

GIS支持下基于支持向量机的滑坡危险性评价

傅文杰

(莆田学院环境与生命科学系, 福建 莆田 351100)

摘要: 以仙游县为例, 探讨了将地理信息系统技术 (GIS) 和支持向量机 (SVM) 算法应用于滑坡灾害危险性评价的基本思路和技术路线。主要内容包括 SVM 的基本原理和方法、滑坡灾害危险性评价指标的选取和量化、SVM 模型的建立以及具体的实现过程。实践证明该方法是一种较好的滑坡灾害危险性评价方法。

关键词: 支持向量机; GIS 滑坡; 危险性评价

中图分类号: P642.2/P208 文献标识码: A 文章编号: 1000-0690(2008)06-0838-04

中国是世界上滑坡地质灾害最严重的国家之一, 据初步统计, 中国每年因崩塌、滑坡造成的经济损失均在 200 亿元人民币左右, 并且, 随着近年来人类活动的深度和广度的不断增加, 上述统计数字还在呈加速增长的趋势^[1]。

近几年来, 地理信息系统 (GIS) 在地质灾害研究中的应用引起了广泛的关注^[2-8]。

支持向量机 (support vector machine, 简称 SVM)^[9-10] 是近几年在统计学习理论的基础上发展起来的一种新的模式识别方法, 它在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势, 并能够推广应用到函数拟合等其他机器学习问题中。目前 SVM 已经成为国际上机器学习领域新的研究热点, 它已广泛地用于趋势分析和模式识别以及遥感图像的分类^[11, 12]。

只依靠 GIS 技术建立一个地区地质资料库, 不能进一步分析评价, 只能称作是空间数据库, 并不能达到地质灾害危险性区划目的^[13-15]。孤立的支持向量机方法由于数据获取的困难和偏差, 在实际应用中也会大受局限。如将 GIS 和 SVM 技术结合, 对未来可能发生的地质灾害的等级和范围进行预测, 可以早采取减灾防灾措施和对策, 使这类灾害造成的损失减少到最低程度。

本文以莆田市仙游县的滑坡地质灾害危险性评价为例, 介绍 GIS 结合 SVM 技术在滑坡地质灾害危险性评价中的应用。

1 SVM 基本原理

支持向量机的完整的数学描述见文献 [9, 10], 这里只对这个算法进行简单描述。支持向量的基本思想可用图 1 说明。图中, 实心点和空心点代表两类样本, H 为分类线, H_1 , H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线, 它们之间的距离叫做分类间隔。假设训练样本集为 $(x_i, y_i) i=1, \dots, n, x \in R^d$, 其中 x 为输入向量, i 为样本数, y 是输入向量所属的类别, 对于两类的分类问题, $y \in \{+1, -1\}$ 。D 维空间中线性判别函数的一般形式为 $g(x) = w \cdot x + b$ 分类面方程为:

$$w \cdot x + b = 0 \quad (1)$$

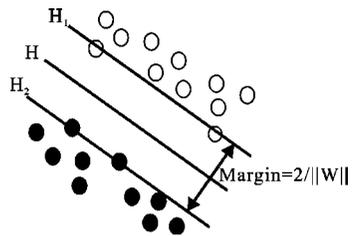


图 1 支持向量机的基本思想

Fig 1 Basis of support vectormachines

我们将判别函数进行归一化, 使两类样本都满足 $g(x) \geq 1$ 即使离分类面最近样本的 $g(x) = 1$ 这样分类间隔就等于 $2/\|w\|$, 因此, 使间隔最大等价于使 $\|w\|^2$ 最小; 而要求分类线对所有样本

收稿日期: 2008-01-02 修订日期: 2008-05-11

基金项目: 福建省科技厅青年人才项目“基于支持向量机的地质灾害危险性评价 GIS 系统” (2006F3111) 资助。

作者简介: 傅文杰 (1967-), 男, 福建莆田人, 博士, 高工, 研究方向为遥感技术及 GIS 应用。E-mail: fwj@163.com

正确分类, 要求它满足约束条件:

$$y_i [(w \cdot x_i) + b] - 1 \geq 0 \quad i = 1, \dots, n \quad (2)$$

考虑到有些训练样本是线性不可分的, Vapnik和 Cortes等人引入了非负的松弛变量, 将 (2) 式放宽为:

$$y_i [(w \cdot x_i) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad i = 1, \dots, n \quad (3)$$

