

一种基于地理环境相似度的滑坡负样本可信度度量方法

缪亚敏^{1,2,3}, 朱阿兴^{1,2,3,4,5*}, 杨琳⁴, 白世彪^{1,2,3}, 刘军志^{1,2,3}

(1. 南京师范大学虚拟地理环境教育部重点实验室, 南京 210023; 2. 江苏省地理环境演化国家重点实验室培育建设点, 南京 210023; 3. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023; 4. 中国科学院地理科学与资源研究所, 资源与环境信息系统国家重点实验室, 北京 100101; 5. 威斯康辛—麦迪逊大学地理系, 美国 麦迪逊 53706)

摘要: 滑坡负样本在统计型滑坡危险度制图中具有重要作用, 能抑制统计模型对滑坡危险度的高估。当前滑坡负样本采样方法采集的负样本可信度未知, 在负样本采样过程中, 极有可能将那些潜在滑坡点错选为负样本, 这些假的负样本会降低负样本集的质量和训练样本集的质量, 进而影响统计模型的精度。本文基于“地理环境越相似、地理特征越相似”的地理学常识, 认为与正样本有着相似地理环境的点极有可能是未来发生滑坡的点; 与正样本的地理环境越不相似的点, 则越有可能是负样本。基于此假设提出一种基于地理环境相似度的负样本可信度度量方法, 将该方法应用于滑坡灾害频发的陇南山区油房沟流域, 对油房沟进行滑坡负样本可信度评价制图; 使用油房沟流域的滑坡发生初始面来验证该方法的有效性。结果发现: 滑坡发生初始面上所有栅格点的负样本可信度平均值为0.26, 超过95%的栅格点的负样本可信度都小于0.5, 说明本文提出的负样本可信度度量方法合理。

关键词: 负样本可信度; 地理环境; 相似度; 滑坡危险度制图

1 引言

滑坡危险度(landslide susceptibility)是指在地形地貌、地质环境、覆被等基础孕灾条件下某一区域发生滑坡的可能性(Guzzetti et al, 1999; Guzzetti et al, 2005; Guzzetti et al, 2006), 滑坡危险度评价不考虑“极端气候条件”、“地震活动”等外部诱发因素, 不涉及如“什么时候”或“什么频率”等滑坡时间概率问题, 不涉及滑坡量级问题(Dai et al, 2001; 尹志华, 2011)。通过对滑坡危险度进行评价制图, 推测滑坡发生的潜在可能性, 有针对性地实施相关措施可以有效减少滑坡带来的损失, 对地质灾害区域

的防灾减灾规划具有重要意义(Mathew et al, 2009; Feizizadeh et al, 2014)。

统计模型在滑坡危险度评价中应用最为广泛, 它从训练样本集中获取滑坡危险度与影响滑坡发生的基础孕灾环境要素之间的关系, 将这种关系应用到整个研究区, 实现滑坡危险度空间分布的推测与制图。统计型滑坡危险度推测模型根据其训练样本集的构成可以分为只使用正样本的模型(presence-only method)和使用正负样本的模型(presence-absence method)(Zaniewski et al, 2002; Guo et al, 2005)。只使用正样本的模型, 如双变量统计模型等(Süzen et al, 2004a, 2004b; Schicker et al, 2012),

收稿日期: 2016-02; 修订日期: 2016-05。

基金项目: 国家自然科学基金项目(41431177, 41471178); 江苏省高校自然科学研究重大项目(14KJA170001); 国家重点基础研究发展计划(973计划)项目(2015CB954102) [Foundation: National Natural Science Foundation of China, No.41431177, 41471178; Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, No.14KJA170001; National Basic Research Program of China (973 Program), No.2015CB954102]。

作者简介: 缪亚敏(1991-), 女, 江苏泰州人, 硕士研究生, 主要从事滑坡危险度评价研究, E-mail: miaopaper@163.com。

通讯作者: 朱阿兴(1962-), 男, 浙江长兴人, 教授, 博士生导师, 主要从事地理信息系统、人工智能、模糊数学理论集成及其应用研究, E-mail: axing@njnu.edu.cn。

引用格式: 缪亚敏, 朱阿兴, 杨琳, 等. 2016. 一种基于地理环境相似度的滑坡负样本可信度度量方法[J]. 地理科学进展, 35(7): 860-869. [Miao Y M, Zhu A-X, Yang L, et al. 2016. A method for quantifying the reliability of landslide pseudo-absence samples based on geographic environmental similarity[J]. Progress in Geography, 35(7): 860-869.]. DOI: 10.18306/dlkxjz.2016.07.007

其训练样本集只包含滑坡点(presence samples, 本文简称为“正样本”)。使用正负样本的模型,如多变量统计模型等(Nefeslioglu et al, 2008; Mathew et al, 2009; Das et al, 2010; 方苗等, 2011; Xu, Dai et al, 2012; Xu, Xu et al, 2012),其训练样本集中包含滑坡点与非滑坡点(absence samples, 本文简称为“负样本”)。以往研究发现,只使用正样本的模型会高估滑坡危险度,即将本应是低危险区推测为高危险区(Zaniewski et al, 2002; Engler et al, 2004; Yao et al, 2008),使用正负样本的模型能较为合理地推测滑坡危险度,约束滑坡高危险区的过度扩张。研究表明,负样本的存在可有效抑制滑坡危险度推测值的高估,控制所推测滑坡高危险区的扩张(Chefaoui et al, 2008),对滑坡危险度推测结果的合理性有重要影响。

