

基于粗糙集的范例推理在泥石流 危险性评价中的应用

刘勇健¹, 刘义建²

(1 广东工业大学岩土工程研究所, 广东 广州 510006; 2 湖南省地质矿产勘查开发局 407队, 湖南 怀化 418000)

摘要: 将粗糙集理论和范例推理相结合, 建立了基于粗糙集-范例推理的泥石流危险性评价模型。运用粗糙集理论对范例库中的范例进行属性约简和特征向量权值计算, 用相似度量理论来实现最相似范例检索, 得到泥石流的危险性评价结果。选取了一次泥石流(可能)最大冲出量、泥石流发生频率、流域面积、主沟长度、流域相对高差、流域切割密度、24 h最大降雨量、松散固体物质储量共 8 个因素作为泥石流危险性的影响因子; 选取云南东川和西藏八宿等共 14 条泥石流沟资料进行建模分析。实例研究表明, 基于粗糙集-范例推理的泥石流危险性评价结果与实际状态相吻合, 该方法具有简便、高效、直观、实用的特点。

关键词: 泥石流; 危险性评价; 粗糙集; 范例推理; 相似度量

中图分类号: P642.23

文献标识码: A

泥石流是山区特有的一种不良地质现象, 是由暴雨或上游冰雪消融形成的携带有大量泥沙和石块的间歇性洪流。它具有暴发突然、来势凶猛、历时短暂、破坏力强的特点, 常在瞬间淹没农田、冲毁道路桥梁等建筑物, 对山区生态环境和经济建设带来巨大的破坏性。近年来, 以牺牲环境为代价的发展模式引起了人们的反思, 防灾减灾研究受到了世界各国政府的高度重视。正确合理的泥石流危险性评价可为土地利用、城镇规划、道路选线、投资决策和减灾措施的制定提供重要理论依据。目前, 在国内外有关泥石流危险性评价的研究中, 常采用模糊综合评价法、灰色系统理论、多元回归分析法、神经网络法和 GIS 技术等^[1-5]。这些评价方法都从不同侧面反映了泥石流危险性及其影响因素的不确定性, 为山区泥石流的预防与治理起到了一定的作用。

众所周知, 泥石流的形成发展受到多种因素的影响制约, 其中部分因素可以定量判定, 而多数因素则是不能完全确定的。在对泥石流危险性评价和预

防治过程中, 研究人员已经充分认识到工程地质工作及基础地质资料的重要性。在地质灾害治理过程中, 开始强调理论、方法、技术的一体化, 并注意到对大量的已被研究清楚的泥石流实例在地质灾害预防与治理中的类比和借鉴作用。自 1982 年 Roger Schank 在出版的《Dynamic Memory》书中首次提出基于范例推理(Case-Based Reasoning, 简称 CBR)以来, 范例推理原理已在计划、决策、错误诊断、预报预测等领域得到广泛应用。范例推理原理为泥石流危险性评价研究提供了一条新途径, 本文结合前人的研究成果, 提出基于粗糙集的范例推理的泥石流危险性评价方法。

1 基于粗糙集的范例推理

1.1 CBR 的基本原理

CBR 是近年来人工智能中出现的一种新的推理模式, 它兴起的主要原因是由于基于规则的系统

收稿日期(Received date): 2007-12-03; 改回日期(Accepted): 2008-03-01。

基金项目(Foundation item): 广东省自然科学基金项目(6021462)和广东省岩土工程重点学科基金资助。[Supported by the Natural Science Foundation of Guangdong Province and Geotechnical Engineering Key Subject Foundation of Guangdong Province.]

作者简介(Biography): 刘勇健(1968-)女, 湖南安化人, 博士研究生, 副教授, 主要从事岩土工程研究。[Liu Yongjian(1968-), female, born in Nanhuan, Hunan province in 1968. PH.D. Candidate, associate professor, works mainly on geotechnical engineering.]

