

The Research of Natural and Human Factors Affect in Landslide Susceptibility Assessment

Alu Si¹, Jiquan Zhang^{1,*}, Zhijun Tong¹, Xingpeng Liu¹, Yichen Zhang², Qiuling Lang^{1,3}

¹School of Environment, Northeast Normal University, Changchun 130024, China

²Department of Land and Resources of Jilin Province, Changchun 130000, China

³School of Survey, Changchun Institute of Technology, Changchun 130021, China

滑坡灾害易发性评价中自然与人为因素对评价结果的影响研究

阿鲁思¹, 张继权^{1,*}, 佟志军¹, 刘兴朋¹, 张以晨², 郎秋玲^{1,3}

¹ 东北师范大学环境学院, 长春 130024, 中国

² 吉林省国土资源厅, 长春 130000, 中国

³ 长春工程学院勘测学院, 长春 130021, 中国

Abstract

Natural and human factors has different effect on landslide occurrence and susceptibility analysis. Through using natural factors as the index system and put into Logistic Regression and Random Forest model to conduct susceptibility assessment, the accuracy of the results are 0.7639 and 0.8884, which is slightly higher than the assessment using human factors as the index system. Therefore, it can be concluded that the influence of natural factors on the accuracy of susceptibility assessment is higher than the human factors. Meanwhile, some human factors will lead to reduce the accuracy of assessment.

Keywords: susceptibility assessment, weight of evidence, logistic regression, random forest

***通讯作者:** 张继权 (1965-), 教授, 博士生导师, 主要从事环境灾害风险评价与管理等方向研究。邮箱: zhangjq022@nenu.edu.cn

摘要

自然与人为因素对于滑坡灾害的发生和易发性评价的结果都有不同的影响。通过使用逻辑回归和随机森林模型以自然因素为指标体系进行易发性评价, 其结果准确率为 0.7639、0.8884, 略高于以人为因素为指标体系的评价。因此可以断定, 自然因素对于易发性评价准确度的影响高于人为因素。同时, 部分人为因素会导致易发性评价结果准确度降低。

关键词: 易发性评价, 证据权重, 逻辑回归, 随机森林

1. 引言

滑坡灾害作为最主要的地质灾害类型之一, 其发生深受自然及人为因素的影响[1]。人为因素即指, 在人类社会过程中出现, 对自然地质环境造成的影响的事件。当人为因素的影响超出自然环境自身承受能力时, 则会引起滑坡事件[2], 如: 工程、道路和铁路的建设; 矿产资源与水利资源的开采等等。而自然因素主要由外部和内部因素两部分组成, 即诱发事[3]件(外部因素)及地质环境易发程度(内部因素)。滑坡灾害诱发事件是指可能引起斜坡应力面力学结构遭到

破坏的自然事件，如：降雨、地震等等。而地质环境易发程度指的是由：地形、地貌和地质等一系列自然条件综合形成的孕灾环境在遭到外部因素影响时发生滑坡灾害的可能性或概率。对于人为因素和自然因素的研究可以揭示滑坡灾害发生的可能性或概率等问题。Varnes[4]1984年提出了较为准确和规范的滑坡灾害风险定义，即：“在一定事件内某特定地区发生特定类型滑坡灾害导致生命伤亡和财产损失的期望值”。根据此定义 van Western[5]等人在2008年提出了滑坡灾害风险评价理论框架和技术路线。在此框架中对于易发性分析的描述为，易发性分析属于滑坡灾害发生的空间概率研究，而危险性分析则是在空间概率分析的基础上考虑滑坡发生的时间概率。滑坡灾害易发性的中心思想是基于一句格言“the past and present are keys to the future”，换言之，与以发生滑坡灾害区域具有相似地质环境条件的区域更容易发生滑坡灾害。从上述几点关于滑坡易发性评价内容的叙述可以总结出，滑坡易发性评价研究的是研究区发生滑坡灾害的空间概率，是由该地区特定的地质环境条件决定的，而不考虑滑坡灾害发生的事件概率问题。因此，综合地形、地貌、地质等自然条件的地质环境条件（自然因素的内部因素）是易发性评价中最重要的因素之一。由于滑坡发生的时间概率完全由诱发事件决定，因此诱发事件即自然因素的外部因素是影响危险性的主要因素之一，而在易发性评价中不考虑。由于人类工程活动可以在诱发滑坡灾害的同时也可以强行改变地形、地貌等地质环境条件，所以人类工程活动（人为因素）即为易发性评价中较为重要的因素之一，也可以同诱发事件同时影响滑坡灾害的危险性[6]。

