

DOI: 10.11779/CJGE201903011

# FRPFP 模型滑坡监测预警关联规则挖掘分析 ——以三峡库区江津到奉节段为例

赵久彬<sup>1, 2</sup>, 刘元雪<sup>1, 2</sup>, 刘娜<sup>3</sup>, 胡明<sup>1, 2</sup>

(1. 陆军勤务学院军事设施系, 重庆 401311; 2. 岩土力学与地质环境保护重庆市重点实验室, 重庆 401311; 3. 重庆市地质矿产测试中心, 重庆 400042)

**摘要:** 采用传统的关联规则用于岩土工程监测预警领域的知识发现, 在数据庞大情形下单机机器学习实时性差, 无法获得多因素综合作用的规则。由于未对前后部项进行约束, 得到的关联规则冗余度高, 含有大量不符因果逻辑的规则。基于此, 提出一种前后部项约束关联规则并行化 FRPFP (fore-part and rear-part parallel FP-growth) 算法, 并在大数据分布式处理平台 Spark 下进行实现。通过对三峡库区奉节至江津库段滑坡的孕灾因子统计分类, 采用 7 个滑坡发育基础因子和 4 个滑坡诱导因子作为前部集合, 滑坡前缘、中部、后缘监测点位移参数为后部集合, 采集研究区 25 个滑坡 11 年监测数据。以 FRPFP 算法为模型架构基于关联规则的滑坡监测预警大数据系统, 设计区域滑坡危险性规则挖掘、典型滑坡危险性规则挖掘、滑坡发生原因分析挖掘 3 个功能, 用于库岸滑坡稳定性预测和分析, 为认清库岸滑坡的破坏机制和提升其预报水平提供新的思路。

**关键词:** FRPFP 模型; 关联规则; Spark 平台; 滑坡监测预警; 三峡库区

中图分类号: P642.22

文献标识码: A

文章编号: 1000-4548(2019)03-0492-09

**作者简介:** 赵久彬(1987-), 男, 硕士, 主要从事滑坡监测预警大数据系统研究。E-mail: 459694118@qq.com。

## Association rules of monitoring and early warning by using landslides FRPFP model—Case study of Jiangjin-Fengjie reach in Three Gorges Reservoir area

ZHAO Jiu-bin<sup>1, 2</sup>, LIU Yuan-xue<sup>1, 2</sup>, LIU Na<sup>3</sup>, HU Ming<sup>1, 2</sup>

(1. Army Logistics University of PLA, Chongqing 401311, China; 2. Chongqing Key Laboratory of Geomechanics and Geoenvironment Protection, Chongqing 401311, China; 3. Chongqing Testing Center of Geology and Mineral Resources, Chongqing 400042, China)

**Abstract:** When the traditional association rules are applied to the monitoring and early warning of geotechnical engineering, the machine learning has poor real-time performance and high redundancy of association rules. Aiming at the real-time and logic requirements of the association rules in the case of massive monitoring data of landslides, a fore-part and rear-part parallel FP-growth (FRPFP) algorithm is proposed. Through the statistical classification of landslide disaster factors from Fengjie to Jiangjin of Three Gorges reservoir, 7 basic factors and 4 induced factors are set as the front set, and the displacement parameters at the monitoring points at front, middle and rear edges of the landslide are set as the rear set. In addition, the monitoring data of 25 landslides in the study area for 11 years are collected. Based on the FRPFP algorithm, a large data system of intelligent landslide monitoring and warning based on the association rules is established. The three functions, hazard criterion mining of regional landslides, hazard criterion mining of typical landslides and occurrence analysis, and mining of landslides, are designed, and implemented under the large data-distributed processing platform Spark. The engineering verification shows that the proposed model has good real-time performance and logical rules. It is used to predict and analyze the stability of the landslide on the bank, which provides a new way of thinking for identifying the failure mechanism of the bank landslides and improving the forecast level.

