

基于粗糙集理论滑坡影响因子评价研究 ——以三峡库区青干河流域为例

刘吉平¹, 刘汉青¹, 曾忠平², 汪华斌³

(1. 武汉大学资源环境学院, 武汉 430079; 2. 华中科技大学公共管理学院, 武汉 430079;
3. 华中科技大学土木工程与力学学院, 武汉 430079)

摘要: 运用地理信息系统和遥感技术, 从 Landsat TM 遥感图像获取了一系列滑坡及其影响因子数据。结合粗糙集理论, 对三峡库区秭归县青干河流域滑坡发生的影响因子进行了分析, 提取了一组基本的滑坡影响条件属性因子, 并导出了基于该因子集合的判断滑坡与非滑坡的规则集。研究表明, 所选择的坡度、高程、斜坡类型、植被指数和岩石地层单元等 5 个条件属性因子对滑坡是重要的影响因素(核); 由粗糙集生成的对预测滑坡相对较有价值的 11 条决策规则中, 3 条主要决策规则可作为滑坡影响因子的评价规则。

关键词: 滑坡; 数据挖掘; 粗糙集理论; 空间信息

中图分类号: P642.22

文献标识码: A

文章编号: 1000-3665(2010)05-0118-06

对滑坡因子的判定是一个典型的地理学不确定性问题。现阶段的滑坡因子的选取和分级, 主要依靠经验判定和专家分级的定性指导定量方法, 具有很强的主观性, 而且因子之间没有建立起有机的联系, 易于导致大量的数据冗余, 也不适合大范围的计算机精确模拟、监测、预报及综合分析。对区域滑坡的空间预测, 不管采用哪种定量方法, 其评价指标体系的合理确定是一个非常棘手的问题, 也是在许多相关研究中没有解决的关键问题^[1~2]。因此, 很有必要寻求一种有效的数据分析方法, 确定合理的评价指标体系, 为实现滑坡空间科学预测提供强有力的技术支持。

由波兰学者 Pawlak 提出的粗糙集理论不需要对对象的特征或属性给出某种数学预设, 如统计学中的概率分布, 也不要人为地给出某种关系度量, 如模糊集理论中的隶属度或隶属函数, 而是直接对不完整不精确数据进行分析处理, 从数据本身的内在相互联系得到简明扼要的知识表达形式^[3~4]。由于这一重要特点, 它已经在人工智能、知识与数据发现、模式识别等方面取得了较为成功的应用尝试。将粗糙集理论引入滑坡影响因子分析评价, 将建立一种由数据本身决定决策规则的更加客观的因子评价方法。这对于传统的

经验判定和专家分级的因子选取和分级方法, 是十分有益的改进, 能够推动海量滑坡数据的计算机挖掘和智能知识发现。

本文以三峡库区秭归县青干河流域滑坡为研究对象, 运用地理信息系统和遥感技术, 从数字高程模型 (DEM) 和 Landsat TM 遥感图像获取了一系列滑坡影响因子数据, 结合当地滑坡编目资料和地貌地质条件, 选取出合理的影响因子并进行分级。用粗糙集方法研究探讨影响滑坡发生的因素及其组合规则, 并对其进行定性和定量的分析。

1 研究区概况

研究区位于三峡库区滑坡发生频度较高的典型地区——湖北省秭归县青干河与锣鼓洞河流域, 地理位置及空间范围如图 1 所示。

从图中可见, 在该流域内发育了较为密集的滑坡体(共 27 个)。这一系列的库岸滑坡绝大多数顺向坡发育, 且部分滑坡成群分布^[5]。根据三峡库区滑坡发育情况分析, 库区发生在顺向坡或疏松土质坡岩层的山地地质灾害将近 2000 个, 可见顺层基岩滑坡和堆积层—基岩接触面滑坡是库区滑坡研究的重点对象。对库岸上发育的体积在 $1.0 \times 10^6 \text{ m}^3$ 以上的 64 个大型、特大型滑坡的资料进行分析, 发现滑坡发育与三峡地区分布最广的侏罗系地层密切相关, 其中巴东组是三峡库区典型的易滑地层。发育在顺向坡上的顺层基岩滑坡的滑带的形态主要由地层岩性及产状决定。研究

