

# 多分类支持向量机在泥石流危险性区划中的应用

李秀珍<sup>1,2</sup>, 孔纪名<sup>1,2</sup>, 李朝凤<sup>2</sup>

(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)

**摘要:** 以凉山州安宁河流域 129 个乡镇的泥石流危险性区划资料为依据, 随机选取总样本数的 2/3 和 1/2 作为训练样本, 建立不同数量训练样本下安宁河流域泥石流危险性区划的多分类 SVM 模型, 进行以乡镇为单元的区域泥石流危险性评价研究。评价结果表明, SVM 模型的预测精度随着训练样本数量的增加而提高; 2 个 SVM 模型对测试样本的预测准确率均高于相应的 BP 神经网络模型, 对训练样本的回判准确率高于或接近于 BP 神经网络模型。因此, 支持向量机方法是一种比神经网络方法具有更优精度和更强泛化性能的新机器学习方法, 在泥石流危险性评价实践中具有十分广阔的应用前景和推广应用价值。

**关键词:** 多分类支持向量机; 泥石流危险性区划; BP 神经网络; 凉山州; 安宁河流域

文献标识码: B

文章编号: 1000-288X(2010)05-0128-06

中图分类号: P642.23

## Application of Multi-classification Support Vector Machine in Regionalization of Debris Flow Hazards

LI Xiu-zhen<sup>1,2</sup>, KONG Ji-ming<sup>1,2</sup>, LI Chao-feng<sup>2</sup>

(1. Key Laboratory of Mountain Hazards and Surface Processes, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China; 2. Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China)

**Abstract:** Based on the debris flow data collected from 129 villages and towns in the Anning River valley of Liangshan Prefecture, two multi-classification support vector machine models were built to evaluate debris flow hazards of the villages and towns. 86 samples from the villages and 65 samples from the towns were randomly selected as training samples and the remainders, as testing samples. Results show that the prediction accuracy of SVM model is improved with the increase of training samples and prediction accuracy of the two SVM models are higher than that of BP neural network models. Therefore, support vector machine method is a new machine learning method with higher precision and better generalization performance than neural network method. It has very broad application prospects and promotion and application values in the practice of debris flow hazard assessment.

**Keywords:** multi-classification support vector machine; debris flow hazard regionalization; BP neural network; Liangshan Prefecture; Anning River valley

危险性评价是目前国内外灾害科学研究的热点之一,也是灾害预测预报和防灾减灾工作中的重要内容<sup>[1]</sup>。泥石流危险性评价是根据泥石流危险度划分出各区域(或单沟)泥石流危险等级的方法,其本质属于模式识别问题。

支持向量机(SVM)是 Vapnik 等人提出的一种新的机器学习方法,被许多学者认为是继模式识别和神经网络研究之后机器学习领域的新的研究热点,也是最有可能弥补神经网络容易陷入局部极小值,泛化

能力不强以及只能解决大样本数据等问题的机器学习方法<sup>[2-4]</sup>。该方法是建立在统计学习理论的 VC 维理论和结构风险最小化原理的基础上,根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度, accuracy)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳折衷,以期获得最好的推广能力(generalization ability)<sup>[2-4]</sup>。由于 SVM 在解决小样本、非线性、高维数及局部极小点等实际问题中表现出许多特有的优势,已经被广泛应用于手写识别,

收稿日期: 2009-09-03

修回日期: 2010-04-13

资助项目: 中国科学院山地灾害与地表过程重点实验室开放基金“基于支持向量机的潜在滑坡早期判别方法研究”(110100L104); 中国科学院“西部之光”人才培养计划项目“基于进化支持向量机的滑坡预测模型研究”(08R2140140)

作者简介: 李秀珍(1975—),女(汉族),内蒙古卓资县人,助理研究员,研究方向为地质灾害评价与预测。E-mail: lixiuzhen6688@163.com.

三维目标识别, 人脸识别, 文本图像分类等实际模式识别问题中<sup>[2,9]</sup>。近年来, 有少数研究者已开始尝试将支持向量机方法应用于泥石流等灾害的评价及预测中<sup>[7-13]</sup>, 如原立峰等<sup>[7]</sup>基于 SVM 进行了单沟泥石流危险性评价研究, 王涛等<sup>[8]</sup>将 SVM 方法应用于小流域泥石流输沙量预测, 均取得了良好的效果。但总的说来, SVM 方法在泥石流灾害领域的应用研究深度和广度均十分有限, 如 SVM 在区域泥石流危险性评价中的应用研究, 目前尚未见有文献报道。

本文以凉山州安宁河流域 129 个乡镇泥石流危险性区划基础资料为依据, 随机选取总样本数的 2/3 和 1/2 作为训练样本, 建立不同训练样本下泥石流危险性区划的多分类 SVM 模型, 进行以乡镇为单元的区域泥石流危险性评价研究, 取得了良好的效果。

