

区域泥石流危险度评价的影响因子识别

苏鹏程^{1,2}, 倪长健³, 孔纪名^{1,2}, 汪阳春^{1,2}

(1. 中国科学院 山地灾害与地表过程重点实验室, 四川 成都 610041; 2. 中国科学院 水利部 成都山地灾害与环境研究所, 四川 成都 610041; 3. 成都信息工程学院, 四川 成都 610225)

摘要: 区域泥石流危险度评价是环境保护和减灾的重要内容。针对泥石流危险度评价的多因素影响问题,通过全面收集有代表性的泥石流样本资料,应用免疫进化算法,建立了区域泥石流危险度评价的投影寻踪聚类模型。研究表明,该模型依据样本自身的数据特性寻求最优的投影方向,并通过投影方向计算反映评价样本综合特征信息的投影特征指标。该模型不但可以确定各评价指标的权重,避免权重人为的任意性,还能在区域泥石流不同危险度级别划分的基础上,以平均贡献率的大小揭示不同影响因子对评价结果的影响,从而可清楚地筛选出其中的关键影响因子。此研究从定量的角度进一步深化了对泥石流危险度评价影响因子的认识。

关键词: 泥石流; 投影寻踪聚类; 危险度; 贡献率; 关键因子

文献标识码: A

文章编号: 1000—288X(2009)01—0128—05

中图分类号: P642.23

Factor Identification for Regional Danger Degree of Debris Flow

SU Peng-cheng^{1,2}, NI Chang-jian³, KONG Ji-ming^{1,2}, WANG Yan-chun^{1,2}

(1. Key Laboratory of Mountain Hazards and Surface Processes, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China; 2. Institute of Mountain Hazards and Environment, Chinese Academy of Sciences and Ministry of Water Conservancy, Chengdu, Sichuan 610041, China; 3. Chengdu University of Information and Technology, Chengdu, Sichuan 610225, China)

Abstract: Regional evaluation of debris-flow danger degree is an important part of environment protection and disaster reduction. Based on immune evolutionary algorithm and by collecting typical and complete samples, the projection pursuit cluster model (PPC) is established and applied to treat the multi-factor evaluation of danger degree of debris flow. Projection pursuit can project high dimensional data to low dimensional space. Through studying the main characteristics in the $1 \sim D$ space, the key information in original data can be obtained, which can not only eliminate the subjectivity of the weight of each factor, but also reveal its impacts on the evaluation according to its actual mean contribution rate, based on the classification of different regional danger degrees of debris flow. In this way, the key factors can be clearly selected. The study further deepens the quantitative understanding of the evaluation factors determining regional danger degree of debris flow.

Keywords: debris flow; projection pursuit cluster; danger degree; contribution rate; key factor

泥石流灾害遍及我国广大山区,其发生、发展受到多种不可控制的随机因素的影响。由于它对山区城镇、交通线路、能源基地、矿山、生态环境和人民生活财产的安全具有极大的破坏力,因此,如何将影响泥石流的相关因素进行量化,将相关数据通过特定的方法进行处理,进而对泥石流危险度作出评价,这不仅是泥石流研究的核心问题之一,也是环境保护和减灾对策研究的重要内容。针对影响泥石流危险度

评价的诸多因素的不完整性和不确定性,国内外很多学者进行了这方面的研究,学者们分别从地质、地貌、气候、水文条件以及人为活动等众多影响泥石流发生的因素出发,首先根据专家经验,将作用不大、难以获取的因子删去后,在剩余的影响因子中进行再次遴选,并赋予各因子一定的权重,线性叠加后进行综合评判。由于影响泥石流形成、发生的环境背景因子多达几十种,在现有评价方法中,各个学者对因子选择