**Key words:** FRPFP model; association rule; spark platform; monitoring and early warning of landslide; Three Gorges Reservoir area

**基金项目:** 重庆市基础科学与前沿技术研究专项重点项目 (cstc2015jcyjBX0073); 重庆市国土资源和房屋管理局科技计划项目 (KJ-2018016); 国家自然科学基金项目 (41877219)

收稿日期: 2018-02-05

\*通讯作者 (E-mail: 459694118@qq.com)

## 0 引言

在地理信息系统、数据存储技术及物联网等技术

高速发展的信息时代,将大数据技术、数据挖掘技术与滑坡监测预警技术结合起来进行研究,运用大数据的理论、机制、模型和方法,采用数据挖掘方法应用到滑坡监测数据预测,得出可靠性较高的预警信息,是岩土监测预警大数据化、智能化发展潮流。岩土工程监测数据复杂多样,数据量随着时间增大,传统的数据库技术和迭代串行处理的机器学习方式,在巨量监测数据情形下,实时性能和效率面临严峻挑战。

关联规则是用于发现数据之间隐蔽的相互关系,在岩土工程知识发现中运用较广。李炯等<sup>[1]</sup>使用遥感影像取得三峡库区滑坡面的土地利用变化趋势,采用Apriori算法挖掘了滑坡稳定性与土地利用变化类型之间的规则,该方法仅考虑了土地利用变化因素,而三峡库区的滑坡主要以降雨和库水为诱因;揭奇等<sup>[2]</sup>采用分布式光纤技术获得边坡多场监测数据,利用Apriori算法挖掘多场滑坡多因素作用下的变形规则,该方法考虑了应变、变形、渗流、温度等因素,但在算法实现时将数据转换成布尔型,使得信息熵增大,不利于知识发现;胡畅<sup>[3]</sup>采用关联规则对三峡库区秭归——巴东段顺层滑坡变形规律进行了研究,考虑了降雨量和沿江库岸水位对滑坡发育的作用,得到了单个滑体的变形关联规则,但使用的挖掘方式为迭代串行处理,实时性和效率较低。付杰等<sup>[4]</sup>、刘小珊等<sup>[5]</sup>都利用了关联规则对滑坡的稳定性与环境因素进行了挖掘和分析,但所获得的规则前后项因果混淆,冗余量高,有效比低。

本文按照Map/Reduce并行模型原理,对FP-growth算法进行了并行化开发,提出和设计了前后部项约束关联规则并行化FRPFP算法,并在大数据并行计算分布式处理框架Spark等<sup>[6]</sup>下进行了实现;选取重庆市奉节至江津库段25个的涉水滑坡为研究区域,统计分析了滑坡地层岩性、地质模型等7个孕灾基础因子对于灾害发生的贡献比率;综合考虑了滑坡灾害基础因子与降雨量、库岸水位等诱发因子对于滑坡变形的影响,对研究区区域性滑坡、典型滑坡进行了危险性关联规则挖掘分析,通过合理设置前后部项集合,获得了冗余量低、有效比高的规则,表明该模型和方法具有使用价值,基于规则前后部项约束关联规则的滑坡监测预警大数据系统可以推广使用。

## 1 前后部项约束关联规则并行化算法FRPFP模型

采用关联规则<sup>[7]</sup>进行挖掘,得到的规则数量较大、冗余度高,用户需要对这些规则一一分析是否符合实际意义和专家经验,但用户更希望直接得出诸如:“滑

坡坡度为10~20度、高差100~200m→滑坡加速变形”的规则,这些规则符合逻辑、冗余度低。FP-growth算法是一种挖掘频繁项集关联规则的算法<sup>[8]</sup>,本文进行了并行化设计,提出了基于规则前后部项约束的关联规则挖掘算法FRPFP模型。

### 1.1 基本定义

定义1 规则前部集合 $F=\{F_1, F_2, \dots, F_n\}$ ,为用户希望的关联规则的前部项的集合,满足 $F \subset I$ 且 $F \neq \Phi$ ,其中 $I$ 为所有项的集合。 $F_S=\{F_{S1}, \dots, F_{Sj}, \dots, F_{Sg}\}$ 为 $F$ 所有非空子集的集合,其中 $F_{Sj}(j=1, 2, 3, \dots, g)$ 为 $F$ 的非空子集,称为前部项集。 $C_{Fj}$ 称为前部频繁候选项集,大于等于最小支持度(min\_sup)的 $C_{Fj}$ 称为前部频繁项集,记为 $L_{Fj}$ , $L_F$ 为前部频繁项集集合。

定义2 规则后部集合 $R=\{R_1, R_2, \dots, R_m\}$ ,为用户希望得到的关联规则的后部项的集合,满足 $R \subset I$ 且 $R \neq \Phi$ 。 $R_S$ 为 $R$ 所有非空子集的集合。 $R_S=\{R_{S1}, \dots, R_{St}, \dots, R_{Sn}\}$ , $R_{St}(t=1, 2, 3, \dots, n)$ 为 $R$ 的非空子集,称为后部项集。