## 1 支持向量机基本理论

支持向量机的核心是构造最优超平面。其基本思想可以概括为: 首先通过某种事先选择的非线性映射(核函数)将输入向量映射到一个高维特征空间, 然后在特征空间中寻找最优分类超平面, 使得它能够尽可能多地将两类数据点正确分开, 同时使分开的两类数据点距离分类面最远<sup>[6,7]</sup>(图 1)。

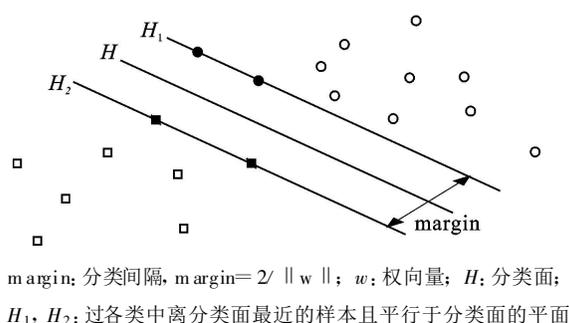


图 1 线性可分情况下的最优分类面

### 1.1 二分类支持向量机

假设存在训练样本  $\{x_i, y_i\}, i = 1, 2, \dots, l; x_i \in R^n, y_i \in \{-1, +1\}, l$  为样本数,  $n$  为输入维数。在线性可分的情况下, 会有一个分类超平面, 使两类样本完全分开:

$$\langle w \cdot x_i \rangle + b = 0 \quad (1)$$

求解分类超平面就是找到给定训练样本的权值  $w$  和阈值  $b$  的最优值, 故可归结为以下二次规划问题:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \text{s. t. } y_i \langle w \cdot x_i \rangle + b \geq 0 \quad (i = 1, 2, \dots, l) \end{cases} \quad (2)$$

通过求解上述二次规划的对偶问题, 即可得到对应的最优  $a^*$ , 对应  $a^* \neq 0$  的样本称为支持向量。选

择  $a^*$  的一个正分量  $a_j^*$ , 并据此计算  $b^* = y_j - \sum_{i=1}^l y_i a_i^* K(x_j, x_i)$ , 由此求得决策函数为:

$$F(x) = \text{sgn} \left[ \sum_{i=1}^l a_i^* y_i \langle x_i \cdot x \rangle + b^* \right] \quad (3)$$

在线性不可分情况下, 可引入松弛变量  $\xi_i$  和惩罚参数  $C$ , 这样式(2)就变为:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t. } y_i \langle w \cdot x_i \rangle + b \geq 1 - \xi_i \quad (\xi_i \geq 0, i = 1, 2, \dots, l) \end{cases} \quad (4)$$

对于线性不可分问题, 可通过引入一个非线性映射函数, 把样本映射到某个高维空间, 在属性空间将其转化为线性分类问题。只要函数  $K(x_i, x)$  满足 Mercer 条件, 都可以作为核函数, 并且  $K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ 。根据 Mercer 条件, 在最优分类面中采用不同的内积函数  $K(x_i, x)$  就可以实现某一非线性变换的线性分类。引入核函数后, 以上各式中的内积都可以用核函数代替。

同理, 通过求解线性不可分条件下的二次规划对偶问题, 可得到线性不可分条件下的分类决策函数为

$$F(x) = \text{sgn} \left[ \sum_{i=1}^l a_i y_i K(x_i, x) + b^* \right] \quad (5)$$

目前, 常用的核函数有: 线性核函数、多项式核函数、径向基核函数和 Sigmoid 核函数。支持向量机本质上是针对二分类问题提出的, 但二分类方法远远不能满足实际问题的需要。随着理论上的研究和应用上的不断推广, 逐渐出现了多分类支持向量机。

### 1.2 多分类支持向量机

多分类支持向量机算法主要是通过组合二分类支持向量机来实现的。目前, 基于支持向量机的多分类方法主要有: 成对分类法、一类对余类法、纠错输出编码方法及确定多类目标函数方法等。实际问题中常用的是成对分类法和一类对余类法<sup>[6]</sup>。本文主要采用成对分类法。

假设训练样本集为  $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ , 其中  $l$  为训练样本集规模,  $x_i \in R^n$  为  $n$  维特征向量,  $y_i \in \{1, 2, \dots, k\}$  为第  $i$  类样本对应的等级。