训练样本集的质量对统计型滑坡危险度制图的精度有着至关重要的影响(Kirkby et al, 1987; Atkinson et al, 1998; Dai et al, 2001)。训练样本集中,正样本以历史滑坡编目数据为依据,在滑坡发生区域进行采样(Süzen et al, 2004b; Nefeslioglu et al, 2008),其质量较高,能充分反映滑坡发生的典型地理环境条件(Xiao et al, 2010);负样本无法直接获取,只能在没有发生过滑坡的地区生成“虚拟”负样本(pseudo-absence samples)(Hirzel et al, 2001; Hirzel et al, 2002),现有“虚拟”负样本采样方法未对所采集负样本的可信度进行定量评价,无法很好地为采集有效的负样本提供依据,采集的负样本中极有可能包含潜在滑坡点(假的负样本),这些点与滑坡点的地理环境相似,只是因为以往缺少诱发条件而未发生滑坡。这些假的负样本会降低训练样本集的质量,进而影响统计型滑坡危险度制图的精度

(Iverson et al, 1998; Guisan et al, 1999; Guo et al, 2005; Yao et al, 2008)。因此,有必要对负样本的可信度进行度量,以改善所采集负样本的质量,提高统计型滑坡危险度制图的精度。

本文基于“地理环境越相似、地理特征越相似”的地理学常识(Hudson, 1992),提出一种基于地理环境相似度的负样本可信度量方法。以滑坡灾害频发的陇南山区油房沟流域为研究区,应用上述方法对油房沟进行滑坡负样本可信度空间分布推测,并使用该研究区的滑坡发生初始面来验证本文方法的有效性。

2 研究思路

统计型滑坡危险度评价制图是在由影响因素构成的环境特征空间中对正负样本集进行分类,获得滑坡危险度与影响因素之间的关系,然后将这种关系从环境特征空间映射到地理空间,实现对滑坡危险度空间分布的推测与制图。因此,为避免在地理空间与环境特征空间之间的反复映射,是否可直接在环境特征空间中采集负样本呢?

根据“地理环境越相似、地理特征越相似”的地理学常识,认为同一类样本在环境特征空间中是靠近的、相邻的,不同类样本在环境特征空间中是分离的。在这样的认知下,与正样本有着相似地理环境的点极有可能是未来发生滑坡的点;与正样本在环境特征空间中越不相似的点就越有可能是非危险区,负样本可在这样的地区采集。基于这一基本假设可对负样本的可信度进行度量,认为某点与正样本在环境特征空间中越不相似,其作为负样本的可信度越高。研究思路如图1。

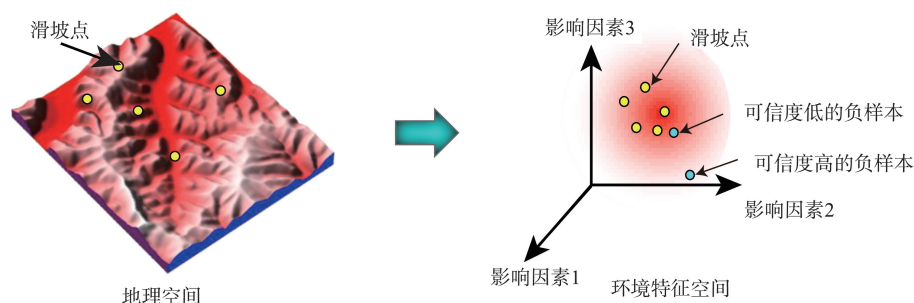


图1 基于地理环境相似度的负样本可信度量方法研究思路

Fig.1 Principle of the method for quantifying the reliability of pseudo-absence samples based on geographic environmental similarity

3 研究方法

3.1 基于地理环境相似度的负样本可信度度量方法

滑坡在某种环境条件下出现越频繁,说明这种环境条件更适合滑坡的发生,因此可以认为,这种环境条件是孕育滑坡的典型地理环境(Clerici et al, 2002; Qi et al, 2003; Zhu et al, 2015)。与滑坡发生的典型地理环境越接近或越相似的地方,越可能发生滑坡;与滑坡发生的典型地理环境越不相似的地方,则越不容易发生滑坡,在这样的区域采集的负样本可信度越高。基于这一思路对负样本的可信度进行度量。

3.1.1 地理环境相似度量

选择能描述滑坡孕灾地理环境的影响因素,将遴选出的影响因素分为2种类型:离散型和连续型。对于离散型影响因素(如岩性),使用滑坡发生频率比法(frequency ratio)表达离散型影响因素与滑坡发生频率之间的关系,如式(1)(Yilmaz, 2009; Mohammady et al, 2012);然后将影响因素与滑坡发生频率的关系归一化,以度量各离散型影响因素与该因素下发生滑坡的典型类别之间的相似度,如式(2)。对于连续型影响因素(如高程),使用核密度估计表达连续型影响因素与滑坡发生频率的关系,如式(3)(张桂铭等, 2013; Zhu et al, 2015);然后将影响因素与滑坡发生频率的关系归一化,以度量各连续型影响因素与该因素下发生滑坡的典型值之间的相似度,如式(6)。

频率比法是一种单因子定量分析模型,通过对样本分类别计算频率,可大致判断出某一离散型影响因素的哪些类别对滑坡发生产生很大影响,是一种常见的双变量统计模型。本文使用频率比法定量刻画离散型影响因素与滑坡发生频率之间的关系,公式如下:

$$S_{ij}' = \frac{p_{ij}/A_{ij}}{\sum_{j=1}^m p_{ij}/A} \quad (1)$$

式中: S_{ij}' 表示影响因素 i 的类别 j 中滑坡发生的频率; p_{ij} 表示影响因素 i 的类别 j 中滑坡发生的频数; A_{ij} 表示影响因素 i 下类别 j 的面积; m 表示影响因素 i 的类别数; A 表示研究区的总面积。