定义3 给定关联规则 $X \rightarrow Y$ ,若 $X \subseteq F$ , $Y \subseteq R$ ,则规则 $X \rightarrow Y$ 为前后部项约束关联规则。

定义4 给定前后部项集集合 $I_{FR}=F_S \times R_S=\{I_{FR1}, \dots, I_{FRt}, \dots, I_{FRd}\}$ ,那么其中的任一项集 $I_{FRt}(t=1, 2, 3, \dots, d)$ 是同时包含规则前部项和规则后部项的项集即 $I_{FRt} \subseteq (F \cup R)$ ,且 $I_{FRt} \cap F \neq \Phi$ , $I_{FRt} \cap R \neq \Phi$ , $I_{FRt}$ 简称为前后部项集。

前后部频繁项集候选项集记为 $C_{FRt}$ ,若某个项集 $I_{FRt}$ 大于等于min\_sup,则称为前后部频繁项集,记为 $L_{FRt}$ 。 $L_{FRt}$ 的集合记为 $L_{FR}$ 。挖掘前后部项的关联规则,关键是求出集合 $L_{FR}$ 及集合 $L_F$ 。

对于前后部项约束规则 $X \rightarrow Y$ ,由置信度定义 $\text{Confidence}(X \rightarrow Y) = \frac{\text{Sup\_count}(XUY)}{\text{Sup\_count}(X)}$ 可知,需要求出 $(X \cup Y)$ 和 $(X)$ 对应的前后部频繁项集 $L_{FRt}$ 和前部项频繁项集 $L_{Fj}$ ,计算它们的支持度,由于每个 $L_{FRt}$ 有且只有一个 $L_{Fj}$ 与之对应,故只能生成一条关联规则,是能够减少冗余规则产生的原因所在。

因此,FRPFP算法的思想在于:通过挖掘出所有的前后部频繁项集 $L_{FRt}$ 和前部频繁项集 $L_{Fj}$ ,来达到求取符合项约束条件的关联规则以及减少关联规则冗余量的目的。整个算法流程均是围绕如何挖掘出所有的 $L_{FRt}$ 和 $L_{Fj}$ 来进行求解。

### 1.2 FRPFP模型算法思想及流程

如图1所示,本文提出的FRPFP算法分为4个步骤,运用2个Map/Reduce并行编程模型来完成。

步骤1 计算频繁1-项集。此步骤通过一次Map/Reduce来完成。采用Map/Reduce里的词频统计思

想, 频繁1-项集的结果采用表 $M_{FR}$ -list来描述。表 $M_{FR}$ -list的元素包括4个域: 项名, 支持度计数, 标记和指针。首先, 项(item)按照支持度计数值(sup-count)从大到小排序; 标记域(label)用来描述某个项是否属于前部项集合 $F$ 或后部项集合 $R$ , 如果该项属于前部项集合, 则标记为f, 如果该项属于后部项集合, 则标记为r; 加入指针域(link)则是使得表 $M_{FR}$ -list具有项头表的功能, 用于链接下一步构建的FP-tree上的节点。

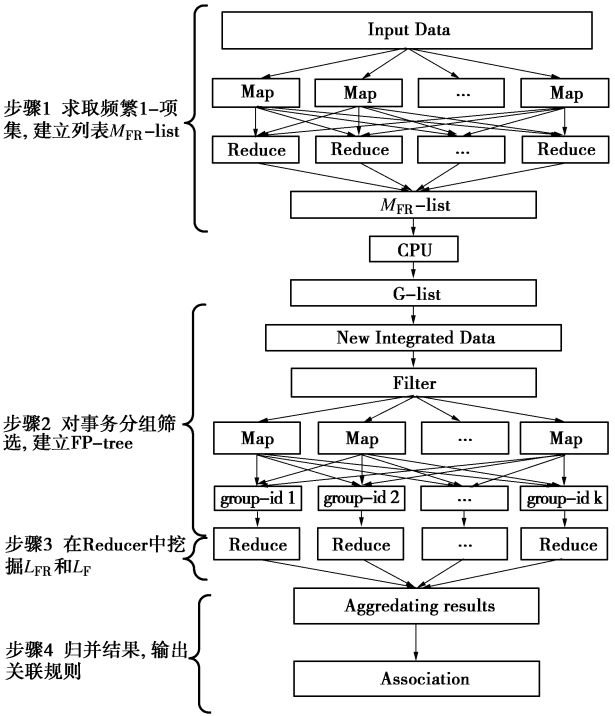


图1 FRPFP 算法流程图

Fig. 1 Flow chart of FRPFP algorithm

步骤2 分组建立FP树。此操作通过一次Map/Reduce来完成。在Map阶段, 首先, 根据项的支持度大小, 对事务集中的所有事务由大到小重新排列。其次, 按照分组列表G-list对事务集进行分组并筛选事务, 压缩分组事务集的空间。在Reduce阶段, 对压缩后的分组事务集, 根据FP-Growth的方法构建分组FP-tree。