成对分类法就是用 2 种不同类别的样本来构造一个子分类器  $f_{ij} (i \neq j)$ , 这样穷尽所有的两两组合, 就可以得到  $C_k^2 = k(k-1)/2$  个子分类器。如对第  $i$  类和第  $j$  类我们解如下的 2 分类问题:

$$\begin{cases} \min_{w^i, b^i, j, \xi_i} \frac{1}{2} \|w^{i,j}\|^2 + C \left[ \sum_t \xi_t^{i,j} \right] \\ \text{s. t. } \left[ (w^{i,j} \cdot x_i) + b^i \right] - 1 + \xi_i^j \geq 0 \quad (y_i = i) \\ \left[ (w^{i,j} \cdot x_i) + b^i \right] - 1 + \xi_i^j \leq 0 \quad (y_i = j, \xi_i^j \geq 0) \end{cases} \quad (6)$$

对未知类别属性的样本  $x$  进行测试时, 分别使用  $k(k-1)/2$  个子分类器对其进行属性判别, 并采取如下投票策略, 如果分类器  $f_i$  判定  $x$  属于第  $i$  类, 则第  $i$  类的得票数加 1, 否则第  $j$  类的得票数加 1, 最终看哪一类的得票数最多,  $x$  就属于哪一类。

## 2 多分类支持向量机在泥石流危险性区划中的应用

凉山州安宁河流域泥石流历史悠久, 活动频繁, 流域内的西昌市、冕宁县、喜德县、德昌县和会理县均分布有大量的泥石流, 而且几乎每年均有泥石流造成不同程度的人员伤亡和财产损失<sup>[14]</sup>。

本文以凉山州安宁河流域 129 个乡镇的泥石流危险性区划基础资料为依据<sup>[14]</sup>, 随机选取一定数量的样本作为训练样本, 通过机器学习, 建立泥石流危险性区划的多分类 SVM 模型, 进行以乡镇为单元的泥石流危险性区划研究。

### 2.1 评价指标的确定

影响泥石流危险性的因素有很多, 依据刘希林<sup>[15]</sup>提出的最新区域泥石流危险性评价成果, 选定 8 个评价因子: 泥石流沟分布密度  $y$ , 岩石风化程度系数  $x_1$ , 断裂带密度  $x_3$ , 大于等于  $25^\circ$  坡地面积百分比  $x_6$ , 洪灾发生频率  $x_8$ , 月降雨量变差系数  $x_9$ , 年平均大于等于 25 mm 大雨日数  $x_{11}$  以及大于等于  $25^\circ$  坡耕地面积百分比  $x_{16}$  为区域泥石流危险性评价的主要评价指标。

为便于比较, 各评价因子的归一化方法也采取刘希林<sup>[15]</sup>提出的分段函数赋值方法进行。凉山州安宁河流域 129 个乡镇 8 个泥石流评价指标归一化后的取值见表 1。

安宁河流域区域泥石流危险性等级可分为 4 级:

低度危险区、中度危险区、高度危险区和极高危险区, 分别用 1, 2, 3, 4 表示(表 1)。

### 2.2 SVM 模型的建立

SVM 模型的建立主要是选择相应的支持向量机参数: 核函数和惩罚系数  $C$ 。就 4 种常用的核函数而言, RBF 径向基函数由于具有较强的非线性映射能力, 得到了最广泛的应用。因此, 本文选择 RBF 函数作为 SVM 的核函数。RBF 核函数有一个参数  $\gamma$ ,  $\gamma$  的大小影响着 RBF 函数的形状,  $\gamma$  越小, 函数的形状越平滑, 得到的分类面也越平坦。惩罚因子  $C$  的大小, 表示对误判样本的惩罚程度的大小,  $C$  越大, 得到的训练样本的分类精度相对来说会有提高, 但这可能会造成所谓的“过度训练”问题。因此, 参数  $C$  和  $\gamma$  选择的合理与否就直接影响 SVM 模型的精度和推广能力。

本文对  $C$  和  $\gamma$  的选择主要采取网格搜索方法, 通过对训练样本进行分组交叉验证寻找给定范围内的 SVM 最优参数。利用网格搜索技术, 将训练数据分成 5 份, 轮流其中 4 份作为训练数据, 1 份作为测试数据, 进行试验, 获得了不同数量训练样本下多分类 SVM 模型的最优参数(表 2)。

### 2.3 评价结果分析

为了更好地说明支持向量机的优良性能, 本文从安宁河流域的 129 个乡镇中随机选取了 2 组训练样本(86 个训练样本和 65 个训练样本), 分别建立了安宁河流域泥石流区域危险性评价的多分类支持向量机模型, 评价结果见表 1(以随机选择的 65 个训练样本为例)和表 2。

此外, 为了对比评价效果, 利用随机选取的 2 组训练样本, 分别建立了凉山州安宁河流域泥石流危险性区划的 BP 神经网络模型, 评价结果见表 1(以随机选择的 65 个训练样本为例)和表 2。

