

文章编号: 2096-1618(2024)01-0101-07

四川省滑坡灾害危险性评价模型比较研究

刘亚彬¹, 刘云腾², 朱真梅³, 刘志红⁴, 范智高¹, 史红人¹

(1. 成都信息工程大学大气科学学院, 四川 成都 610225; 2. 荣县气象局, 四川 自贡 643100; 3. 成都信息工程大学管理学院, 四川 成都 610225; 4. 成都信息工程大学资源环境学院, 四川 成都 610225)

摘要:为探讨四川省滑坡灾害分布状况及对滑坡灾害危险性评价建模效果进行比较,以四川省为研究区,分析地质灾害尤其是滑坡灾害的时空分布,运用信息量法、基于信息量的逻辑回归法、随机森林法3种方法,取坡度、植被覆盖率、地震烈度、断层密度及年累计降雨量5个因子建立危险性评价模型,结合ROC曲线和AUC值对模型性能进行精度比较。研究表明:(1)坡度因子对滑坡灾害的贡献最大,3种评价模型中坡度、地震烈度及降雨3个因子均有较大贡献。(2)滑坡灾害主要分布于地震带附近,在成都平原处灾害发生几率最小。(3)3种评价方法均适用于滑坡危险性区划工作,具有良好的建模效果。相对于其他两种模型,逻辑回归模型在实际情况下更加适合四川省滑坡灾害的危险性评价工作。

关键词:滑坡灾害;信息量方法;随机森林方法;逻辑回归方法;模型评价

中图分类号:P49

文献标志码:A

doi:10.16836/j.cnki.jcui.2024.01.015

0 引言

滑坡灾害通常具有突发性、群发性、区域性和成灾性等特点,对人类的生命和财产安全构成了巨大的威胁^[1-3]。滑坡灾害危险性区划通过对域内不同位置的滑坡易发程度进行评价,为滑坡灾害管理和预警预报工作提供数据和方法支持,是防灾减灾工程中的重要一环^[4]。常用的滑坡危险性区划方法主要有启发式求解法、基于物理确定性模型和工程地质的方法及统计学习方法等^[5-6]。启发式求解法依赖专家经验,具有一定主观性。确定性方法需收集大量的工程地质数据,不适用于大面积的空间范围。统计学习方法通过机器学习从历史滑坡数据中发现规律,假定因子状况与历史滑坡相似的位置也容易发生滑坡,该方法不需要收集大量数据,且具有一定客观性,适用于大范围的区域滑坡危险性评价研究^[7-9]。

目前,滑坡危险性区划中主要使用的统计学习方法包括 Logistic 回归^[10]、分类与回归树(CART)、广义加法模型(GAM)、支持向量机(SVM)及信息量法(IM)^[11-12]等。上述研究均基于单个统计模型。近年来,通过组合多个单一模型来提高正确率和泛化能力的集成学习方法(ensemble learning)在机器学习领域兴起,并成功应用于信号处理、手写体识别、指纹识别和人脸识别等多个领域^[13-14]。

统计现有的方法,本研究采取传统的信息量法、改进的逻辑回归法及机器学习领域的随机森林法3种方法通过ROC曲线比较其准确性。再分析诱发四川地区滑坡灾害的因子所占权重及最终危险性的分布状况。

1 数据来源及地理概况

滑坡灾害致灾因子的选择一直是滑坡灾害预警的重点研究内容。四川省滑坡灾害如此严重,与特殊的地质条件和大气条件密不可分。四川省地形梯度巨大,河流下切强烈,地质构造活跃,断裂发育,地震频发^[15],为滑坡灾害的发育提供了最根本的地质条件;受青藏高原对东亚季风、南亚季风阻挡影响,省内降雨充沛而集中,为地质灾害频发提供动力因素。本研究分别从自然地理概况和地质岩性、地形构造、植被覆盖、气候条件、土地利用情况5类致灾因子概况对四川省自然条件做具体的描述。

四川省位于97°21'E~108°33'E和26°03'N~34°19'N,是中国西南五省之一,东西长1075 km,南北宽921 km,东西边境时差51 min,幅员面积为48.5万km²,与7个省(区、市)接壤,东邻重庆,北连青海、甘肃、陕西,南接云南、贵州,西衔西藏(图1)。是西南、西北和中部地区的重要结合部,是承接华南华中、连接西南西北、沟通中亚南亚东南亚的重要交汇点和交通走廊。

四川省辖21个市、州,181个县(市、区),其中民族自治州3个、民族自治县4个、市辖区43个、县级市14个;有4800个乡镇,其中乡2918个、镇1882个;街道办事处222个。

收稿日期:2023-02-28

基金项目:四川省科技厅重点研发资助项目(2022YFS0542)

