DOI:10.13409/j.cnki.jdpme.2020.03.016

基于SVM-RF的泥石流窗口坝闭塞度判别研究*

焦 亮[1,2,3], 柳金峰[1,2], 游 勇[1,2], 袁 东[1,2,3], 周文兵[1,2,3]

(1. 中国科学院山地灾害与地表过程重点实验室,四川 成都 610041; 2. 中国科学院水利部成都山地灾害与环境研究所,四川 成都 610041; 3. 中国科学院大学,北京 100049)

摘要: 为了研究野外泥石流防治工程中窗口坝的开口闭塞类别,基于量纲分析理论,以室内水槽试验模拟实际工程,分析模型试验与实际工程的相关物理量及对应的相似准数;引入支持向量机和随机森林分类模型,在开源机器学习工具Scikit-Learn中,采用python编程实现算法;以室内水槽试验数据作为支持向量机和随机森林的训练样本,进行机器学习得到分类模型,提出一种用于判别泥石流窗口坝闭塞类型的新方法;将测试结果与经验公式中闭塞度判别值F的分类结果进行正确率对比,结果表明,F值的分类准确率为88%,而支持向量机为92%,随机森林为94%,随机森林分类效果最好,机器学习理论为泥石流窗口坝在实践中的设计提供了新思路。

关键词: 支持向量机; 随机森林; 窗口坝; 闭塞度; 相似分析; 机器学习

中图分类号: P642.23 **文献标识码:** A **文章编号:** 1672-2132(2020)03-0439-08

Research on the Occlusion of Debris Flow Window-frame Dam based on SVM and RF Methods

JIAO Liang[1,2,3], LIU Jinfeng[1,2], YOU Yong[1,2], YUAN Dong[1,2,3], ZHOU Wenbing[1,2,3]

(1. Key Laboratory of Mountain Hazards and Earth Surface Process, CAS, Chengdu 610041, China;
2. Institute of Mountain Hazards and Environment, Chinese Academy of Science, Chengdu 610041, China;
3. University of Chinese Academy of Science, Beijing 100049, China)

Abstract: In order to investigate the window-frame block categories of the field debris flow prevention engineering projects, the laboratory flume experiment is conducted to simulate the actual condition, using dimensional analysis method to ensure the similarity criterion of model test and the actual engineering. This research introduces the basic theory of support vector machine and random forest and realizes the algorithm in python language environment through the open source machine learning tool Scikit-Learn. Making the laboratory flume experiment data as the training sample of support vector machines and random forests then, got the learning classification model, put forward a new method for identifying debris flow dam block type. The test results compared with the empirical formula of discriminant value F block degree of accuracy of the classification, the results show that the F value of classification accuracy is 88%, and the support vector machine (SVM) was 92%, the random forest

* **收稿日期:**2018-06-03;**修回日期:**2018-08-16

基金项目:国家自然科学基金面上项目(41772343)、中国科学院"西部青年学者"项目(2018)资助

作者简介:焦 亮(1995-),男,硕士研究生。主要从事山地灾害实验与防治工程研究。Email:jl@imde.ac.cn

通讯作者:柳金峰(1979-),男,研究员,博导,博士。主要从事山地灾害实验与防治工程研究。Email:liujf@imde.ac.cn

was 94%. The random forest classification effect is best. The machine learning theory provides a new idea on the debris flow window-frame dam design in practice.

Keywords: support vector machine; random forest; window-frame dam; extent of closing; similarity analysis; machine learning

引　言

窗口坝是一种开口尺寸介于筛子坝和缝隙坝间的新型透过型坝，其在实体坝上开一些孔洞，起透水输砂、拦粗排细、调节流量的功能[1]。传统泥石流拦挡坝多采用非透过型坝，易出现坝体满库、淤积、破坏等现象，坝体提前达到使用极限，失去拦挡效果。窗口坝因能起拦蓄泥石、分离水土、稳固沟岸的作用而逐渐应用于实际工程中[2]，如云南东川大桥河坝[3]。窗口坝开口闭塞度是反映其拦挡泥石流性能的重要参数，主要体现在当窗口坝开口逐渐被泥石流堵塞时，坝体不能及时排泄粒径较小的泥石流，将会快速达到满库状态，威胁到周边区域安全。窗口坝开口尺寸的设计方法多参考水工设计，较依赖于设计者的经验，缺乏相应的理论依据和技术支持[4]，赵彦波等[5]、刘曙亮等[6]通过室内水槽试验、初步探索了坝体开口尺寸对泥石流的拦挡作用，提出了闭塞度判别式的经验公式。工程设计中，需要快速判别不同参数下坝体开口闭塞类型，这实则是一个多分类问题。

