020, 238: 106227.

(上接第 707 页)

[29] 赵玉清,杨慧丽,张悦,等. 基于特征组合与 SVM 的小粒种咖啡缺陷生豆检测[J]. 农业工程学报, 2022, 38(14): 295 - 302.
ZHAO Yuqing, YANG Huili, ZHANG Yue, et al. Detection of defective Arabica green coffee beans based on feature combination and SVM[J]. Transactions of the CSAE, 2022, 38(14): 295 - 302. (in Chinese)

[30] 严萌,于雅雯,王玲静,等. 基于多特征联合稀疏表达的 SOM - K-means 非侵入负荷辨识[J]. 电力建设, 2023, 44(5): 61 - 71.
YAN Meng, YU Yawen, WANG Lingjing, et al. SOM - K-means non-intrusive load identification based on multi feature joint sparse expression[J]. Electric Power Construction, 2023, 44(5): 61 - 71. (in Chinese)

[31] 安晓飞,王培,罗长海,等. 基于 K-means 聚类和分区寻优的秸秆覆盖率计算方法[J]. 农业机械学报, 2021, 52(10): 84 - 89.
AN Xiaofei, WANG Pei, LUO Changhai, et al. Corn straw coverage calculation algorithm based on K-means clustering and zoning optimization method[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10): 84 - 89. (in Chinese)

[32] 刘东,杨丹,张亮亮,等. 基于 BOA - SVM 模型的区域洪水灾害风险评估与驱动机制[J]. 农业机械学报, 2023, 54(10): 304 - 315.
LIU Dong, YANG Dan, ZHANG Liangliang, et al. Regional flood disaster risk assessment and driving mechanism based on BOA - SVM model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 304 - 315. (in Chinese)

[33] 马亚丽,牛最荣,王兴繁,等. 缺资料地区产流径流时空特性分析及其关系研究——以洮河流域为例[J]. 水资源与水工程学报, 2023, 34(1): 58 - 65.
MA Yali, NIU Zuirong, WANG Xingfan, et al. Spatio-temporal characteristics of water yield and runoff and their relationship in regions with scarce data: a case study of Taohe River basin[J]. Journal of Water Resources and Water Engineering, 2023, 34(1):58 - 65. (in Chinese)

[34] 林康聆,周研来,陈华,等. 基于自组织映射神经网络的水库洪水过程分类[J]. 水科学进展, 2022, 33(6): 934 - 943.
LIN Kangling, ZHOU Yanlai, CHEN Hua, et al. Classification of reservoir flood processes based on self-organizing map neural networks[J]. Advances in Water Science, 2022, 33(6): 934 - 943. (in Chinese)

[35] 郑彦辰,李建柱,荣佑同,等. 降雨时空分布量化及其在洪水过程分类中的应用[J]. 水利学报, 2022, 53(5): 560 - 573.
ZHENG Yanchen, LI Jianzhu, RONG Youtong, et al. Quantification of rainfall spatial and temporal distribution characteristics on the flood hydrograph and its application in flood type classification[J]. Journal of Hydraulic Engineering, 2022, 53(5): 560 - 573. (in Chinese)

[36] 周炜兴,孙昭华,周坤,等. 三峡水库蓄水前后长江中下游流量频率分布特征及其对洪水造床作用的影响[J]. 湖泊科学, 2022, 34(2): 616 - 629.
ZHOU Weixing, SUN Zhaohua, ZHOU Kun, et al. Characteristics of discharge frequency and their effects on flood channel formation in the middle and lower reaches of the Yangtze River before/after the impoundment of the Three Gorges Reservoir [J]. Journal of Lake Sciences, 2012, 34(2): 616 - 629. (in Chinese)