

# 数据挖掘在岩质边坡稳定性预测中的应用

张治强<sup>1)</sup> 蔡嗣经<sup>1)</sup> 马平波<sup>2)</sup>

1) 北京科技大学土木与环境工程学院, 北京 100083 2) 东北大学资源与土木工程学院, 沈阳 110006

**摘要** 提出了一种预测岩土边坡稳定性的方法. 利用数据挖掘的方法从边坡实例数据中挖掘出知识, 在此基础上建立岩质边坡稳定性预测的智能模型系统. 该系统可以根据获得的现场测试和监测资料, 对边坡的稳定性进行较好的预测和估计.

**关键词** 数据挖掘; 边坡; 稳定性

**分类号** TD 325\*.1; TU 457

岩土边坡的失稳破坏一直是岩土工程建设中的主要灾害之一, 正确评价边坡的稳定性和其破坏类型, 对于确保生产建设和人民生命财产安全具有重要意义. 目前判断边坡的稳定性主要是用极限平衡法和有限元法. 但极限平衡法由于不考虑岩土体的变形机制, 其安全系数误差较大; 有限元法的应力应变一般是按未破坏时的边界条件计算出来的, 而实际情况并非如此, 任何超过抗剪或抗拉强度的应力状态都是不稳定的, 一旦发生局域破坏, 应力将会重新调整, 边坡稳定性安全系数也随之而发生变化<sup>[1]</sup>.

有的学者将专家系统和神经网络引入对边坡的稳定性研究中<sup>[2,3]</sup>, 但结果不令人满意. 专家系统由于其掺杂了人的经验等主观因素, 所以效果不理想. 神经网络尽管具有非线性、并行性、鲁棒性等特点, 但由于它很容易陷入局部极小点和收敛速度慢的局面, 极大地影响了它的使用. 本文从智能岩石力学的角度出发, 提出用数据挖掘法来预测岩土边坡的稳定性问题.

## 1 数据挖掘

数据库中的知识发现 KDD (Knowledge Discovery in Database) 是近年来随着人工智能和数据库技术的发展而出现的一门新兴的技术. 数据挖掘 (Data Mining) 是 KDD 中最重要的处理阶段, 它主要基于人工智能、机器学习和统计学技术,

通过对原始数据的高自动化分析处理, 作出归纳性的推理, 得到数据对象间的关系模式. 这些关系模式反映了数据的内在特性, 是对数据所包含信息的更高层次的抽象. 因此它可以从大型数据库或数据仓库中提取人们感兴趣的知识, 这些知识是隐含的、事先未知的潜在有用信息<sup>[4-12]</sup>.

数据挖掘过程是在一些事实或观察数据的集合中寻找模式的决策支持过程. 关联规则是当前数据挖掘研究的主要模式, 侧重于确定数据中不同领域之间的联系, 找出满足给定支持度和置信度阈值的多个域之间的依赖关系. 一条关联规则是如下形式的蕴涵式:  $X \rightarrow Y (S\%, C\%)$ , 即关联规则  $X \rightarrow Y$  在交易集合  $D$  中成立, 且关联规则的支持度为  $S\%$ , 关联规则的置信度为  $C\%$ . 挖掘关联规则问题就是在给定的交易集合  $D$  中产生所有满足最小支持度 (minsupp) 和最小置信度 (minconf) 的关联规则的过程. 挖掘关联规则问题可以分为两个子问题:

(1) 寻找所有这样的项的集合, 它们的支持度超过用户给定的最小支持度, 这个项的集合称为频繁集 (也称为大项目集).

(2) 应用频繁集产生关联规则.