支持向量机(Support vector machine，SVM)是以VC维理论和结构风险、经验风险最小化理论为基础，在学习精度、能力上有较好的适应性[7-9]。SVM主要通过核函数进行非线性变换，将原始数据映射到高维特征空间，寻找最优分类超平面，使得在低维空间原本不可分的数据变得可分；SVM分类是一个凸二次规划问题，得到的解是全局最优解，所以广泛应用于小样本、非线性、高维数等实际问题，如遥感图像[10]、Internet网页信息[11]、人脸识别[12]、文本图像[13]等问题；随机森林(Random Forest，RF)是抽样组合决策的统计学习理论，利用多棵树对样本训练并预测的一种分类器，通过Bootstrap重抽样方法抽取多个样本进行决策树建模，然后将建模的决策树组合，最后投票得出分类结果[14]。随机森林分类方法不会随着决策树的增加而产生过度拟合问题，引入随机性来提高预测精度[15]，因而被广泛应用。将支持向量机和随机森林理论应用于泥石流窗口坝闭塞类型判别中，为坝体参数的设计拓宽思路。

1　支持向量机理论

1.1　基本概念

支持向量机主要步骤是先选择合适的核函数把样本数据映射到高维空间，然后在高维空间寻找支持向量来构造最优分类超平面，超平面应尽量将样本数据正确分开，同时保证分开的样本数据距离分类面最远，方程描述为：

$$\boldsymbol{w}^{\mathrm{T}}x+b=0 \tag{1}$$

式中，$\boldsymbol{w}=(w_1;\ w_2;\ ...;\ w_d)$和$b$为约束参数。

当超平面将训练样本分类正确时，即$y_i=1$，$(x_i,y_i)\in D$，则有：

$$\begin{cases}\boldsymbol{w}^{\mathrm{T}}x_i+b\geqslant 1, y_i=1\\ \boldsymbol{w}^{\mathrm{T}}x_i+b\leqslant -1, y_i=-1\end{cases} \tag{2}$$

满足式(2)的训练样本称为支持向量，定义两个异类支持向量到超平面的距离之和为：

$$\gamma=\frac{2}{\|\boldsymbol{w}\|} \tag{3}$$

寻找超平面就是要找到能满足式(1)中约束的参数$\boldsymbol{w}$和b，使得γ最大，即：

$$\begin{cases}\min\limits_{w,b}=\dfrac{1}{2}\|\boldsymbol{w}\|^2\\ s.t.\ y_i(\boldsymbol{w}^{T}x_i+b)\geqslant 1, i=1,2,..,m\end{cases} \tag{4}$$

1.2　核函数

样本通过核函数进行映射，设$\boldsymbol{\phi}(x)$为x映射后的特征向量，于是得到划分超平面的函数为：

$$f(x)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(x)+b \tag{5}$$

式中，$\boldsymbol{w}$和b为模型参数，此时有：

$$\begin{cases}\min\limits_{w,b}\dfrac{1}{2}\|\boldsymbol{w}\|^2\\ s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(x_i)+b)\geqslant 1, i=1,2,...,m\end{cases} \tag{6}$$

式(6)通过求解对偶问题得到划分超平面所对应的模型：

$$f(x)=\sum_{i=1}^{m}\alpha_i y_i k(x,x_i)+b \tag{7}$$

式中，$k(x,x_i)$为核函数。

1.3 软间隔支持向量机

在高维空间中，所有样本必须划分正确称为硬间隔，但划分训练样本的核函数一般很难确定[16]，因此，需要在一定程度上允许支持向量机在一些样本上出错，称为软间隔，如图1中⊕和⊖表示满足一定精度下允许出错的点。