显然, 当划分出现错误时 ξ_i 就会大于零。因此在求分类平面的同时, $\sum_i \xi_i$ 的值也希望愈小愈好。所以原本目标函数是求 $\|w\|^2/2$ 的最小值, 会变成求目标函数 $\|w\|^2/2 + C(\sum_i \xi_i)$ 的最小值; 其中 $C > 0$ 是一个常数, 是可调的参数, 控制对错样本惩罚的程度, C 越大表示对错误的惩罚越重。这是一个二次规划问题, 求解这个对偶问题得到最优分类决策函数:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i \cdot x) + b \right\} \quad (4)$$

其中, $K(\cdot, \cdot)$ 是核函数, $\text{sgn}(\cdot)$ 是符号函数, n 为训练样本的数目。

对于线性问题, 式 (4) 中 $K(\cdot, \cdot)$ 就是两向量的点积运算; 而对于非线性问题, SVM 的核心思想就是通过引入非线性映射 Φ , 将输入向量映射到高维空间来构造最优分类面。目前应用较多的核函数主要有 3 类: 多项式核函数, 径向基函数核函数, Sigmoid 函数。

2 应用实例

2.1 研究区地质环境概况

仙游县地处戴云山脉东坡, 地势总体自西北向东南倾斜, 地貌主体呈向东南开口的马蹄形, 海拔高程为 10~1 803.2 m, 坡度 25~40° 居多, 局部相对高程大, 地面坡降大, 天然陡坡地形较发育。主要地貌类型有中低山、丘陵和台地平原, 是一个地质条件复杂的地区。由于人多地少, 且山地面积占全县总面积的 80% 以上, 山体斜坡普遍发育, 山区村民建房、修路及水利水电工程建设等工程活动较频繁, 地表破坏较为严重, 为地质灾害的发育创造了有利条件。根据研究区特殊的地理地貌及地质环境条件特点, 本区的地质灾害类型主要有滑坡、崩塌、泥石流、地面塌陷等 4 种。

2.2 基础资料的收集与处理

影响滑坡活动的因素很多, 但可以将其大体分为两类: 一类是滑坡本身所具有的内部特征, 称为内因, 如地层岩性、坡体结构、有效临空面等都属于

内部条件; 另一类, 滑坡发育过程中起决定作用的外部因素, 称为外因, 如地下水、地表水、地震等都是外部条件。选取指标时要充分考虑这两类条件。结合研究区地质环境特征, 在借鉴大量已有研究成果和实地考察的基础上, 经过筛选, 建立研究区滑坡地质灾害危险性评价的指标体系。在指标选取过程中充分考虑资料的可获得性、研究尺度的大小、研究范围大小、研究精度的要求等因素。所选指标为: 高程指标、坡度指标、坡向指标、岩土体指标、降雨指标及植被覆盖指标等 6 个指标。对各评价指标分别规定量化方法。收集该区 1:10 000 地形图、地质图、年降雨量等值线图及土地利用现状等图件, 利用 MAPGIS 数字化成电子地图, 在获取大量野外现场调查资料的基础上, 分类录入相应图件的属性库。然后根据制定的指标体系, 从基础图件中提取相应的信息, 量化形成单因素分区图。这一步的工作仅仅依靠 MAPGIS 本身的分析功能尚不能完成, 必须进行二次开发, 扩展其功能, 才能生成上述各评价指标专题图 (图 2)。

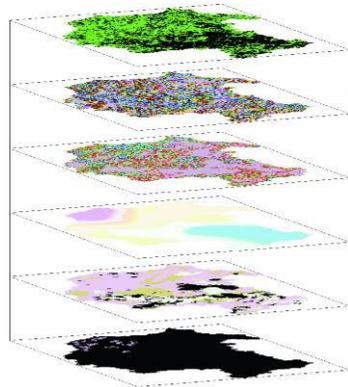


图 2 评价指标专题图叠加示意图

Fig 2 Overlap graph of evaluation index maps

2.3 SVM 运用于仙游县滑坡灾害危险性评价

将 SVM 方法应用于滑坡灾害危险性评价一般可以分为三步, 即: 样本库建立、评价单元的划分、SVM 模型选择及评价。这三个步骤说明如下:

1) 样本库的建立。在仙游县境内共发生滑坡 224 处, 占灾害总数的 73.68%。根据野外调查结果, 选取其中 32 处滑坡作为极高危险点, 95 处滑坡作为高危险点, 91 处作为中危险点, 另选 120 处作为低危险点, 总计 338 点组成总的样本点, 从以上 6 个专题图中采集样本点数据, 形成训练样本和