表 1 研究区 129 个乡镇泥石流危险性区划指标归一化后取值及各模型评价结果

县市名	编号	乡镇名	泥石流沟分布密度 $y$	岩石风化程度系数 $x_1$	断裂带密度 $x_3$	$\geq 25^\circ$ 坡地面积百分比 $x_6$	洪灾发生频率 $x_8$	月降雨量变差系数 $x_9$	年平均 $\geq 25$ mm 大雨日数 $x_{11}$	$\geq 25^\circ$ 坡耕地面积百分比 $x_{16}$	危险性等级	SVM 评价结果	BP 评价结果
	1	回龙乡	1.00	0.61	0.00	0.94	0.30	1.00	1.00	0.31	3	3	3
	2	惠安乡	1.00	0.60	0.00	0.94	0.32	1.00	1.00	0.31	3	3	3
	3	森荣乡	0.36	0.67	0.00	0.94	0.27	1.00	1.00	0.31	2	2	2
冕宁县	4	柘槽乡	1.00	0.62	0.67	0.94	0.50	1.00	1.00	0.31	3	3	3
	5	复兴镇	0.87	0.60	0.00	0.94	0.67	1.00	1.00	0.31	3	3	3
	6	富强乡	1.00	0.58	1.00	0.94	0.56	1.00	1.00	0.31	4	4	4
	7	泸沽乡	1.00	0.49	0.28	0.94	0.60	1.00	1.00	0.31	3	3	3
	8	洛瓦乡	1.00	0.57	1.00	0.94	0.60	1.00	1.00	0.31	4	4	4
	9	河边乡	1.00	0.67	0.00	0.94	0.40	1.00	1.00	0.31	3	3	3

续表 1

冕宁县	10	铁厂乡	1.00	0.58	0.32	0.94	0.60	1.00	1.00	0.31	3	3	3
	11	沙坝乡	1.00	0.50	0.00	0.94	0.60	1.00	1.00	0.31	3	3	3
	12	东临乡	1.00	0.53	0.00	0.94	0.60	1.00	1.00	0.31	3	3	3
	13	成功乡	0.48	0.69	0.79	0.94	0.60	1.00	1.00	0.31	3	3	3
	14	大桥乡	1.00	0.51	0.15	0.94	0.42	1.00	1.00	0.31	3	3	3
	15	宁源乡	0.44	0.64	0.42	0.94	0.30	1.00	1.00	0.31	2	2	2
	16	曹古乡	0.00	0.60	0.73	0.94	0.36	1.00	1.00	0.31	2	2	2
	17	南山营乡	0.85	0.67	0.00	0.94	0.40	1.00	1.00	0.31	3	3	3
	18	漫水湾乡	1.00	0.49	0.00	0.94	0.83	1.00	1.00	0.31	3	3	3
	19	彝海子乡	1.00	0.58	0.00	0.94	0.44	1.00	1.00	0.31	3	3	3
20	环城乡	1.00	0.56	0.43	0.94	0.48	1.00	1.00	0.31	3	3	3	
喜德县	21	冕山乡	1.00	0.53	0.00	0.71	1.00	1.00	1.00	1.00	4	4	4
	22	拉克乡	1.00	0.50	1.00	0.71	1.00	1.00	1.00	1.00	4	4	4
	23	贺波洛乡	1.00	0.49	0.00	0.71	0.97	1.00	1.00	1.00	4	4	4
	24	则约乡	0.00	0.49	0.34	0.71	0.71	1.00	1.00	1.00	2	2	2
	25	博洛拉达	0.00	0.49	0.00	0.71	0.71	1.00	1.00	1.00	2	2	2
	26	两河口乡	1.00	0.49	0.00	0.71	0.83	1.00	1.00	1.00	3	4*	3
	27	巴久乡	0.00	0.49	0.00	0.71	0.71	1.00	1.00	1.00	2	2	2
	28	红莫乡	1.00	0.49	0.08	0.71	0.91	1.00	0.90	1.00	4	4	4
	29	鲁基乡	0.34	0.56	0.00	0.71	0.91	1.00	0.90	1.00	2	2	2
	30	桃源乡	1.00	0.49	0.83	0.71	0.91	1.00	0.90	1.00	4	4	4
	31	热河依达乡	1.00	0.49	0.00	0.71	0.91	1.00	1.00	1.00	4	4	4
	32	李子乡	1.00	0.49	1.00	0.71	0.67	1.00	1.00	1.00	4	4	4
	33	东河乡	1.00	0.49	0.00	0.71	0.86	1.00	1.00	1.00	4	4	4
	34	阿洛乡	0.48	0.49	0.38	0.71	0.91	1.00	1.00	1.00	3	3	3
西昌市	35	民胜乡	0.00	0.60	0.84	0.86	0.85	1.00	1.00	0.60	2	2	2
	36	琅环乡	1.00	0.55	0.00	0.86	1.00	1.00	1.00	0.60	4	4	4
	37	太和乡	0.00	0.62	0.00	0.86	0.85	1.00	1.00	0.60	2	2	2
	38	开元乡	0.00	0.58	1.00	0.86	0.53	1.00	1.00	0.60	2	2	2
	39	九龙乡	0.00	0.55	0.85	0.86	0.85	1.00	1.00	0.60	2	2	2
	40	小庙乡	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	41	川兴乡	0.85	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	42	佑君镇	1.00	0.57	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	43	黄连关乡	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	44	经久乡	0.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	2	2	2
	45	西溪乡	0.60	0.49	0.15	0.86	0.74	1.00	1.00	0.60	3	3	3
	46	大箐乡	0.00	0.49	0.00	0.86	0.64	1.00	1.00	0.60	2	2	2
	47	凉山州螺髻山种畜场	0.00	0.49	0.17	0.86	0.53	1.00	1.00	0.60	2	2	2
	48	荞地乡	1.00	0.53	0.98	0.86	0.74	1.00	1.00	0.60	4	4	4
	49	阿七乡	1.00	0.67	0.00	0.86	0.85	1.00	1.00	0.60	4	4	4
	50	黄水乡	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
德昌县	51	阿月乡	1.00	0.50	0.00	0.91	0.67	1.00	1.00	0.49	3	3	3
	52	银鹿乡	0.00	0.67	0.71	0.91	0.56	1.00	1.00	0.49	2	2	2
	53	王所乡	1.00	0.57	0.39	0.91	0.78	1.00	1.00	0.49	4	4	4
	54	六所乡	1.00	0.60	0.00	0.91	0.38	1.00	1.00	0.49	3	3	3
	55	巴洞乡	1.00	0.60	0.89	0.91	0.33	1.00	1.00	0.49	3	3	3
	56	丹桂乡	0.33	0.69	0.29	0.91	0.42	1.00	1.00	0.49	2	2	2
	57	永郎乡	1.00	0.57	0.61	0.91	0.77	1.00	1.00	0.49	4	4	4
	58	宽裕乡	1.00	0.54	0.76	0.91	0.52	1.00	1.00	0.49	4	4	4
	59	南山乡	1.00	0.69	0.00	0.91	0.38	1.00	1.00	0.49	3	3	3