步骤3 挖掘 $L_{FR}$ 和 $L_F$ 。在步骤2的Reducer建立的各分组FP-tree上, 挖掘出所有的前后部频繁项集和前部频繁项集, 并分别放入集合 $L_{FR}$ 和 $L_F$ 中。

步骤4 输出关联规则。将各组挖掘结果归并至同一计算节点, 遍历 $L_{FR}$ , 针对每个 $L_{FRi}$ 若存在对应的前部频繁项集 $L_{Fj}$ , 则根据每个前后部频繁项集的置信度  $Confidence(X \rightarrow Y) = P(B|A) = \frac{P(X \cup Y)}{P(X)}$ , 若大于等于最小置信度 (min\_conf), 则输出强关联规则。若不存在对应的前部频繁项集 $L_{Fj}$ , 则不考虑。

## 2 库岸滑坡数据监测与处理

### 2.1 研究区介绍

本文选取重庆市奉节至江津库段, 重庆辖区 15 个县(区)范围内 25 个的涉水滑坡为研究区域, 干流长为 441.50 km。以低山丘陵地貌为主, 主要为侏罗系地层, 广泛分布砂岩和泥岩<sup>[9]</sup>。气候特点为湿度大、降雨充沛, 夏季主要以暴雨形式出现。该研究区内年平均平均降雨量为 996.7~1309.9 mm, 一年之内降雨不均匀, 3 月至 9 月夏季降雨较多。库区内地下水丰富, 常与降雨与水库水位相关, 主要补给为大气降雨, 水库上涨时库水倒灌入滑坡体。库区水位为 145~175 m, 变幅高达 30 m, 水位变动会使水对库岸坡体反复冲刷、浸泡, 使涉水滑坡稳定性的受到干扰<sup>[10]</sup>。

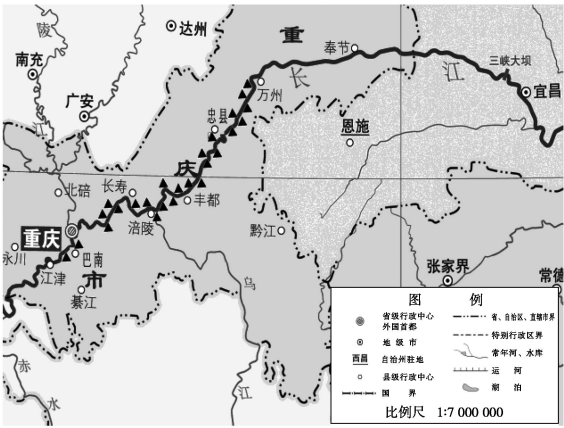


图2 研究区 25 个滑坡分布

Fig. 2 Distribution of 25 landslides in study area

### 2.2 库岸滑坡发育因子分类

结合地质灾害专家经验知识, 本文选取地貌、地层、坡度、高差、坡体体积、地下水、地质模型等 7 个因素作为孕灾的基础因子<sup>[11]</sup>, 选取三日最大降雨量、月降雨量、库岸水位、库岸水位降速等作为孕灾的诱发因子<sup>[12]</sup>。基础因子采用贡献率大小进行分类, 诱发因子由监测数据产生, 采用 K-means 聚类方法分类<sup>[13]</sup>。

为了分析孕灾基础因子对滑坡稳定性的作用, 本文参考了乔建平<sup>[14]</sup>利用地层岩性对滑坡危险性贡献率的方法, 统计具有典型意义的滑坡 230 余处资料, 采集并计算 7 类 21 个孕灾基础因子的贡献比率。计算过程如下:

设单个孕灾基础因子  $x_i \in X, i=1,2,\dots,n$ , 其中  $X$  为一类因子集合, 求出  $x_i$  的贡献占比:

$$D = n / m \quad (1)$$

式中,  $n$  为含有  $x_i$  因子的滑坡数量,  $m$  为该类因子  $X$  滑坡总数。对  $D$  采用等差数列划分为高、中、低 3 类,