对影响因素 i 的类别 j 中滑坡发生频率进行归一化,即可得到离散型影响因素 i 中类别 j 与该影响因素 i 下发生滑坡的典型类别之间的相似度,归

一化表达式如下:

$$S_{ij} = \frac{S_{ij}'}{\max(S_{ij}')} \quad (2)$$

式中: S_{ij} 表示影响因素 i 的类别 j 与影响因素 i 下发生滑坡的典型类别之间的相似度。

核密度估计是一种由样本估计总体的概率密度函数的方法(Porter et al, 2012)。一方面,核密度估计从样本本身出发来获取数据的分布特征,以估计总体的概率密度函数,不需要引入对数据分布的先验假设,因此用来估计的概率密度函数的形状是任意的;另一方面,相比于直方图等其他密度估计方法,通过核密度估计得到的概率密度函数更为平滑连续,并且不会依赖于选取的区间长度(张桂铭等, 2013)。设有 n 个滑坡点,其所在的影响因素 x 的值为: $x_1, x_2, x_3, \dots, x_n$, 则对影响因素 x 的核密度估计基本表达式如下:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x-x_i}{h}\right) \quad (3)$$

式中: $f(x)$ 为影响因素 x 与滑坡发生频率之间关系的概率密度函数; $k(\cdot)$ 为核函数; h 为带宽,其值会影响核密度估计曲线的形状和平滑度; $x-x_i$ 是影响因素 x 与滑坡点所在的影响因素 x 的值 x_i 之间的差距。以经常使用的核函数之一——高斯核函数来估计核密度曲线,采用大拇指“rule of thumb”法则(Silverman, 1986)来计算带宽 h , 公式如下:

$$k\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}} \quad (4)$$

$$h = \sigma \left(\frac{4}{3n}\right)^{0.2} \quad (5)$$

式中: σ 是 n 个滑坡点所在影响因素 x 值的标准差。

对计算得到的概率密度函数进行归一化,即可得到单个影响因素与该因素下发生滑坡的典型值之间的相似程度,归一化表达式如下:

$$S_x = \frac{f(x)}{f_{\max}(x)} \quad (6)$$

式中: S_x 表示影响因素 x 与该因素下滑坡发生典型值的相似度; $f_{\max}(x)$ 表示 $f(x)$ 的最大值。

通过综合单个影响因素对该因素下滑坡发生典型值的相似度,可以度量与滑坡发生典型地理环境的综合相似度,公式如下:

$$S = f(s_1, s_2, \dots, s_k, \dots, s_n) \quad (7)$$

式中: S 表示综合地理环境相似度; s_k 表示第 k 个影响因素与该因素下滑坡发生典型值的相似度; f

表示综合方法,可以通过平均值法、最小限制因子法、线性加权法进行综合。

基于以上公式,即可推算出研究区每个栅格点与滑坡发生典型地理环境的相似度,生成与滑坡发生典型地理环境相似度的空间分布图。

3.1.2 负样本可信度度量

基于“地理环境越相似、地理特征越相似”的地理学常识,认为与滑坡点地理环境相似的点极有可能是潜在滑坡点,这样的点若被选为负样本,其可信度较低。例如,某栅格点与滑坡点的地理环境相似度为0.8,则该点极有可能是潜在滑坡点,不宜选作负样本,其负样本可信度应较低。由于与滑坡点的地理环境相似度的值域范围是 $[0, 1]$,本文使用式(8)度量负样本可信度,负样本可信度值域为 $[0, 1]$,值越大,负样本可信度越高;经计算,该栅格点的负样本可信度为0.2。

$$Reliability_{ij} = 1 - S_{ij} \quad (8)$$

式中: S_{ij} 是 (i, j) 位置上的栅格点与滑坡点的地理环境相似度; $Reliability_{ij}$ 是该栅格点被选为负样本的可信度。

基于本文提出的负样本可信度度量方法,以正样本和影响因素为输入数据,可计算研究区中每一栅格点被选为负样本的可信度,从而得到研究区的负样本可信度空间分布图。

3.2 验证方法

滑坡发生初始区域是指滑坡的发源区,不包含滑坡运移和沉积区,是斜坡稳定性遭到破坏、发生变形的区域(Nefeslioglu et al, 2008),滑坡发生初始

区域发生滑坡的可能性大,滑坡危险度高,因此滑坡发生初始区域中一般不宜进行负样本采样,其负样本可信度应较低。本文通过滑坡发生初始区域的负样本可信度验证基于地理环境相似度的负样本可信度度量方法的有效性。

4 应用案例

4.1 研究区与滑坡编目数据

油房沟流域位于中国陇南山区,面积49.74 km²,地势北高南低,平均高程2000 m以上,区内地势起伏较大,河谷深切,平均坡度20°以上(图2)。流域位处青藏高原、黄土高原和四川盆地之间,地质构造复杂,地震活动频发。区内泥盆系、志留系地层分布广泛,岩性多为千枚岩、板岩、粉砂岩、泥岩等。该区域亚热带季风气候显著,气候温暖潮湿,受到山地地形效应的影响,全年降水量达到400~900 mm(谌文武等, 2006),降水多集中于5-9月,且以暴雨的形式出现。受特殊地质环境和气候环境的影响,境内滑坡灾害发生频繁,造成极大的人员伤亡,严重制约着该区工农业生产和社会经济的发展(Bai et al, 2012)。

根据室内遥感解译与野外滑坡检核识别本研究区中的滑坡体,使用滑坡发生初始面的顶部中心点来代表滑坡灾害点,既可获得滑坡发生前的地形地貌条件,也可兼顾不同大小的滑坡(Atkinson et al, 2011),还可避免因滑坡边界的不确定性而带来的误差。基于这一滑坡点采样方法,油房沟流域共