收稿日期: 2009-11-24; 修订日期: 2010-03-05

基金项目: 自然科学基金资助项目(40772170)

作者简介: 刘吉平(1957-), 男, 博士, 副教授, 从事遥感科学的教学和科研工作。

通讯作者: 汪华斌, E-mail: huabin@mail.hust.edu.cn

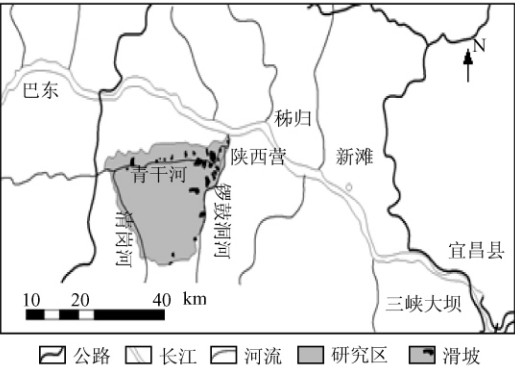


图 1 研究区地理位置及滑坡空间分布示意图
Fig.1 Geographic location of the study area showing the distribution of landslides

区的顺向坡滑坡中,地层岩性构成了滑体的物质基础,而岩层面则往往与潜在的滑动面一致,岩层产状与滑坡发育有密切的相关性^[6-7]。

2 研究方法

运用粗糙集理论和方法,从滑坡实例样本数据中挖掘出各因子对滑坡发生作用的知识或规则。本文使用由波兰 Poznan 科技大学开发的决策分析软件 ROSE (Rough Set data Explorer,ROSE2 是它的新版)进行粗糙集分析,该软件已在许多领域中得到应用^[8]。其中空间分析和因子提取还用到 ArcGIS 和 ERDAS IMAGEN 两个软件。本文中只提取规则而不对未知对象进行分类识别。

从大量样本的属性数据中获得关于样本分类(如发生滑坡单元和不发生滑坡单元)的知识,即获得进行分类的规则,其工作过程包括选择待研究对象的样本集,提取样本集的条件属性和决策属性,对属性进行约简,最后采用修正的 LEM2 算法生成对样本的判别性规则(LEM2 算法是一种用于决策的归纳算法)。工作流程如图 2 所示。

3 数据分析处理

3.1 滑坡影响因子的选择及其图层信息提取

分析研究区内滑坡发生的可能影响因素,并考虑实际样本数据的可得性,选取坡度、高程、斜坡类型、植被指数、岩石地层单元 5 种可能的影响因子。其中斜坡类型反映的是地形坡面产状与地层产状的相互关系,其计算公式为:

$$TOBIA = \cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\beta - \sigma) \quad (1)$$

式中:θ、σ——斜坡坡度、坡向(°);

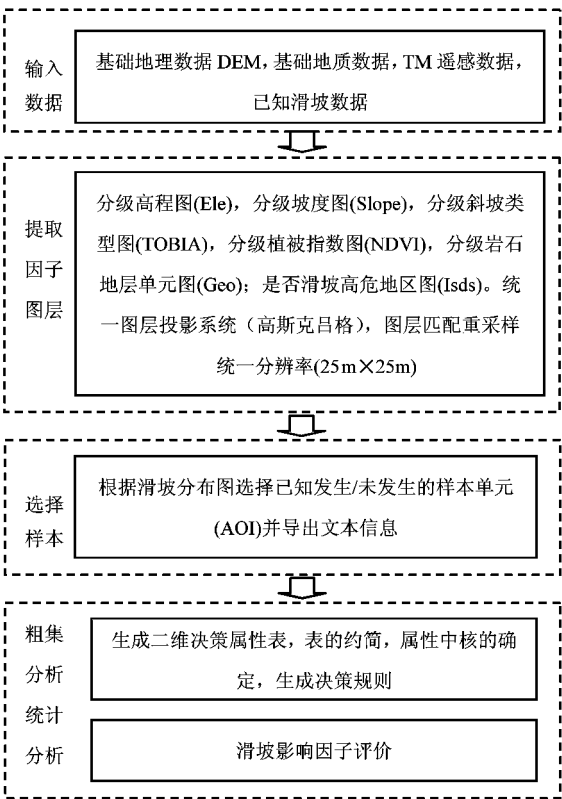


图 2 基于粗糙集的滑坡影响因子评价研究方法流程图
Fig.2 Flow chart of assessment of impact factors for landslides with rough set method