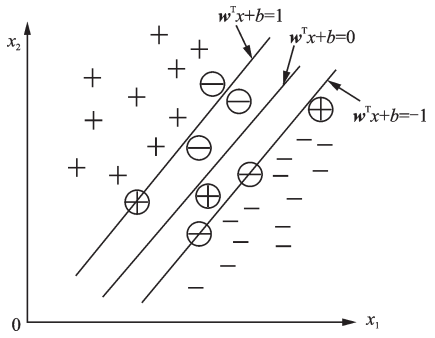

图1 样本软间隔

Fig.1 The soft interval of sample

在软间隔下，出错的样本点应尽可能少，于是得到软间隔支持向量机：

$$\begin{cases}\min\limits_{w,b,\zeta_i}\dfrac{1}{2}\|\boldsymbol{w}\|^2+C\sum\limits_{i=1}^{m}\zeta_i\\ s.t\ y_i(\boldsymbol{w}^T x_i+b)\geqslant 1-\zeta_i\\ \zeta_i\geqslant 0,i=1,2,\cdots,m\end{cases} \tag{8}$$

式中，$\zeta_i(\zeta_i\geqslant 0)$表示松弛变量，用来表征每个样本不满足约束的程度；$C$表示惩罚因子，表示对错判样本的惩罚程度。

式(8)是一个凸二次规划问题，可通过拉格朗日乘子法得到拉格朗日函数，令函数对$\boldsymbol{w},b,\zeta_i$的偏导为0，化简回代即可求解对偶问题，其中拉格朗日乘子采用SMO算法[17]进行求解。

2 随机森林

随机森林(Random Forest，RF)是一种抽样组合决策的统计学习理论。随机产生的森林由一棵棵决策树组成，每棵决策树之间没有关联，森林通过每一棵决策树分别对新输入的样本进行判断，然后分类，决策树较多的类别即为样本属性。建立每一棵决策树时，需注意两点：采样与完全分裂。随机森林对输入的数据需要进行行、列采样，行采样采取有放回的方式，也即采样可能有重复的样本，才会在训练的时候每一棵树的输入样本都不是全部的样本，相对不易出现过拟合；列采样从N个特征中选择m个($m\ll N$)，采取完全分裂来建立决策树，这样得到的决策树在每一个叶子节点上无法继续分裂或里面的所有样本都指向同一个分类。由大数定律可知，分类树数目越多，泛化误差值不断收敛[15]。模型通过构建大量的训练集分类差异来提高外推能力，设通过N轮训练，得到的分类模型序列为$\{h_1(X),h_2(X),...,h_k(X)\}$，通过多数投票法来决定该系统的分类效果，函数表达式为：

$$\mathrm{H}(x)=\arg\ \max_{Y}\sum_{i=1}^{k}I(h_i(x)=Y) \tag{9}$$

式中，$H(x)$为组合分类模型；$h_i(x)$为单个决策树分类模型；Y为目标变量；$I(.)$为示性函数。

3 窗口坝模型试验的相似分析

窗口坝对泥石流的运动调控是一个复杂的动力学过程，坝体开口闭塞度是反映窗口坝对泥石流拦挡性能的重要参数，其影响因素主要有：①坝体开口特征尺寸：开口高度h，开口宽度b，开口数量n，模型试验和实际工程有着相同的开口数，这里不考虑开口数量；②沟道条件：沟床纵比降j，j主要影响泥石流运动速度和拦砂坝有效库容，模型试验与实际工程的沟道比降设置相同，这里不考虑此因素；③越坝前泥石流物理属性：泥石流容重γ_c，水的容重γ_w，块石的重度γ_s，采用泥石流土体体积浓度C_v，即$C_v=(\gamma_c-\gamma_w)/(\gamma_s-\gamma_w)$化为无量纲数进行研究；泥石流规模$V_m$，泥石流土体颗粒特征粒径$d_r$；④越坝前泥石流运动情况：泥石流流速$\nu_0$和泥深$h_0$。任何一个物理定理总可以表示为确定的函数关系[18]，于是得到：

$$\psi=f(h,b,C_v,V_m,d_r,\nu_0,h_0,t,\mathrm{g}) \tag{10}$$

式中，ψ为坝体开口闭塞度；h和b分别为窗口坝开口高度和宽度；C_v为泥石流土体体积浓度；V_m为泥石流规模；d_r为泥石流土体颗粒特征粒径；ν_0为泥石流流速；h_0为泥深；t为运动时间，g为重力加速度。