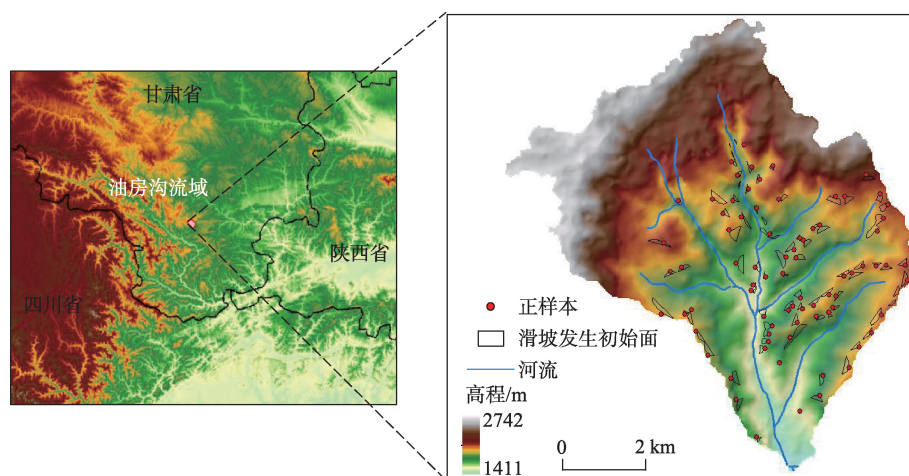


图2 油房沟流域地理位置和滑坡正样本分布图

Fig.2 Location of the Youfang catchment and the spatial distribution of landslide presence samples

识别出79个滑坡正样本(图2)。

4.2 影响因素

根据研究区的地质环境特征和前人已有研究成果,选取高程、坡度、坡向、平面曲率、剖面曲率、距河流的距离、距道路的距离、岩性、距断层线的距离、土地利用类型等10个影响因素(表1和图3)(Bai et al, 2015),用于定量描述研究区内的滑坡孕灾环境。

对研究区1:50000地形图数字化,生成30 m分辨率的数字高程模型(Digital Elevation Model, DEM),在ArcGIS10.1软件中基于DEM派生出一系列地形因子,包括:高程、坡度、坡向、平面曲率、剖面曲率,以描述研究区的地形地貌条件。对研究区1:50000地形图数字化,获得矢量格式的河流和道路数据,河流和道路对滑坡的影响是根据距其距离表达。对研究区1:25000地质图数字化,获得地层分布和断层构造数据。同时代的地层可由不同成

因的岩石类型组成,根据相似的物质组成和物理力学特征,可将研究区的地质层组合并划分为3类岩性单元,第一类为粉砂岩、泥岩、薄砂砾岩岩组,第二类为厚石灰岩、板岩岩组,第三类为千枚岩、板岩、薄石灰岩岩组。断层构造对滑坡的影响是以距

表1 影响因素与数据源
Tab.1 Predisposing factors of landslide and data sources

类别	影响因素	数据源	比例尺
地形地貌	高程	地形图	1:50000
	坡度		
	坡向		
	平面曲率		
	剖面曲率		
排水状况	距河流的距离	地质图	1:25000
人类活动	距道路的距离		
地质条件	岩性		
	距断层线的距离	TM影像	
土地利用	土地利用类型		