区域滑坡灾害易发性评价研究中，薛东剑[7]等人，采用多源遥感数据作为指标体系进行了地质灾害的风险评价。同时，在区域尺度上，白世彪[8]等人，运用逻辑回归模型进行了地质灾害的危险性评价。由于选取不同类型数据作为评价指标，导致相同模型可能计算出不同结果的情况或者由于选择单一类型的数据作为易发性评价指标导致在

某些区域评价失真。因此本研究中，选取10类数据作为评价指标，通过证据权重法及逻辑回归模型对各个指标之间、指标与滑坡之间的关系进行分析；通过逻辑回归、随机森林模型与受试者工作特征曲线分析各类指标对易发性评价结果与评价精度的影响。从而反映出自然因素中的地质环境条件（内部因素）与人为因素在滑坡易发性评价中起到的作用及影响，为易发性评价及指标体系选取提供可参考的研究。

2. 研究区与数据

2.1 研究区概况

因吉林省东部山区与西部平原地质环境的差异，导致东部山区滑坡灾害分布远远多于西部平原。本文研究区坐落于通化县东北角，为代表性的东部山区地质环境区域。整个研究区地势由南向北沉降，形成东南高西北低的地势地貌，最高点为1175m，最低点为560m。全年降雨主要集中在夏季，年均降水量为870mm，6-8月三个月的降水量占年总降水量的60%以上。气候条件属中温带湿润气候区，年平均气温5.5℃。

2.2 数据类型

选择适当的数据作为指标是一项艰巨的挑战，在选择能够满足特定的易发性评价的指标时，没有任何标准可以遵循[9]。在易发性评价研究中选取指标时，往往采用专家调研法的同时参考以往的研究进行指标的选取。基于本文的研究目的，结合以往相关的研究中经常采用的指标，从中选取高程、坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度和地形湿度指数作为表示基本的地质环境条件的指标，其后，选择土地利用类型数据中的建筑物、道路和铁路、农田、林区等目标信息作为表示人为因素的指标，为进行对比选择坡度坡长指数作为表示地质环境条件的指标。

3 证据权重法

证据权重法是Agterberg提出的一种数学统计方法。证据权重法其基本思路是将分别赋以不同权重值的指标图层进行空间叠

加和综合分析^[10]。

证据权重法中存在和不存在该指标的权重由下述公式计算所得：

$$W^+ = \ln[P(D_i|B)/P(D_i|\bar{B})]$$

$$W^- = \ln[P(\bar{D}_i|B)/P(\bar{D}_i|\bar{B})]$$

式中， W^+ 和 W^- 表示存在或不存在该指标的权重值。 D_i 是存在第*i*个指标的面积， \bar{D}_i 是不存在第*i*个指标的面积， B 和 \bar{B} 是发生和未

发生滑坡的面积。使用 C_i 来表示第*i*个指标与滑坡灾害的相关性，由下述公式计算：

$$C_i = W_i^+ - W_i^-$$

C_i 值越高表明该指标与滑坡相关性越强，等于零表示无相关性，大于零表示该指标对于滑坡的发生有积极影响，小于零表示对于滑坡的发生有消极的影响即阻碍滑坡发生。图1至图10为是个指标的相关性图。

