

斜坡单元支持下的滑坡易发性评价支持向量机模型

武雪玲¹ 任 福² 牛瑞卿¹ 彭 令¹

(1 中国地质大学(武汉)地球物理与空间信息学院,武汉市鲁磨路 388 号,430074)

(2 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)

摘 要:针对传统滑坡预测手段数据源有限、数据更新周期长、难以发现隐藏在复杂滑坡系统中的规律等问题,本文以三峡库区为研究对象,从多源空间数据中提取滑坡孕灾环境和影响因素等信息,采用数字地形水文分析方法划分斜坡单元,对评价因子进行重采样,进而构建两类支持向量机模型。分析了多源影响因素与滑坡易发性的定量关系,并生成滑坡易发性分区图。采用成功率曲线和误差率评价预测结果,模型预测精度达到 98.21%,与野外调查实际情况吻合较好。

关键词:滑坡易发性;斜坡单元;支持向量机;三峡库区
中图分类号:P208

自 2008 年 10 月完成首次高水位蓄水以来,受库水位周期性高达 30 m 消落的影响,三峡库区进入了一个库岸再造、斜坡失稳等地质灾害相对多发的不稳定时期^[1]。调查显示,三峡库区已识别的较大规模的滑坡、崩塌体共 2 490 处,严重威胁三峡大坝的正常运行和库区居民的生命财产安全。对库区滑坡灾害大规模采用监测预警是不现实的,因此,探索研究准确高效的滑坡预测方法势在必行。

在遥感技术和地理信息系统的支持下,国内外学者对滑坡灾害定量预测已经进行了一系列相关研究^[2-7]。尽管这些方法提高了滑坡灾害空间预测的精度和效率,仍存在两个亟待解决的问题:① 滑坡评价单元的选择;② 准确高效预测模型的构建。

本文以我国重大地质灾害防治工作典型示范地区——三峡库区为研究对象,选择三峡库区泄滩乡长江干流岸坡为研究区,从多源空间数据中提取滑坡孕灾环境和影响因素等信息,采用地形分析模型划分斜坡评价单元,进而构建了两类 SVM 模型,定量预测滑坡易发性。

1 支持向量机

滑坡成因类型多样,影响因素及作用过程复杂,目前仍然无法精确确定影响滑坡灾害的全部影响因素及对应的影响程度。支持向量机(support vector machine, SVM)^[7-8]基于结构风险最小原则,利用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分类最优超平面,将两类数据正确分开。

假设存在一个线性可分的训练样本向量集合 $\{x_i\} (i=1, 2, \dots, N)$, $y_i=0$ 或 1 (0 表示非滑坡或者不易发, 1 表示滑坡或者易发), N 为样本数。两类 SVM 的目标是找到一个 N 维特征空间上的分类超平面,该平面能以最大间隔区分滑坡易发区和稳定区,最优代价函数的数学表达式为^[9]:

$$L = \|w\|^2/2 - \sum_{i=1}^N \lambda_i (y_i ((w \cdot x_i) + b) - 1) \quad (1)$$

其中, $\|w\|^2/2$ 的约束条件为:

$$y_i ((w \cdot x_i) + b) \geq 1 \quad (2)$$

式中, $\|w\|^2$ 是超平面法向量的欧几里德范数; b 是常数。

收稿日期:2013-10-26。

项目来源:国家自然科学基金资助项目(41271455/D0108);中央高校基本科研业务费专项资金资助项目(CUGL120207);中国博士后科学基金资助项目(2011M501261);民政部减灾和应急工程重点实验室开放基金资助项目(LDRE20120207)。

当训练样本向量集合线性不可分时,引入松弛因子 ξ_i ,式(2)调整为:

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i \quad (3)$$

2 研究区概况及数据源

研究区位于湖北省秭归县泄滩乡,地理坐标为东经 $110^{\circ}32' \sim 110^{\circ}39'$,北纬 $30^{\circ}59' \sim 31^{\circ}02'$,区域总面积 50 km^2 ,图1为研究区地理位置示意图。该区地处中国地形第二阶梯向第三阶梯过渡的地带,川东褶皱与鄂西山地会合部位,为中低山和侵蚀峡谷地貌,相对高程达 $80 \sim 1220 \text{ m}$ 。区内地层岩性多样、构造复杂,易滑地层巴东组广泛分布,且大部分属于第四系堆积层滑坡,不仅提供了滑坡灾害发育的物质条件,还控制空间孕灾环境的差异,受新构造运动的影响,山原期夷平面快速抬升,河流强烈下切,形成高陡岸坡,卸荷效应显著,滑坡灾害广泛发育且频繁发生,是我国地质灾害防治工作的重点。

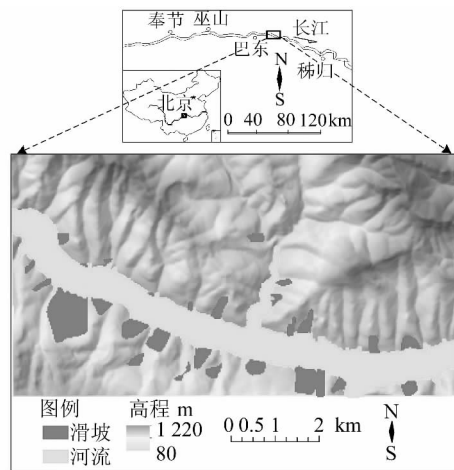


图1 研究区地理位置示意图

Fig.1 Location of the Study Area

本文采用的主要数据源包括:①我国环境减灾卫星影像 HJ-1A 1景(2011-04-13,轨道号为 7/76),用于提取土地利用、植被指数等地表覆盖信息;② $1:10\,000$ 数字化地形图,主要用于提取地形地貌和水文特征等信息;③ $1:50\,000$ 数字化地质图,用于提取断层、斜坡结构和工程岩组等地质信息;④其他历史滑坡存档资料、野外调查资料及部分航片等,用于历史滑坡解译和空间定位。研究区内共解译出滑坡 28 处,总面积为 3.163 km^2 ,其中,最小滑坡面积为 0.008 km^2 ,最大滑坡面积为 0.854 km^2 。

3 研究方法

3.1 评价因子

滑坡灾害的发生受时空多源要素影响,根据研究区的具体特点与已有研究成果,提取坡度、坡向、曲率、坡高、地层、斜坡结构、断层、库水影响和植被指数(normalized difference vegetation index, NDVI)作为预测变量。结合本文研究区地形地质资料的比例尺和卫星遥感影像的空间分辨率,对所有评价因子重采样为 $28.5 \text{ m} \times 28.5 \text{ m}$ 。采用自然断点方法,将所有连续变量离散化为分类变量,统计分析研究区内历史滑坡与各评价因子之间的相关性,各评价因子的频率分析见表1,比率越大,滑坡易发性高。坡度和坡向与滑坡发生直接相关,40.74%的滑坡分布在 $15^{\circ} \sim 30^{\circ}$ 的缓倾坡,对应的比率为 1.7,该坡度范围的岸坡前缘地形较缓而后缘地形较陡,如果坡体两侧分布有冲沟或走向与坡向相近的断层条件,则容易形成稳定性较低的六面滑坡体。50%的滑坡发生在 NE 和 S 方向,对应的比率均大于 1。曲率因子描述地形的复杂性,曲率分类 $0 \sim 0.19$ 、 $0.19 \sim 0.30$ 和 $0.30 \sim 0.40$ 对应的滑坡易发性较高。坡高是斜坡单元高程的平均值,55.56%的滑坡分布在 $225.30 \sim 354.28 \text{ m}$ 。地层、断层和斜坡结构是重要的滑坡孕灾地质因子。侏罗系和三叠系中统巴东组地层滑坡易发,其中,三叠系巴东组地层是调查中库区滑坡发育最多的地层,岩性为灰岩、白云岩、泥灰岩、砂页岩互层等,48.15%的滑坡发生在巴东组易滑地层,33.33%的滑坡分布于距断层 150 m 的范围内。斜坡结构是根据岸坡坡度、坡向与下伏地层倾角的关系划分的,顺向坡对应的比率最大,滑坡易发性高。85%的滑坡分布在库水位波动区和强影响区。NDVI 能够定量反映植被覆盖密度,其值越大,表示植被覆盖密度越高,55.56%的滑坡分布在 NDVI $0.2 \sim 0.3$ 区域。