通信作者:刘云腾. E-mail:592206414@qq.com

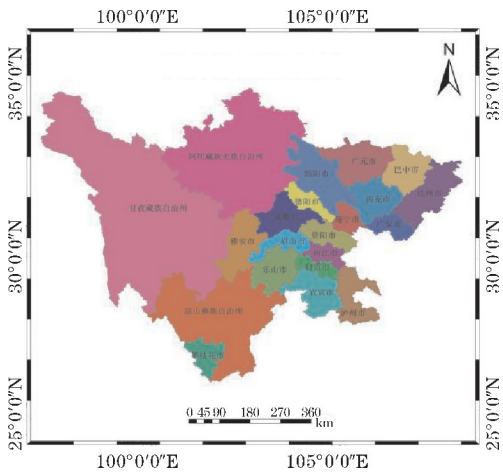


图1 四川省地理位置及行政区划

全省共有 56 个民族,少数民族人口 412 万人,彝族 212 万人、藏族 127 万人。2004 年末省总人口 8724.6 万人,其中乡村人口为 6796.88 万人;人口自然增长率 2.8%。

1.1 数据来源及处理

1.1.1 降雨量数据

1981–2020 年共收集数据雨量站 971 个。因采用精细化研究,所以对获得的降雨数据进行剔除误测、缺测数据,共获得 921 个雨量站的有效小时降雨量数据。再对获得的小时数据通过 Excel 进行数学计算与整理,得到川南地区的日、月、季、年的累计降雨数据,为后续的滑坡危险性评价中的降雨因子计算打下基础。

1.1.2 地形数据

通过国家开放下载的 2020 年数字高程数据,选取空间分辨率为 15 m,再通过 ENVI 和 ARCGIS 软件对数据进行拼接、裁剪,获得川南地区的高程数据,利用 ARCGIS 对高程数据进行计算,得到川南地区的坡度数据。

1.1.3 植被覆盖度数据

采用中国科学院资源环境科学与数据中心网站选取 30 m 空间分辨率下载,通过 ARCGIS 裁剪,获得川南地区的植被覆盖度数据。

1.1.4 地震烈度及断层数据

采用四川省地质环境监测总站制作的四川省地震烈度栅格数据,空间分辨率为 3000 m。断层为四川省地质环境监测总站提供的矢量数据,再计算密度得出。

为方便后续计算,将 5 个数据的分辨率通过 ARCGIS 统一转化为 100 m。

1.2 地理概况

1.2.1 地质构造

地震是滑坡灾害发生的第二大诱因。四川省地震频发,省内地震烈度纵跨 3 ~ 11 级,如图 2 ~ 3 所示;川东地区、川西南地区地震烈度较低,大部分地区为 3 度、6 度;川中及川西大部分地区为 7 度。省内有 3 条主要地震带:盆地西北缘龙门山断裂带、川西鲜水河断裂带、川南安宁河断裂带,地震烈度由外向内从 8 ~ 11 度增加,其中龙门山断裂带核心区域地震烈度最大为 11 度,鲜水河断裂带和安宁河断裂带中心区域地震烈度为 10 度。断层沿着地震带密切分布,集中在四川高海拔地区。

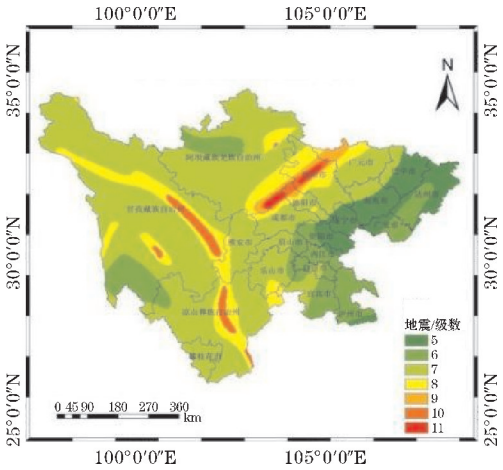


图2 四川省地震烈度图

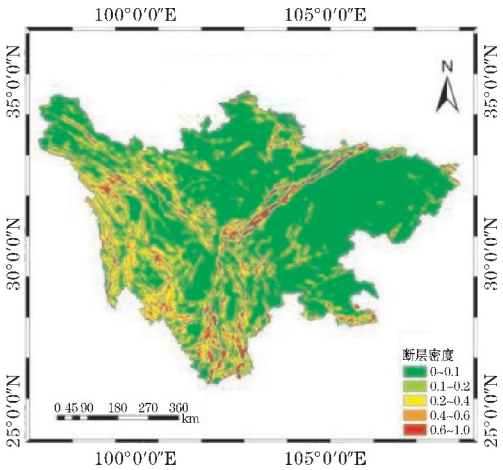


图3 四川省断层密度图

1.2.2 地形条件

地形条件是控制滑坡发育的重要因素,其中以坡度和高程为重。四川省介于中国自西向东 3 个台阶的一、二台阶的过渡地带(图 4),西部为青藏高原东南边缘,东部为四川盆地。地貌形态类型多样,拥有平原、丘陵、山地和高原,西高东低,高差悬殊,河流纵横,切割强烈,山丘广布,平原狭小。境内最高点是西部大雪山主峰贡嘎山,海拔 7556 m,最低点在东部邻水县御临河出境处,海拔 186.77 m^[16]。