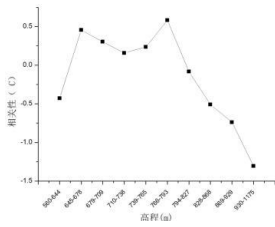


图 1

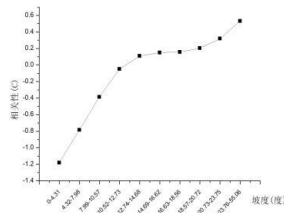


图 2

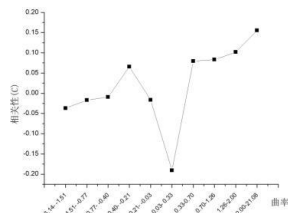


图 3

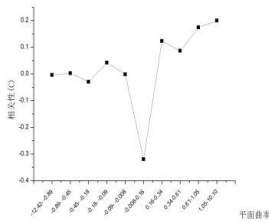


图 4

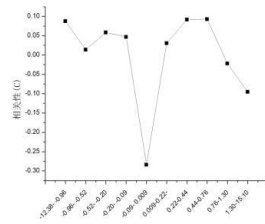


图 5

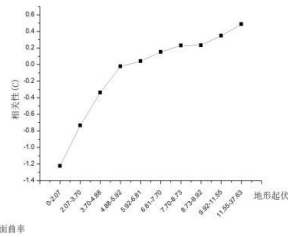


图 6

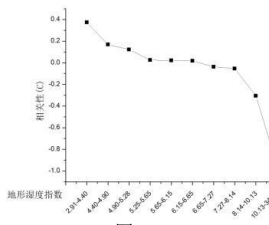


图 7

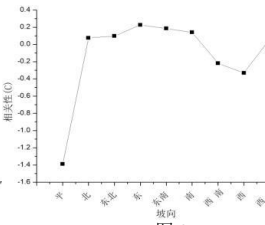


图 8

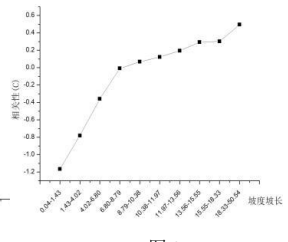


图 9

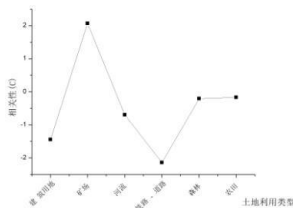


图 10

图1至图10，高程、坡度、曲率、平面曲率、剖

面曲率、地形起伏度、地形湿度指数、坡向、坡度坡长指数和土地利用类型的指标相关性。

坡度、地形起伏度、坡度坡长指数等三个指标随着各自指标值变大其与滑坡灾害的相关性变得越来越强，这三个指标在各自指标值低时对滑坡发生起着负面作用，随着值的提升在最大值时纷纷对滑坡发生起着积极作用。而地形湿度指数指标随着其值变

大其与滑坡灾害的相关性变得越来越小，在最大值时对滑坡发生起着消极作用。曲率、平面曲率、剖面曲率等三个指标在某一阶段表现出极为明显的负面相关性，其余阶段均表现出与滑坡发生无特别明显的相关性。高程、坡向与土地利用类型等三个指标的相关性表现出不规律的趋势。其中土地利用指标在建筑用地、矿场、道路等三个子因素上表现出几位明显的相关性。

4.逻辑回归模型

逻辑回归模型是在滑坡灾害易发性评价中最为常用的方法之一^[11]。逻辑回归模型用来研究因变量 Y 与各个自变量 $X_1, X_2 \dots X_n$ 之间的联系。逻辑回归模型的公式如下所示：

$$P = 1 / (1 + e^{-z})$$

式中，P 代表每个像元的滑坡灾害易发性，Z 的表达式如下：

$$Z = \alpha + \beta_1 x_1 + \dots + \beta_n x_n$$

式中 $\beta_1, \beta_2 \dots \beta_n$ 是 $X_1, X_2 \dots X_n$ 指标的系数， α 是常数由计算所得，因此逻辑回归模型的公式通过逻辑回归模型计算十个指标权重结果如下表 1 所示。

表 1 逻辑回归模型指标系数

指标	系数	指标	系数
高程	0.1098	土地利用	
坡度	0.2131	建筑用地	-0.0151
坡向		矿场	0.0017
北	0.0056	河流	-0.0006
东北	0.0039	森林	-0.0008
东	0.0052	铁路、道路	0.1490
东南	0.0032	农田	-0.0118
南	0.0046	坡度坡长	0.1469
西南	0.0026	剖面曲率	-0.0292
西	0.0046	地形湿度	-0.2617
西北	0.0073	地形起伏度	0.2074
曲率	-0.0714		
平面曲率	-0.0655	常数	-0.0077

从表 1 中我们可以发现，与滑坡灾害最为相关的指标为坡度、地形湿度指数和坡度坡长指数，系数值分别为 0.2131、-0.2617 和 0.2074。具有最想相关性指标集中在坡向与