续表 1

会理县	60	苍田乡	1.00	0.57	0.11	0.38	0.71	1.00	1.00	0.36	3	3	3
	61	下村乡	1.00	0.56	0.71	0.38	0.44	1.00	1.00	0.36	3	3	3
	62	六华乡	0.00	0.49	0.00	0.38	0.44	1.00	1.00	0.36	1	1	3*
	63	三地乡	0.00	0.49	0.00	0.38	0.38	1.00	1.00	0.36	1	1	3*
	64	龙泉乡	0.00	0.69	0.10	0.38	0.48	1.00	1.00	0.36	2	2	2
	65	益门乡	1.00	0.51	1.00	0.38	0.44	1.00	1.00	0.36	3	3	3
冕宁县	66	回坪乡	1.00	0.58	0.00	0.94	0.30	1.00	1.00	0.31	3	3	3
	67	哈哈乡	1.00	0.68	0.00	0.94	0.27	1.00	1.00	0.31	3	3	3
	68	宏模乡	1.00	0.55	0.00	0.94	0.56	1.00	1.00	0.31	3	3	3
	69	林里乡	1.00	0.56	0.65	0.94	0.67	1.00	1.00	0.31	4	4	4
	70	石龙乡	1.00	0.49	0.92	0.94	0.56	1.00	1.00	0.31	4	4	3*
	71	后山乡	1.00	0.63	0.69	0.94	0.56	1.00	1.00	0.31	4	3*	4
	72	白蛉乡	0.00	0.68	0.00	0.94	0.56	1.00	1.00	0.31	2	2	2
	73	先锋乡	1.00	0.57	0.00	0.94	0.40	1.00	1.00	0.31	3	3	3
	74	河里乡	0.00	0.68	0.11	0.94	0.40	1.00	1.00	0.31	2	2	2
	75	大坪乡	1.00	0.60	0.77	0.94	0.33	1.00	1.00	0.31	3	3	3
	76	泽远乡	0.00	0.66	0.00	0.94	0.60	1.00	1.00	0.31	2	2	2
	77	拖乌乡	0.63	0.62	0.00	0.94	0.36	1.00	1.00	0.31	2	2	3*
	78	冶勒乡	0.00	0.66	0.78	0.94	0.36	1.00	1.00	0.31	2	2	2
	79	樟木沟乡	0.00	0.69	0.00	0.94	0.27	1.00	1.00	0.31	1	2*	2*
喜德县	80	深沟乡	1.00	0.53	0.00	0.71	0.55	1.00	1.00	1.00	3	3	3
	81	则莫乡	1.00	0.49	0.00	0.71	0.75	1.00	1.00	1.00	3	3	3
	82	依兹觉乡	1.00	0.49	0.00	0.71	0.85	1.00	1.00	1.00	3	4*	4*
	83	喜德	1.00	0.49	0.00	0.71	0.95	1.00	1.00	1.00	4	4	4
	84	且拖乡	1.00	0.49	0.00	0.71	0.83	1.00	1.00	1.00	3	4*	3
	85	沙马拉达乡	1.00	0.49	0.52	0.71	0.83	1.00	1.00	1.00	4	4	4
	86	波振乡	1.00	0.49	0.71	0.71	0.71	1.00	1.00	1.00	4	4	4
	87	依洛乡	0.60	0.49	0.00	0.71	0.71	1.00	1.00	1.00	3	3	3
	88	米市乡	0.54	0.49	0.47	0.71	1.00	1.00	1.00	1.00	3	3	3
	89	瓦哈乡	0.56	0.49	0.65	0.71	0.91	1.00	1.00	1.00	3	3	3
	90	洛莫乡	0.00	0.49	0.00	0.71	0.91	1.00	1.00	1.00	2	2	2
	91	西河乡	1.00	0.49	0.63	0.71	0.86	1.00	1.00	1.00	4	4	4
	92	北山乡	1.00	0.49	0.14	0.71	0.86	1.00	1.00	1.00	4	4	4
	93	额泥乡	1.00	0.49	0.00	0.71	0.91	1.00	1.00	1.00	4	4	4
	94	洛哈乡	1.00	0.49	0.00	0.71	0.91	1.00	1.00	1.00	4	4	4
西昌市	95	响水乡	0.43	0.59	1.00	0.86	0.64	1.00	1.00	0.60	3	3	3