α、β——岩层倾角、倾向(°)。

一般随着 TOBIA 指数的增大(坡度产状与地层产状趋向一致),滑坡发生的可能性也随之增加。

植被指数反映了植被发育程度。本文选择应用最广的归一化植被指数并利用 Landsat5 TM 数据进行计算,公式为:

$$NDVI = (TM4 - TM3) / (TM4 + TM3) \quad (2)$$

其中 TM4、TM3 分别为 TM 的近红外波段和红光波段。

坡度、坡向信息从 DEM 数据中计算得到。岩石地层单元信息来自地质图数据。以上 5 个因子为条件属性。用于分类的决策属性则是已知地形单元是否发生了滑坡,即其样本为已知的滑坡单元和非滑坡单元,它来自现有滑坡空间分布图。岩性和滑坡信息为定性数据,其余为定量数据。TM 影像的分辨率为 30m,DEM 为 25m,其余图层数据的比例尺为 1:5 万。

对上述属性信息数据中的连续变化属性值,需要根据经验或统计分析划分为若干适当的区间,并用离散值表示。处理过程如下:

首先按 1:5 万的 DEM 的 25m 分辨率,将各图层的分辨率重采样为 25m × 25m 的栅格图层。然后将各图层的不同投影统一变换为高斯克吕格投影,并作各栅格图层的匹配。再根据文献 [5] 给出的定量属性数据的分级方案(表 1)进行分级,得到条件属性因子和决策属性因子共 6 个因子图层。

表 1 滑坡因子分级方案

Table 1 Grades of impact factors for landslides

因子	分级区间	类别	因子	分级区间	类别
滑坡状态	有	0	斜坡类型	0 ~ 639	1
	无	1		639 ~ 888	2
植被类型 (NDVI 转换为整数)	61 ~ 108	1	(TOBIA 原值 区间为 [-0.139, 1] 转换为整数)	888 ~ 989	3
	108 ~ 148	2		989 ~ 1 089	4
	> 148	3		1 089 ~ 1139	5
高程 (DEM)	80 ~ 300	1	岩石地层 单元 (GEO)	Tj1	1
	300 ~ 500	2		Tb1	2
	500 ~ 700	3		Tb2	3
	> 700	4		Tb3	4
斜坡坡度 (Slope)	0 ~ 5°	1		Tb3 + 4	5
	5° ~ 28°	2		Tb5	6
	28° ~ 38°	3		Jt	7
	38° ~ 50°	4		Jn	8
	> 50°	5			

注:根据经验,某些属性的某一值域范围对判断滑坡是没有意义的,如负值,故略去。

3.2 滑坡影响因子的粗糙集分析

通过对上述 6 个因子的粗糙集分析,找出对滑坡的发生有不同程度的影响的因子组合,从而为滑坡预测选择关键性的滑坡影响因子并提供规则。

粗糙集对不确定性或不完整信息进行知识挖掘的基本思想,是通过分类的方法对不同概念构成的集合之间的关系进行分析,从中找出概念的属性之间的依赖关系。所选定的研究对象的全体称为论域,如前述提取的属性图层中栅格单元的集合。本文中只选择了研究区部分栅格单元的集合作为论域,即图层中所有已发生滑坡的栅格单元(6 475 个)和部分未发生滑坡的单元(6 069 个),总共 12 544 个有效单元构成论域。没有选取栅格图层中所有的单元作为论域,是为了在保证样本具有代表性的前提下减小计算量。