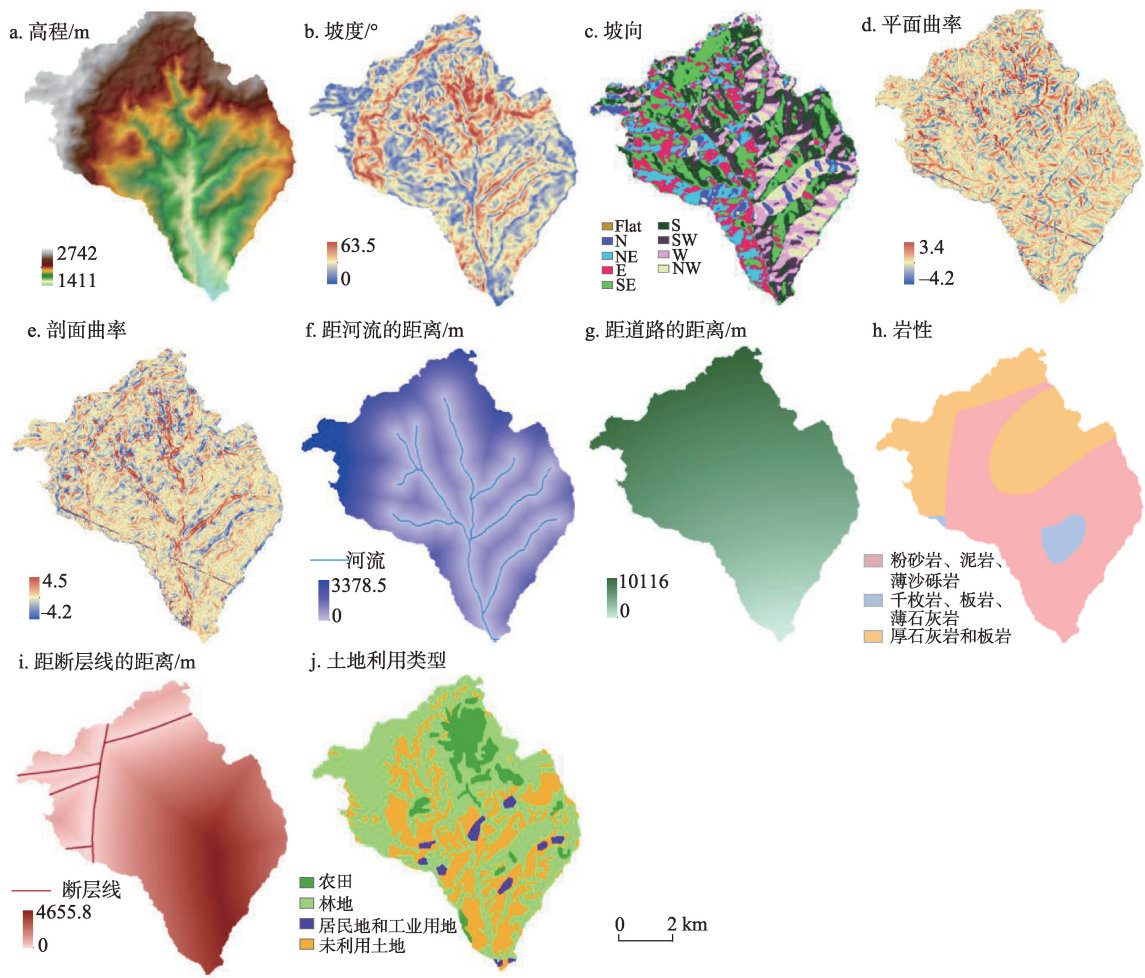


图3 油房沟流域滑坡影响因素数据
Fig.3 Predisposing factors of landslide in the Youfang catchment

断层线的距离表达。对研究区的TM影像进行遥感解译,获得土地利用类型,并将其划分为农田、林地、居民地与工业用地、未利用土地4类。

本研究区中不同比例尺的数据源会派生出不同分辨率的影响因素数据,为方便空间分析与计算,需要保证影响因素数据尺度的一致性。考虑到研究区的尺度和数据源的尺度,选择30 m作为所有影响因素数据的分辨率,既能充分利用高精度的地形图,又能兼顾低精度的遥感影像。

5 结果与分析

5.1 负样本可信度的空间分布推测

根据研究区历史滑坡编目采集的79个滑坡灾害点和遴选出的滑坡影响因素,定量化表达各影响因素与该因素下发生滑坡的典型值之间的相似度。对于离散型影响因素,其相似度关系如图4;对于连续型影响因素,其相似度关系如图5。影响因素与该因素下发生滑坡的典型值越相似,滑坡发生可能性越大。从图4-5上可以看出,油房沟流域较容易发生滑坡的区域具有如下环境特征:海拔约在1800~2000 m之间,坡度约在10°~25°之间,坡向为西坡或西北坡,平面曲率和剖面曲率约在-0.2~0.2之间,距河流的距离约在300~500 m之间,距道路的距离约为3500~5500 m之间,距断层线的距离约为2000~4000 m之间,岩性为千枚岩、板岩、薄石灰岩,土地利用类型为未利用土地、居民地和工业用地。

在研究区中,滑坡易发生区域多为山体斜坡,坡度较陡,相对高差较大,可形成临空面;岩性多为较软的千枚岩、板岩、薄石灰岩等,可为滑坡的形成

提供大量的物质基础;土地利用类型多为居民地和未利用荒地,地表涵养水源能力弱,容易被侵蚀,较容易发生滑坡。滑坡是各种内外影响因素综合作用的产物,不同区域地质环境中发育的滑坡,主导因素不同,地形地貌、排水状况等对滑坡发生影响的程度也不同。进一步研究发现,地形地貌条件、岩性、土地利用类型等对滑坡发生有着更大的影响,而距道路、河流、断层线的距离等因素对滑坡的影响力相对较小,导致滑坡没有发生在常识所认为的离河流、道路很近的区域。

通过综合各个影响因素对该因素下发生滑坡的典型值的相似度,获得整个研究区中每个栅格点与易于发生滑坡的地理环境的综合相似度。基于地理环境相似度,推算整个研究区中每个栅格点被选为负样本的可信度,从而得到研究区的负样本可信度空间分布图(图6)。图中负样本可信度的值域为[0, 1],值越大说明该点被选为负样本时,其可信度越高。

基于图6发现,研究区中推测的负样本可信度较低的区域位于河流沿岸的山体斜坡面上,从地形地貌上分析,该区域坡度较陡,相对高差较大,且多分布于河流沿岸,河流不断地掏蚀坡脚会促使临空面生成;从地质条件上分析,该区域的岩性多为较软的粉砂岩、泥岩、薄砂砾岩、薄石灰岩等,为滑坡的形成提供大量的物质基础;从土地利用类型上分析,该区域多为居民地和未利用荒地,地表涵养水源能力弱,容易被侵蚀,因此较容易发生滑坡。研究区中推测的负样本可信度较高的区域主要位于研究区北部,从地形地貌上分析,该区域海拔较高,地形较为平坦,相对高差较小,不易形成临空面;从地质条件上分析,该区域的岩性多为较硬的厚石灰