选取3个具有独立量纲的基本量：时间t、泥石流土体颗粒特征粒径d_r、重力加速度g，根据π定理可知，9个自变量可通过量纲分析变成6个无量纲自变量，即：

$$\prod = f(\pi_1,\pi_2,\pi_3,\pi_4,\pi_5,\pi_6) \tag{11}$$

引入泥石流运动学中的无量纲数可将式(11)化为：

$$\frac{\psi_1-\psi_2}{\psi_1} = f\left(\frac{h}{d_r},\frac{b}{d_r},\frac{\gamma_c-\gamma_w}{\gamma_s-\gamma_w},\frac{V_m}{{d_r}^3},\frac{\nu_0}{\sqrt{g.h_0}}\right) = f\left(\frac{h}{d_r},\frac{b}{d_r},C_v,\frac{V_m}{{d_r}^3},F_r\right) \tag{12}$$

泥石流运动是一个相互耦合作用的过程，在确定的试验条件下，$(V_m/{d_r}^3)$为一常数，用β_0表示，因此可将式(12)进一步写成如下形式：

$$\frac{\psi_1-\psi_2}{\psi_1} = f\left(\frac{h}{d_r},\frac{b}{d_r},C_v,\frac{V_m}{{d_r}^3},F_r\right) = \beta_0 \cdot f_1\left(\frac{h}{d_r}\right)\cdot f_2\left(\frac{b}{d_r}\right) f_3(C_v).f_4(F_r) \tag{13}$$

式中，$(\psi_1-\psi_2)/\psi_1$为无量纲因变量，表征过坝前后闭塞度削减率量化指标；β_0为随试验条件变化的常数；(h/d_r)和(b/d_r)为窗口坝相对开度，表征坝体开口与泥石流颗粒组成关系；C_v和$(V_m/{d_r}^3)$分别为泥石流土体体积浓度和泥石流相对规模，均表征泥石流属性；F_r为弗劳德数，表征泥石流越坝前的运动性质。

要使野外实际工程具有可模拟性，室内水槽模型试验需要满足以下条件：

(1)窗口坝相对开度(h/d_r)，(b/d_r)相同；

(2)泥石流土体体积浓度C_v相同；

(3)泥石流的弗德劳数F_r相同。

4　模型试验和算法实现

4.1　窗口坝室内水槽试验

窗口坝坝型之一的3D模型如图2所示。

为使试验更加接近野外工程真实情况，窗口坝均采用混凝土坝，模型固定在全长2 m、宽20 cm、高30 cm，坡度为9°的水槽中，水槽如图3所示。

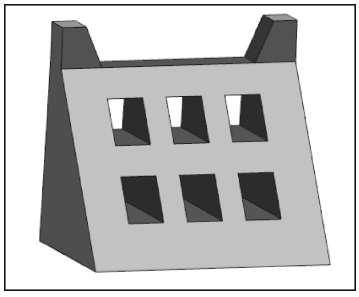

图2　窗口坝模型

Fig.2　Window-frame dam model

实验中窗口坝的开口出现全堵塞、不堵塞和部分堵塞类型，拍摄到的部分情况如图4、5所示；定义窗口坝的闭塞度为开口被堵塞面积与开口总面积的比值。

影响窗口坝闭塞度的指标较多，但主要因素为泥石流容重，窗口坝开口宽度、高度、开口数量、高宽比、开口面积等，为了避免小数被大数影响，对数据进行(0,1)归一化。泥石流容重只设计了三个值，归一化后对数据影响较大，将原容重均缩小5倍，化为0～1即可。闭塞度是一个数值，通过表1对

图3　室内水槽

Fig.3　Indoor water gullet experiment

图4　窗口坝部分闭塞

Fig.4　The Partial occlusion of Window-Frame Dam

图5　窗口坝全部闭塞

Fig.5　The Total occlusion of Window-Frame Dam

表1　闭塞度类别

Table 1　The type of occlusion

开口闭塞	[0,25%]	(25%, 50%]	(50%, 75%]	(75%, 100%]
类别	1	2	3	4

数值进行“标签化”,即闭塞值所属类别,使得经验公式和机器学习理论能在同一平台上对比分类结果正确率。实验中通过控制变量法测量不同影响参数下窗口坝的闭塞情况,得到表2所示结果。