论域被决策属性划分为滑坡和非滑坡两类。论域还可以被其他属性或属性集合来划分,比如只按斜坡类型属性划分,论域被分为 5 类(5 个集合,每个称为一个等价类)。若其中某个类(比如斜坡类型 3 对应的样本集合)的所有样本都包含在滑坡这一类别集合中,则就斜坡类型 3 这一属性值来说,该等价类称为滑

坡这个概念(集合)关于斜坡类型属性的下近似。在这种情况下,显然有理由推断斜坡类型属性中取值为 3 的因素是与滑坡有确定的因果关系的。由此得到的规则称为明确规则。如果斜坡类型 3 的集合中,有部分元素包含在滑坡集合中,还有部分元素则不包含在滑坡集合中,这时类型 3 的集合称为滑坡概念(集合)关于斜坡类型属性的上近似。此时称滑坡概念对于斜坡类型属性而言是粗糙的(即不精确)。这时显然斜坡类型属性值 3 与滑坡是否发生的因果联系是不确定的。这样导出的规则称为近似规则。如果斜坡类型 3 的集合中,没有元素包含在滑坡集合中,则该属性值与滑坡是无关的。有多个属性时,用多个属性的交集对样本进行划分,然后有上述类似的分析和概念,但过程要复杂很多。

滑坡影响因子粗糙集分析的方法和过程如下:

(1) 将所确定论域中各单元的属性制成二维决策属性表,其形式如表 2 所示。

表 2 滑坡二维决策属性表格式

Table 2 Decision attributes table of landslides

(单元用栅格 行号标识, 共 12544 个)	条件属性				决策属性	
	坡度 slope	高程 Ele	斜坡类型 TOBIA	植被指数 NDVI	岩石地 层单元 Geo	是否滑坡 高危区 Isds
单元号(0,0)	2	1	3	4	7	0
单元号(1,0)	1	2	2	1	3	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
单元号(n,n)	2	3	4	4	5	0

(2) 用条件属性的集合对滑坡和非滑坡两类单元进行近似性分析。这一分析是度量这两个类别对条件属性的粗糙度。ROSE2 中所用的度量不是隶属度而是近似性,即条件属性集合分别关于滑坡和非滑坡类别的下近似除以其上近似。该值越小则相应类别概念越粗糙。计算结果滑坡的粗糙度为 0.005 7,非滑坡则为 0.067 3。这表明就所选论域来说,这两个集合对于所使用的属性来说都是很粗糙的。

(3) 求取二维决策属性表的核。核是条件属性集合的一个唯一子集,它对于产生与使用条件属性全集进行分类的相同结果来说,核中的每个属性都是不能删除的,即核是保证条件属性集合的任一子集的分类,等价于条件属性全集的分类的必要条件。但它不是充分条件。计算结果是 5 个条件属性都包含于核中,也就是所有的条件属性都不能删除。这样,决策表中的条件属性不需要进行整列属性的约简。所谓约简是指这样一个属性子集,它与一个包含它的属性集有相同

的分类结果,并且它不能与它所包含的任何一个属性子集有相同的分类结果(这里的子集是真子集)。属性集的约简不具唯一性,即在一个约简中,除了其中的核属性外,别的属性可能与另外的非核属性可以替代。

(4)生成粗糙集下的决策规则。ROSE 系统支持三种规则生成方法:基本最小覆盖法、扩展最小覆盖法、满意度描述法,三种方法各有优劣,考虑本次研究确定性数据较少、粗糙度很大的特点,确定用扩展的最小覆盖法进行关于“是否滑坡”决策规则的生成。最小覆盖法也是一种属性约简的方式。

3.3 决策规则的有效性分析及其筛选

分析后生成了 86 条决策规则,确定规则 49 条,其编号为 rule 1 - rule 49,其中 isds = 0 的为 rule 1 - rule 41, isds = 1 的为 rule 42 - rule 49;近似规则 37 条,编号为 rule 50 rule 86。关于每条规则的输出内容,除了规则本身外,还包含对规则的一些辅助说明。如满足规则条件属性的单元数与满足规则推理结论的单元数;在与规则推理结论一致的已知类别中,有多少单元是