挖掘关联规则常采用的 Apriori 算法, 由 Agrawal 在 1994 年提出. Apriori 算法采用迭代方法, 通过对交易数据进行多遍扫描以生成关联规则. 本文关联规则的生成就是通过 Apriori 算法完成的. 根据数据挖掘的实现原理可知, 对某些比较杂乱的数据库, 并不能保证挖掘得到的规则的正

收稿日期 2002-04-19 张治强 男, 31 岁, 博士研究生

\* 教育部博士后基金资助项目 (No.13416)

确性. 用户如何参与到挖掘过程中, 利用用户的背景知识和指导作用来加快挖掘的进程, 并且保证发现的知識的有效性, 將相关领域的知識融入数据挖掘系统中是一个重要但没有很好解决的问题. 另外, 如何清理脏数据、做好数据预处理工作, 也是相当重要的.

## 2 边坡稳定性和破坏方式的预测

与岩质边坡稳定性有关的因素概括起来有地形、岩体和其他影响因素三大类. 其中, 地形是稳定性的外部表现; 岩体是稳定性的核心, 它是

岩性和地质构造及外力地质作用的综合反映; 而其他影响因素则是通过地形和岩体来影响边坡稳定性的外部作用. 本文选取了几个露天矿已知的 120 个岩质边坡的资料, 输入此模型进行分析. 需要挖掘的数据库见表 1 (这里只列 20 个样本), 检验情况见表 2.

### 2.1 建立边坡稳定性预测的智能模型

首先由现场测试和监测获得地形、岩体、外在影响三大因素的特征. 地形因素包括坡高、坡角等, 岩体因素包括结构特征、结构面发育程度、地下水情况等, 外在影响因素有植被覆盖率、坡角开挖值等. 然后将其中主要因素输入数据挖掘

表 1 用于挖掘的样本  
Table 1 Samples for data mining

编号	$\sigma_s$ /MPa	$\alpha$ (°)	A	B	C	D	$c$ /MPa	$\theta$ (°)	$h$ /m	$n$	$\beta$ (°)
1	106.3	50	平行	5H	薄层-镶嵌	圆弧	5.0	37.5	496	1.20	39.5
2	78.0	70	垂直	4H	层状	圆弧	2.2	39.0	496	1.10	37.5
3	38.2	70	斜交	5H	块状-镶嵌	平面-圆弧	3.8	37.5	494	1.25	37.0
4	154.9	50	斜交	3H	似层状	双滑块折线	5.7	36.0	480	1.10	42.0
5	154.8	47	斜交	3H	块状	双滑块折线	5.0	38.0	292	1.15	45.0
6	67.7	62	斜交	3H	块状	双滑块折线	4.5	36.0	365	1.15	46.0
7	67.7	62	斜交	3H	似层状	圆弧	6.4	35.0	382	1.15	46.0
8	67.7	62	斜交	3H	层状-块状	折线形	6.0	39.0	645	1.15	37.0
9	72.0	65	斜交	4H	层状-块状	折线形	7.2	38.0	630	1.20	50.0
10	64.2	65	斜交	7H	层状	圆弧	6.8	35.0	608	1.20	55.0
11	46.2	45	斜交	7H	层状	圆弧	6.8	35.0	200	1.20	55.0
12	64.8	45	斜交	5H	层状-块状	折线形	9.0	39.0	375	1.25	49.0
13	64.8	45	斜交	5H	层状-块状	折线形	7.0	37.0	231	1.25	52.5
14	59.0	80	斜交	2H	层状-块状	折线形	4.8	37.0	218	1.20	39.5
15	82.1	60	斜交	5H	块状	双滑块折线形	4.1	38.0	318	1.20	48.0
16	82.1	50	斜交	5H	块状	双滑块折线形	4.2	47.0	115	1.20	57.5
17	82.1	45	垂直	5H	块状	双滑块折线形	2.9	34.0	123	1.20	52.5
18	82.1	45	垂直	5H	块状	双滑块折线形	4.0	36.0	110	1.20	57.5
19	147.4	67	斜交	5H	块状-镶嵌	平面-圆弧	9.9	36.0	198	1.20	48.0
20	147.4	45	斜交	5H	块状-镶嵌	平面-圆弧	8.5	36.0	142	1.20	52.5

注:  $\sigma_s$  为岩石单轴抗压强度;  $\alpha$  为结构面倾角; A 为结构面与边坡面的关系; B 为地下水条件; C 为岩体结构类型; D 为可能的破坏类型;  $c$  为内聚力;  $\theta$  为内摩擦角;  $h$  为边坡高度;  $n$  为安全系数;  $\beta$  为边坡角.