土地利用类型指标中，而在土地利用指标中与滑坡发生相关性最高的是铁路与道路次级指标。

5.随机森林模型

随机森林指的是利用多棵树对样本进行训练并预测的一种分类器^[12]。随机森林模型使用 bootstrap sample 方法，从训练数据集中提取多个不同的指标作为样本集，并使用每个样本集构建决策树模型，然后结合 N 棵树的预测结果，通过投票获得最终的预测结果。随机森林的构建过程大致如下：

(a)从原始训练集中使用 Bootstrapping 方法随机有放回采样选出 m 个样本，共进行 n 次采样，生成 n 个训练集。(b)对于 n 个训练集，我们分别训练 n 个决策树模型。(c)对于单个决策树模型每次分裂时根据基尼指数选择最好的特征进行分裂。(d)每棵树都一直这样分裂下去，直到该节点的所有训练样例都属于同一类。(e)将生成的多棵决策树组成随机森林。对于回归问题，由多棵树预测值的均值决定最终预测结果。在本文中结合使用随机森林模型来验证不同类型数据对于易发性评价结果的影响。

6.受试者工作特征曲线

受试者工作特征曲线(receiver operating characteristic curve, 简称 ROC 曲线)是根据一系列不同的二分类方式(分界值或决定阈)，以真阳性率为纵坐标，假阳性率为横坐标绘制的曲线^[13]。ROC 曲线下的面积值在 1.0 和 0.5 之间。在 AUC>0.5 的情况下，AUC 越接近于 1，说明诊断效果越好。AUC=0.5 时，说明诊断方法完全不起作用，无诊断价值。AUC<0.5 不符合真实情况，在实际中极少出现。

本文使用逻辑回归模型与随机森林模型，在选用不同数量的指标情况下共建立了 6 个模型(其中逻辑回归模型三个分别是模型 1、模型 2 和模型 3，随机森林分别是模型 4、模型 5、模型 6)，并运用 ROC 法计算各个模型易发性结果的准确性。ROC 结果如图 11 与图 12 所示。

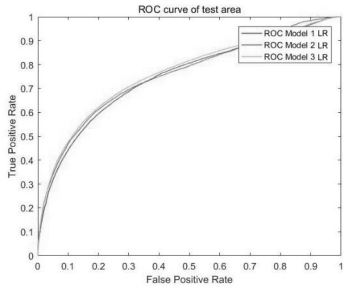


图 11.逻辑回归模型算法 ROC 曲线图

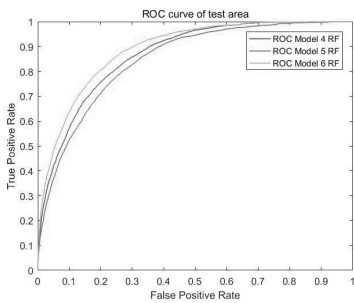


图 12.随机森林模型算法 ROC 曲线图

其中模型 1 与模型 4 在进行易发性评价时选用了高程、坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度、地形湿度等八个指标，模型 2 和模型 5 在这八个指标的基础上加入了土地利用数据作为指标进行易发性分析，模型 3 与模型 6 则是用坡度坡长指数替换土地利用数据作为指标进行滑坡易发性分析（表 2 为 6 个模型的 AUC 值）。

表 2 模型 AUC 值

	模型1	模型2	模型3
测试区AUC值	0.7502	0.7513	0.7639
	模型4	模型5	模型6
测试区AUC值	0.8683	0.8466	0.8884

7.讨论

结合证据权重法的结果与逻辑回归模型所算得的各个指标的权重结果，可以反映出自然因素指标与人为因素指标对于滑坡发生等影响。在证据权重法中指标与滑坡的相关程度随着该指标值的增加而递增或递减的，该指标在逻辑回归模型中具有极高系数，表明它对于滑坡的发生具有极大影响。

如坡度、地形起伏度、地形湿度指数和坡度坡长指数等四个自然因素指标，其中坡度、地形起伏度与坡度坡长指数显示从对滑坡产生有负面影响开始随着其值的提高慢慢的转变为积极影响，其在逻辑回归模型中的权重值分别为 0.2131、0.2074 和 0.1469，而地形湿度指数在证据权重法中显示出相反趋势因此其在逻辑回归模型中的权重为 -0.2617。从结果中可以看出这四个指标的重要性远远超过其余指标。通过证据权重法算出的曲率、平面曲率、剖面曲率、坡向、高程等指标的权重显示出并无明显规律，同时，其在逻辑回归中的权重系数也较低，意味着这些指标对滑坡发生具有较小的影响。值得一提的是在土地利用类型指标中，道路、铁路次级指标在两个模型中都显示出对滑坡灾害的发生有明显的积极影响（逻辑回归系数：0.1490），其余几个次级指标则显示出对于滑坡灾害的发生并无太大影响。