各不相同^[1-7]。随着计算机技术的发展,结合 GIS 和 RS 技术,通过图象的计算机处理提取评价指标进行评价^[8,10]。文献[7]对区域泥石流危险度评价问题进行了比较系统的阐述,应用灰色关联度分析的方法确定了泥石流区域危险度评价的指标体系,并以主要因子和次要因子这样定性分析的方式对指标的重要性进行了描述。上述方法的提出拓展了泥石流危险度区划这一问题的应用研究,但是该方法对影响因子权重的确定存在经验或人为的因素,使得评价结果具有一定的任意性,另外,也未就各因子之间的相对重要性进行阐述。本研究首次将投影寻踪技术应用于区域泥石流危险度评价关键因子的识别,不但可根据样本自身的数据结构确定各指标的权重,而且还可以定量的方式找出其中的关键影响因子。深化了对区域泥石流危险度评价指标体系的认识,同时为指标体系的精简和构建更为简洁的评价模型奠定了基础。

投影寻踪技术是 20 世纪 70 年代由国际统计界发展起来的一类新型的探索性数据分析方法,其基本思路是将高维数据向低维空间进行投影,通过低维投影特征信息来研究高维数据的结构特征。由于其具有稳健性、抗干扰性和准确度高等优点,因而在优化与环境评价等领域得到了很广泛的应用^[11-15]。

1 投影寻踪聚类模型

投影寻踪是一种处理多因素复杂问题的统计方法,迄今为止,它在理论研究和实际应用中都取得了重大突破。其基本思想是:借助计算机技术将高维数据通过组合,投影到低维(1~3 维)空间上,并通过极小化某个投影指标,寻求出能反映原高维数据结构或特征的投影,在低维空间上对数据结构进行分析,以达到研究和分析高维数据的目的。投影寻踪聚类则是依据投影寻踪思想建立的综合评价模型,它可广泛应用于多因素影响的综合评价分析问题^[16-18]。

设第 i 个样本的第 j 个因素为 x_{ij}^0 ($i=1, \dots, n; j=1, \dots, m; n$ 为样本个数, m 为因素个数),建立投影寻踪聚类模型步骤如下:

(1) 数据规一化。由于各因素 x_{ij}^0 的量纲不尽相同或数值范围相差较大,因此对数据进行规一化处理:

$$x_{ij} = (x_{ij}^0 - x_{j \max}) / (x_{j \max} - x_{j \min}) \quad (1)$$

式中: $x_{j \max}, x_{j \min}$ ——表示第 j 个因子的样本最大值和最小值。

(2) 线性投影及构造综合特征值。把高维的数据信息通过投影的方法转化到低维空间,不但形象直观,而且便于运用常规的方法进行分析处理。投影实质上就是从不同的角度去观察数据,寻找能够最大程

度地反映数据特征和最能充分挖掘数据信息的最优投影方向。这里选用线性投影,即将高维数据投影到线性空间进行研究。将每个样本点 x_{ij} 作为 m 维空间的一点(或一个向量),设 $\vec{a} = (a_1, a_2, \dots, a_m)$ 为 m 维空间的某个特殊方向(单位向量),则向量 x_{ij} 与 \vec{a} 的点积

$$z_i = \sum_{j=1}^m a_j x_{ij} \quad (2)$$

即点 x_{ij} 在 \vec{a} 方向(一维空间)的投影。

(3) 构造目标函数。首先定义类间距离与类内密度:

类间距离用样本序列的投影特征值方差计算

$$s(a) = \left[\sum_{i=1}^n (z_i - \bar{z}_a)^2 / n \right]^{1/2}$$

式中: \bar{z}_a —— n 个样本的投影综合特征值在 a 方向上的均值,即 $\bar{z}_a = (\sum_{i=1}^n z_i) / n$ 。 $s(a)$ 愈大,投影散布愈开。

类内密度则定义为任意两点间的距离 $r_{ik} = |z_i - z_k|$ ($k=1, \dots, n$),并定义 $d(a) = \sum_{i=1}^n \sum_{k=1}^n (R - r_{ik}) f(R - r_{ik})$ 。其中, $f(R - r_{ik})$ 为单位阶跃函数,当 $R > r_{ik}$ 时, $f(R - r_{ik}) = 1$,反之为 0; R 为密度窗宽,其取值范围为 $\max(r_{ik}) + \frac{m}{2} R \cdot 2^{m^{18}}$ 。 $d(a)$ 愈大,分类愈显著。