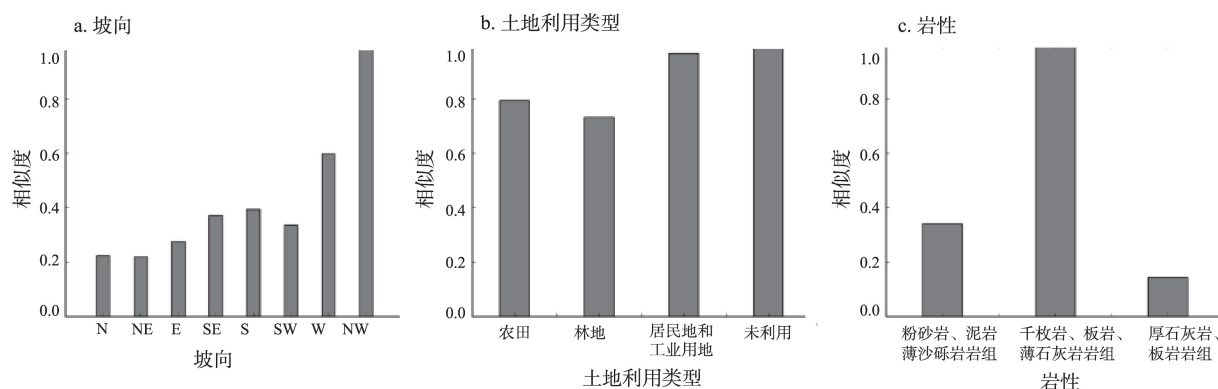


图4 各离散型影响因素与该因素下发生滑坡的典型类别之间的相似度关系图

Fig.4 Similarity between discrete predisposing factor types and the typical type under which landslides occurred

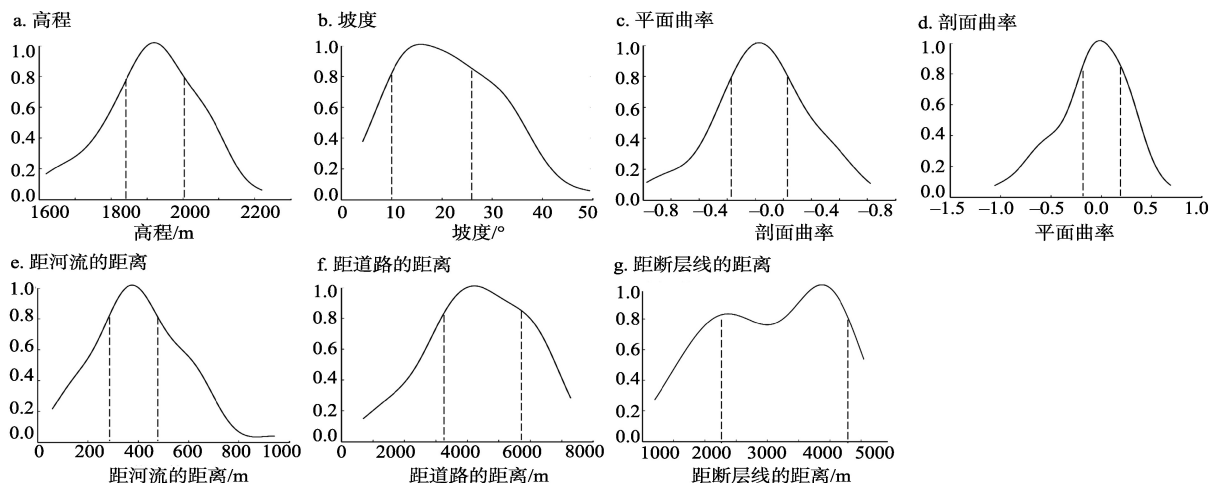


图5 各连续型影响因素与该因素下发生滑坡的典型值之间的相似度关系图

Fig.5 Similarity between continuous predisposing factor values and the typical value under which landslides occurred

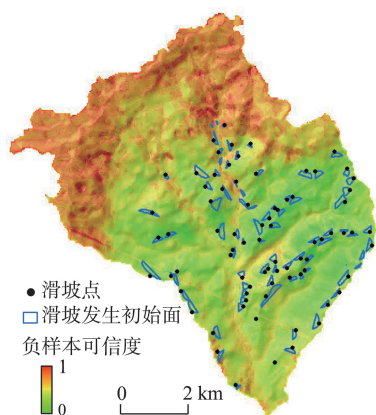


图6 基于地理环境相似度的负样本可信度空间分布图

Fig.6 Spatial distribution of the reliability of pseudo-absence samples based on the geographic environmental similarity

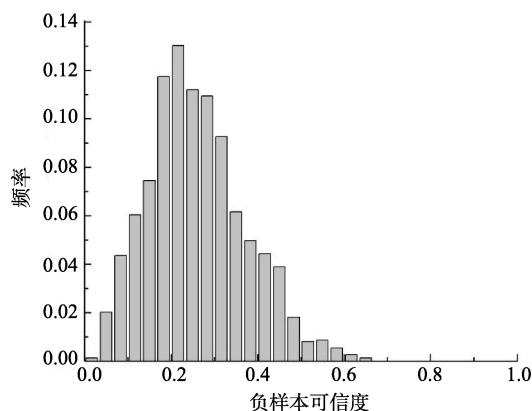


图7 滑坡发生初始区域的负样本可信度频率分布图

Fig.7 Frequency histogram of the reliability of grids in the landslide initial zones chosen as pseudo-absence samples

岩、板岩等,岩土体强度较大;从土地利用类型上分析,该区域多为林地,土壤保水能力强,较难被侵蚀,因此发生滑坡的可能性较小。

5.2 结果验证

滑坡发生初始区域发生滑坡的可能性大,滑坡危险度高,因此滑坡发生初始区域中不宜进行负样本采样,该区域的负样本可信度应较低。统计滑坡发生初始区域中1493个栅格点上的负样本可信度发现(图7),所有栅格点上的负样本可信度平均值为0.26,标准差为0.11,超过95%的栅格点的负样本可信度都小于0.5,这说明滑坡发生初始区的负样本可信度推测值都较低,表明本文提出的负样本可信度度量方法有效。