在知识获取上存在困难,对于处理过的问题没有记忆,导致推理效率低下,没有异域联想能力,整体性能较为脆弱,而 CBR 恰好能克服这些缺陷。

在范例推理中,把当前要求解的新问题称为目标范例(target case),而把记忆的问题称为源范例(base case)^[6-9]。范例推理就是由目标范例的提示而获得记忆中的源范例,并由源范例来指导目标范例求解的一种策略。基于范例推理中知识表示是以范例为基础,范例的获取比规则获取容易,从而克服了知识获取上的瓶颈。在长期的实践中,研究人员已经对大量泥石流危险性进行了深入的研究,获得了比较明确的结论。所有这些有明确结论的泥石流将构成泥石流危险性评价的源范例,为实现范例推理奠定了基础。通过泥石流的目标范例与源范例之间的类比,根据目标范例与源范例之间的相似性程度,检索到最相似范例,完成泥石流目标范例的危险性评价。

CBR 具有信息的完全表达,增量式学习,形象思维的准确模拟,知识获取较为容易等优点。但 CBR 对噪音数据较为敏感,错误数据容易影响系统检索效率和求解效果。而范例库中往往含有一些对范例影响甚微的特征项和冗余信息,这对范例库的存储和检索带来了不必要的负担。本文将粗糙集理论和相似度理论引入范例推理中,为高效准确地检索最相似范例进行了一次新的尝试。

1.2 粗糙集理论

粗糙集理论(Rough Set 简称 RS)是分析不完整、不精确信息系统的有力工具^[10-11]。粗糙集方法把研究对象视作一个信息系统,通过不可分辨关系和不可分辨类确定给定问题的近似域,从而找到问题的内在规律。根据知识表达中不同属性的重要性,明确有用信息和冗余信息,简化输入信息的表达空间的维数。在保持分类能力不变的前提下,通过知识约简,直接从训练数据中导出分类规则。RS 的基本定义如下:

定义 1 (决策系统) 一个信息系统可以表示为 $S = (U, R, V, f)$ 。其中, U 为对象的非空有限集合,即论域; R 是属性集合; R 为属性的非空有限集合; $V = \bigcup_{r \in R} V_r$, V_r 为属性 r 的值域; $f: U \times R \rightarrow V$ 是一个信息函数,指定 U 中各对象的属性惟一值。对信息系统 S , 若属性集 $R = C \cup D$, 子集 C 和 D 分别称为条件属性集和决策属性集, $C \cap D = \phi$, 则该信息系统称为一个决策系统。

定义 2 (属性的约简) 对于给定的知识表达系统 $S = (U, R, V, f)$, 设 R 是一个等价关系族, $a \in R$, 如果, $ind(R) = ind(R \setminus \{a\})$, 则称 a 在 R 中是可以被约去的属性。

定义 3 (正域) 设 $p, Q \subseteq R$, 若 $ind(p) \subseteq ind(Q)$, 即 Q 可由 p 完全推出, 记为 $P \rightarrow Q$ 。而在粗糙集理论中, 有些因素并不是完全可推出的, Q 的 p 正域记为 $pos_p(Q)$

$$pos_p(Q) = \bigcup_{x \in U} \bigcap_{R \in p} (x)$$
 (1)

定义 4 (属性重要度和属性权值) 对于给定的知识表达系统 $S = (U, R, V, f)$, $R = C \cup D$, 条件属性的重要性可以用以下公式来进行判断

$$\sigma_c(r) = \frac{card(pos_c(D)) - card(pos_{c-\{r\}}(D))}{card(U)} \quad (2)$$