由于聚类分析是根据待评价样本的数据特性将样本进行分类和评价,因此可以用分类指标作为目标函数。根据分类原则,使多元数据在一维空间散布的类间距离 $s(a)$ 和类内密度 $d(a)$ 同时取得最大值。因此,将目标函数 $Q(a)$ 定义为类间距离与类内密度的乘积,即

$$Q(a) = s(a) \cdot d(a) \quad (3)$$

(4) 优化投影方向。模型建立的关键是找到能反映系统特征的最优投影方向,根据上述分析可知,当式(3)取得最大值时所对应的 \vec{a} 就是最优投影方向向量。所以,此问题可转化为在满足一定的约束条件的情况下,求解出 $Q(a)$ 的最大值。即式(4)这样一个优化问题。函数优化的方法很多,本研究采用免疫进化算法^[19]对其进行优化。

$$\begin{cases} \max Q(a) \\ a = 1 \end{cases} \quad (4)$$

(5) 综合评价。根据优化得到的 \vec{a} ,由式(1)便可计算出反映各评价指标综合信息的投影特征量 z_i ,以 z_i 为基础进行区域泥石流危险度分析。

2 实例研究

2.1 区域泥石流危险度评价因子的选取

区域泥石流危险度评价是泥石流危险性分区的

基本依据,其核心内容是评价指标(影响因子)的选取和评价模型的建立。目前各个学者在评价指标体系的选取上还存在很大的不一致性,区域泥石流危险度评价模型总体上仍属于经验模型的范畴。为进一步分析影响区域泥石流危险度各因子的重要性,并确定各因子的贡献率,从而为评价指标体系的优选提供基础理论依据。本研究在前人研究成果的基础上,采用投影寻踪聚类模型,对影响区域泥石流危险度各因子的重要性进行了定量研究。

为便于将本方法与已有的研究结果进行比较,基于文献[7]的研究,选择我国泥石流灾害最为严重的四川省凉山州和云南省昭通地区为例,以县为基本的行政评价单元。泥石流危险度评价指标构成:(1)泥石流沟分布密度 x_1 (条/10³ km²),它表明了区域泥石流的发育历史以及目前的活动状况,同时预示着将来的发展趋势;(2)岩石风化程度系数 x_2 (取倒数),由于泥石流的固体物质大多直接来源于风化破碎后的松散岩体,故该因子能较好地反映出—个地区泥

流形成的可能性的大小;(3)断裂带密度 x_3 (km/10³ km²),它表明了一个地区地层岩石的破碎情况;(4)大于等于 25° 坡地面积百分比 x_4 (%),该因子表明地区的斜坡失稳情况;(5)洪灾发生频率 x_5 (%),选择其是因为山区洪水灾害与泥石流常常相伴发生;(6)月降雨量变差系数 x_6 ,这一因子反映了一个地区降雨量在年内各月的分配情况;(7)年平均大于等于 25 mm 大雨日数 x_7 (d),—个地区大雨日数越多,表明激发泥石流的可能性越大;(8)大于等于 25° 坡耕地面积百分比 x_8 (%),这一因子反映了人类活动对泥石流发育的影响。

上述 8 个指标的侧重点不同,但对泥石流的产生均有重要影响,其中岩石风化程度系数和断裂带密度为地质指标,泥石流沟分布密度和 25° 坡地面积百分比为地貌指标,洪灾发生频率、月降雨量变差系数和年平均 25 mm 大雨日数为水文气象指标, 25° 坡耕地面积百分比为人类活动指标。所选的样本及对应的区域泥石流危险度状况见表 1 评价因子的取值部分。