与该条规则匹配的;在所有推定为与该规则结论一致的单元中,有多少单元是由该条规则决定的。

然后对决策规则系统采取 10 重交叉有效性分析,结果表明规则生成的有效性为强。其中推论结果为 isds = 1 (即判断单元属于滑坡类型)的决策规则平均准确率为 89.66%,说明本研究使用的基于滑坡影响因子类别系统生成的决策规则,覆盖了绝大部分滑坡发生单元,也说明这些影响因子对评价滑坡是有效的。

对 89 条决策规则以规则匹配频度为依据(匹配频度是符合规则条件的单元数除以与该规则推论结果一致的单元总数),提取了 11 条匹配频度较高的规则。其中明确规则有 3 条,近似规则 8 条,近似规则选取了匹配频度最高和较高的(37.83% ~ 16.53%)的前 8 条规则。明确规则中发生滑坡的不粗糙单元极少(67/6 475,约 1.0%),因此选取了 rule 42 这一相对频度最高(0.39%)的规则作为例子。但由于其匹配频度太低,实际意义不大。表 3 列出了 3 条明确规则和 1 条近似规则。

表 3 4 条主要规则的信息
Table 3 Information contained in 4 main rules

规则 编号	规则条件					结论	类型	匹配频度 (%)	符合规则的单元		粗糙 程度
	高程	坡度	地层单元	斜坡类型	植被指数				非滑坡(个)	滑坡(个)	
1	4	-	-	-	-	0	明确	3.03	184	0	0
3	3	-	3,4,6,7	-	-	0	明确	2.84	172	0	0
4	2	1,5	8	-	-	1	明确	0.39	0	25	0
57	1	2,3	3,7,8	3,4,5	-	0 或 1	近似	37.83	1712	2 690	1.57

3.4 单个滑坡影响因子与滑坡关系的统计分析

滑坡影响因子的敏感性分析对于滑坡预测研究很重要^[9-10]。因此以下对每个滑坡影响因子分级与滑坡发生与否的相关性进行统计学的分析和评价。

3.4.1 方法原理

设单个影响因子的论域为集合 C , $A, B \subseteq C$ 是由决策因子 $D = \{d_1, d_2\}$ 将影响论域 C 划分开的,即 $A = \{c: isds = 0, c \in C\}$, $B = \{c: isds = 1, c \in C\}$ 。则 $A \cap B$ 为空集, $A \cup B = C$ 。在本研究中, $C = A + B = 12\ 544$,其中 $A = 6\ 069$ 为 C 中所有非滑坡单元总数, $B = 6\ 475$ 为 C 中所有滑坡单元总数。由于 A, B 还因每个影响因子的分级而划分成若干子集,可表示为:

$A = \{A_1^l, A_2^l, \dots, A_{n_l}^l\}$, $B = \{B_1^l, B_2^l, \dots, B_{n_l}^l\}$, 每个子集中对应的元素个数为 $a_1, a_2, \dots, a_{n_l}$ 和 $b_1, b_2, \dots, b_{n_l}$, 其中 l 为影响因子编号, n_l 为第 l 个影响因子的分级级数。

设 $\alpha_i = a_i/A$ 是集合 A 中以某个因子划分出的子

集 $A_i (i \leq n)$ 的元素个数 a_i 与 A 的比值, $\beta_i = b_i/B$ 是集合 B 中相同因子划分的子集 $B_i (i \leq n)$ 的元素个数 b_i 与 B 的比值。定义每个因子第 i 分级对滑坡发生与否的相关性用因子第 i 分级的似然比 β_i/α_i 来表达, $\beta_i/\alpha_i \in [0, +\infty]$ 。当 a_i, b_i 值都过小时,因样本数量过少而 β_i/α_i 相关性不可采信,忽略。由上可知, ① β_i/α_i 在 $[1, +\infty]$ 上值越大,相应的因子分级与滑坡发生的相关性就越大;② β_i/α_i 在 $[0, 1]$ 范围内的指数值越小,表示相应因子分级与滑坡不发生的相关性大;③ $\beta_i/\alpha_i = 0$, 且 a 足够大,表示该因子类别直接独立决定了对应单元不发生滑坡;④ $a = 0$ 且 b 足够大时, $\beta_i/\alpha_i = +\infty$, 表示该因子类别独立决定了对应单元发生滑坡;⑤ β_i/α_i 越趋近于 1,说明分级指标体系越粗糙,决策越不明确。

