

2018年度重庆市出版专项基金资助项目

DOI:10.11721/cqnuj20190324

三峡地区资源环境生态研究

山地斜坡灾变韧性大数据分析模型与应用研究^{*}

——以城口县为例

孙德亮^{1,2,3}, 吴健平^{1,2}, 文海家⁴, 薛蒙蒙⁴

(1. 华东师范大学 地理信息科学教育部重点实验室; 2. 华东师范大学 地理科学学院, 上海 200241;
3. 重庆师范大学 地理与旅游学院, 重庆 401331; 4. 重庆大学 土木工程学院, 重庆 400045)

摘要:【目的】山地斜坡因抵抗变形破坏的能力不足而常遇变形破坏致灾,因此需要对山地斜坡抵抗变形破坏的能力进行评价。【方法】采用典型实例,构建山地斜坡灾变韧性评价模型。选取高程、坡向、归一化植被指数、坡位、微地貌、岩性、坡度、曲率等13个因子作为山区斜坡灾变韧性的评价因子,结合研究区638个历史滑坡点,建立山地斜坡韧性评价地理空间数据库;进一步在地理空间数据库中,分别随机选取70%的灾变区域与非灾变区作为训练数据集,剩下的30%作为验证数据集;对样本数据进行逻辑回归模型训练,利用训练好的逻辑回归模型,对整个研究区斜坡的灾变韧性进行仿真分析,划分为低、较低、中、较高、高5种韧性等级。【结果】对训练数据集、验证数据集以及区域仿真结果,采用受试者工作特征曲线进行合理性检验,得到受试者工作特征曲线的下面积值分别为0.876、0.867和0.873,新近10个斜坡变形破坏点分布于中等-低韧性区域。【结论】基于历史数据挖掘分析构建的山地斜坡灾变韧性评价模型,具有较高的稳定性以及良好的斜坡灾变韧性评价可靠性。

关键词:山地斜坡;灾变韧性;地理空间数据;大数据挖掘;逻辑回归模型

中图分类号:U41;X43;X9

文献标志码:A

文章编号:1672-6693(2019)03-0064-08

全球的山地面积约占总面积的1/4,而中国山地面积则占中国陆地总面积的2/3。因复杂的地形、地质条件,加上环境、气候的变化和人类工程活动的影响,斜坡区域常发生滑坡、崩塌、泥石流等地质灾害。不同斜坡区域抵抗灾变破坏的能力如何,是确保山地环境人类工程活动与生活环境安全必需关注的问题。已有文献多从滑坡灾害的易发性角度开展相关研究。例如,将层次分析法、信息量法、支持向量机等广泛地应用于滑坡的易发性评价^[1-3]。Ashish等人^[4]利用RS和GIS采用加权评价系统研究了迪克朗河流域的滑坡易发性;徐涛等人^[5]运用GIS信息量法和MORPAS 3.0中信息量计算模块,对玉林到铁山港高速公路的崩塌、滑坡、岩溶塌陷灾害进行了危险性评价;Bahareh等人^[6]利用GIS技术和逻辑回归模型、支持向量机模型、人工神经网络模型这3种模型对伊朗马赞达兰省萨里县的多丹格分水岭地区进行滑坡易发性评价,不仅对研究区的滑坡有了定量评价,还对比了3种建模方法的优缺点。经过学习研究,当样本数据量较大时,逻辑回归模型较为简单,而且评价效果较为理想^[7-9]。

山地斜坡抵抗灾变破坏的能力,即为灾变韧性^[10-11],现有文献未见斜坡灾变韧性研究成果。山地斜坡灾变韧性受多因素耦合影响,包括地形地貌、地质条件、环境因素、工程活动影响等多方面,斜坡灾变破坏因子众多、机理复杂,一般难以得到一个可靠的数学分析模型。通过斜坡变形破坏的历史数据及其灾变因子,构建相应的地理空间数据库,采用大数据挖掘的分析方法,可以构建山地斜坡灾变韧性评价模型。

三峡库区山地面积占陆地面积的86%以上,为中国山地地质灾害发生数量最多、损失程度最大的地区之一;据不完全统计,在21世纪初就有各种类型、大小地质灾害点2万余处^[12]。随着社会经济发展与工程建设加剧,近10多年来区内斜坡灾害也不断出现新的案例。因此,论文选取库区典型的山区县域——城口县为研究区,以

* 收稿日期:2019-03-14 修回日期:2019-05-15 网络出版时间:2019-05-27 17:13

资助项目:国家重点研发计划(No. 2018YFC1505501);国家自然科学基金(No. 41807498);重庆市高技术产业技术开发专项([2016]1270)

第一作者简介:孙德亮,男,讲师,博士研究生,研究方向为GIS开发,E-mail: sundeliang2002@aliyun.com;通信作者:吴健平,男,教授,博士生导师,E-mail: jpwu@geo. ecnu. edu. cn