表 1 区域泥石流危险性的评价因子取值及评价结果

序号	市县名	区域泥石流危险度评价影响因子								区域危险度评价结果 ^[1]	投影特征值
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8		
1	德昌	37.6	1.8	78.7	66.0	82.0	0.98	10.7	18.5	级(极度危险)	2.032
2	宁南	31.1	1.8	115.0	54.0	56.0	0.92	10.5	7.3	级(极度危险)	1.983
3	甘洛	43.2	1.9	111.6	63.2	83.0	0.81	6.0	31.3	级(极度危险)	2.153
4	巧家	25.0	2.0	120.2	56.0	66.7	0.89	8.2	33.1	级(极度危险)	2.184
5	会东	23.6	1.9	135.3	45.3	44.0	0.98	12.0	6.9	级(高度危险)	1.850
6	普格	35.7	1.8	86.6	56.5	46.0	0.84	15.1	5.4	级(高度危险)	1.816
7	布拖	19.0	1.9	51.6	45.8	83.0	0.88	9.4	10.4	级(高度危险)	1.686
8	美姑	29.6	1.8	6.3	31.0	100.0	0.83	5.3	16.5	级(高度危险)	1.649
9	永善	25.6	2.0	46.6	52.4	62.5	0.88	4.6	15.0	级(高度危险)	1.637
10	木里	3.8	1.9	90.3	44.5	100.0	1.14	4.5	5.2	级(显著危险)	1.285
11	鲁甸	9.4	2.0	74.7	30.5	28.1	0.85	8.4	8.0	级(显著危险)	1.352
12	绥江	11.2	1.7	47.7	46.1	30.8	0.91	6.9	28.7	级(显著危险)	1.530
13	昭通市	26.7	1.9	49.8	22.1	25.0	0.88	4.4	8.8	级(显著危险)	1.342
14	水富	11.2	1.7	47.7	46.1	30.8	0.91	6.8	28.7	级(显著危险)	1.530
15	黑水	24.0	1.8	20.3	44.5	25.0	0.77	2.7	13.5	级(中度危险)	1.120
16	南坪	19.1	1.8	79.7	33.9	42.9	0.80	1.7	5.7	级(中度危险)	1.160
17	小金	19.1	1.8	2.8	48.5	14.3	0.83	1.3	22.7	级(中度危险)	0.917
18	马尔康	6.0	1.7	8.2	76.0	16.7	0.86	2.5	1.6	级(轻度危险)	0.614
19	松潘	7.9	1.8	66.0	61.6	14.3	0.71	1.4	19.7	级(轻度危险)	0.871
20	壤塘	5.5	1.7	8.1	91.7	16.7	0.93	1.8	1.3	级(轻度危险)	0.550

2.2 投影寻踪模型计算

将表 1 中 20 个县(市)的 8 个区域泥石流危险度影响因子作为样本序列 x_{ij} , $i = 20$, $j = 8$ 。由表 1 样本资料自身数据特征建立区域泥石流危险度评价的

投影寻踪聚类模型,将 20 个样本的相应指标值代入投影寻踪聚类模型进行评价模型确定,其中 $n = 20$, $m = 8$ 。采用分段函数赋值法^[9]使得不同区域的泥石流具有可比性,按照投影寻踪建模步骤,计算得到最优

投影方向向量为:

$$\vec{a} = (0.495\ 2, 0.085\ 7, 0.407\ 9, 0.027\ 6, 0.392\ 1, \\ 0.082\ 3, 0.568\ 3, 0.311\ 1)$$

于是,根据式(2)可计算各样本相应的投影特征值,结果见表 1。投影特征值与原多维数据的空间特征是一致的。因此,经投影变化计算得出的特征值的大小反映的就是危险度的大小,并且两者正相关。将此 20 个样本的投影特征值由小到大进行排序,根据相邻投影特征值之间差值的分布状况,确定以(1.950, 1.600, 1.250, 0.900, 0.550, 0.200)作为分类标准,将区域泥石流危险度划分为以下 5 个等级,即 Ⅰ级(极度危险区)、Ⅱ级(高度危险区)、Ⅲ级(显著危险区)、Ⅳ级(中度危险区)和 Ⅴ级(轻度危险区)。