3.4.2 结果分析

用上述统计分析方法计算各滑坡影响因子分级与滑坡的相关性,得到表 4。从表 4 可以看出,除高程、

斜坡类型和植被指数三个因子中存在对滑坡决策有意义的分级外(表中用加粗斜体标识),其余皆显示出很大的不确定性(其似然值接近于1)。这一方面说明单一因子对于滑坡分类来说必然是粗糙度大,另一方面也说明因子分级的合理性可能存在一些问题。

表4 滑坡影响因子分级与滑坡的相关性

Table 4 Correlation between grading of impact factors and landslide

因子	分级	元素个数		因子分级似然比		相关性
		a_i	b_i	$\alpha_i(\%)$	$\beta_i(\%)$	
Ele	1	3 017	3 933	49.71	60.74	1.22
	2	2 127	2 088	35.05	32.25	0.92
	3	741	454	12.21	7.01	0.57
	4	184	0	3.03	0	0
Slope	1	292	75	4.81	1.16	0.24
	2	4 667	5 571	76.90	86.04	1.12
	3	946	717	15.59	11.07	0.71
	4	163	104	2.69	1.61	0.60
Geo*	5	1	2	—	—	—
	1	0	0	—	—	—
	2	621	378	10.23	5.84	0.57
	3	985	1 148	16.23	17.73	1.09
	4	336	140	5.54	2.16	0.39
	5	0	0	—	—	—
	6	387	363	6.38	5.61	0.88
	7	1 183	1 388	19.49	21.44	1.10
TOBIA	8	2 567	3 068	42.30	47.38	1.12
	1	72	55	1.19	0.85	0.72
	2	886	974	14.60	15.04	1.03
	3	650	1 693	10.71	26.15	2.44
NDVI	4	1 881	989	30.99	15.27	0.49
	5	2 580	3 764	42.51	58.13	1.37
	1	4 958	5 474	81.69	84.54	1.03
	2	1 106	1 998	18.22	30.86	1.69
	3	5	3	—	—	—

注:Geo*:geo=1,5出现的单元超出了其他图层的空间范围,故该数据被剔除。被忽略的项目记为“—”,其中还包括因样本少而忽略的项目。

4 结论及问题

(1)所选择的5个条件属性因子对滑坡都是必要的影响因素(滑坡粗糙集中的核)。

(2)对预测滑坡相对较有价值的决策规则是明确规则 RULE 1(高程大于700m时不发生滑坡)和明确规则 RULE 3(高程为500~700m而地层为Tb2、Tb3、Tb3+4和Jt时不发生滑坡)。但这两条规则都是预测为不发生滑坡的,且匹配频度不高,而预测为发生滑坡的规则匹配频度更低。

(3)大量的规则是近似规则,而且粗糙度较大。

上述结果说明,坡度、高程、斜坡类型、植被指数、

岩石地层单元5个影响因子是滑坡的关键属性,但不是预测滑坡的完备属性集。因此进一步的研究中还需增加更全面的可能影响因子。规则粗糙度过大的原因,一是可能受数据分辨率所限而采用的单元大小,不满足实际滑坡体的空间尺度要求,比如坡度没有反映出滑坡的实际地形地貌状态,对地层产状空间变化性的描述精度不够等;二是可能对各属性的分级不尽合理,即没有很好地反映出滑坡产生的机理和规律。粗糙集分析本身不能直接提供对合理选择单元大小和属性因子分级的帮助,这些工作依赖于细致的滑坡地质成因分析。

但也可以看出,由于滑坡与其影响因子之间的关系存在不确定性,使用粗糙集方法进行分析是有价值的。一方面,可通过条件属性的核的概念为滑坡预测筛选出关键性影响因子;另一方面,利用发现的规则提供预测滑坡的一种方法。