网络出版地址: <http://kns.cnki.net/kcms/detail/50.1165.N.20190527.1713.002.html>

90 m 精度栅格建立研究区斜坡灾变韧性评价的地理空间数据库,采用逻辑回归模型,对研究区斜坡灾变韧性进行区划评价分析。

1 研究区概况

城口县位于长江上游地区、重庆市东北部,地处渝川陕三省交界处,如图 1 所示。处于北亚热带山地气候带,具有山区立体气候的特征,年均气温 13.8℃,年均降水量 1 261.4 mm^[12]。城口县属米仓山、大巴山中山区,是中国典型的山地县域。区内位于大巴山南麓,山岭由北西向南东展布,山脉受地质构造和岩性的控制,属大巴山弧形断褶皱带,由西北—东西走向的雁列式褶皱和冲断层组成。境内计有第四系、三迭系、二迭系等 7 个系,37 个组、群的地层。

研究区北部为汉江流域的仁河水系,南部为嘉陵江流域的前河水系。地下水主要以岩溶水的形式在地下深处运动,资源总量为 6.22 亿 m³,占境内水资源总量的 20%。

研究区的复杂山地地貌,多变的岩性及褶皱、断层等地质条件,加上气候、水文环境等,为滑坡、崩塌等地质灾害的发育提供了良好的天然环境,遇到夏季降雨集中,造成研究区地质灾害频发。

2 地理空间数据库

2.1 影响因子

山地斜坡灾变韧性受多因素耦合影响,主要斜坡自身因素为地质构造、岩土体类型和结构等,外部因素为水文条件、气候条件和人类工程活动等,诱发因素主要有降雨、地震和人类工程活动等。例如 Reichenbach 等人^[13]分析了 1983—2016 年 565 篇涉及滑坡易发性研究论文,用于滑坡易发性评价的因子共有 596 个,其中常用的有 23 个,可归为地质、水文、土地覆盖、地貌和其他共 5 种类型。在文献[6-17]的基础上,结合研究区具体情况^[14],选取高程、坡向、归一化植被指数(NDVI)、坡位、微地貌、岩性等 13 个因子作为斜坡灾变韧性评价因子。

2.2 数据来源

上述 13 个斜坡灾变韧性评价因子,基础的地形数据为 ASTER 卫星 30 m 精度数字高程模型,岩性与断层等地质数据采用 1:20 万地质图道路与河流数据来自于国家基础地理数据,降雨数据来源于国家气象信息网,NDVI 采用 NASA 的 MODIS 遥感成品数据;历史斜坡破坏数据采用研究区 2016 年之前地质灾害资料。

2.3 数据处理

研究区高程、坡度、坡向、坡位、微地貌、曲率 5 个因子是由数字高程模型(DEM)经 GIS 软件处理获取。倾坡类型则可利用坡向与地层倾向叠加分析经重分类获取。均岩性、断层从 1:20 万地质图提取。对坡向、坡位、微地貌、倾坡类型、岩性等 5 个非定量因子量化处理并重分类。对断层、道路、河流按 200 m 梯度多级缓冲且重分类,其他 5 个定量指标直接采用原始数据,处理结果见表 1。

为降低数据离散性,对 13 个因子归一化处理,对因子线性变换,使数值归一到[0,1]区间,归一化公式为:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

式中: X* 为归一化后的数据; X 为原始数据; X_{min} 为数据最小值; X_{max} 为数据最大值。

所有的 13 个因子均转化为 90 m×90 m 栅格数据,建立城口县山地斜坡灾变韧性评价因子的地理空间数据库,栅格单元为 386 968 个。研究区中 13 个山区地带韧性评价因子专题数据,如图 2 所示。