按此分类标准,样本 1—4 隶属 Ⅰ级,为极度危险区;样本 5—10 隶属 Ⅱ级,为高度危险区;样本 11—

14 隶属 Ⅲ级,为显著危险区;样本 15—17 隶属 Ⅳ级,为中度危险区;样本 18—20 隶属 Ⅴ级,为轻度危险区,此分类结果与前期的泥石流危险度区划结果^[7]是完全一致的,说明该模型的计算结果合理,该方法有效可行。

3 结果分析

由(2)式可计算出每个样本各个指标的贡献率 $w_{ij} = z_{ij} / \sum_{i=1}^n z_i$ ($i=1, \dots, n$, 为类内样本个数; z_{ij} 为样本 i 的第 j 个指标的投影分量值),据此再分别求出各指标的总体和各危险度等级所对应的平均贡献率(表 2)。贡献率是源于数据自身驱动,反映的是泥石流危险度评价指标体系中各指标之间的相对重要性,而不表明该因子与泥石流发生二者之间的密切程度。

表 2 区域泥石流危险度影响因子的平均贡献率

危险度 分级	x_1 平均 贡献率	x_2 平均 贡献率	x_3 平均 贡献率	x_4 平均 贡献率	x_5 平均 贡献率	x_6 平均 贡献率	x_7 平均 贡献率	x_8 平均 贡献率	累计平均 贡献率
Ⅰ级	22.81	3.09	13.84	1.13	18.48	3.84	26.32	10.49	100
Ⅱ级	24.48	3.73	9.99	1.05	20.61	4.60	28.89	6.65	100
Ⅲ级	14.65	4.50	12.23	1.06	16.51	5.79	34.06	11.20	100
Ⅳ级	32.26	5.89	8.15	1.60	16.30	6.96	14.46	14.38	100
Ⅴ级	15.80	8.92	9.41	4.10	16.09	11.74	24.37	9.57	100
贡献率平均	21.55	4.90	10.96	1.60	17.84	6.17	26.83	10.15	100

从各影响因子的总体和分级平均贡献率来看,泥石流沟分布密度、年平均大于等于 25 mm 大雨日数以及洪灾发生频率的平均贡献率最大,3 者的总体累计值达到 66% 左右,故此 3 指标是区域泥石流危险度评价指标体系中重要影响因子,也即关键影响因子;断裂带密度和大于等于 25° 坡耕地面积百分比的平均贡献率在 10% 左右,属区域泥石流危险度评价指标体系中一般影响因子;月降雨量变差系数、岩石风化程度系数和大于等于 25° 坡地面积百分比的平均贡献率基本上小于 10%,属泥石流危险度评价指标体系中次要影响因子。重要因子、一般因子和次要因子的划分反映了不同因子对区域泥石流危险度评价影响的层次性。

3 类影响因子对区域泥石流危险度大小的影响作用不同,主要体现在平均贡献率上。重要影响因子的累计平均贡献率随区域泥石流危险度增大而明显增加,次要影响因子的累计平均贡献率随区域泥石流危险度增大而明显减小,一般影响因子的累计平均贡献率随区域泥石流危险度增大无明显的变化趋势。因此,对区域泥石流危险度评价结果起主导作用的是

主要影响因子和次要影响因子。从一定意义上讲,区域泥石流危险度的变化是评价体系中重要因子和次要因子的平均贡献率之间的转化。

就区域泥石流危险度评价指标体系的各指标物理属性而言,水文气象指标的累计平均贡献率最大,地貌指标的累计平均贡献率次之,地质指标的累计平均贡献率位居第三,人类活动指标的平均贡献率最小。这一分析结果反映了基于物理属性上的各个因子的相对重要性,其与实际情况也是吻合的。

现行的评价方法在大的地貌背景环境相似的情况下采用相同的评价因子和模型,利用投影寻踪聚类模型可以进一步分析不同区域之间评价因子对危险度影响的差异。

4 结语

将投影寻踪技术应用于区域泥石流危险度评价因子的识别,从平均贡献率的角度,以定量的方式把评价指标体系分为 3 类:重要影响因子、一般影响因子和次要影响因子,同时讨论了这 3 类因子随危险度的变化规律。以指标的物理属性进行分类,揭示了各