表2　不同控制参数量化下窗口坝闭塞度

Table 2　The occlusion of Window-Frame under different control parameters

坝体编号	容重γ	b	h	开口个数	h/b	开口面积	闭塞度
1	0.3	0	0	0	0	0	1
2	0.3	0.250 0	0.444 4	1	0.667 5	0.444 4	1
3	0.3	0.250 0	0.666 7	0.666 7	1	0.444 4	1
4	0.3	0.333 3	0.222 2	1	0.250 0	0.296 3	1
5	0.3	0.333 3	0.444 4	0.500 0	0.500 0	0.296 3	1
6	0.3	0.333 3	0.666 7	0.333 3	0.750 0	0.296 3	1
7	0.3	0.333 3	0.666 7	0.666 7	0.750 0	0.592 6	0.383 0
8	0.3	0.333 3	0.666 7	0.666 7	0.750 0	0.592 6	0.813 0
9	0.3	0.416 7	0.666 7	0.500 0	0.600 0	0.555 6	0.119 0
10	0.3	0.500 0	0.333 3	1	0.250 0	0.666 7	0.667 0
11	0.3	0.500 0	0.333 3	1	0.250 0	0.666 7	0.3
12	0.3	0.500 0	0.666 7	0.500 0	0.500 0	0.666 7	0
13	0.3	0.500 0	1	0.500 0	0.750 0	1	0
14	0.3	0.500 0	0.888 9	0.500 0	0.667 5	0.888 9	0
15	0.3	0.583 3	0.666 7	0.333 3	0.427 5	0.518 5	0
16	0.3	0.666 7	0.666 7	0.333 3	0.375 0	0.592 6	0
17	0.3	1	0.333 3	0.333 3	0.125 0	0.444 4	0
18	0.3	1	0.444 4	0.333 3	0.167 5	0.592 6	0
19	0.36	0	0	0	0	0	1
20	0.36	0.250 0	0.444 4	1	0.667 5	0.444 4	1
21	0.36	0.250 0	0.666 7	0.666 7	1	0.444 4	1
22	0.36	0.333 3	0.222 2	1	0.250 0	0.296 3	1
23	0.36	0.333 3	0.444 4	0.500 0	0.500 0	0.296 3	1
24	0.36	0.333 3	0.666 7	0.333 3	0.750 0	0.296 3	1
25	0.36	0.333 3	0.666 7	0.666 7	0.750 0	0.592 6	0.708 0
26	0.36	0.416 7	0.666 7	0.500 0	0.600 0	0.555 6	0.333 0
27	0.36	0.500 0	0.333 3	1	0.250 0	0.666 7	0.583 0
28	0.36	0.500 0	0.666 7	0.500 0	0.500 0	0.666 7	0
29	0.36	0.500 0	1	0.500 0	0.750 0	1	0
30	0.36	0.500 0	0.888 9	0.500 0	0.667 5	0.888 9	0
31	0.36	0.583 3	0.666 7	0.333 3	0.427 5	0.518 5	0
32	0.36	0.666 7	0.666 7	0.333 3	0.375 0	0.592 6	0
33	0.36	1	0.333 3	0.333 3	0.125 0	0.444 4	0
34	0.36	1	0.444 4	0.333 3	0.167 5	0.592 6	0
35	0.4	0	0	0	0	0	1
36	0.4	0.250 0	0.048 9	1	0.667 5	0.444 4	1
37	0.4	0.250 0	0.060 0	0.666 7	1	0.444 4	1
38	0.4	0.333 3	0.043 3	1	0.250 0	0.296 3	1

续表

坝体编号	容重γ	b	h	开口个数	h/b	开口面积	闭塞度
39	0.4	0.333 3	0.061 1	0.500 0	0.500 0	0.296 3	1
40	0.4	0.333 3	0.075 6	0.333 3	0.750 0	0.296 3	1
41	0.4	0.333 3	0.106 7	0.666 7	0.750 0	0.592 6	1
42	0.4	0.416 7	0.144 4	0.500 0	0.600 0	0.555 6	1
43	0.4	0.500 0	0.207 8	0.500 0	0.500 0	0.666 7	1
44	0.4	0.500 0	0.146 7	1	0.250 0	0.666 7	1
45	0.4	0.500 0	0.207 8	0.500 0	0.500 0	0.666 7	1
46	0.4	0.500 0	0.312 2	0.500 0	0.750 0	1	0.296 0
47	0.4	0.500 0	0.277 8	0.500 0	0.667 5	0.888 9	1
48	0.4	0.583 3	0.231 1	0.333 3	0.427 5	0.518 5	1
49	0.4	0.583 3	0.231 1	0.333 3	0.427 5	0.518 5	1
50	0.4	0.666 7	0.302 2	0.333 3	0.375 0	0.592 6	0.667 0
51	0.4	1	0.302 2	0.333 3	0.167 5	0.592 6	0.125 0