测试样本数据。

2) 评价单元的划分。滑坡灾害危险性评价以栅格数据的单个栅格单元大小为最小的计算单位。评价单元通常分为规则单元和不规则单元两种。本研究区采取规则单元 $100\text{ m} \times 100\text{ m}$ 为格网大小,利用开发的栅格划分模块将研究区分割成 182 311个格网。每个网格看作一个点,利用加权平均方法,分别从以上 6个评价因子专题图中采集数据。

3) SVM 模型选择及评价结果。选择合适的核函数参数和误差惩罚因子 C 对 SVM 的性能至关重要。统计学理论目前对这些问题给出了一些建议和解释,但还没有给出实际可行的方案,因此,在使用支持向量机进行分类和预测时,如何选择适当的参数就成了一个非常重要的问题。Vapnik 等人的研究表明, SVM 的性能与所选用的核函数的类型关系不大,而核函数的参数和误差惩罚因子 C 是影响 SVM 性能的主要因素^[16]。因此,作者在设计 SVM 分类器时选择 RBF 函数作为核函数,而对于 RBF 需要选择的参数只有核参数 γ 和惩罚因子 C 。本文采用交叉比对算法 (cross-validation) 来选取这两个参数为 $\gamma = 0.5$ $C = 32$ 。最后应用该模型对研究区滑坡灾害危险性进行评价,最终形成的滑坡灾害危险性评价结果图 (图 3)。

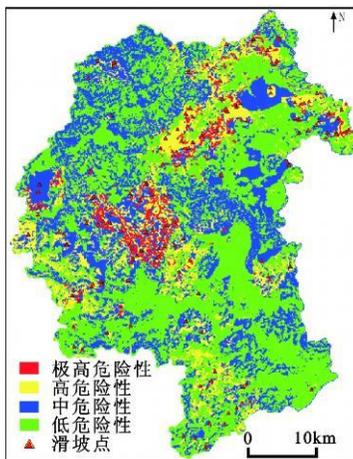


图 3 滑坡灾害危险性评价结果图

Fig 3 Result of landslide hazard evaluation

图 3 看出,由评价结果图统计得出极高危险性、高危险性、中危险性和低危险性 4 个级别所占比例分别为 3.31%、14.98%、32.21% 和 49.50%。极高危险区和高危险区基本分布在研究区中部及西北部,呈东北向带状展布,其两侧为中危险分布

区,低危险区则大致沿北西走向展布,该分区结果基本反映了仙游县境内地质灾害发育现状。通过实地验证和莆田市国土局提供的仙游县地质灾害易发区规划图对比,结果基本吻合。

1) 极高危险区和高危险区地质灾害评价。极高危险区和高危险区面内共发育滑坡灾害点 128 处,占全县滑坡灾害总数 57.1%。结合地质环境可以看出该区地质灾害具有点多、面广、密度高、危害性大的特点;其分布状况在宏观上大致受北东向构造控制;与该区所处的地形地貌条件复杂,构造发育,岩土体工程地质性能较差密切相关。一方面是高易发区内岩石风化强烈,形成的残坡积层结构松散,厚度大,具有上硬下软的双层结构特点;富含粘土矿物及泥质,吸水性强,干时易张裂,遇水易软化,属易滑地层。另一方面是该分区内人口较密集,人类工程活动强烈;各种各样的不合理工程活动形成的高陡临空面,为地质灾害的产生创造了有利条件,这样在暴雨或地震的诱发下就极易导致灾害的产生。

2) 中危险区地质灾害评价。中危险区内共发育滑坡灾害点 90 处,占全县滑坡灾害总数的 40.2%。该区地质灾害发育程度相对较弱,区内发育北东、北西及近东西向构造。分布的地层主要为晚侏罗系和早白垩系的火山岩组、第四系残坡积层及冲洪积层;而区内残坡积层发育地带岩土体结构条件较差,是地质灾害发育的主要地段。虽然区内人口密度相对较低,但是人类工程活动依然较频繁;其所发育的地质灾害虽然与地形地貌、地层岩性、地质构造和水文地质条件等内在因素密切相关,但是人为因素在这些灾害中大多占主导地位。因此,该分区的自然环境条件相对复杂以及人类不合理工程活动所造成的破坏为地质灾害的发育提供了有利环境。

3) 低危险区地质灾害评价。低危险区内仅发育地质灾害 6 处,占全县总灾害数 2.7%。此区相对危险性较小,但人口密集,人类工程活动剧烈,在暴雨或地震影响下局部地段诱发地质灾害。

3 结 论

本文将 GIS 技术与支持向量机有机结合起来进行滑坡地质灾害危险性评价,从理论方法和技术路线上实现了滑坡地质灾害危险性评价,具备了较强的评价能力和扩展能力。支持向量机在解决