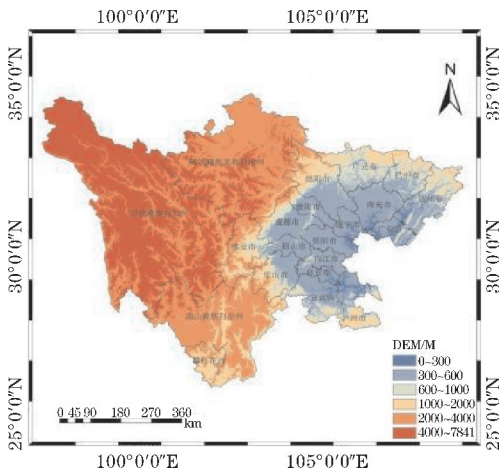


图 4 四川省高程图

除龙泉山脉外的川东盆地地区、若尔盖湿地保护区,大部分区域坡度在 5° 以内,盆地北部、南部丘陵地区坡度主要在 $5^{\circ} \sim 25^{\circ}$,盆地西部山地地区落差很大,平均坡度在 $25^{\circ} \sim 40^{\circ}$,少数区域坡度达 40° 以上;川西高山高原地区局部地区坡度在 5° 以内,但大部分地区平均坡度在 $25^{\circ} \sim 40^{\circ}$ 。根据坡度的分布,参考前人研究,本文将坡度分为 11 级,分级标准如下: $0^{\circ} \sim 5^{\circ}$, $5^{\circ} \sim 10^{\circ}$, $10^{\circ} \sim 15^{\circ}$, $15^{\circ} \sim 20^{\circ}$, $20^{\circ} \sim 25^{\circ}$, $25^{\circ} \sim 30^{\circ}$, $30^{\circ} \sim 35^{\circ}$, $35^{\circ} \sim 40^{\circ}$, $40^{\circ} \sim 60^{\circ}$, $60^{\circ} \sim 80^{\circ}$, $80^{\circ} \sim 100^{\circ}$,见图 5。

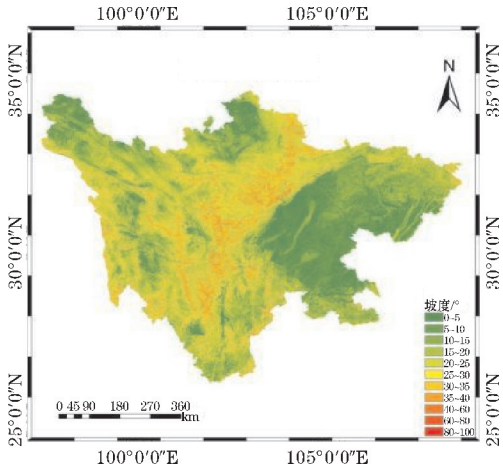


图 5 四川省坡度图

1.2.3 植被条件

植被覆盖度对滑坡灾害的形成是把双刃剑,四川省植被类型丰富,由于复杂的地形地貌、土壤和气候条件,植被覆盖度有明显的水平分布和垂直分布差异,将植被覆盖度作为研究四川省滑坡发育的影响因素十分必要。年均植被覆盖东多西少,东部以常绿林和平原草地为主,西部以高山亚高山草甸为主,包括中国亚热带区域的所有栽培植物,并有不少南亚热带和温带的种类。根据《水土保持技术规范》把植被覆盖度分为 4 级,如图 6 所示:川东植被覆盖度相对较高以高覆盖为

主;川中过渡带地区植被覆盖度很高,以高覆盖和极高覆盖为主;川西高山高原植被覆盖度较低,以中覆盖和较低覆盖为主。总体四川省以高覆盖度植被为主,面积约为 239108.5 km^2 ,占全省面积的 50%;中覆盖度植被面积次之,约为 127191.875 km^2 ,占全省面积的 26%,较低覆盖度植被面积约为 45047.8125 km^2 ,约占全省面积的 9%,低覆盖度植被面积约为 60678.4375 km^2 ,约占全省面积的 13%;极高覆盖度植被面积只 10511.4375 km^2 ,约占全省面积的 2%。