注：表中泥石流容重参数为原试验容重做缩小5倍处理，其余控制参数进行(0,1)归一化

4.2 算法实现和分类准确率对比

支持向量机和随机森林算法实现目前已经得到很好的解决，开源机器学习工具Scikit-Learn是基于python的一个机器学习模块[19]，在数据挖掘、数据分析有着广泛的应用[20]。在python语言环境中，通过命令from sklearn import SVM 和from sklearn.ensemble import Random Forest Classifier即可调用机器学习工具里的SVM模型和RF模型，避免了繁琐的编程，模型建立的好坏主要取决于参数的选择，通过命令random.permutation调用库实现51组数据进行随机打乱，选择41组作为训练样本，10组作为测试样本，设置不同参数对数据进行训练和测试。根据分类准确率对SVM和RF参数进行择优选择，见表3、表4。

表3中，C表示错误项的惩罚系数；rbf表示径向基函数；auto表示γ值默认为样本特征数的倒数；ovr(one-versus-rest)表示通过组合多个二分类器时采用一对多法，即训练时依次把某个类别的样本归为一类，其它剩余样本归为另一类；表4中，bootstrap为True表示有放回的采样；criterio的计算属性为gini(基尼不纯度)；n_jobs=1表示不并行计算；n_estimators=10表示决策树的个数；min_samples_leaf表示叶子节点最少的样本数；min_samples_split表示根据属性划分节点时每个划分最少的样本数。

为了判别SVM模型的分类性能，采用赵彦波等[4]经验判别公式计算F值进行对比：

$$F=\begin{cases}[\min(b,h)/d_{95}]^2\times\sqrt{e}\times\sqrt{1/m} & (m<1)\\ [\min(b,h)/d_{95}]^2\times\sqrt{e}\times\sqrt{m} & (m\geqslant 1)\end{cases} \tag{14}$$

式中，F为临界条件判别系数；d_{95}为颗粒级配曲线中累计百分含量95%对应的颗粒粒径；e为开口总面积与有效高度以下横断面面积之比；m为开口宽度与开口高度之比；当$F\leqslant 0.95$时，窗口坝全闭塞；当$0.95<F\leqslant 1.5$时，窗口坝部分闭塞；当$F>1.5$时，窗口坝不闭塞。

以试验数据作为样本，对三种方法进行10次分类效果测试，每次测试均从51个样本中随机选取41个作训练样本，10个作测试样本；将经验公式计算的F值所对应的窗口坝闭塞类型与支持向量机、随

表3 SVM模型参数设置

Table 3 The model parameter setting of SVM

主要参数	惩罚因子C	核函数	核函数参数γ	Decision_function
参数值	430	rbf	auto	ovr

表4 随机森林参数设置

Table 4 The model parameter setting of RF

主要参数	bootstrap	criterion	n_estimators	n_jobs	min_samples_leaf	min_samples_split
参数值	True	gini	10	1	1	2