小样本、非线性及高维模式识别问题中表现出的特有优势,使得支持向量机能够成为解决滑坡地质灾害危险性评价问题的有力工具。实验结果表明,支持向量机具有良好的模式识别和泛化能力,适用于对地质灾害危险性评价进行快速有效的评价。本文的研究成果,为研究区地质灾害危险性评价区划提供了一定的参考依据,也为其它地区地质灾害的调查与区划工作提供理论依据和工作方法。

应用 GIS 和 SVM 技术进行滑坡危险性评价研究还是一个新领域,有很大的发展空间,如核函数的类型和参数的选择;优化 SVM 训练模型,降低计算量,提高判别效率等。

参考文献:

- [1] 李天斌,陈明东. 滑坡实时跟踪预报[M]. 成都: 成都科技大学出版社, 1999
- [2] 闫满存,王光谦,刘家宏. GIS支持的澜沧江下游区泥石流灾害危险性评价[J]. 地理科学, 2001, 21(4): 334~339
- [3] 唐川,朱大奎. 基于GIS技术的泥石流风险评价研究[J]. 地理科学, 2002, 22(3): 300~304
- [4] 闫满存,王光谦,刘家宏. 基于GIS的澜沧江下游区滑坡灾害危险性分析[J]. 地理科学, 2007, 27(3): 365~369
- [5] 阮沈勇,黄润秋. 基于GIS的信息量法模型在地质灾害危险性区划中的应用. 成都理工学院学报, 2001, 28(1): 89~92
- [6] 王文俊,向喜琼,黄润秋,等. 区域崩塌滑坡的易发性评价[J]. 中国地质灾害与防治学报, 2003, 14(2): 31~38
- [7] 计会凤,蒋涛,张军. 基于GIS的滑坡地质灾害损失评估建库方法研究[J]. 遥感信息, 2005, 1: 50~55
- [8] 张像源,周萌. 基于专家评分模型和GIS的滑坡预警分析开发研究[J]. 中国地质灾害与防治学报, 2006, 17(2): 107~110
- [9] Vapnik V N. The Nature of Statistical Learning theory[M]. New York: Springer Verlag, 1995
- [10] Vapnik V N. Statistical Learning theory[M]. New York: Wiley, 1998
- [11] 马勇,丁晓青. 基于层次型支持向量机的人脸检测[J]. 清华大学学报(自然科学版), 2003, 43(1): 35~38
- [12] 李侃,高春晓,刘玉树. 基于SVM的空间数据库的层次聚类分析[J]. 北京理工大学学报, 2002, 22(4): 485~488
- [13] 向喜琼,黄润秋. 基于GIS的人工神经网络模型在地质灾害危险性区划中的应用[J]. 中国地质灾害与防治学报, 2000, 11(3): 23~27
- [14] 陈宁生,张飞. 2003年中国西南山区典型灾害性暴雨泥石流运动堆积特征[J]. 地理科学, 2006, 26(6): 701~705
- [15] 原立峰. 基于SVM的泥石流危险度评价研究[J]. 地理科学, 2007, 27(2): 296~300
- [16] Vapnik V N. (张学工译). 统计学习理论的本质[M]. 北京: 清华大学出版社, 2000

Landslide Hazard Evaluation Based on GIS and SVM

FU Wen-jie

(Putian University, Putian, Fujian 351100)

Abstract The development of Geographic Information System (GIS) technology provides a new technical method for the evaluation of landslide risk. Support Vector Machine is the hotspot of machine-learning industry research and has been applied successfully in many areas. By taking Xianyou County as an example, a new method for landslide hazard evaluation based on GIS and Support Vector Machines (SVM) is presented in this paper. It includes the basic principles and methods of SVM, selection and quantification of landslide hazard evaluation index, foundation of SVM model and the way to realize it. According to the actual situation of the research area, the quantification method has been stipulated separately for each of six selected appraisal indexes including the elevation, the gradient, the slope, the gneiss, the rainfall and the vegetation. The system of landslide geology disaster risk appraisal has been established, the special chart of each appraisal index has been obtained by the use of the geographic information system spatial analysis function. From the results of the appraisal, the extremely high-risk danger and the high-risk danger areas are basically located in the central and northwestern parts of the study area, the secondary risk in both sides of them. This distribution result has basically reflected the present situation of geological disaster in the research area. This method can be put in practice in geology hazard investigation.

Key words Support Vector Machine; Geographic Information System; landslide; hazard evaluation