3.2 斜坡单元

模型单元的选取是滑坡灾害空间预测的前提,其影响预测变量的处理和预测模型的构建,是滑坡易发性结果评价的基本绘图单元。本文选取斜坡单元作为基本评价单元,采用 DEM 水文分析模型划分斜坡单元,关键步骤包括正反地形 DEM 无洼地填充、水流方向提取、汇流累积量计算、河网提取、正反流域多边形合并等。为保证数据精度,生成的初始斜坡单元需经过人工修编才能满足滑坡易发性评价的实际需要。基于此,整

表 1 滑坡评价因子频率分析

Tab. 1 Frequency Analysis of the Conditioning Factors

评价因子	分类	斜坡单元	所占百分比(A_i)	历史滑坡	所占百分比(B_i)	比率 $R = B_i/A_i$
坡度/(°)	<15	16	5.73	0	0.00	0.00
	15~30	67	24.01	11	40.74	1.70
	30~47	82	29.39	7	25.93	0.88
	47~60	82	29.39	8	29.63	1.01
	>60	32	11.47	1	3.70	0.32
坡向	N	7	2.51	3	11.11	4.43
	NE	36	12.90	6	22.22	1.72
	E	37	13.26	3	11.11	0.84
	SE	38	13.62	2	7.41	0.54
	S	44	15.77	6	22.22	1.41
	SW	51	18.28	2	7.41	0.41
	W	45	16.13	3	11.11	0.69
	NW	21	7.53	2	7.41	0.98
	曲率	<0.19	22	7.89	6	22.22
0.19~0.30		59	21.15	8	29.63	1.40
0.30~0.40		106	37.99	12	44.44	1.17
0.40~0.52		69	24.73	1	3.70	0.15
>0.52		23	8.24	0	0.00	0.00
坡高/m	<225.30	58	20.79	5	18.52	0.89
	225.30~354.28	74	26.52	15	55.56	2.09
	354.28~472.07	77	27.60	5	18.52	0.67
	472.07~610.43	56	20.07	1	3.70	0.18
	>610.43	14	5.02	1	3.70	0.74
地层	三叠系巴东组二段	4	1.43	0	0.00	0.00
	三叠系巴东组三段	33	11.83	3	11.11	0.94
	三叠系巴东组四段	33	11.83	10	37.04	3.13
	三叠系九里岗组	7	2.51	3	11.11	4.43
	侏罗系千佛崖组	79	28.32	8	29.63	1.05
	侏罗系沙溪庙组上段	34	12.19	0	0.00	0.00
	侏罗系沙溪庙组下段	61	21.86	1	3.70	0.17
	侏罗系遂宁组	14	5.02	0	0.00	0.00
	侏罗系桐竹园组	14	5.02	2	7.41	1.48
断层/m	<150	50	17.92	9	33.33	1.86
	150~250	62	22.22	4	14.81	0.67
	250~350	49	17.56	6	22.22	1.27
	350~450	63	22.58	7	25.93	1.15
	>450	55	19.71	1	3.70	0.19
斜坡结构	飘倾坡	2	0.72	0	0.00	0.00
	伏倾坡	37	13.26	3	11.11	0.84
	顺向坡	42	15.05	6	22.22	1.48
	横向坡	118	42.29	11	40.74	0.96
	逆斜坡	15	5.38	2	7.41	1.38
	逆向坡	65	23.30	5	18.52	0.79
	波动区	2	0.72	1	3.70	5.17
库水	强影响区	76	27.24	22	81.48	2.99
	中影响区	35	12.54	2	7.41	2.61
	弱影响区	166	59.50	2	7.41	1.03
NDVI	<0.10	12	4.30	3	11.11	0.88
	0.10~0.20	25	8.96	9	33.33	1.37
	0.20~0.30	89	31.90	15	55.56	1.86
	0.30~0.40	100	35.84	0	0.00	0.67
	>0.40	53	19.00	0	0.00	1.27