式中 $card(\cdot)$ 为集合的基, $r \in C$ 为范例的某一属性。 $\sigma_c(r)$ 越大, 表明属性 r 对决策越重要。属性重要度 $\sigma_c(r)$ 值范围为 $[0, 1]$, 当 $\sigma_c(r) > 0$ 时, 属性 r 是必要的。本文参考文献[9], 根据属性的重要度值按比例确定范例条件属性的 $r \in C$ 权重值 w_r , 计算公式为

$$w_r = \frac{\sigma_c(r)}{\sum_{r \in C} \sigma_c(r)} \quad (3)$$

1.3 相似度计算

从 CBR 系统中高效地完成最相似范例的检索是 CBR 的关键, 这对新问题的求解效率和准确性有较大影响。以往的研究大都采用欧氏距离计算相似度。由于距离测度的相似度量没有从系统的组成要素及相互关系角度分析系统的相似性, 它忽视了组成要素特性的相似性, 从而导致算出的相似系数不能很好地反映范例间的相似程度。为了克服距离测度相似度的不足, 本文运用相似度理论, 全面比较两个相似系统之间的各特性的相似性, 定量刻画两个相似系统之间的相似程度。

本文将具相似特性的单元定义为相似元, 设相似度为相似元的数量以及每个相似元对系统相似度影响权值等因素的函数^[12]。设系统 A 由 K 个要素组成, 系统 B 由 L 个要素组成, 系统 A 和 B 之间有 n 个相似要素, 即 n 个相似元, 记为 $\lambda(e_i)$, ($i = 1, 2, \dots, n$)。每一相似元对相似系统相似程度的影响权重为 w_i , 则系统 A 与 B 的相似度可以定义为

$$Q_m(A, B) = \frac{n}{k + l - n} \sum_{i=1}^n w_i \cdot \lambda(e_i) \quad (4)$$

式中 k, l 分别表示相似系统 A 和 B 的组成要素的数量, n 为相似系统 A 和 B 共有的相似要素的个数, $n/(k+l-n)$ 表示 A 和 B 间相似元的多少对系统相似度的影响。 $\lambda(e_i)$ 为第 i 个相似元的数值 (本文取

两相似元的比值, $0 < \lambda(e_i) \leq 1$), $w_i \cdot \lambda(e_i)$ 表示考虑了权重时, 第 i 个相似要素的相似度对系统相似度的综合影响。

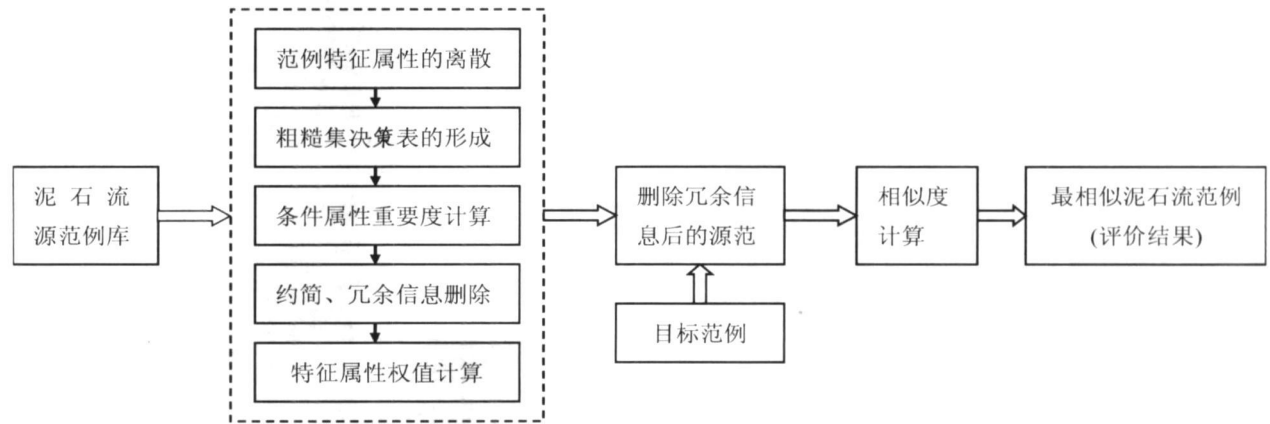


图 1 基于粗糙集-相似度的范例推理流程图

Fig. 1 Sketch map of Case-Based Reasoning method to debris flow hazard assessment based on rough set