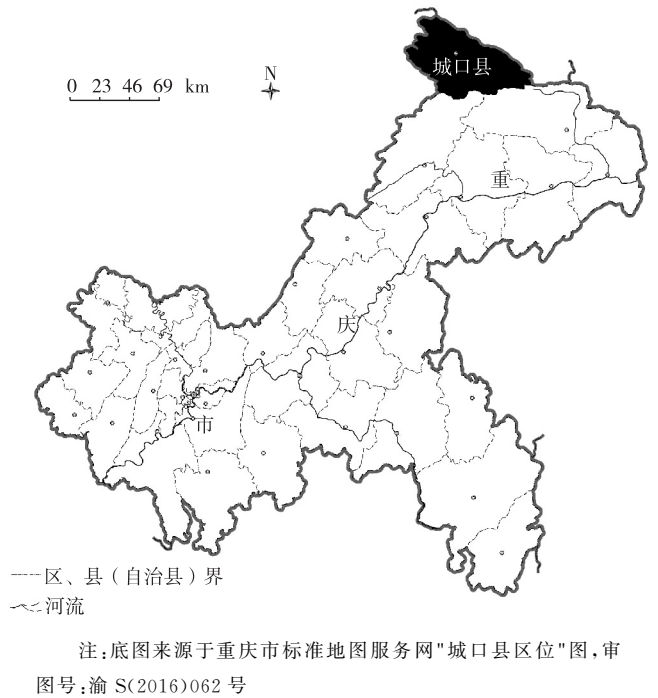


图 1 城口县位置图
Fig. 1 Location of study area, Chengkou County

表 1 山区斜坡韧性影响因子分类

Tab.1 Impact factors classification of mountain slope resilience

影响因子	分级数量	分类标准
高程/m	6	1. <500; 2. ≥500~1 000; 3. ≥1 000~1 500; 4. ≥1 500~2 000; 5. ≥2 000~2 500; 6. ≥2 500
坡向/(°)	9	1. 平面; 2. 北; 3. 东北; 4. 东; 5. 东南; 6. 南; 7. 西南; 8. 西; 9. 西北
NDVI	9	1. <2000; 2. ≥2 000~3 000; 3. ≥3 000~4 000; 4. ≥4 000~5 000; 5. ≥5 000~6 000; 6. ≥6 000~7 000; 7. ≥7 000~8 000; 8. ≥8 000~9 000; 9. ≥9 000
坡位	6	1. 山谷; 2. 下坡; 3. 平坡; 4. 中坡; 5. 上坡; 6. 山脊
微地貌	10	1. 峡谷、深流; 2. 中坡处水系、浅谷; 3. 高地水系、水源; 4. U 型山谷; 5. 平原; 6. 空旷斜坡; 7. 上斜坡、台地; 8. 局部山谷中的山脊; 9. 在平原中坡处山脊, 小山; 10. 山顶, 山脊高处
岩性	7	1. Q_{bz1} ; 2. $P, \epsilon_3, \epsilon_2, \epsilon_1$; 3. T_{1d-j}, N_{h2}, O ; 4. T_{1d} ; 5. T_{ij}, S ; 6. Z, δ, β_{μ} ; 7. S_2
坡度/(°)	8	1. <10; 2. ≥10~20; 3. ≥20~30; 4. ≥30~40; 5. ≥40~50; 6. ≥50~60; 7. ≥60~70; 8. ≥70
曲率	5	1. <10; 2. ≥10~20; 3. ≥20~30; 4. ≥30~40; 5. ≥40~50; 6. >50
倾坡类型	5	1. 顺向; 2. 切向外倾; 3. 切向; 4. 逆向; 5. 水平地形
距道路/m	16	1. <200; 2. ≥200~400; 3. ≥400~600; 4. ≥600~800; 5. ≥800~1 000; 6. ≥1 000~1 200; 7. ≥1 200~1 400; 8. ≥1 400~1 600; 9. ≥1 600~1 800; 10. ≥1 800~2 000; 11. ≥2 000~2 200; 12. ≥2 200~2 400; 13. ≥2 400~2 600; 14. ≥2 600~2 800; 15. ≥2 800~3 000; 16. ≥3 000
距水系/m	16	1. <200; 2. ≥200~400; 3. ≥400~600; 4. ≥600~800; 5. ≥800~1 000; 6. ≥1 000~1 200; 7. ≥1 200~1 400; 8. ≥1 400~1 600; 9. ≥1 600~1 800; 10. ≥1 800~2 000; 11. ≥2 000~2 200; 12. ≥2 200~2 400; 13. ≥2 400~2 600; 14. ≥2 600~2 800; 15. ≥2 800~3 000; 16. ≥3 000
距断层/m	11	1. <1000; 2. ≥1 000~2 000; 3. ≥2 000~3 000; 4. ≥3 000~4 000; 5. ≥4 000~5 000; 6. ≥5 000~6 000; 7. ≥6 000~7 000; 8. ≥7 000~8 000; 9. ≥8 000~9 000; 10. ≥9 000~10 000; 11. ≥10 000
降雨量/mm	7	1. <900; 2. ≥900~1 000; 3. ≥1 000~1 100; 4. ≥1 100~1 200; 5. ≥1 200~1 300; 6. ≥1 300~1 400; 7. ≥1 400

3 边坡灾变韧性区划评价

3.1 逻辑回归模型 LR

LR 模型是一种多元统计分析模型,该模型适用于多变量控制,可处理二分类问题,预测变量与结果的关系利用最大似然法来构建^[17-18]。在 LR 模型中,山区斜坡抵抗灾变破坏的能力和自变量之间的关系为:

$$P=\frac{1}{1+e^{-Y}}, \tag{2}$$

$$Y=C_0+C_1 X_1+C_2 X_2+\cdots+C_{13} X_{13}, \tag{3}$$

其中 P 代表山地斜坡抵抗灾变的能力,即韧性大小,输出范围从 0 到 1。当值为 1 时,表示该点的韧性很好,对灾变的抵抗能力好,斜坡未产生变形、破坏;值为 0 则表示该点的斜坡韧性差,灾变抵抗能力低,发生了斜坡破坏^[19-20]。 Y 是各评价因子的加权线性组合, C_0 为模型的截距, $C_i(i=1,2,\cdots,13)$ 是相关性系数, n 为斜坡韧性评价因子的数目, $X_i(i=1,2,\cdots,13)$ 代表前述 13 个评价因子。