本文使用两种模型结合不同数量的指标进行易发性评价，通过分析模型结果的准确性可以揭示加入或移除何种指标时其对易发性评价结果准确性的影响。模型 1 与模型 4 其结果 AUC 值为 0.7502 和 0.8683。当将土地利用（人为因素指标）作为指标加入到两个模型时（模型 2 与模型 5），发现其时逻辑回归模型预测结果准确度增加了 0.0011，而随机森林预测准确度则减少了 0.0117。发现土地利用指标的加入并未明显增加预测结果的准确性，反而在随机森林模型预测时起到反作用，使结果准确性降低。而用坡度坡长指数指标代替土地利用数据进行易发性评价时，两个模型（模型 3 和模型 6）结果的预测准确性得到了明显的提高，分别提高了 0.0137 和 0.0201。发现坡度坡长数据作为自然因素指标的代表，其的加入明显增加了两个模型滑坡易发性评价的准确性。

8.结论

描述自然因素的地质环境条件的指标，其对于滑坡的发生及滑坡易发性评价准确性的影响都很大。而人为因素指标对于滑坡的发生及易发性评价结果的影响则较为复杂。本文中代表人为因素的土地利用指标中的道路、铁路次级指标则表现出对于滑坡发

生有积极的影响,其他次级指标则并无太大影响。总结上述部分,可以发现自然因素即地质环境条件是影响滑坡易发性评价最为重要的指标,其对于提高评价结果预测的准确性有极大帮助,而人为因素指标则会影响整体评价的准确性,但是部分次级指标可能会对预测准确性的提高起到一定帮助作用。

Acknowledgements

This study was supported by the The Key Scientific and Technology Research and Development Program of Jilin Province (20180201033SF) and The Key Scientific and Technology Research and Development Program of Jilin Province (20180201035SF)

致谢

本研究得到了吉林省重点科技攻关项目(20170204035SF);吉林省重点科技研发项目(20180201033SF)的资助。

参考文献

[1] 王佳佳.三峡库区万州区滑坡灾害风险评估研究.武汉:中国地质大学,2015:1-166.

[2] Pradhan B, Lee S. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software*,2010,25(6) : 747-759.

[3] 殷志强,陈红旗,褚宏亮,徐永强,赵无忌.2008年以来中国5次典型地震事件诱发地质灾害主控因素分析.地学前缘,2013,20(6):289-302.

[4] Varnes D J. Slope international association of engineering. *Landslide Hazard zonation: a review of principle and practice*.1984, 1-63.

[5] Van Western C J, Castellanos E et al. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An

overview. *Engineering Geology*. 2008, 102(3-4):112-131.

[6] 刘绪.长白山天池火山喷发诱发崩塌滑坡灾害危险性评价.长春:吉林大学,2016: 1-88.

[7] Xue D J, He Z w, Tao S, et al. Application of Multi-source Remote Sensing Data in Geological Disaster Risk Assessment// Meeting of Risk Analysis Council of China Association for Disaster Prevention. 2010:103.

[8] Bai S B, Zhang F Y, Dong L f, et al. Area Scale Landslide Hazard Assessment Using Logistic Regression: a Case Study in Bailongjiang River Basin, China// Meeting of Risk Analysis Council of China Association for Disaster Prevention. 2008:395

[9] Wang L, Sawada K, Moriguchi S. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Computer& Geoscience*. 2013, 57: 81-92.

[10] Van Western C J. The modelling of landslide hazards using GIS. *Surveys in Geophysics*. 2000, 21: 241-255.

[11] Bai S B, Wang J, Lu G N, Zhou P G, et al. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, china. *Geomorphology*. 2010, 115: 23-31.

[12] Cutler D R, Edwards T C, Beard K H, et al. Random forests for classification in ecology. *Ecology*. 2007, 88: 2783-2792.

[13] Trigila A, Iadanza C. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology*. 2016, 249: 119-136.