1.4 基于粗糙集-相似度的范例检索

基于粗糙集-相似度的范例推理过程如图 1 所示, 首先运用粗糙集理论对范例特征属性进行离散归一化, 形成初始决策表, 计算各条件属性的重要度, 进行属性约简, 删除冗余信息; 然后, 根据重要度计算特征项权重值; 最后, 计算新问题和范例库中所有源范例的相似度, 并根据相似度排序检索出最相似范例。

2 CBR 在泥石流危险性评价中的应用

泥石流危险性评价是灾情评估、预测、防灾救灾决策的基础。根据研究范围, 可将其分为点评价和面评价, 即单沟泥石流危险度评价和区域泥石流危险度评价。单沟泥石流危险性评价是指对一条泥石流沟或相邻近、具有统一动力活动过程和破坏对象的几条泥石流沟或沟群进行评价。区域泥石流危险度评价是对一个流域、一个地区或更大的自然、行政区域内的泥石流灾害进行评价, 本文主要研究泥石流危险性的点评价。

2.1 泥石流危险性影响因素分析与选择

大量的研究表明, 形成泥石流必须具备三个基本条件: 一是有利的地貌单元, 二是有丰富的松散固体碎屑作为补给物质, 三是适当的降雨或融雪形成

的水源激发^[1-5]。泥石流的发生受到很多因素的控制, 可将其归纳为四个方面, 即地质条件, 地形地貌条件, 气候水文条件, 人为活动条件。在选择评价指标时, 除考虑影响因素的物理意义外, 还需各指标的易获性、独立性等要素。本文选择一次泥石流 (可能) 最大冲出量 $c_1 (\times 10^4 \text{ m}^3)$, 泥石流发生频率 c_2 (次/100 a), 流域面积 $c_3 (\text{km}^2)$, 主沟长度 $c_4 (\text{km})$, 流域相对高差 $c_5 (\text{km})$, 流域切割密度 $c_6 (\text{km}/\text{km}^2)$, 24 h 最大降雨量 $c_7 (\text{mm})$, 松散固体物质储量 $c_8 (\times 10^4 \text{ m}^3)$, 共 8 个因素作为泥石流危险度评价的影响因子。

2.2 范例库的建立

范例库是 CBR 中的重要组成部分, 它是 CBR 中主要的知识库。范例库可以从某个知识丰富的数据库中抽取, 也通过各种知识获取 (knowledge acquisition) 方法得到。笔者从文献 [1, 12] 整理出 12 条研究全面深入、并结论明确的泥石流实例为源范例。蒋家沟流域位于云南省东北部、小江流域内, 为最具代表性的一条泥石流沟, 规模巨大、暴发频繁、危害严重。流域内设有中国科学院东川泥石流观测站, 有丰富的观测与研究资料可以利用。东川小白泥沟与东川达朵沟泥石流流速快、暴发频率高, 属高度危险性泥石流。东川深沟与怒江州六库芭蕉河属中度危险度, 南涧县城的观音寺沟与保山的王官

屯河属低危险度泥石流沟。

位于怒江流域西藏自治区东部的八宿县,县城后山分布有拉曲、日吉村沟、若弄巴沟等大小 5 条泥石流沟。区内受构造作用和寒冻风化作用影响,地形起伏较大、沟谷发育、岩石极破碎、植被不发育,遇

暴雨极易形成泥石流。随着近年来灾害性天气频繁,城区泥石流灾害时有发生,对人民生命财产和生态环境构成严重威胁。本文将泥石流沟划分为为高危险性 (I 级)、中危险性 (II 级)、低危险性 (III 级), 泥石流范例的基本数据如表 1。