由于本文选取了 13 个影响因子,影响因子之间可能存在着高度相关关系,即多重共线性问题。若不对此加以处理,会导致模型失真或难以准确预测。因此,在模型计算之前,对所有影响因子进行多重共线性诊断。容忍度(Tolerance)和方差膨胀因子(VIF)是两个常用的共线性诊断指标,二者互为倒数。当 VIF 值大于 10(即容忍度小于 0.1),表明模型存在很强共线性问题^[21-22]。利用 SPSS 统计软件可获得本文 13 个影响因子的共线性诊断结果,结果表明本文所有影响因子的 VIF 值均小于 10,即所有影响因子之间均不存在多重共线性。

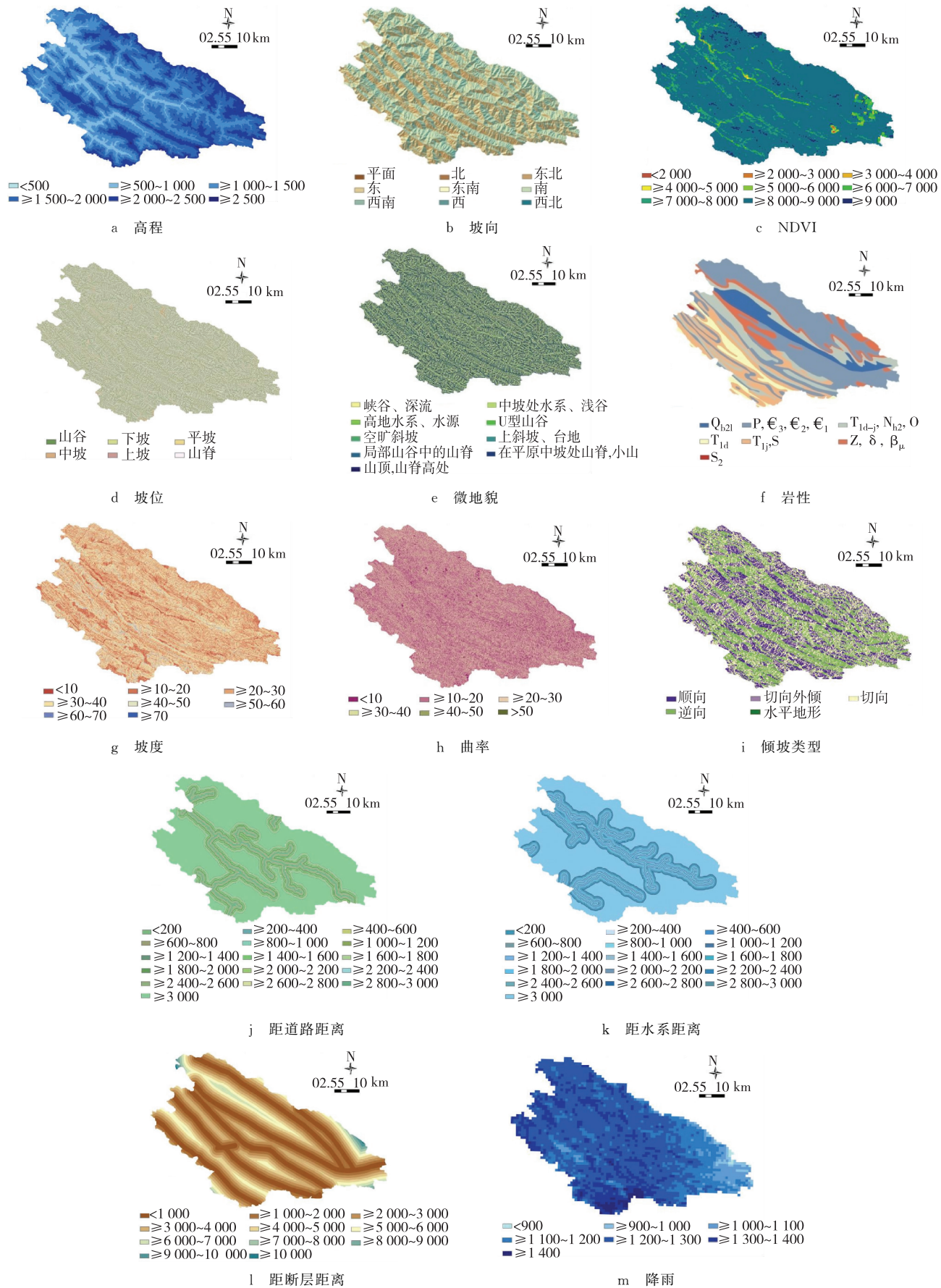


图 2 斜坡灾害韧性评价因子示意图

Fig. 1 The thematic layers of resilience assessment factors

3.2 训练模型

据截止 2016 年地灾管理部门的数据,研究区实际斜坡发生斜坡破坏灾害共 638 处,与之相对应的栅格单元为 607 个,选取这 607 个栅格单元点作为斜坡发生脆性破坏、抵抗灾变能力为 0 的样本。由于实际上,斜坡受到地灾的影响,发生不等程度的破坏是在一定的范围内,因此对地灾点进行 600 m 缓冲,认为地灾点 600 m 以外的区域未受到地灾影响发生变形,赋值为 1。随机选取破坏点与未受影响区 70% 的数据构建训练数据集,包含 425 个栅格破坏点,包含 270 543 个未受影响栅格;另外的 30% 为测试数据集,其中斜坡破坏点验证集包含 182 个栅格,未受影响区集包含 115 908 个栅格。在未受影响区的样本点随机选取 3 000 个与训练的破坏点合成一个训练样本,总数为 3 425 个栅格。选取 10 组样本进行逻辑回归训练,经统计取均值后得到模型截距和相关性系数见表 2。

表 2 评价因子的相关性系数
Tab.2 Correlation coefficients of impact factors

评价因子	相关系数	评价因子	相关系数	评价因子	相关系数
截距(C ₀)	-1.560 62	微地貌(C ₅)	-0.529 93	距道路距离(C ₁₀)	1.841 76
高程(C ₁)	4.721 22	岩性(C ₆)	0.496 59	距水系距离(C ₁₁)	0.648 51
坡向(C ₂)	-0.259 80	坡度(C ₇)	-0.400 58	距断层距离(C ₁₂)	1.318 05
NDVI(C ₃)	2.645 77	曲率(C ₈)	-0.219 38	多年平均降(C ₁₃)	-2.503 14
坡位(C ₄)	0.075 50	倾坡类型(C ₉)	0.035 47		