6 结论与讨论

针对当前滑坡危险度制图中采集负样本的可信度未知这一问题,基于“地理环境越相似、地理特征越相似”的地理学常识,本文提出了一种基于地理环境相似度的滑坡负样本可信度度量方法,实现对滑坡负样本可信度的直观定量认识。将本文方法应用于滑坡灾害频发的陇南山区油房沟流域,对其进行滑坡负样本可信度空间分布推测。滑坡发生初始面的滑坡危险度较高,一般不宜进行负样本采样,其负样本可信度应较低。使用油房沟流域的滑坡发生初始面的负样本可信度来验证滑坡负样本可信度度量方法的有效性,结果发现,滑坡发生

初始面中超过95%的栅格点的负样本可信度都小于0.5,说明本文方法是一种有效的负样本可信度度量方法。

由于本文提出的方法仍处在初步实验探索阶段,还需要大量的研究工作来完善,具体包括以下2个方面:

(1) 影响因素的选择。影响因素的选择关系到对滑坡孕灾地理环境的表达能力。一方面,影响因素选择的完备程度会影响对滑坡孕灾地理环境的描述程度,不同的滑坡类型有不同的主导影响因素,不同区域地质环境中发育的滑坡,地形地貌、土地利用等对其影响的程度也不同。因此,需要针对研究区的地质环境特征选择适宜适量的影响因素来描述滑坡发生的地理环境条件。另一方面,影响因素的选择要与研究区的尺度相适应。在不同尺度下,选择可以反映研究区滑坡孕灾环境的影响因素。通常,在中小型流域的研究尺度下,地形地貌、地质条件、土地利用、排水状况等是滑坡发生的重要孕灾环境。本文以影响因素对滑坡的重要性、影响因素的可量化表达程度、数据源的质量和可获取性等作为基本依据,根据以往研究中广泛使用的影响因素和研究区的地质环境特征来选择研究区的滑坡发生影响因素,有一定的主观性,因此有必要对影响因素的选择进行深入研究。

(2) 统计型滑坡危险度制图的应用与验证。本文对负样本可信度度量方法进行探索,并以滑坡发生初始面作为验证样本对该方法进行了初步验证。在下一步研究中,需要将提出的负样本可信度度量方法应用于负样本采样,并在基于统计模型的滑坡危险度推测中进行验证。此外,负样本可信度未知这一问题在很多研究领域同样存在(如物种生境适宜度制图),因此,本文提出的方法还可以在更多不同研究领域中进行应用与检验,以便对其加以改进。

参考文献(References)

湛文武, 赵志福, 刘高, 等. 2006. 兰州—海口高速公路甘肃段工程地质问题研究[M]. 兰州: 兰州大学出版社: 19-22. [Chen W W, Zhao Z F, Liu G, et al. 2006. Lanzhou-Haikou gaosu gonglu Gansuduan gongcheng dizhi wenti yanjiu[M]. Lanzhou, China: Lanzhou University Press: 19-22.]

方苗, 张金龙, 徐琪. 2011. 基于GIS和Logistic回归模型的兰

州市滑坡灾害敏感性区划研究[J]. 遥感技术与应用, 26(6): 845-854. [Fang M, Zhang J L, Xu Z. 2011. Landslide susceptibility zoning study in Lanzhou City based on GIS and logistic regression model[J]. Remote Sensing Technology and Application, 26(6): 845-854.]

尹志华. 2011. 基于RS和GIS技术对区域滑坡进行高效快速敏感性评价的模型研究: 以北川县为例[D]. 成都: 成都理工大学. [Yin Z H. 2011. Rapid and efficient regional landslide susceptibility assessment model based GIS and RS technology: A case study in Beichuan County[D]. Chengdu, China: Chengdu University of Technology.]

张桂铭, 朱阿兴, 杨胜天, 等. 2013. 基于核密度估计的动物生境适宜度制图方法[J]. 生态学报, 33(23): 7590-7600. [Zhang G M, Zhu A-X, Yang S T, et al. 2013. Mapping wildlife habitat suitability using kernel density estimation [J]. Acta Ecologica Sinica, 33(23): 7590-7600.]

Atkinson P M, Massari R. 1998. Generalised linear modelling of susceptibility to landsliding in the Central Apennines, Italy[J]. Computers & Geosciences, 24(4): 373-385.

Atkinson P M, Massari R. 2011. Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy [J]. Geomorphology, 130(1-2): 55-64.

Bai S B, Lu P, Wang J. 2015. Landslide susceptibility assessment of the Youfang catchment using logistic regression [J]. Journal of Mountain Science, 12(4): 816-827.

Bai S B, Wang J, Zhang Z G, et al. 2012. Combined landslide susceptibility mapping after Wenchuan earthquake at the Zhouqu segment in the Bailongjiang Basin, China[J]. Catena, 99: 18-25.

Chefaoui R M, Lobo J M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance [J]. Ecological Modelling, 210(4): 478-486.

Clerici A, Perego S, Tellini C, et al. 2002. A procedure for landslide susceptibility zonation by the conditional analysis method[J]. Geomorphology, 48(4): 349-364.

Dai F C, Lee C F, Zhang X H. 2001. GIS-based geo-environmental evaluation for urban land-use planning: A case study[J]. Engineering Geology, 61(4): 257-271.

Das I, Sahoo S, van Westen C, et al. 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India)[J]. Geomorphology, 114(4): 627-637.