个研究区被划分为 279 个斜坡单元,利用生成的斜坡单元对预测因子进行重采样,9 个评价因子

对应的条件属性和历史滑坡对应的决策属性构成初始决策表,该二维表包含 279 行、10 列,作为滑

坡空间预测的输入变量。

3.3 滑坡易发性预测

从初始决策表中随机选取 80% 的滑坡和非滑坡数据形成训练样本集,其余 20% 作为检验样本。选择径向基函数为核函数的两类 SVM 模型,采用交叉验证法确定最优参数组合,通过最优参数组合构建两类 SVM 模型,将整个样本数据集输入模型,计算滑坡易发性指数(landslide susceptibility index, LSI),取值范围 0~1,对应滑坡易发性从低到高。为了便于区分滑坡的易发程度,根据滑坡样本分布直方图的特点,采用自然断点法将连续的滑坡易发性指数划分为滑坡易发性分区:0~0.06、0.06~0.25、0.25~0.61、0.61~0.89 和 0.89~1,分别对应稳定区、低易发区、中易发区、高易发区和危险区,滑坡易发性分区结果见图 2。从预测结果的总体空间分布形态来看,滑坡灾害呈线状分布,主要分布于库水波动区和强影响区,即长江干流及支流沿岸,高易发区主要分布在靠近秭归段长江左岸,其余地方为零星点状分布。

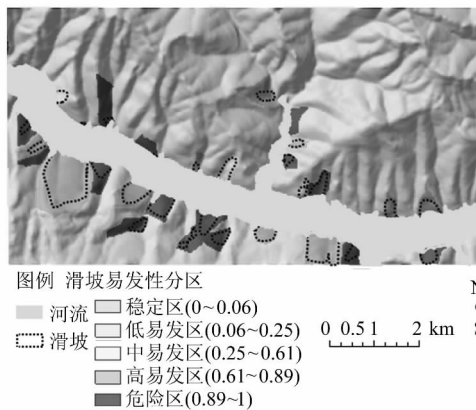


图 2 滑坡易发性分区图

Fig. 2 Landslide Susceptibility Zoning Map

4 结果分析

成功率曲线能定量衡量模型预测的准确度,取值范围为 $[0.5, 1]$,曲线越陡,则模型预测精度越高,曲线下面积为 1 表示一种理想状态,即模型预测结果与滑坡实际空间分布完全一致。通过比较滑坡易发性分区结果与已知滑坡的空间分布情况,分析模型预测精度,基于斜坡单元的两类 SVM 的成功率曲线见图 3,曲线下面积为 0.982 1,即预测准确率为 98.21%。

滑坡空间预测的分类误差是一个成本问题,即“危险区”与“稳定区”的错分不等价。因此,定

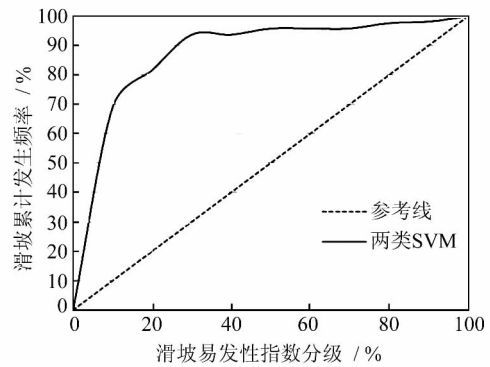


图 3 成功率曲线

Fig. 3 Success Rate Curve

义 A、B 错误, A 错误指将危险区划分为稳定区, B 错误指将稳定区划分为危险区, A 错误率 = A 错误数/滑坡总数, B 错误率 = B 错误数/非滑坡总数。基于格网单元和斜坡单元的两类 SVM 的预测结果见表 3,基于格网单元的整体精度(98.84%)略高于基于斜坡单元的精度(98.21%),但基于斜坡单元的 A 错误率(10.71%)明显低于基于网格单元的 A 错误率(11.84%)。因此,与基于网格单元的两类 SVM 相比,基于斜坡单元的模型具有 A 错误率低和计算效率高的特点。