表 2 用于检验的样本  
Table 2 Samples for testing

编号	$\sigma_s$ /MPa	$\alpha$ (°)	A	B	D	$c$ /MPa	$\theta$ (°)	$h$ /m	$n$	$\beta$ (°)	F	G
1	67.7	65	平行	5H	折线形	6.0	34	462	1.15	43	层状	层状
2	72.0	65	斜交	4H	圆弧	7.0	37	154	1.2	50	块状-镶嵌	块状-镶嵌
3	64.2	65	斜交	3H	圆弧	6.4	35	138	1.20	52	块状-镶嵌	块状-镶嵌
4	82.1	50	斜交	3H	圆弧	4.1	36	100	1.2	57	层状	层状
5	147.4	45	斜交	7H	圆弧	9.0	37	137	1.2	54	层状	层状

注: F 为岩体结构类型(实际值); G 为岩体结构类型(预测值); 其他符号同表 1.

系统进行分析,接着将分析结果输入到识别边坡稳定性的专家系统,给出边坡稳定性的风险估计.针对可能发生的破坏方式,由专家系统给出合理的防治措施.在此之后的现场监测结果用来校验其合理性,并对某些情况进行适当的调整.最后将各种数据存入数据库中,为以后的分析积累实例.整个分析过程是自适应的、智能的.

此分析过程不同于先前仅仅依靠专家系统进行识别边坡稳定性的方法.先前的分析过程一般是由专家或程序员输入规则到专家系统中,这些规则是“唯专家的”,带有很强的主观性;而且,这些规则是很难获得的,大大地限制了专家系统的应用.如将数据挖掘和专家系统综合起来,便能很好地解决这些问题.数据挖掘是从数据中获取尚未被人发现的知识,是“唯数据的”,是客观的.另外,将数据挖掘所获取的各类知识用于专家系统,就能很好地解决专家系统的瓶颈问题——知识获取.

为了方便对用户所给的交易数据库 $D$ 进行数据挖掘,首先须将数据库中的数量属性分解为若干个布尔属性.可采用的方法有等分值域法、用户交互式分段法以及聚类法.这三种方法各有优缺点,等分值域法简单易行,但不利于正确规则的生成.用户交互式分段法由用户参与分段,有利于生成正确的规则,但主观因素太强.聚类法分段较准确,但使用起来稍显复杂.

数据库中影响边坡破坏类型的因素有10个,基本上考虑了各种因素.

## 2.2 模型挖掘过程及挖掘结果

由于数据挖掘对数据质量和数据量多少要求很高,要想获得影响因素和被影响因素之间准确的关系,就必须具备充足的数据样本,而且要保证数据质量,不要出现太多的噪声数据.由表2可知,数据挖掘能够较准确地进行岩质边坡的稳定性和破坏类别的预测.在进行数据挖掘的过程中,由用户指定最小支持度  $\text{minsupp}$  和最小置信度  $\text{minconf}$ .如果发现挖掘得到的规则过多或过少,此时就需要用户调整  $\text{minsupp}$  和  $\text{minconf}$  的大小,重新进行挖掘.得到规则后,将其输入专家系统,既可为边坡稳定性预测专家提供决策支持.

用本文提出的算法在支持度 8%, 置信度 99% 的前提下,发现了117条规则,根据专家经验进行过滤处理后,得到若干有用的规则,例如:

如果岩石单轴抗压强度取值在 80~90 MPa 间,地下水条件为 5H,岩体结构类型为块状,安全系数为 1.2,则边坡破坏形式为双滑块折线型破坏,且支持度为 17%,置信度为 100%.