Engler R, Guisan A, Rechsteiner L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data

- [J]. *Journal of Applied Ecology*, 41(2): 263-274.
- Feizizadeh B, Roodposhti M S, Jankowski P, et al. 2014. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping[J]. *Computers & Geosciences*, 73: 208-221.
- Guisan A, Weiss S B, Weiss A D. 1999. GLM versus CCA spatial modeling of plant species distribution[J]. *Plant Ecology*, 143(1): 107-122.
- Guo Q H, Kelly M, Graham C H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California[J]. *Ecological Modelling*, 182(1): 75-90.
- Guzzetti F, Carrara A, Cardinali M, et al. 1999. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central Italy[J]. *Geomorphology*, 31(1-4): 181-216.
- Guzzetti F, Reichenbach P, Ardizzone F, et al. 2006. Estimating the quality of landslide susceptibility models[J]. *Geomorphology*, 81(1-2): 166-184.
- Guzzetti F, Reichenbach P, Cardinali M, et al. 2005. Probabilistic landslide hazard assessment at the basin scale[J]. *Geomorphology*, 72(1-4): 272-299.
- Hirzel A H, Hausser J, Chessel D, et al. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data[J]. *Ecology*, 83(7): 2027-2036.
- Hirzel A H, Helfer V, Metral F. 2001. Assessing habitat-suitability models with a virtual species[J]. *Ecological Modelling*, 145(2-3): 111-121.
- Hudson B D. 1992. The soil survey as paradigm-based science [J]. *Soil Science Society of America Journal*, 56(3): 836-841.
- Iverson L R, Prasad A M. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States[J]. *Ecological Monographs*, 68(4): 465-485.
- Kirkby M J, Naden P S, Burt T P, et al. 1987. *Computer simulation in physical geography*[M]. Chichester, UK: Wiley.
- Mathew J, Jha V K, Rawat G S. 2009. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method[J]. *Landslides*, 6(1): 17-26.
- Mohammady M, Pourghasemi H R, Pradhan B. 2012. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster-Shafer, and weights-of-evidence models[J]. *Journal of Asian Earth Sciences*, 61: 221-236.
- Nefeslioglu H A, Gokceoglu C, Sonmez H. 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps[J]. *Engineering Geology*, 97(3-4): 171-191.
- Porter M D, Reich B J. 2012. Evaluating temporally weighted kernel density methods for predicting the next event location in a series[J]. *Annals of GIS*, 18(3): 225-240.
- Qi F, Zhu A-X. 2003. Knowledge discovery from soil maps using inductive learning[J]. *International Journal of Geographical Information Science*, 17(8): 771-795.
- Schicker R, Moon V. 2012. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale[J]. *Geomorphology*, 161-162: 40-57.
- Silverman B W. 1986. *Density estimation for statistics and data analysis*[M]. Boca Raton, RL: CRC Press.
- Süzen M L, Doyuran V. 2004a. A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate[J]. *Environmental Geology*, 45(5): 665-679.
- Süzen M L, Doyuran V. 2004b. Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchment, Turkey[J]. *Engineering Geology*, 71(3-4): 303-321.
- Xiao C C, Tian Y, Shi W Z, et al. 2010. A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of Shenzhen[J]. *Science China Technological Sciences*, 53(S1): 75-84.
- Xu C, Dai F C, Xu X W, et al. 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China[J]. *Geomorphology*, 145-146: 70-80.
- Xu C, Xu X W, Dai F C, et al. 2012. Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China[J]. *Computers & Geosciences*, 46: 317-329.
- Yao X, Tham L G, Dai F C. 2008. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China[J]. *Geomorphology*, 101(4): 572-582.
- Yilmaz I. 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey) [J]. *Computers & Geosciences*, 35 (6): 1125-1138.

Zaniewski A E, Lehmann A, Overton J M. 2002. Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns[J]. *Ecological Modelling*, 157(2-3): 261-280.

Zhu A-X, Zhang G M, Wang W, et al. 2015. A citizen data-based approach to predictive mapping of spatial variation of natural phenomena[J]. *International Journal of Geographical Information Science*, 29(10): 1864-1886.

A method for quantifying the reliability of landslide pseudo-absence samples based on geographic environmental similarity

MIAO Yamin^{1,2,3}, ZHU A-Xing^{1,2,3,4,5*}, YANG Lin⁴, BAI Shibiao^{1,2,3}, LIU Junzhi^{1,2,3}

(1. Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing 210023, China; 2. State Key Laboratory Cultivation Base of Geographical Environment Evolution of Jiangsu Province, Nanjing 210023, China; 3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 4. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 5. Department of Geography, University of Wisconsin-Madison, Madison WI 53706, USA)

Abstract: Landslide pseudo-absence samples play an important role in applying statistical methods for landslide susceptibility mapping. It can constrain the overestimation of landslide susceptibility. The reliability of pseudo-absence samples is unknown in existing methods of pseudo-absence sample generation in landslide susceptibility mapping. The absence samples they generate may contain sites very susceptible to landslides but have not yet failed in the past simply because of the lack of triggering factors. These false absence samples lower the quality of absence samples and in turn lower the quality of the entire training sample set, and then decrease the accuracy of the mapping using statistical methods. In this article, a new method of measuring the reliability of pseudo-absence samples is proposed. The basic principle of the method is that the more similar the geographic environment is between two locations, the closer the geographic features are between them. Accordingly, the more similar the geographic environment of a location is with landslide presence samples, the more likely is the location prone to landslides and the less reliable is this location as an absence sample. According to the inference above, the reliability of a grid as an absence sample can be calculated by the similarity of its geographic environment with the typical environmental conditions under which landslides occurred. The Youfang catchment in southern Gansu Province—an area with high risk for landslides—was used as the study area to apply the proposed method and map the reliability of each grid chosen as pseudo-absence sample. The landslide initial zones in the Youfang catchment were used to validate the effectiveness of the proposed method. The results show that the mean reliability of grids in the landslide initial zones chosen as pseudo-absence samples is 0.26 and the reliability of more than 95% of the grids is lower than 0.5. This indicates that the proposed method of measuring the reliability of pseudo-absence samples is effective.

Key words: reliability of pseudo-absence samples; geographic environment; similarity; landslide susceptibility mapping