最终得到的逻辑回归模型的各评价因子的线性组合 Y 为:
$$Y = -1.560\ 62 + 4.721\ 22X_1 - 0.259\ 80X_2 + 2.645\ 77X_3 + 0.075\ 50X_4 - 0.529\ 93X_5 + 0.496\ 59X_6 - 0.400\ 58X_7 - 0.219\ 38X_8 + 0.035\ 47X_9 + 1.841\ 76X_{10} + 0.648\ 51X_{11} + 1.318\ 05X_{12} - 2.503\ 14X_{13}$$
 (4)
其中 $X_i(i=1,2,\cdots,13)$ 分别为前述 13 个因子所在地理空间栅格单元的数值。将(4)式代入(2)式,即为山地斜坡灾变韧性评价的逻辑回归模型。

斜坡灾变与否可以视为二分类问题,常采用混淆矩阵来分析模型的预测精度。即将无灾变看作正类(positive)、有灾变视为负类(Negative),实例是正类且被预测为正类,称为真正类(True positive, TP);实例是负类但被预测为正类,称之为假正类(False positive, FP);如果实例是负类且被预测为负类,称为真负类(True negative, TN);实例是正类但被预测成负类,则称之为假负类(False negative, FN)。将本文的训练样本数据代入逻辑回归模型中,计算得到样本数据的抵抗灾变破坏的能力,值用 P 表示。以 0.5 为阈值,P 大于 0.5,表示抵抗灾变破坏能力大,预测结果为无灾变;P 小于 0.5,表示抵抗灾变破坏能力小,预测结果有灾变。从表 3 可以看出,训练样本的总体预测精度达到 90.45%,效果较好。

3.3 研究区韧性区划与检验

将上述训练获得的逻辑回归模型应用于研究区地理空间数据库,仿真计算得到研究区山地斜坡抵抗灾变破坏的能力 P,经反复试算,按专家经验法,将其划分为 5 个等级,依次为很低($P < 0.55$)、低($0.55 \leq P < 0.80$)、中($0.80 \leq P < 0.93$)、高($0.93 \leq P < 0.98$)、很高($0.98 \leq P$),即可在 GIS 中得到山地斜坡灾变韧性等级评价图,见图 3。

表 3 观测值与预测值的对比表
Tab.3 Comparison of observed and predicted values

	预测		总计	精度/%
	1	0		
实际	1	2 890(TP) 110(FN)	3 000(TP+FN)	96.33
	0	217(FP) 208(TN)	425(FP+TN)	48.94
合计			3 425	90.45