数据挖掘得出的规则与现场监测结果以及常用的专家经验基本相符,说明经过数据挖掘系统处理后得到的结果具有很高的质量,可以应用到其他边坡工程实例中.

## 3 结论

数据挖掘比传统的分析方法更方便实用,能很好地挖掘出地质体各种信息间复杂的关系,在地质灾害预测中具有广泛的应用前景.本文建立的预测模型,数据获取方便,准确度高,对工程现场具有一定的指导意义.

## 参 考 文 献

- 冯夏庭,张治强.长江三峡工程永久船闸高边坡开挖变形的智能预测[J].辽宁工程技术大学学报,1999,18(5): 463
- 卢才金,胡厚田,徐建平,等.改进的BP网络在岩质边坡稳定性评判中的应用[J].岩土力学与工程学报,1999,18(3): 303
- 冯夏庭,王泳嘉.深部开采诱发的岩爆及其防治策略的研究进展[J].中国矿业,1998,7(5): 42
- Agrawal R, Srikant R. Fast algorithm for mining of association rules [A]. In: Proceeding of the 11th International Conference on Data Engineering [C]. Santiago,1995. 25
- Srikant R, Agrawal R. Mining quantitative association rules in large relational tables [A]. In: Proceeding of the ACM SIGMOD Conference on Management of Data [C]. Washington, 1996. 11
- 杨炳儒,张德政.关于知识发现系统的扩展性研究[J].北京科技大学学报,2000,22(1): 84
- 万军,涂序彦.基于知识的广义优化方法[J].北京科技大学学报,1999,21(4): 413
- 杨炳儒,刘发升.数据挖掘与数据库中知识发现[J].北京科技大学学报,1999,21(2): 202
- 冯夏庭,马平波.基于数据挖掘的地下硐室围岩稳定性判别[J].岩石力学与工程学报,2000,20(3): 306
- 马平波,冯夏庭.基于数据挖掘的深部采场岩爆知识的自动获取[J].东北大学学报,2000,21(6): 630
- 张朝晖,陆玉昌.发掘多值属性的关联规则[J].软件学报,1998,9(11): 801
- 胡侃,夏绍纬.基于大型数据仓库的数据挖掘[J].软件学报,1998,9(1): 53

10 Kachanov L M. On the Time to Failure under Creep Con-

dition [M]. Izv Akad Nauk USSA Otd Tekhn Nauk, 1958

### Inner Variable in Crack Healing Processing

ZHANG Yongjun<sup>1)</sup>, XUE Ling<sup>2)</sup>, HAN Jingtao<sup>1)</sup>

1) Material Science and Engineering School, University of Science and Technology Beijing, Beijing 100083, China

2) The Chinese Society for Metals, Beijing 100711, China

**ABSTRACT** According to the thermodynamics of continuous medium, the dissipation inequation of crack healing is derived from the second law of thermodynamics. An inner variable *H* is defined which describes the process of crack healing. Based on variable *H*, the crack healing prosess can be explained, the evolutionary equation and constitutive equations of the process can be established, thus making it possible to analyze the process quantitatively.

**KEY WORDS** crack; healing; inner variable; dissipation inequation

\*\*\*\*\*

(上接第 105 页)

### Application of Data Mining to Predicting the Stability of Rock Slope

ZHANG Zhiqiang<sup>1)</sup>, CAI Sijing<sup>1)</sup>, MA Pingbo<sup>2)</sup>

1) Civil and Environmental Engineering School, University Science and Technology Beijing, Beijing 100083, China 2) School of Resource & Civil Engineering, Northern University, Shenyang 110006, China

**ABSTRACT** A method was proposed for assessing the stability of rock slope, of which the knowledge about the stability of rock slopes is extracted from data cases of rock slopes and the integrated intelligent system is built for assessment. The system can be used to predict and assess the stability of any rock slope with the information measured from the rock slope.

**KEY WORDS** data mining; rock slope; stability