表 2 预测结果比较

Tab. 2 Comparison of the Prediction Results

	预测值		整体精度/%
	0	1	
斜坡单元实际值	0	249	98.21
	1	3	
格网单元实际值	0	53 978	98.84
	1	448	

5 结 语

采用斜坡单元作为滑坡灾害易发性评价的基本单元,能够综合体现滑坡孕灾环境和诱发因子的作用,提高滑坡空间预测的精度和效率。SVM 基于最优理论、统计和函数分析方法实现最大化分割,在小样本学习、学习效率和预测能力等方面都优于当前其他的机器学习算法,适合分析复杂、非线性的滑坡系统。与基于格网的预测结果相比,基于斜坡单元的两类 SVM 模型具有整体精度稳定(98.21%)、A 类错误率低(10.71%)和预测效率高(数据量小)等优点,与野外调查数据相比,预测结果与实际情况较为一致。因此,将斜坡单元和 SVM 相应结合应用于区域滑坡易发性分

析,可以为三峡库区滑坡灾害预警与评估提供重要依据。

参 考 文 献

- [1] 陈德基, 满作武. 三峡工程几个重大地质问题的研究与论证[J]. 中国工程科学, 2011, 13(7):43-50
- [2] Bui D T, Lofman O, Revhaug I, et al. Landslide Susceptibility Analysis in the Hoa Binh Province of Vietnam Using Statistical Index and Logistic Regression[J]. Nat Hazards, 2011, 59:1 413-1 444
- [3] 王尚庆, 徐进军, 罗勉. 三峡库区白水河滑坡险情预警方法研究[J]. 武汉大学学报·信息科学版, 2009, 34(10):1 218-1 221
- [4] 牛瑞卿, 韩舸. 利用数据挖掘的滑坡监测数据处理流程[J]. 武汉大学学报·信息科学版, 2012, 37(7):869-872
- [5] Akgun A, Sezer E A, Nefeslioglu H A, et al. An Easy-to-Use MATLAB Program (MamLand) for the Assessment of Landslide Susceptibility Using a Mamdani Fuzzy Algorithm[J]. Comput Geosci, 2012, 38(1):23-34
- [6] Liu J P, Zeng Z P, Liu H Q, et al. A Rough Set Approach to Analyze Factors Affecting Landslide Incidence[J]. Comput Geosci, 2011, 37: 1 311-1 317
- [7] Scholkopf B, Smola A, Williamson R C, et al. New Support Vector Algorithms[J]. Neural Comput, 2000, 12:1 207-1 245
- [8] Vapnik V. Nature of Statistical Learning Theory [M]. New York:Wiley, 1995
- [9] Yao X, Tham L G, Dai F C. Landslide Susceptibility Mapping Based on Support Vector Machine: A Case Study on Natural Slopes of Hong Kong, China [J]. Geomorphology, 2008, 101: 572-582

第一作者简介:武雪玲,博士,副教授,现主要从事滑坡灾害预测预报方面的研究。

E-mail:snowforeting@163.com

Landslide Spatial Prediction Based on Slope Units and Support Vector Machines

WU Xueling¹ REN Fu² NIU Ruiqing¹ PENG Ling¹

(1 Institute of Geophysics and Geomatics, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

(2 School of Resources and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: Landslides are major natural geological disasters in China, and large-scale engineering activities induce and aggravate the occurrence of catastrophic landslides. Traditional spatial analytical techniques cannot easily discover patterns, trends, and relationships that can be hidden deep within complicated landslide hazard systems due to limited data source and long update cycle. Focusing on the Three Gorges, a variety of environment and triggering factors for landslide occurrence were calculated or extracted from the multi-source spatial data. Secondly, the study area was partitioned into slope units derived semi-automatically from a digital elevation model to resample the conditioning factors. Finally, a two-class SVM was trained and then used to map landslide susceptibility with the best accuracy of 98.21%. To evaluate the models, the susceptibility maps were validated by comparing them with the existing landslide locations according to success rate curve and error rates.

Key words: landslides susceptibility; slope unit; support vector machine; Three Gorges

About the first author: WU Xueling, Ph D, associate professor, research interest is landslide hazard prediction.

E-mail: snowforeting@163.com