机森林分类结果对比，计算出每次试验分类准确率，见表5。

表5　分类方法对比

Table 5　Comparison of classification methods

次数	方法	测试样本										准确率/%
		1	2	3	4	5	6	7	8	9	10	
1	真实值	4	4	4	3	4	3	4	4	4	4	
	RF	4	4	4	2	4	3	4	4	4	4	90
	SVM	4	4	4	2	4	3	4	4	4	4	90
	经验公式	4	4	4	2	4	2	4	4	4	4	80
2	真实值	1	4	4	1	4	4	4	1	4	4	
	RF	1	4	4	1	4	4	4	1	4	4	100
	SVM	1	4	4	1	4	4	4	1	4	4	100
	经验公式	1	4	4	1	4	4	4	1	4	4	100
3	真实值	1	4	4	4	4	1	4	4	4	4	
	RF	1	4	4	4	4	1	4	4	4	4	100
	SVM	1	4	4	4	4	1	4	4	4	4	100
	经验公式	1	4	4	4	4	1	4	4	4	4	100
4	真实值	4	2	4	1	4	1	4	1	2	4	
	RF	4	2	4	1	4	1	4	1	2	4	100
	SVM	4	2	4	1	4	1	4	1	1	4	90
	经验公式	4	3	4	1	4	1	4	1	3	4	80
5	真实值	4	1	2	3	1	1	4	3	4	4	
	RF	2	1	3	3	1	1	4	3	4	4	80
	SVM	2	1	3	3	1	1	4	3	4	4	80
	经验公式	4	1	3	3	1	1	4	2	4	4	80
6	真实值	2	3	4	1	4	4	4	1	4	4	
	RF	2	3	4	1	4	4	4	1	4	4	100
	SVM	1	3	4	1	4	4	4	1	4	4	90
	经验公式	3	3	4	1	4	4	4	1	4	4	90
7	真实值	1	4	4	4	4	1	4	2	3	2	
	RF	1	4	4	4	4	1	4	3	3	2	90
	SVM	1	4	4	4	4	1	4	3	3	2	90
	经验公式	1	4	4	4	4	1	4	3	2	2	80
8	真实值	4	4	1	4	4	4	3	4	4	4	
	RF	4	4	1	4	4	4	3	4	4	4	100
	SVM	4	4	1	4	4	4	3	4	2	4	90
	经验公式	4	4	1	4	4	4	2	4	4	4	90
9	真实值	1	4	4	4	4	1	1	4	1	4	
	RF	1	4	4	4	4	1	1	2	1	4	90
	SVM	1	4	4	4	4	1	1	2	1	4	90
	经验公式	1	4	4	4	4	1	1	4	1	4	100
10	真实值	3	4	4	4	2	1	1	4	4	4	
	RF	3	4	4	4	4	1	1	4	4	4	90
	SVM	3	4	4	4	2	1	1	4	4	4	100
	经验公式	2	4	4	4	3	1	1	4	4	4	80

将10次试验的分类结果进行统计，得到经验公式的分类准确率为88%，随机森林为94%，支持向量机为92%。

4.3　结果分析

从表5可知，在进行的10次实验中，三种方法均在第2类和第3类闭塞度易出现误判，如第1次试验中的样本4，第5次实验中的样本3等；但在第1类和第4类上分类效果理想，如第2次、3次试验，3种分类结果准确率均为100%；因此，三种分类方法的好坏主要体现在第2类和第3类样本的划分上。

从分类统计结果可知，经验公式的分类准确率为88%，低于机器学习算法，其主要原因是经验公式判别窗口坝闭塞度类别由$\min(b,h)/d_{95}$和F共同决定，在完全闭塞和不闭塞时，F值能准确的判断窗口坝闭塞类型，但当$\min(b,h)/d_{95}$在不完 全闭塞下，仅通过F值进行判断，得到的闭塞度划分区间精度不够准确，出现偏差，影响判别结果，且当$\min(b,h)$小于d_{95}时，即高宽比和面积较大，F值判别失效，经验公式确定的窗口坝闭塞度还不够完善。

机器学习中的RF和SVM模型分类准确率均达到90%以上，主要原因是调用了开源机器学习工具Scikit-Learn，通过不断进行机器学习，优化相关参数；SVM在高维空间寻找到最优分类超平面，当输入10组测试样本时，超平面能快速判别样本类别，随机森林理论将Bagging方法和随机选择特征分裂相结合，使得该算法能较好地容忍异常值和噪声，因此在三种方法中正确率最高。

5　结　论

将机器学习理论支持向量机和随机森林应用到泥石流窗口坝开口闭塞度类型判别中是一种新的尝试，根据模型试验和分类方法准确率对比，得到以下结论：

(1)传统经验公式计算F的分类准确率为88%，支持向量机模型分类准确率为92%，而随机森林模型分类准确率为94%，分类准确率最高，因此，通过支持向量机和随机森林理论来建立闭塞度类别与泥石流容重、坝体开口参数的非线性函数模型是可行的，模型通过数据的高维变化处理，相比采用原始数据拟合出的经验公式更为优化，窗口坝闭塞类型判别更为准确。

(2)机器学习理论除了具有较高的分类正确率外,分类模型不必知道窗口坝闭塞类型与影响参数的函数关系,只需将测试样本属性输入训练好的分类模型便能快速得到结果,更适用于工程的快速判别中,在一定程度上降低了工程的设计难度,为窗口坝参数的工程设计提供一个新思路。

参考文献:

[1] 周必凡,李德基,吕孺人,等.泥石流防治指南[M].北京:科学出版社,1991.
Zhou B F, Li D J, Lyu R R, et al. Guidelines for debris flow control[M]. Beijing: Science Press, 1991. (in Chinese)