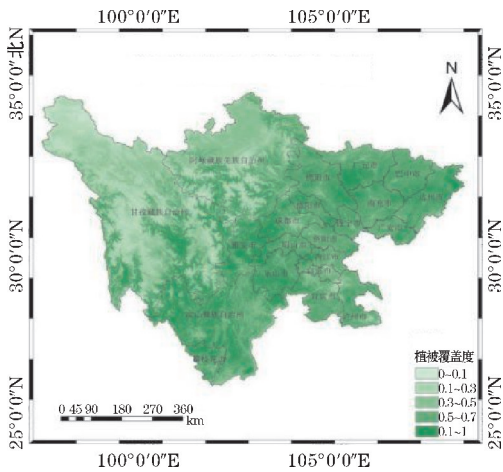


图 6 四川省植被覆盖度分布图

1.2.4 气候条件

气候条件,尤其是降雨条件对滑坡灾害的发生非常重要,根据对四川省 50 年的降雨特征和滑坡灾害的特征分析,绝大多数滑坡灾害都是由降雨诱发的。

四川省幅员辽阔,涵盖 6 个温度带,由东向西呈阶梯状分布;省内年均降水量在 $130 \sim 1500 \text{ mm}$,水平分布差异性很大,东多西少,在川东北和盆西南出现两个强降水中心,强降水中心年均降水量在 $1400 \sim 1600 \text{ mm}$,如图 7 所示。本文以 200 mm 为间隔,将年均降雨量等间隔的分为 8 级。

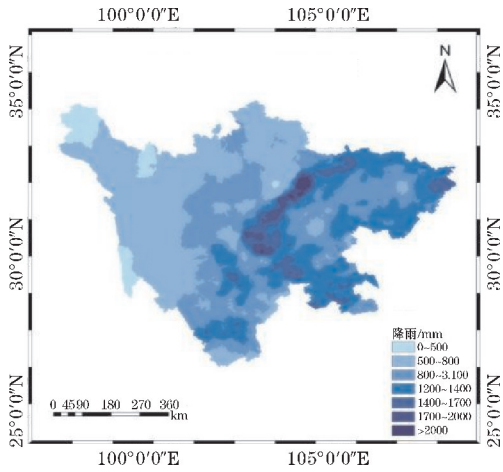


图 7 四川省 2020 年年累计降雨分布图

1.2.5 滑坡灾害概况

四川省滑坡灾害发生有两个主要特征,一是条带状分布的特点,这与四川省地震烈度分布有强烈的呼应性,沿龙门山断裂带、安宁河断裂带和龙泉山脉呈条带状分布的特点尤其明显;二是川东北地区滑坡灾害有群发性特征,体现为川东北地区某区域内滑坡灾害发生密度远高于其他地区。具体分布状况可见图8。

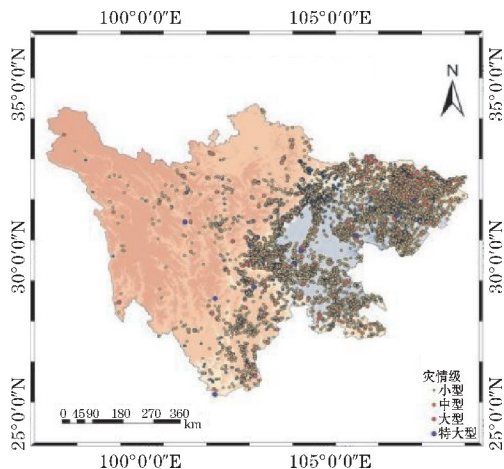


图8 四川省滑坡灾害分布图

2 模型介绍

2.1 信息量法

信息量模型是进行区域滑坡灾害评价的一种方法,一般适用于中小型比例尺度的滑坡灾害评价及预测。理论基础是信息论,指通过计算一个因子的熵值对该因子对研究区危险性的影响程度。本研究通过分析各因子的信息熵来研究各因子对川南地区滑坡灾害的危险性程度。其中信息量模型最关键的是对评价目标各因子进行合理性分类,一般选用相关标准来对评价因子进行分类,其计算公式为

$$I = \sum_{i=1}^n I(x_i, H) = \sum_{i=1}^n \ln \frac{A_i/A}{M_i/M}$$

式中 x_i 指在评价因子内所划分的因子等级, $I(x_i, H)$ 指该等级因子对于滑坡灾害的信息量, A 为研究区域总面积, A_i 为研究区域内含有因子 x_i 的面积, M 为该区域内所含有全部滑坡灾害的数量, M_i 在 x_i 评价因子等级区域内所含有的滑坡灾害点数量, I 为所计算得来的最终信息量, n 为影响因子的数量。