分类指标的重要性的排序。这一方法得出的结论其作用在于:(1)可以从评价指标自身的数据特征出发,按平均贡献率的大小对影响区域泥石流危险度的诸多因子进行排序,比较客观、定量地反映各个评价指标对于危险度的作用,为今后统一评价指标体系提供基础理论依据。(2)从构建评价模型的角度来说,通过各因子重要性的排序及各因子对区域泥石流危险度的贡献大小,可以优选出代表区域泥石流危险度的重要因子,进而对现有的评价指标体系作出一定的精简,从而构建更为简洁的数学模型。因此本研究深化了对区域泥石流危险度评价方法的认识,同时为指标体系的精简奠定了基础。

[参 考 文 献]

- [1] 刘希林. 泥石流危险度区划的探讨[J]. 灾害学, 1989, 4(4): 3-9.
- [2] 钟敦伦, 谢洪, 韦方强. 长江上游泥石流危险度区划研究[J]. 山地研究, 1994, 12(2): 65-70.
- [3] 谢又予, 伍永秋. 北京密云县泥石流危险区及沟谷危险度的初步研究[C]//北京市科学技术协会. 首都圈自然灾害与减灾对策. 北京: 气象出版社, 1991: 166-170.
- [4] Fannin R J, Rollerson T P. Debris flows: Some physical characteristics and behaviors[J]. Canadian Geotechnical Journal, 1993, 30: 71-81.
- [5] Fell R. Landslide risk assessment and acceptable risk[J]. Canadian Geotechnical Journal, 1994, (31): 261-272.
- [6] Glade T. Linking debris-flow hazard assessments with geomorphology[J]. Geomorphology, 2005, 66(14): 189-213.
- [7] 刘希林, 莫多闻. 泥石流风险评价[M]. 成都: 四川科学技术出版社, 2003.
- [8] 汤家法, 谢洪. GIS技术支持下的泥石流危险度区划研究[J]. 四川测绘, 1999, 22(3): 120-122.
- [9] 唐川, 朱静. GIS支持下的滇西北地区泥石流灾害评价[J]. 水土保持学报, 2001, 15(6): 84-87.
- [10] 刘家宏, 王光谦. 基于遥感图象的泥石流地面活动程度评价[J]. 地理科学, 2003, 23(4): 454-459.
- [11] Friedman J H, Turkey J W. A Projection Pursuit Algorithm for Exploratory Data Analysis[J]. IEEE Transactions on Computer, 1974, 23(9): 881-890.
- [12] 庄世坚. 用投影寻踪技术评价环境质量[J]. 环境保护, 2000(2): 25-26.
- [13] 金菊良, 汪明武, 魏一鸣. 用投影寻踪分类模型进行环境监测优化布点[J]. 安全与环境学报, 2004, 4(4): 10-12.
- [14] 胡庆芳, 尚松浩, 温守光, 等. 基于投影寻踪的径向基函数网络在参考蒸发量预测中的应用[J]. 水利科学, 2006, 37(9): 1151-1154.
- [15] 赵小勇, 付强, 邢贞相. 投影寻踪等级评价模型在土壤质量变化综合评价中的应用[J]. 土壤学报, 2007, 44(1): 164-168.
- [16] 倪长健, 王顺久. 边坡稳定性评价的投影寻踪聚类模型[J]. 岩石力学与工程学报, 2004, 23(16): 2687-2689.
- [17] 王顺久, 张欣莉. 投影寻踪聚类模型及其应用[J]. 长江科学院院报, 2002, 19(6): 53-55, 61.
- [18] 张欣莉, 丁晶. 投影寻踪分类模型评定相似流域[J]. 水科学进展, 2001, 12(3): 140-144.
- [19] 倪长健, 丁晶. 免疫进化算法[J]. 西南交通大学学报, 2003, 38(1): 87-91.