	96	琅环乡	0.00	0.55	0.00	0.86	0.95	1.00	1.00	0.60	2	2	2
	97	月华乡	1.00	0.49	0.00	0.86	1.00	1.00	1.00	0.60	4	4	3*
	98	礼州乡	1.00	0.49	0.00	0.86	0.95	1.00	1.00	0.60	4	4	3*
	99	兴胜乡	0.00	0.49	0.00	0.86	0.95	1.00	1.00	0.60	2	2	2
	100	锅盖梁镇	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	101	西乡	1.00	0.50	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	102	西郊乡	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	103	裕隆乡	0.00	0.53	0.00	0.86	0.85	1.00	1.00	0.60	2	2	2
	104	马道乡	0.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	2	2	2
	105	四合乡	0.63	0.49	0.85	0.86	0.53	1.00	1.00	0.60	3	3	4*
	106	高枳乡	1.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	107	海南乡	0.00	0.49	0.00	0.86	0.85	1.00	1.00	0.60	2	2	2
	108	邛海	0.00	0.49	0.00	0.86	0.64	1.00	1.00	0.60	2	2	2
	109	大兴乡	1.00	0.49	0.00	0.86	0.64	1.00	1.00	0.60	3	3	3
	110	磨盘乡	0.59	0.60	1.00	0.86	0.85	1.00	1.00	0.60	3	4*	3
	111	中坝乡	1.00	0.58	0.00	0.86	0.85	1.00	1.00	0.60	3	3	3
	112	羊角坝乡	1.00	0.61	0.00	0.86	0.85	1.00	1.00	0.60	4	3*	4
	113	洛古坡乡	0.00	0.49	0.63	0.86	0.64	1.00	1.00	0.60	2	2	2

续表 1

	114	前山乡	0.45	0.66	0.16	0.91	0.56	1.00	1.00	0.49	2	2	2
	115	麻栗乡	1.00	0.54	0.00	0.91	0.67	1.00	1.00	0.49	3	3	3
	116	方家乡	1.00	0.53	0.61	0.91	0.78	1.00	1.00	0.49	4	4	4
	117	五一乡	1.00	0.57	0.36	0.91	0.78	1.00	1.00	0.49	4	4	4
	118	小高乡	1.00	0.53	0.09	0.91	0.47	1.00	1.00	0.49	3	3	3
德	119	茨达乡	1.00	0.68	0.00	0.91	0.52	1.00	1.00	0.49	3	3	3
昌	120	老碾乡	1.00	0.67	0.00	0.91	0.60	1.00	1.00	0.49	3	3	3
县	121	铜厂乡	0.00	0.68	1.00	0.91	0.34	1.00	1.00	0.49	2	2	2
	122	格坪乡	0.00	0.67	0.00	0.91	0.43	1.00	1.00	0.49	2	2	2
	123	金沙乡	1.00	0.66	0.00	0.91	0.43	1.00	1.00	0.49	3	3	3
	124	凉峰乡	0.38	0.66	0.52	0.91	0.52	1.00	1.00	0.49	2	2	2
	125	乐跃乡	1.00	0.54	0.00	0.91	0.51	1.00	1.00	0.49	3	3	3
	126	锦川乡	1.00	0.56	0.00	0.91	0.56	1.00	1.00	0.49	3	3	3
会	127	云甸乡	1.00	0.52	0.08	0.38	0.63	1.00	1.00	0.36	3	3	3
理	128	白果湾乡	1.00	0.53	0.64	0.38	0.40	1.00	1.00	0.36	3	3	3
县	129	六民乡	0.00	0.49	0.00	0.38	0.38	1.00	1.00	0.36	1	1	3*