注:1 代表正类,无灾变;0 代表负类,有灾变发生

表 4 是经统计分析得到的斜坡灾变韧性 5 个等级分布区及历史斜坡因地灾发生破坏点数据分布表,可以看

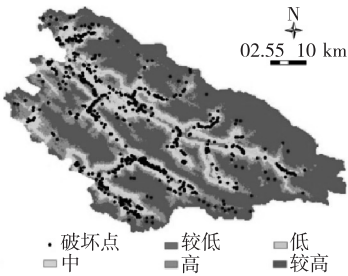


图 3 山区斜坡灾变韧性评价图
Fig.3 The spatial distribution of historical slope damage vs. damage resistance mapping of mountain slope

出随着韧性等级的提高,该等级面积呈递增趋势,而且破坏点在逐渐减少,密度也随之减少;其中,低-较低韧性区面积仅占 19.45%,但历史破坏灾害点占 70.68%,而较高-高韧性区面积占 64.17%,历史灾害破坏点仅占 13.68%。

应用受试者工作特征曲线(Receiver operating characteristic, ROC 曲线)对评价结果进行检验。ROC 曲线下面积(Area under curve, AUC 值)可以定量的表示模型预测的准确率^[23-24]。本文利用训练和验证数据集交叉验证。结果如图 4 所示。研究区训练数据、验证数据、全体数据集验证结果的 AUC 值分别为 0.876,0.867 和 0.873,三者比较接近,说明该模型具有良好的稳定性和可靠性。

表 4 斜坡灾变韧性分级统计

Tab.4 The statistical of resilience classes
zones of mountain slope

韧性等级	栅格单元/个	百分比/%	破坏点/个	破坏点占比/%	破坏点密度
低	4 226	7.14	252	41.52	0.059 6
较低	7 280	12.31	177	29.16	0.024 3
中	9 688	16.38	95	15.65	0.009 8
较高	17 458	29.51	61	10.05	0.003 5
高	20 504	34.66	22	3.63	0.001 1

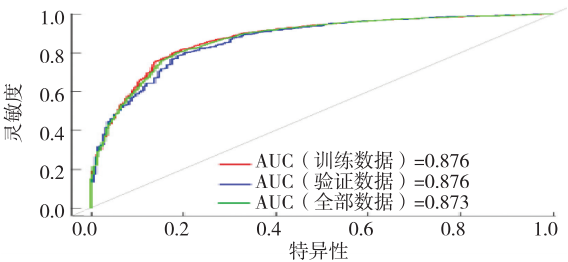


图 4 训练数据、验证数据和区域仿真的 ROC 曲线

Fig.4 The ROC curve of the training dataset, the validation dataset and the simulated results

4 实例分析

据现场调查,2017—2018 上半年,城口范围内新增斜坡或变形破坏点 10 个,如表 5 所示。

表 5 新增地质灾害点统计

Tab.5 The detail of 10 new slope deformation/damage events of the study area

序号	地灾点名称	发生时间	面积/ ($\times 10^4\text{ m}^2$)	体积/ ($\times 10^4\text{ m}^3$)	威胁 人口/人	威胁 房屋/间	威胁财产/ 万元	稳定状况	备注
1	风竹坪滑坡	2017-09-27	0.8	5	25	10	200	稳定性较差	加强监测和治理
2	廖家河坝	2017-09-27	0.5	1	250	200	2 000	稳定性较差	加强监测和治理
3	不稳定斜坡 黄家院子	2017-09-27	0.8	3.2	38	24	200	稳定性差	搬迁避让
4	滑坡								
4	林家湾不 稳定斜坡	2017-09-28	0.84	2.5	50	39	100	潜在不稳定	群测群防、工程 治理、立警示牌
5	柏树包不 稳定边坡	2017-09-28	0.96	6	6	12	100	潜在不稳定	群测群防、工程 治理、立警示牌
6	庄子上滑坡	2017-04-07	2.96	40	101	171	1 600	稳定性差	工程治理
7	沙塘不 稳定斜坡	2017-02-21	3.6	18.05	44	40	1 000	稳定性差	工程治理
8	青树湾滑坡	2017-08-06	3	60	20	30	200	稳定性差	工程治理
9	肖家院子 滑坡	2018-04-15	0.2	0.6	102	80	300	稳定性差	工程治理
10	狮子岩危岩	2018-01-16	0.04	0.39	52	66	330	稳定性差	工程治理

通过表 5 的分析可知,新增的地质灾害点大部分分布在斜坡韧性等级低、较低、中的地带。其中庄子上滑坡、狮子岩危岩分布在韧性中等地带,廖家河坝不稳定斜坡、黄家院子滑坡、林家湾不稳定斜坡、柏树包不稳定边坡、庄子上滑坡、青树湾滑坡位于韧性等级较低,沙塘不稳定斜坡、肖家院子滑坡位于斜坡韧性较低的斜坡地带。可以看到近两年来发生的斜坡失稳大都发生在韧性较差的地带,表明仿真结果与实际的斜坡破坏的空间位置是重叠的,山区斜坡灾变韧性模型具有较强的预测能力。