2.2 基于信息量的逻辑回归法

研究采用以信息量模型为基础,叠加逻辑回归模型进行计算评价的方法,其优点是可以充分考虑到地质条件、地理环境对其研究区滑坡灾害产生的影响与

贡献。该组合方法避免了信息量模型过于简单导致权重分配不合理的状况,降低了以往评价方法的主观程度。其计算公式为

$$I = \sum_{i=1}^n I(x_i, H) = \sum_{i=1}^n \ln \frac{A_i/A}{M_i/M}$$

式中 x_i 指在评价因子内所划分的因子等级, $I(x_i, H)$ 指该等级因子对于滑坡灾害的信息量, A 为研究区域总面积, A_i 为研究区域内含有因子 x_i 的面积, M 为该区域内所含有全部滑坡灾害的数量, M_i 在 x_i 评价因子等级区域内所含有的滑坡灾害点数量, I 为所计算得来的最终信息量, n 为影响因子的数量。

$$P = \frac{\exp(Z)}{1 + \exp(Z)}$$

$$P = \frac{\exp(A + B_1 x_1 + \cdots + B_i x_i)}{1 + \exp(A + B_1 x_1 + \cdots + B_i x_i)}$$

$$= \frac{1}{1 + e^{-(A + B_1 + \cdots + B_i)}}$$

式中 P 为研究区域内 Z (滑坡灾害) 的发生概率; A 为逻辑回归方程截距; B 为逻辑回归模型中代表的评价因子计算获得的逻辑回归数,其分级 B_i 为该评价因子分级的逻辑回归数。

2.3 随机森林

随机森林(RF)的计算流程,如图9所示。RF是由很多决策树组成的集成学习分类模型。首先采用 Bootstrap 的方法,假设总样本个数为 N ,总特征(评价因子)个数为 P 。在总样本 N 中随机抽取 1 个样本,且抽取的样本集 N 与原始数据大小相同,Bootstrap 采样是以有放回的采样方式,增加了采样的随机性。然后采用最佳分裂方式,每棵决策树最终都完整成长而不会减枝,这样可以得到差错较少的决策树,确保随机森林对基础数据分类训练的准确度。弱分类器可以相互补充,最大限度地减少单个分类器的误差,从而提高整体准确度。输入特征值数目, p 的数值小于 P ,决策树中的每一个节点从该样本集中总特征 P 中选择 p 个,再根据基尼系数从 p 个特征值中选择最优值。经过实施 T 次步骤后得到 T 棵决策树,最后构建出随机森林(RF)模型。

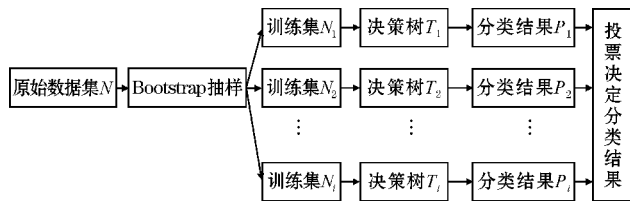


图9 随机森林算法流程图

由于随机森林的样本数量和特征个数是随机挑选,会导致模型分类偏差的发生。因此对结果的均分选择,随机森林的方差减少,且比模型偏差量少,就能有效减小过度拟合。随机森林输出的平均概率值计算公式为

$$H_p = \max \left\{ H_i = \frac{\sum_{j=1}^N H_{ij}}{T} \mid i \in I \right\}$$

式中 H_i 是指在时间 i 内是否发生的概率值, H_p 是随机森林模型最终输出分类结果所对应的概率值, H_{ij} 是第 j 棵决策树在时间 i 内是否存在的概率值, I 是所有决策树分类的集合, T 代表模型决策树的总数量。

3 建模结果分析

3.1 相关性分析

为避免选取的评价因子间关联过大导致的数据冗

余现象,需要对各因子的相关性进行分析。利用 SPSS 软件通过计算各因子间的相关系数 R 来进行相关性分析,根据相关标准当各评价因子间的相关系数 R 小于0.3时,可以认为这几个因子不相关,能被作为评价因子进行下一步处理。通过表 1 可知,本研究所选取的 6 个评价因子之间的相关系数均小于0.3,可作为后续危险性评价的因子。

3.2 计算结果

3.2.1 信息量模型

通过上述公式计算得出 6 个因子的信息量,结果如表 2。

表 1 相关性分析表

评价因子	坡度/°	年降雨量/mm	植被覆盖度	土壤类型	地震烈度	断层密度
坡度	1	0.175	0.029	0.344	-0.014	0.178
年降雨量	0.175	1	0.214	0.263	0.14	-0.084
植被覆盖度	0.029	0.214	1	0.234	0.111	0.034
土壤类型	0.344	0.263	0.234	1	0.06	0.246
地震烈度	0.127	0.14	0.111	0.06	1	0.056
断层密度	0.178	-0.084	0.034	0.246	0.056	1