注: 1—65 为训练样本, 66—129 为测试样本, 带“\*”为错判样本。

表 2 SVM 模型和 BP 模型的预测精度比较

项目	SVM 模型		BP 模型	
	86 个训练样本/ 43 个测试样本	65 个训练样本/ 64 个测试样本	86 个训练样本/ 43 个测试样本	65 个训练样本/ 64 个测试样本
测试样本预测准确率	93.02%	90.63%	88.37%	87.50%
训练样本回判准确率	98.84%	98.46%	98.84%	96.92%
模型参数	C=512, $\gamma=0.03125$ C=2048, $\gamma=0.03125$		网络结构: 8-3-4	网络结构: 8-5-4

由表 1, 表 2 可见, SVM 模型与 BP 模型的预测精度均随训练样本数量的增加而有不同程度的提高; 利用不同数量训练样本建立的 SVM 模型对测试样本的预测准确率均高于相应的 BP 神经网络模型, 对训练样本的回判准确率高于或接近于 BP 神经网络模型。

### 3 结论

本文以安宁河流域 129 个乡镇的泥石流危险性区划基础资料为依据, 随机选择总样本数的 2/3 和 1/2 作为训练样本, 剩余的作为测试样本 (86 个训练样本/43 个测试样本, 65 个训练样本/64 个测试样本), 进行机器学习, 建立不同训练样本下安宁河流域泥石流危险性区划的多分类 SVM 模型, 进行以乡镇为单元的区域泥石流危险性评价。评价结果表明, SVM 模型的预测精度随典型训练样本数量的增加而有不同程度提高; 利用不同数量训练样本建立的 SVM 模型对测试样本的预测准确率均高于相应的 BP 神经网络模型, 对训练样本的回判准确率高于或接近于 BP 神经网络模型。因此, 在安宁河流域泥石

流区域危险性评价中, 支持向量机方法比神经网络具有更优的预测精度和更强的泛化性能, 可以在今后的泥石流危险性评价实践中加以推广应用。

#### [ 参 考 文 献 ]

[1] 刘希林, 唐川. 泥石流危险性评价[M]. 北京: 科学出版社, 2004: 1-19.

[2] 边肇祺, 张学共. 模式识别[M]. 北京: 清华大学出版社, 2000: 284-303.

[3] 邓乃杨, 田英杰. 数学挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004: 49-223.

[4] Nello Cristianini, John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods[M]. Beijing: China Machine Press, 2005: 93-122.

[5] 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.

[6] 张红涛, 胡玉霞, 毛罕平. 基于 SVM 的储粮害虫图像识别分类[J]. 农机化研究, 2008(8): 36-38.

(下转第 157 页)

流作用共同影响, 雨量的南北坡的季节变化与大气环流的季节变化一致。

(3) 本文虽利用相对高海拔的气候考察站的数据, 弥补了常规气象台站海拔偏低的不足, 但对整个研究区而言样本数仍偏少, 且分布不均匀, 如果加入质量控制后的高密度的自动雨量站数据建模, 山区降水模拟精度将会进一步提高。近年来安徽省气象部门在山区布设了大量的自动雨量站, 弥补了固定台站的不足, 但目前已有自动雨量站资料时间尺度尚不够, 这项工作有待后续研究。

致谢 安徽省气象台首席预报员郑缓缓正研级高工对本文提出的有益建议和帮助, 在此深表敬意和感谢!