表 1 泥石流沟范例的基本数据
Table 1 Basic data of base cases of debris flow gullies

序号	泥石流沟名	最大冲出量 c_1 ($\times 10^4 \text{ m}^3$)	发生频率 c_2 (次 /100 a)	流域面积 C_3 (km^2)	主沟长度 c_4 (km)	流域相对高差 c_5 (km)	切割密度 c_6 (km / km^2)	最大降雨量 c_7 (mm)	松散固体物质储量 c_8 ($\times 10^4 \text{ m}^3$)	危险性等级 d
1	蒋家沟	195. 10	1 500	47. 10	12. 00	2. 19	23 80	102. 00	75 000	I
2	小白泥沟	60. 00	1 000	12. 49	6. 00	1. 67	20 50	100. 40	14 000	I
3	达朵沟	12. 00	50	58. 07	13. 75	2. 19	11 20	102. 00	3 811	I
4	深 沟	9. 00	50	28. 32	9. 05	2. 80	15 60	100. 40	7 000	II
5	芭蕉河	8. 20	13	4. 10	3. 80	1. 67	6. 60	82. 00	150	II
6	王官屯河	1. 60	20	1. 93	2. 19	0. 32	4. 20	109. 40	72	III
7	观音寺沟	0. 40	20	0. 15	1. 13	0. 23	8. 20	69. 20	65	III
8	拉 曲	19. 68	30	91. 60	19. 00	2. 59	18 50	40. 20	12 900	II
9	日吉村沟	0. 46	20	0. 75	2. 18	0. 98	5. 30	40. 20	21. 10	III
10	若弄巴沟	0. 18	20	0. 63	1. 92	1. 06	5. 10	40. 20	11. 00	III
11	3号沟	0. 09	20	0. 14	0. 98	0. 53	3. 57	40. 20	1. 20	III
12	4号沟	1. 18	20	1. 07	2. 62	1. 28	4. 67	40. 20	8. 60	III
13	Z1 沟	23. 00	40	35. 70	10. 29	1. 40	9. 20	87. 00	4 000	I
14	Z2 沟	2. 05	30	1. 20	1. 38	1. 43	4. 81	43. 00	25. 00	III

2.3 数据预处理——属性约简及重要度计算

泥石流沟危险性评价时,将收集到的范例集合作为论域 $U = \{U_1, U_2, \dots, U_{12}\}$, 取表 2 中前 12 个样本为源范例, 设后 2 个样本为目标范例。泥石流危险性的影响因子作为条件属性, 记 $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$, 泥石流的危险程度为决策属性, 记 $D = \{d_1, d_2, d_3\}$, 其中, d_1 = 高危险性、 d_2 = 中危险性、 d_3 = 低危险性, 属性集 $R = C \cup D, C \cap D \neq \phi$ 。决策表的表头为各属性, 每一行表示论域中的一个事例, 每一列表示属性及属性值。将各属性离散化后形成初始决策, 如表 2。

由表 2 依据条件属性划分等价类如下:
 $U/\{c_1\} = \{\{1\}, \{2, 3, 8\}, \{4, 5, 6, 12\}, \{7, 9, 10, 11\}\}, U/\{c_2\} = \{\{1, 2\}, \{3, 4\}, \{5, 6, 7, 8, 9, 10, 11, 12\}\}$
 $U/\{c_3\} = \{\{13, 4, 8\}, \{2, 5, 6, 7, 9, 10, 12\},$

$\{11\}\}, U/\{c_4\} = \{\{1, 3, 4, 8\}, \{2, 5, 6, 7, 9, 10, 12\}, \{11\}\}$
 $U/\{c_5\} = \{\{1, 3, 4, 8\}, \{2, 5, 6, 7, 9, 10, 11, 12\}\}, U/\{c_6\} = \{\{1, 3, 4, 8\}, \{2, 5, 6, 7, 9, 10, 11, 12\}\}$
 $U/\{c_7\} = \{\{1, 2, 3, 5, 6\}, \{5, 7\}, \{8, 9, 10, 11, 12\}\}, U/\{c_8\} = \{\{1, 2, 3, 4, 8\}, \{5, 6, 7, 9\}, \{10, 11, 12\}\}$
决策属性分类为: $U/\{D\} = \{\{1, 2, 3\}, \{4, 5, 8\}, \{6, 7, 9, 10, 11, 12\}\}$
可得 D 的 C 正域为: $pos_C(D) = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\} \cup \{7\} \cup \{8\} \cup \{9\} \cup \{10\} \cup \{11\} \cup \{12\}$
从表 2 中可知, 属性 c_3 和 c_4, c_5 和 c_6 在分类表中作用是一致的, 考虑影响指标的易获性, 删除冗余信息 c_4 和 c_6 , 保留 c_3 和 c_5 。