表 2 信息量计算表

评价因子	分级	灾情点占比	子集面积占比	信息量
年降雨量/mm	0 ~ 500	0.001	0.044	-3.752
	500 ~ 800	0.067	0.381	-1.736
	800 ~ 1100	0.334	0.292	0.135
	1100 ~ 1400	0.385	0.215	0.584
	1400 ~ 1700	0.113	0.051	0.805
	1700 ~ 2000	0.073	0.014	1.630
	>2000	0.027	0.004	2.001
坡度/°	0 ~ 5	0.244	0.163	0.403
	5 ~ 10	0.219	0.150	0.379
	10 ~ 15	0.174	0.141	0.211
	15 ~ 20	0.133	0.140	-0.047
	20 ~ 25	0.103	0.132	-0.250
	25 ~ 30	0.068	0.116	-0.532
	30 ~ 35	0.036	0.086	-0.877
	35 ~ 40	0.016	0.046	-1.041
	40 ~ 100	0.006	0.026	-1.391
	3	0.258	0.089	1.064
地震烈度	6	0.141	0.137	0.031
	7	0.388	0.609	-0.452
	8	0.113	0.121	-0.067
	9	0.056	0.013	1.449
	10	0.032	0.025	0.254
植被覆盖度	11	0.011	0.005	0.794
	0 ~ 0.1	0.0001	0.027	-5.052
	0.1 ~ 0.3	0.027	0.202	-2.004
	0.3 ~ 0.5	0.138	0.226	-0.491
	0.5 ~ 0.7	0.532	0.307	0.548
断层密度	0.7 ~ 1	0.303	0.238	0.241
	0 ~ 0.1	0.744	0.612	0.196
	0.1 ~ 0.2	0.105	0.152	-0.367
	0.2 ~ 0.4	0.088	0.164	-0.621
	0.4 ~ 0.6	0.039	0.053	-0.320
	0.6 ~ 1.5	0.024	0.019	0.204

在降雨与地震烈度中信息量呈现雨量、震级越大所占权重越大的情况,说明大量的降水与地震会明显影响滑坡灾害的发生。

坡度则是与降雨呈现相反的趋势,坡度越小灾害越容易发生,这与通常理解相反。这可能是因为高坡度地区的观测点较少,也可能是因为选取影像分辨率较高的缘故。

植被覆盖率与断层密度也遵从密度越大滑坡灾害发生率越高的情况。

3.2.2 逻辑回归模型

将处理好的滑坡点数据按照 7 : 3 设为训练样本与验证样本,再在坡度小于 5°的区域内通过 ARCGIS 软件建立与滑坡点数量相同的随机点,并按照相同的方法将其划分为训练样本与验证样本。

获得 8064 个训练样本后,将其导入 SPSS 软件中,通过以是否发生滑坡灾害为因变量,坡度、地震烈度、断层密度、植被覆盖度、年累计降水、土壤类型为自变量进行逻辑回归运算。

B 值越大,意味着该因子对研究区滑坡灾害危险性的贡献越大(表 3)。

其中,降雨、地震烈度及坡度对四川地区滑坡灾害危险性的贡献较大,植被覆盖度和断层密度贡献较小。

3.2.3 随机森林模型

与逻辑回归模型相同,将数据按 7 : 3 划分为训练样本和测试样本,随机森林模型计算结果得出,对于滑坡灾害年降雨量的权重占比最大,其次是坡度和地震因子,植被覆盖度、断层占比较小。如图 10 和表 4 所示随机森林模型得出较准确 R^2 测试结果为0.83。

表 3 逻辑回归计算表

评价因子	<i>B</i>	标准误差	瓦尔德	自由度	显著性	exp(<i>B</i>)
坡度	0.96	0.037	58.52	1	0	1.378
降雨	0.81	0.021	446.692	1	0	1.573
植被覆盖度	0.35	0.03	5.93	1	0.01	1.077
地震烈度	0.74	0.031	331.794	1	0	1.75
断层密度	0.17	0.06	1.753	1	0.19	1.082
常量	0.15	0.019	90.667	1	0	1.199

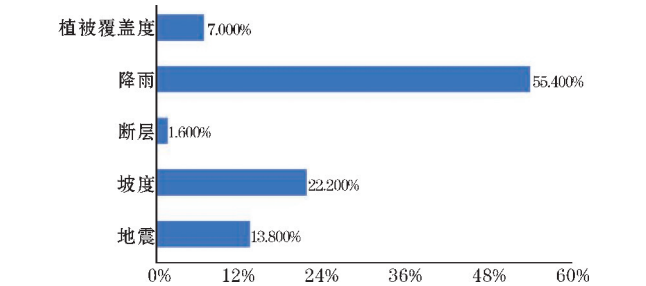


图 10 随机森林模型计算结果

表 4 交叉验证结果

数据	MSE	RMSE	MAE	MAPE	<i>R</i> ²
训练集	0.085	0.291	0.182	88.5	0.86
交叉验证集	0.108	0.329	0.208	104.5	0.79
测试集	0.106	0.325	0.205	96.3	0.83