#### [ 参 考 文 献 ]

- [ 1 ] 万晔, 司徒群, 李子海. 滇西北地理环境特征与区域水土流失相互作用机制研究[ J ]. 水土保持通报, 2003, 23(5): 1-5.
- [ 2 ] Ninyerola M, Pons X, Roure J M. A methodological approach of climatological modeling of air temperature and precipitation through GIS techniques[ J ]. Int. J. Climatol., 2000, 20(14): 1823-1841.
- [ 3 ] Wothing G, Bouvier C H, Danloux J, Fritsch J M. Regionalization of extreme precipitation distribution using the principal components of the topographical environment[ J ]. J. Hydrol., 2000, 233: 86-101.
- [ 4 ] Weisse A K, Bois P. Topographic effects on statistical characteristics of heavy rainfall and mapping in the French Alps [ J ]. J. Appl. Meteorol., 2001, 40(4): 720-740.
- [ 5 ] Sevruck B, Nevenic M. The geography and topography effects on the areal pattern of precipitation in a small prealpine basin[ J ]. Wat. Sci. Tech., 1998, 37: 163-170.
- [ 6 ] Marquinez J, Lastra J, Garcia P. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis[ J ]. J. Hydrol., 2003, 270: 1-11.
- [ 7 ] 孙鹏森, 刘世荣, 李崇巍. 基于地形和主风向效应模拟山区降水空间分布[ J ]. 生态学报, 2004, 24(9): 1910-1915.
- [ 8 ] 周长艳, 李跃清, 彭俊. 九寨沟、黄龙风景区的降水特征及其变化[ J ]. 资源科学, 2006, 28(1): 113-119.
- [ 9 ] 蔡福, 于慧波, 矫玲玲, 等. 降水要素空间插值精度的比较: 以东北地区为例[ J ]. 资源科学, 2006, 28(6): 73-79.
- [ 10 ] 张红平, 周锁铨. 山地降水的空间分布特征研究综述[ J ]. 陕西气象, 2004(6): 27-30.
- [ 11 ] 周锁铨, 薛根元, 周丽峰, 等. 基于 GIS 降水空间分析的逐步插值方法[ J ]. 气象学报, 2006, 64(1): 100-111.
- [ 12 ] 王菱. 华北山区坡地方位和海拔高度对降水的影响[ J ]. 地理科学, 1996, 16(2): 150-158.
- [ 13 ] 傅抱璞. 地形和海拔高度对降水的影响[ J ]. 地理学报, 1992, 47(4): 302-314.
- [ 14 ] 马蔼乃. 地理科学与地理信息科学论[ M ]. 武汉: 武汉出版社, 2000.
- [ 15 ] 亚热带东部丘陵山区农业气候资源及其合理利用研究课题协作组. 中国亚热带东部丘陵山区农业气候资源研究[ M ]. 北京: 科学出版社, 1989.
- [ 16 ] 张宏群, 安裕伦. 贵州省喀斯特山区水土流失遥感信息模型的建立[ J ]. 水土保持通报, 2003, 23(5): 39-42.
- [ 17 ] 周锁铨, 孙琪, 肖桐松, 等. 长江中上游区基于 GIS 的不同时间尺度降水插值方法探讨[ J ]. 高原气象, 2008, 27(5): 1021-1034.
- [ 18 ] 冯强, 叶汝杰, 王昂生, 等. 中尺度地形对暴雨降水影响的数值模拟研究[ J ]. 中国农业气象, 2004, 25(4): 1-4.
- (上接第 133 页)
- [ 7 ] Yuan Li-feng, Zhang You-shu. Debris flow hazard assessment based on support vector machine[ J ]. Wuhan University Journal of Natural Sciences, 2006, 11(4): 897-900.
- [ 8 ] 王涛, 刘兴年, 黄尔. 基于 SVM 方法的小流域泥石流输沙量预测[ J ]. 水利水电科技进展, 2008, 28(2): 1-3, 64.
- [ 9 ] Pijush Samui. Slope stability analysis: a support vector machine approach [ J ]. Environmental Geology, 2008(56): 255-267.
- [ 10 ] 马文涛. 基于 PSO 和 LSSVM 的边坡稳定性评价方法[ J ]. 岩土力学, 2009, 30(3): 845-848.
- [ 11 ] 罗战友, 杨晓军, 龚晓南. 基于支持向量机的边坡稳定性预测模型[ J ]. 岩石力学与工程学报, 2005, 24(1): 144-148.
- [ 12 ] 马文涛. 土坡稳定性分析的 SVM 方法[ J ]. 烟台大学学报: 自然科学与工程版, 2006, 19(1): 47-51.
- [ 13 ] 马志江, 陈汉林, 杨树锋. 基于支持向量机理论的滑坡灾害预测: 以浙江庆元地区为例[ J ]. 浙江大学学报: 理学版, 2003, 30(5): 592-596.
- [ 14 ] 李秀珍, 刘希林, 苏鹏程. 四川凉山州安宁河流域泥石流危险性评价[ J ]. 防灾减灾工程学报, 2005, 25(4): 426-430, 457.
- [ 15 ] 刘希林, 莫多闻. 泥石流风险评价[ M ]. 成都: 四川科学技术出版社, 乌鲁木齐: 新疆科技卫生出版社, 2003.