表 2 初始决策表
Table 2 Original decision making table

U_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	d
u_1	0	0	0	0	0	0	0	0	0
u_2	1	0	1	1	1	1	0	0	0
u_3	1	1	0	0	0	0	0	0	0
u_4	2	1	0	0	0	0	0	0	1
u_5	2	2	1	1	1	1	1	1	1
u_6	2	2	1	1	1	1	0	1	2
u_7	3	2	1	1	1	1	1	1	2
u_8	1	2	0	0	0	0	2	0	1
u_9	3	2	1	1	1	1	2	1	2
u_{10}	3	2	1	1	1	1	2	2	2
u_{11}	3	2	2	2	1	1	2	2	2
u_{12}	2	2	1	1	1	1	2	2	2

去除 c_1 分类结果为: $pos_{(C-\{c_1\})}(D) = \{1, 2, 6, 8, 9, 11\} = \{1\} \cup \{2\} \cup \{6\} \cup \{8\} \cup \{9\} \cup \{11\}$

据式 (2)得条件属性 c_1 的重要度: $\sigma_{c_1}(r) = \frac{card(pos_C(D)) - card(pos_{C-\{r\}}(D))}{card(U)} = \frac{12-6}{12} = 0.5$

同理, 可求得其他因子的重要度

$$\begin{aligned}\sigma_{c_2}(r) &= 1/6 \quad \sigma_{c_3}(r) \\ &= 1/6 \quad \sigma_{c_5}(r) \\ &= 1/6 \quad \sigma_{c_7}(r) \\ &= 1/6 \quad \sigma_{c_8}(r) \\ &= 1/3\end{aligned}$$

2.4 特征项权值计算

运用粗糙集理论去除冗余条件属性 (c_4, c_6), 约简后保留泥石流危险性评价特征向量为 $C = \{c_b, c_2, c_3, c_5, c_7, c_8\}$, 根据公式 (3)计算特征权值得, $w_1 = 0.344, w_2 = w_3 = w_5 = w_7 = 0.111, w_8 = 0.222, w_4 = w_6 = 0$. 计算结果表明, 属性 ($c_b, c_2, c_3, c_5, c_7, c_8$)在泥石流危险性评价中是决策分类中必要的属性, 其中, 一次泥石流 (可能)最大冲出量 c_1 对分类影响最大, 其次是松散固体物质储量 c_8 。运用公式 (4)计算相似度时, 取 $n=6, k=l=8$

2.4 相似度计算及最佳范例检索

在评价泥石流沟 Z1沟和 Z2沟的危险性时, 将 Z1沟和 Z2沟作为目标范例, Z1沟目标范例向量 $A_{Z1} (13.00, 40, 35.70, 10.29, 1.40, 9.20, 87.00, 4000)$, Z2沟目标范例向量 $A_{Z2} (2.05, 30, 1.20, 1.38, 1.43, 4.81, 43.00, 25.00)$, 按公式 (4)计算目标范例与源范例间相似度 $Q_m(A, B)$:

以 Z1与源范例 1(u_1)的相似度计算为例

$$\begin{aligned}Q_m(A, B) &= \frac{n}{k+l-n} \sum_{i=1}^n w_i \cdot \lambda(e_i) \\ &= \frac{6}{8+8-6} [0.333 \times \lambda(e_1) + \\ &\quad 0.111 \times \lambda(e_2) + 0.111 \times \lambda(e_3) + 0.111 \times \lambda(e_5) + \\ &\quad 0.111 \times \lambda(e_7) + 0.222 \times \lambda(e_8)] \\ Q_1(A, B) &= \frac{6}{10} [0.334 \times \frac{13}{195} + 0.111 \times \frac{40}{1500} + \\ &\quad 0.111 \times \frac{35.70}{47.10} + 0.111 \times \frac{1.40}{2.19} + 0.111 \times \frac{9.20}{102.00} + \\ &\quad 0.222 \times \frac{4000}{75000}] \\ &= 0.1721\end{aligned}$$

同理, 可计算出 Z1、Z2与其他各源范例间的相似度, 如表 3

表 3 目标范例与源范例的相似度

Table 3 Similarity degree between target cases and base cases

目标范例	$Q_1(1, B)$	$Q_2(1, B)$	$Q_3(1, B)$	$Q_4(1, B)$	$Q_5(1, B)$	$Q_6(1, B)$	$Q_7(1, B)$	$Q_8(1, B)$	$Q_9(1, B)$	$Q_{10}(1, B)$	$Q_{11}(1, B)$	$Q_{12}(1, B)$
Z1	0.1721	0.2209	0.5049	0.4427	0.2789	0.1321	0.1058	0.4608	0.1199	0.1187	0.0910	0.1454
Z2	0.0767	0.1010	0.1479	0.1513	0.3373	0.3291	0.1950	0.1966	0.3512	0.2671	0.1543	0.3865

依据相似度值由大到小排序,可得目标范例 Z_1 与源范例库中 u_3 (达朵沟)范例特征属性相似度最大,属高危险性泥石流,目标范例 Z_2 与源范例库中 u_{12} (4号沟)范例特征属性相似度最大,属低危险性泥石流。可参照源范例 u_3 (达朵沟泥石流)和 u_{12} (4号沟泥石流)采取相应的预防与整治措施。如表 1 所示,位于云南东北部的 Z_1 沟和 Z_2 沟, Z_1 沟地形陡峻,降雨充沛,植被稀少,崩塌、滑坡较发育,可动固体物质丰富,泥石流屡发成灾,危险性高。 Z_2 沟地形较平缓,降雨量较小,上游的风化产物来源相对较少,泥石流规模较小。本文的范例推理结果与 Z_1 沟和 Z_2 沟的实际危险性程度一致。

3 结语

范例推理是根据目标范例来检索范例库中的相似范例,并由相似范例来指导求解的策略。如何从范例库中检索和选择最相似的范例决定了范例推理系统的学习与推理性能,范例间的相似性度量是检索的关键。本文针对泥石流危险性影响因素的复杂多变性和相当强的不确定性,将粗糙集理论和范例推理有机结合起来,运用粗糙集理论中对范例库中的属性进行约简,删除决策表中冗余信息,并依据属性重要度计算特征向量的权值,运用相似度量理论来实现最相似范例检索,提高了目标范例检索的精确度和效率。通过云南东川泥石流和西藏八宿泥石流实例研究表明,基于粗集-范例推理的泥石流危险性评价结果与实际状态一致,表明该方法具有简便、高效、直观、实用的优点。这一新的研究思路和方法同样适合于对其他类型的灾害地质评价。

参考文献 (References)