3.2.4 结果验证及分析

选取 ROC 曲线对 3 种模型的精度进行验证,计算结束需对得出的结果进行验证,其中 AUC(area under curve)作为评判标准,AUC 是指 ROC 曲线下的面积,一般认为当 ROC 曲线0.5 ~ 1.0,认为验证结果与所计算的结果成正相关,ROC 在 0 ~ 0.5则被认为是负相关。

当结果>0.6,可认为该结果较为准确,本研究结果经过验证得出 ROC 曲线面积为0.857。3 种模型的精度均>0.81,具有代表性,如图 11 ~ 14 所示。但通过对滑坡灾害点的叠加得出逻辑回归模型具有较好的实际意义。比较符合滑坡灾害的分布规律和现状。

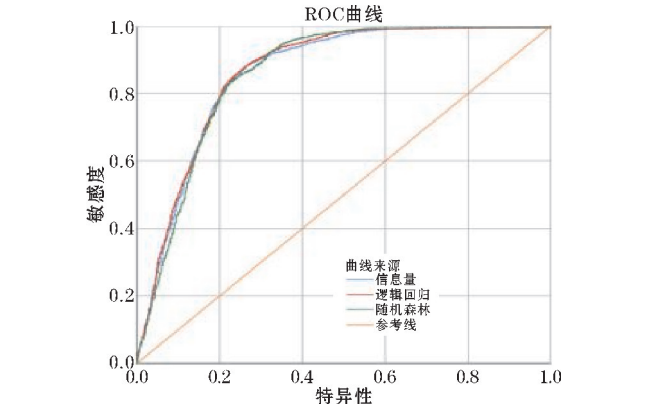


图 11 ROC 曲线验证结果

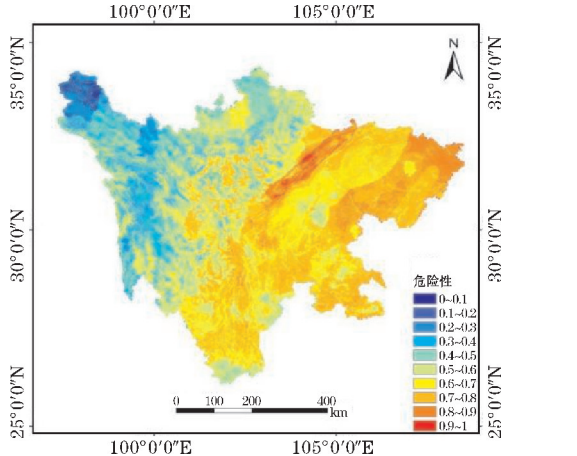


图 12 信息量模型危险性评价结果

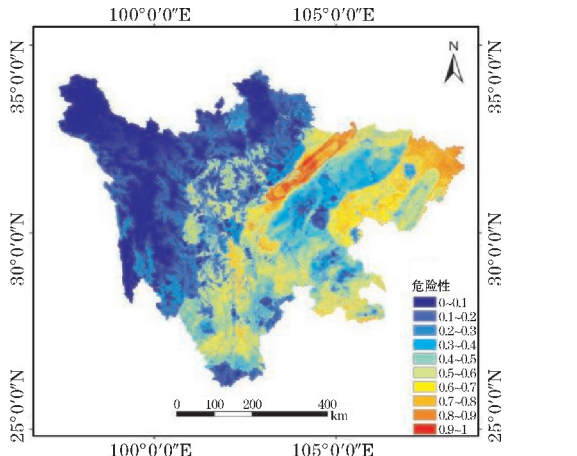


图 13 逻辑回归模型危险性评价结果

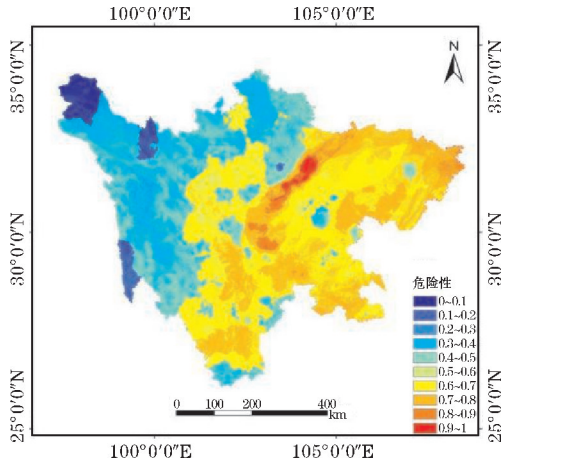


图 14 随机森林模型危险性评价结果

4 结论

使用坡度、年累计降雨量、地震烈度、断层密度、植被覆盖率5种因子作为四川省滑坡灾害危险性的评价因子,并通过信息量模型、逻辑回归模型及随机森林模型进行建模和对比后得出以下结论:

(1)对于滑坡灾害,3种模型的评估效果均较好,其权重占比最高的3个因子均为年累计降雨量、地震烈度和坡度,不同的是逻辑回归模型坡度权重占比最高,随机森林模型为年总降雨量占权重最高。从实际效果看,逻辑回归的模型精度高于其他两种模型,比较适合四川省的滑坡灾害危险性评价。

(2)以逻辑回归模型为例,按照现有评判标准,0~0.4为低危险性,0.4~0.7为中危险性,0.7~1为高危险性,四川省中低危险性占总面积的60.33%,中危险性为30.45%,高危险性为9.22%,总体上看,以龙门山地震带为界,西部较低,东部较高,在龙门山地震带上危险性较高,另外在成都平原危险性均为0~0.1,说明无危险性。

参考文献:

- [1] 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.
- [2] 赵艳南,牛瑞卿.基于证据权法的滑坡危险性区划探索[J].地理与地理信息科学,2010,26(6):19-23.
- [3] 李亭,田原,郭伦,等.基于随机森林方法的滑坡灾害危险性区划[J].地理与地理信息科学,2014,30(6):25-30.
- [4] Xiao C C, Tian Y, Shi W Z, et al. A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of Shenzhen [J]. Science China (Technological Sciences),

2010,53(S1):75-84.

- [5] 司康平,田原,汪大明,等.滑坡灾害危险性评价的3种统计方法比较——以深圳市为例[J].北京大学学报(自然科学版),2009,45(4):639-646.
- [6] 王丽婷,丁晓青,方驰.基于随机森林的人脸关键点精确定位方法[J].清华大学学报(自然科学版),2009,49(4):543-546.
- [7] 陈朝亮.基于GIS的内江市地质灾害易发性评价研究[D].绵阳:西南科技大学,2019.
- [8] 张晓东.基于遥感和GIS的宁夏盐池县地质灾害风险评价研究[D].北京:中国地质大学,2018.
- [9] 叶帮苹,冯汉中,刘志红,等.基于Logistic模型的四川山洪流域危险性评价[J].成都信息工程大学学报,2020,35(5):573-578.
- [10] 胡德勇,李京,陈云浩,等.GIS支持下滑坡灾害空间预测方法研究[J].遥感学报,2007(6):852-859.
- [11] Tian Y, Xiao C C, Liu Y, et al. Effects of raster resolution on landslide susceptibility mapping: A case study of Shenzhen [J]. Science in China (Series E: Technological Sciences), 2008(S2):188-198.
- [12] 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[J].成都理工学院学报,2001(1):89-92.
- [13] 柏廷臣,王劲峰.结合多分类器的遥感数据专题分类方法研究[J].遥感学报,2005(5):555-563.
- [14] 唐伟,周志华.基于Bagging的选择性聚类集成[J].软件学报,2005(4):496-502.
- [15] 许向宁.高地地震烈度区山体变形破裂机制地质分析与地质力学模拟研究[D].成都:成都理工大学,2006.
- [16] 唐晓玲,钱江澎.四川省丹巴县建设街滑坡特征及形成机理[J].四川地质学报,2013,33(3):340-343.

A Comparative Study on Landslide Hazard Risk Assessment Models in Sichuan Province

LIU Yabin¹, LIU Yunteng², ZHU Zhenmei², LIU Zhihong⁴, FAN Zhigao¹, SHI Hongren¹

(1. College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China; 2. Rong County Meteorological Bureau, Zigong 643100, China; 3. College of Management, Chengdu University of Information Technology, Chengdu 610225, China; 4. College of Resources and Environment, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: To explore the distribution of landslide disasters in Sichuan Province and compare the modeling effect of landslide hazard risk assessment, this paper takes Sichuan Province as the research area to analyze the temporal and spatial distribution of geological disasters, especially landslide disasters, and uses three methods: information quantity method, information-based logistic regression method, and random forest method, to establish a risk evaluation model based on five factors: slope, vegetation coverage, seismic intensity, fault density and annual cumulative rainfall. This paper combines the ROC curve and AUC value to accurately compare the model performance. The results show that: (1) The slope factor has the greatest contribution to landslide disasters, and the three factors of the slope, seismic intensity and rainfall in the three evaluation models have a greater impact. (2) Landslide disasters are mainly distributed near seismic zones, and the probability of disasters is the smallest in the Chengdu Plain. (3) The three evaluation methods are suitable for landslide hazard zoning and have good modeling effects. Among the three models, the logistic regression model is more suitable for the risk assessment of landslide disasters in Sichuan Province under actual circumstances.

Keywords: landslide disasters; informative approach; random forest approach; logistic regression methods; model evaluation