[2] 李德基.透水型拦挡坝在泥石流防治中的应用[J].中国地质灾害与防治学报,1997,8(4):60-66.
Li D J. The application of permeable dam in debris flow control[J]. The Chinese journal geological hazard and control, 1997,8(4):60-66.(in Chinese)

[3] 康志成,罗德富,张 军,等.中国泥石流灾害与防治[M].北京:科学出版社,1996.
Kang Zh Ch, Luo D F, Zhang J, et al. China debris flow injury prevention[J]. Beijing: Science Press, 1986. (in Chinese)

[4] 赵彦波,游 勇,柳金峰,等.泥石流窗口坝闭塞类型及其临界条件实验研究[J].防灾减灾工程学报,2015,35(2):256-262.
Zhao Y B, You Y, Liu J F, et al. Experimental study on blocked types and critical conditions of window frame dam Preventing debris flow[J]. Journal of Disaster Prevention and Mitigation Engineering, 2015, 35(2):256-262. (in Chinese)

[5] 赵彦波,游 勇,柳金峰,等.泥石流窗口坝调节泥砂粒径试验研究[J].长江科学院院报,2016,33(3):9-13.
Zhao Y B, You Y, Liu J F, et al. Experimental study on the function of window-frame dam in changing sediment size distribution for debris flow prevention[J]. Journal of Yangtze River Scientific Research Institute, 2016,33(3):9-13.(in Chinese)

[6] 刘曙亮,游 勇,柳金峰,等.窗口坝拦截泥石流性能试验研究[J].长江科学院院报,2015,32(8):40-44.
Liu Sh L, You Y, Liu J F, et al .Experimental study on perfo-rmance of window-frame dam intercepting debris flow[J]. Journal of Yangtze River Scientific Research Institute, 2015,32(8):40-44.(in Chinese)

[7] 边肇祺,张学工.模式识别[M].2版.北京:清华大学出版社,2000.
Bian Zh Q, Zhang X G. Schematic identification[M].2rd ed. Beijing: Tsinghua University Press, 2000. (in Chinese)

[8] Andrew A M. An introduction to support vector machines and other kernel-based learning methods[J]. Kybernetes, 2000, 32(1):1-28.

[9] 邓乃扬.数据挖掘中的新方法[M].北京:科学出版社,2004.
Deng N Y. A new way of numerical drilling[M]. Beijing: Science Press, 2004.(in Chinese)

[10] 张锦水,何春阳,潘耀忠,等.基于SVM的多源信息复合的高空间分辨率遥感数据分类研究[J].遥感学报,2006,10(1):49-57.
Zhang J Sh, He Ch Y, Pan Y Zh, et al. The high spatial resolution RS image classification based on SVM method with the multi-source data[J]. Journal of Remote Sensing, 2006,10(1):49-57. (in Chinese)

[11] 李晓黎,刘继敏,史忠植.基于支持向量机与无监督聚类相结合的中文网页分类器[J].计算机学报,2001,24(1):62-68.
Li X L, Liu J M, Shi Zh Zh. A chinese web page classifier based on support vector machine and unsupervised clustering[J]. Chinese Journal of Computers, 2001,24(1):62-68. (in Chinese)

[12] 王雪峰,周国标.基于SVM的人脸识别方法研究[J].上海应用技术学院学报(自然科学版),2006,6(2):104-107.
Wang X F, Zhou G B.A Research on the SVM method for facial recognition[J]. Journal of Shanghai Institute of Technology (Natural Science), 2006,6(2):104-107. (in Chinese)

[13] 张红涛,胡玉霞,毛罕平,等.基于SVM的储粮害虫图像识别分类[J].农机化研究,2008(8):36-38.
Zhang H T, Hu Y X, Mao H P, et al. Image recognition and classification of the stored-grain pests based on support vector machine[J]. Agricultural Research,2008(8):36-38. (in Chinese)

[14] Breiman L. Random forests, machine learning 45[J]. Journal of Clinical Microbiology, 2001, 2:199-228.

[15] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
Fang K N, Wu J B, Zhu J P, et al. Concerning forest method research[J]. Statistics and Information Forum, 2011,26(3):32-38.(in Chinese)

[16] 周志华.机器学习[M].北京:清华大学出版社,2016.
Zhou Zh H. Machine learning[M]. Beijing: Tsinghua University Press, 2016. (in Chinese)

(本文责编:池营营)

(下转第466页)