5 结论

1) 山区斜坡灾变韧性的 大小决定了在遭受地灾时斜坡对地灾的抵抗能力,经过研究可以确定韧性是受到高程、坡度、坡位、坡向、曲率、岩性等 13 个因子影响的,选取这些因素构建地理空间数据库,可构建山区斜坡灾变韧性的分析模型。

2) 运用统计分析和 ROC 曲线对逻辑回归模型进行检验,训练数据集、验证数据集以及区域仿真的 ROC 曲线 AUC 值分别为 0.876,0.867 和 0.873;综合表明该模型具有较高的可靠性与稳定性。

3) 近两年新发生的斜坡灾变也集中在韧性中-低的地带,这表明基于历史数据挖掘分析构建的山地斜坡灾变韧性评价模型,具有较高的稳定性,以及良好的斜坡灾变韧性评价可靠性预测功能。

参考文献:

- [1] XIE P, WEN H J, MA C C, et al. Application and comparison of logistic regression model and neural network model in earthquake-induced landslides susceptibility mapping at mountainous region, China[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 501-523.
- [2] WEN H J, XIE P, XIAO P, et al. Rapid susceptibility mapping of earthquake-triggered slope geohazards in Lushan County by combining remote sensing and the AHP model developed for the Wenchuan earthquake[J]. Bulletin of Engineering Geology and the Environment, 2016, 73(3): 1-13.
- [3] ARNONE E, FRANCIPANE A, NOTO L V, et al. Strategies investigation in using artificial neural network for landslide susceptibility mapping: application to a Sicilian catchment[J]. Journal of Hydroinform, 2014, 16(2): 502-515.
- [4] PANDEY A, DABRAL P P, CHOWDARY V M. Landslide hazard zonation using remote sensing and GIS: a case study of Dikrong river basin, Arunachal Pradesh, India[J]. Environmental Geology, 2008, 54(7): 1517-1529.
- [5] 徐涛, 张峰, 张桂林, 等. 基于 GIS 的玉林-铁山港公路地质灾害危险性评价[J]. 桂林理工大学学报, 2012, (01): 55-62.
- XU T, ZHANG F, ZHANG G L, et al. GIS-based risk assessment for geological disasters of Yulin-Tieshangang highway[J]. Journal of Guilin University of Technology, 2012, 32(1): 55-62.
- [6] KALANTAR B, PRADHAN B, NAGHIBI S A, et al. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 49-69.
- [7] 文海家, 胡东萍, 王桂林. 汶川县地震滑坡易发性 LR 与 NN 评价比较研究[J]. 土木工程学报, 2014, 47(S1): 17-23.
- WEN H J, HU D P, WANG G L. A comparative study on the susceptibility mapping of earthquake triggered landslide by neural network and logistic regression model, Wenchuan County[J]. China Civil Engineering Journal, 2014, 47(S1): 17-23.
- [8] 谢朋, 文海家, 胡东萍. 山区公路沿线地震滑坡易发性区划研究[J]. 中国公路学报, 2018, 31(02): 106-114.
- XIE P, WEN H J, HU D P. Research on susceptibility mapping of earthquake-induced landslides along highway in mountainous region[J]. China Journal of Highway and Transport, 2018, 31(2): 106-114.
- [9] HARP E L, KEEFER D K, SATO H P, et al. Landslide inventories: the essential part of seismic landslide hazard analyses[J]. Engineering Geology, 2011, 122(1): 9-21.
- [10] Wikipedia: Resilience (engineering and construction) [DB/OL]. https://en.wikipedia.org/wiki/Resilience_(engineering_and_construction)
- [11] HERRERA M, ABRAHAM E, STOIANOV I. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks [J]. Water Resources Management, 2016, 30(5): 1685-1699.
- [12] 文海家, 张永兴, 柳源. 三峡库区地质灾害防减灾战略关键问题[J]. 中国地质灾害与防治学报, 2004, 15(2): 22-28.
- WEN H J, ZHANG Y X, LIU Y. Some crucial problems of controlling geological hazard in the section of the Three Gorges Reservoir [J]. The Chinese Journal of Geological Hazard and Control, 2004, 15(2): 22-28.
- [13] REICHENBACH P, ROSSI M, MALAMUD B D, et al. A review of statistically-based landslide susceptibility models [J]. Earth-Science Reviews, 2018, 180: 60-91.
- [14] 王慧, 文海家, 胡东萍, 等. 山地环境地质灾害易发性县级区划研究: 以重庆城口县为例[J]. 工程地质学报, 2015, 23(S): 578-583.
- WANG H, WEN H J, HU D P, et al. Research on geological hazard susceptibility in mountain area: a case of study in Chengkou County, Chongqing[J]. Journal of Engineering Geology, 2015, 23(S): 578-583.
- [15] 文海家, 李洋, 薛靖元, 等. 基于大数据挖掘的山区公路沿线滑坡易发性小区划[J]. 自然灾害学报, 2018, 17(4): 159-165.
- WEN H J, LI YA, XUE J Y, et al. Landslides susceptibility microzoning along highway in mountainous region based on mining the big data[J]. Journal of Natural Disasters, 2018, 17(4): 159-165.
- [16] WANG L J, GUO M, SAWAD A K, et al. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence

- and artificial neural network[J]. *Geosciences Journal*, 2016, 20(1):117-136.
- [17] WANG H, WEN H J, HU D P, et al. Research on geological hazard susceptibility in mountain area: a case of study in Chengkou County, Chongqing[J]. *Journal of Engineering Geology*, 2015, 23(S):578-583.
- [18] 王艳龙, 杜立志, 何赛, 等. 逻辑回归模型在边坡稳定性分析中的应用[J]. *世界地质*, 2018, 37(3):945-951.
- WANG Y L, DU L Z, HE S, et al. Application of logistic regression model in slope stability analysis[J]. *Global Geology*, 2018, 37(3):945-951.
- [19] REGMI N R, GIARDINO J R, Mc DONALD E V, et al. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado[J]. *Landslides*, 2014, 11(2):247-262.
- [20] JENNINGS B J, VUGRIN E D, BELASICH D K. Resilience certification for commercial buildings: a study of stakeholder perspectives[J]. *Environment Systems & Decisions*, 2013, 33(2):184-194.
- [21] AYYUB B M. Systems resilience for multihazard environments: definition, metrics, and valuation for decision making[J]. *Risk Analysis*, 2014, 34(2):340-355.
- [22] 高惠璇. 处理多元线性回归中自变量共线性的几种方法: SAS/STAT 软件(6.12)中 REG 等过程增强功能的使用[J]. *数理统计与管理*, 2000, 20(5):49-55.
- GAO H X. Some method on treating the collinearity of independent variables in multiple linear regression[J]. *Statistics and Management*, 2000, 20(5):49-55.
- [23] 李亨, 田原, 邬伦, 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.
- LI T, TIAN Y, WU L, et al. Landslide susceptibility mapping using random forest[J]. *Geography and Geo-Information Science*, 2014, 30(6):25-30.
- [24] 王冬燕. Logistic 回归与决策树分类效能的 ROC 曲线比较[J]. *智能计算机与应用*, 2014, 4(5):34-36.
- WANG D Y. The ROC curves comparing of classification performance between logistic regression and decision tree[J]. *Intelligent Computer And Applications*, 2014, 4(5):34-36.

Resources, Environment and Ecology in Three Gorges Area

Damage Resistance Mapping of Mountain Slopes Based on Geospatial Big Data Mining

SUN Deliang^{1,2,3}, WU Jianping^{1,2}, WEN Haijia⁴

(1. Key Laboratory of Geographic Information Science, East China Normal University;

2. School of Geography Science, East China Normal University, Shanghai 200241;

3. College of Geography and Tourism, Chongqing Normal University, Chongqing 401331;

4. School of Civil Engineering, Chongqing University, Chongqing 400045, China)

Abstract: [Purposes] Due to the insufficient ability to resist deformation and damage, disasters often occur at mountain slopes. However, a method of assessing the ability of mountain slopes to resist deformation damages is still not available. [Methods] Therefore, using typical example, this study mainly focuses on developing a mapping model for the damage resistance of mountain slopes. To be specific, 13 assessment factors were firstly selected as the evaluation factors for the mountain slope damage resistance, which are the elevation, aspect, normalized difference vegetation index, slope position, microtopography, lithology, slope gradient, curvature, type of slope, distance from roads, distance from waters, distance from faults, and average annual rainfall. Afterwards, a geospatial database was built to assess the damage resistance of mountain slopes by combining 638 historical landslides in the study area. Meanwhile, 70% of the disaster and non-disaster zones in the geospatial databases were randomly selected as the training dataset, and the remaining 30% were used as the validation dataset. Besides, logistic regression model training was performed for the sample dataset, which contributes to the simulation analysis of the damage resistance of mountain slopes throughout the study area. Finally, the simulation results were classified into 5 resilience classes: low, relatively low, medium, relatively high, and high. [Finding] It should be noted that low and relatively low resilience zones accounted for only 19.45% of the total area but 70.68% of the historical disasters. Whereas the relatively high and high resilience zones accounted for 64.17% of the total area and only 13.68% of the historical disasters. In addition, an examination based on the receiver operating characteristic curve showed that the area under curve values of the training dataset, validation dataset, and regional simulation results were 0.876, 0.867, and 0.873, respectively. It is found that over the last 1.5 years, the 10 deformation damage locations were mainly distributed in the relatively low and low resilience zones. [Conclusions] Thus it can be concluded that the present assessment model for damage resistance of mountain slopes based on a data mining analysis of historical data has high stability and reliability in the assessment of mountain slope damage resistance.

Keywords: mountain slopes; damage resistance; geospatial data; data mining; logistic regression model