- [1] Liu Xilin, Tang Chuan. Risk Assessment on Debris flows[M]. Beijing: Science Press, 1995. [刘希林, 唐川. 泥石流危险性评价[M]. 北京: 科学出版社, 1995.]
- [2] Kuang Lehong, Xu Linrong, Liu Baochen. Debris flow hazard assessment based on extension method[J]. *China Railway Science*, 2006, 27(5): 1~6 [匡乐红, 徐林荣, 刘宝琛. 基于可拓方法的泥石流危险性评价[J]. 中国铁道科学, 2006, 27(5): 1~6]
- [3] Liu Jinfeng, Ou Guoqian. New option on debris flows hazard assessment[J]. *Journal of Geological Hazards and Environment Preservation*, 2004, 15(1): 5~8 [柳金峰, 欧国强. 泥石流危险性评价的新思路[J]. 地质灾害与环境保护, 2004, 15(1): 5~8]
- [4] Li Kuang, Tang Chuan. Progress in research on debris flow hazard Assessment[J]. *Journal of Catastrophology*, 2007, 22(1): 106~111 [李阔, 唐川. 泥石流危险性评价研究进展[J]. 灾害学, 2007, 22(1): 106~111]
- [5] Kang Zhicheng, Ma Genai, Li Zhuofeng. Study on Debris Flows of China[M]. Beijing: Science Press, 2004: 1~4 [康志成, 马谰乃, 李焯芬. 中国泥石流研究[M]. 北京: 科学出版社, 2004: 1~4]
- [6] Liu Muyu, Feng Xiating. Evaluation of slope stability based on case-based reasoning integrated with neural network[J]. *Rock and Soil Mechanics*, 2005, 26(2): 193~197 [刘沐宇, 冯夏庭. 基于神经网络范例推理的边坡稳定性评价方法[J]. 岩土力学, 2005, 26(2): 193~197]
- [7] Sakino M., Gobbardes E. Rough Sets Reduction Techniques for Case-based Reasoning[J]. *Proceedings of ICCBR*, 2001: 466~482
- [8] Zhang Wenxiu, Wu Weizhi. Rough Sets Theory and Method[M]. Beijing: Science Press, 2001. [张文修, 吴伟志. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.]
- [9] Ji Sai, Shen Xing, Shen Chao. A retrieval method of CBR based Rough Set and Similarity Measurement[J]. *Computer Engineering and Applications*, 2006, (13): 172~174 [季赛, 沈星, 沈超. 基于粗糙集和相似度量的 CBR 检索方法[J]. 计算机工程与应用, 2006, (13): 172~174]
- [10] Lu Yijuan, Ni Zhiwei, Hu Caiping. Technique used in CBR based on similarity rough set[J]. *Computer Engineering*, 2003, 29(19): 50~53 [卢亦娟, 倪志伟, 胡彩平. 基于相似粗糙集的范例推理技术[J]. 计算机工程, 2003, 29(19): 50~53]
- [11] Pawlak Z. Rough Sets and Decision Analysis[J]. *Information Sciences*, 2000, 38(3): 132~144.
- [12] Zhang Chengjie, Wang Changming, Wang Gangcheng et al. An application of gray correlation method in evaluating debris flow hazard[J]. *Jilin Geology*, 2005, 24(4): 111~115 [张成杰, 王常明, 王钢城. 灰色关联法在泥石流危险性评价中的应用[J]. 吉林地质, 2005, 24(4): 111~115]

Application of Case-based Reasoning Method to Assess Debris Flow Hazard Based on Rough Set

LU Yongjian¹, LU Yijian²

(1. Institute of Geotechnical Engineering, Guangdong University of Technology, Guangzhou 510006, China;

2. No. 407 Brigade Geology and Mineral Resources Exploration and Development Bureau of Hunan Province, Huaihua 418000, China)

Abstract To assess debris flow hazard, A case-based reasoning model (CBR), combining with rough set is established in the paper. The rough set is used to reduce the features of the case and calculate features weights, while the similarity degree measurement, which improves retrieval speed and accuracy, is used to retrieve the similar cases. Thus the above methods can obtain the criticality of debris flow. In the paper, eight factors including the maximum expulsion amount of debris flow once occurrence frequency, drainage area, length of the main gully, relative altitude, cutting density, maximum quantity of rainfall in 24 hours and amount of loss solid matter, are chosen as the influence factors of debris flow hazard. 14 debris flows of Dongchuan in Yunnan Province and Basu in Tibet are applied to set up model of debris flow hazard assessment. The research shows that the results of debris flow hazard assessment are the same as the practical states, and this method is simple, effective, visual and practical.

Key words debris flow; hazard assessment; rough set; case based reasoning; similarity degree measurement