参考文献:

- [1] 王念秦,王永锋,罗东海,等.中国滑坡预测预报研究综述[J].地质论评,2008,54(3):354-361.
- [2] 段伟国,孙世国,刘玉福.滑坡预测技术发展概论[J].北方工业大学学报,2008,20(1):90-94.
- [3] Pawlak Z. Rough Sets - Theoretical Aspects of Reasoning about Data [M]. Dordrecht: Kluwer Academic Publishers,1991:56-63.
- [4] 史忠植.知识发现[M].北京:清华大学出版社,2002:143-168.
- [5] 曾忠平,汪华斌,张志,等.地理信息系统/遥感技术支持下三峡库区青干河流域滑坡危险性评价[J].岩石力学与工程学报,2006,25(S1):2777-2784.
- [6] 汪华斌,吴树仁,汪稳.长江三峡库区滑坡灾害危险性评价[J].长江流域资源与环境,1998,7(2):186-191.
- [7] 白世彪,阎国年,盛业华,等.基于GIS的长江三峡库区滑坡影响因子分析[J].山地学报,2005,23(1):63-70.
- [8] 波兰 Poznan 科技大学.粗糙集实验系统 ROSE 使用说明[M].1999.
- [9] 陈晓利,冉洪流,祁生文.1976年龙陵地震诱发滑坡的影响因子敏感性分析[J].北京大学学报:自然科学版,2009,45(1):104-110.
- [10] 兰恒星,伍法权,周成虎,等.基于GIS的云南小江流域滑坡因子敏感性分析[J].岩石力学与工程学报,2002,21(10):1500-1506.

(下转第132页)

王连成,林黎,林建旺,等. 地热回灌系统工程建设及
新技术应用研究 [R]. 天津:天津地热勘查开发设计
院,2007.

Application of acidifying & fracturing technology to carbonate rock reservoir

WANG Lian-cheng¹, LI Ming-lang², CHENG Wan-qing¹, JIANG Guo-sheng¹

(1. *Tianjin Geothermal Exploration Development Designing Institute, Tianjin 300250, China;*

2. *Tianjin Bureau of Geology and Minerals Exploration and Development, Tianjin 300191, China*)

Abstract: This paper describes the application of technology of acidifying & fracturing to carbonate rock reservoir. The enhanced effect, technical elements, suitable conditions, technical cruces are discussed with a combination of the practical projects. The results indicate that acidifying & fracturing can increase the permeability of the reservoir and recuperate the reinjection ability of the carbonate-rock reservoir.

Key words: carbonate rock reservoir; acidifying & fracturing; increase in production of geothermal well; permeability

责任编辑:汪美华

(上接第 122 页)

Assessment of impact factors for landslides based on rough sets theory: A case study on the Qingganhe River of the Three Gorges area

LIU Ji-ping¹, LIU Han-qing¹, ZENG Zhong-ping², WANG Hua-bin³

(1. *College of Resource and Environment Science of Wuhan University, Wuhan 430079, China;*

2. *College of Public Administration of Huazhong University of Science and Technology,*

*Wuhan 430079, China; 3. School of Civil Engineering and Mechanics of Huazhong University
of Science and Technology, Wuhan 430079, China*)

Abstract: Landslides can be affected by a series of environmental factors. The present study aimed to analyze landslide susceptibility in the Qingganhe River of Three Gorges area in China. Spatial information techniques, including RS and GIS with an innovative data mining approach, namely, the rough sets theory, are incorporated to the study for landslide factors analysis. In a case study, the cores (i. e., slope, elevation, slope type, vegetation index, and rock type) and reductions of the data attributes were obtained after data mining using the rough sets theory. Rules of all the impact factors, which can make contributions to landslide occurrence, were generated on the basis of landslide knowledge. All of the 11 rules were classified into straight rules and approximate rules, while three main rules were extracted to be the decision-making rules for landslide predictions in view of importance.

Key words: landslides; data mining; rough sets theory; spatial information

责任编辑:汪美华