即区间等差数列:

$$d = \frac{D_{\max} - D_{\min}}{3} \quad (2)$$

由小至大取三级划分点:

$$\begin{aligned} a_1 &= D_{\min}, \\ a_2 &= D_{\min} + d, \\ a_3 &= D_{\max} - d, \\ a_4 &= D_{\max}. \end{aligned} \quad (3)$$

设低贡献指数区因子为  $x_i \in [a_1, a_2)$ , 中贡献指数区因子为  $x_j \in [a_2, a_3)$ , 高贡献指数区因子为  $x_k \in [a_3, a_4]$ 。为了求 3 个等级基础因子对滑坡发育的贡献比率, 首先对基础因子贡献占比  $D_i, D_j, D_k$  进行均值化处理, 求出各区间的平均贡献占比, 最后对 3 个平均贡献占比进行归一化处理, 得到各基础因子的贡献比率, 如表 1 到表 7 所示。

表 1 地层岩性分类

Table 1 Classification of strata and lithology		
类别	地层岩性	贡献比率
B1	J <sub>2</sub> S	65.9%
B2	J <sub>3</sub> S	27.1%
B3	J <sub>2</sub> X J <sub>2</sub> XS J <sub>3</sub> P J <sub>1-2</sub> Z J <sub>1</sub> Z T <sub>3</sub> Xj T <sub>2</sub> b T <sub>1j</sub> P <sub>2</sub> T <sub>1</sub> d	7.0%

表 2 坡度分类

Table 2 Classification of slope gradient		
类别	坡度范围/(°)	贡献比率/%
E1	10~20	59.5
E2	0~10, 20~30	32.0
E3	>30	8.5

表 3 高差分类

Table 3 Classification of height difference		
类别	高差范围/m	贡献比率/%
F1	0~100	65.9
F2	100~200	27.1
F3	>200	7.0

表 4 地貌分类

Table 4 Classification of geomorphology		
类别	地貌类型	贡献比率/%
G1	低山	60.7
G2	丘陵	27.4
G3	中山	11.9

表 5 灾体体积分类

Table 5 Volume classification of disaster bodies		
类别	灾体体积/m <sup>3</sup>	贡献比率/%
H1	490	65.9
H2	1268	27.1
H3	2254	7.0

表 6 地下水分类

Table 6 Classification of groundwater conditions		
类别	地下水情况	贡献比率/%
I1	地下水丰富	60.7
I2	地下水较丰富	27.4
I3	地下水一般	11.9

表 7 地质模型分类

Table 7 Classification of geological models for landslides		
类别	高差范围	贡献比率/%
J1	顺层	56.0
J2	水平	37.0
J3	斜交、逆层	7.0

2.3 监测方法及数据处理

研究区内 25 个涉水滑坡以地表绝对位移监测、降雨量监测、库岸水位监测为主, 辅以地表巡查, 所有监测点数据实时采集, 连续解算<sup>[15]</sup>。每个滑坡采集位于滑坡前缘、中部、后缘各 2 个绝对位移监测数据, 库水位监测数据和降雨量监测数据, 监测周期为每月监测一次<sup>[16]</sup>。截至 2016 年 12 月共 11 年, 共产生 19650 条监测数据, 存储于 Hadoop 集群<sup>[17]</sup>的 master 节点中, 关系型数据采用 mysql 数据库存储, 非关系型数据如宏观监测报告、照片等, 存储在 Hadoop 集群自带的 Hbase 非关系型数据库中。

(1) 降雨和库水数据聚类分析

根据专家经验知识, 降雨和库水是导致研究区内涉水滑坡的诱发因子<sup>[18]</sup>, 降雨和库水通过软化、饱水加载、静水压力、动水压 4 个作用造成滑坡变形。本文选取三日最大降雨量、月降雨量、库水位、库水位降速监测作为诱发因子数据, 采用 K-means 算法对诱发因子数据进行聚类并分类。由于降雨越大对滑坡发育诱发作用越大, 月平均降雨量、三日最大降雨量数据均按照从小到大的顺序聚类为 3 类, 聚类中心最大的类为滑坡贡献最大类; 库水位、水位降速均按照升序的顺序聚类为 3 类, 聚类中心最大的类为滑坡贡献最大类。表 8~11 举例列出了研究区内, 重庆市忠县某滑坡三日最大降雨量、月降雨量、水位降速和库水位分类情况。

表 8 三日最大降雨量分类

Table 8 Classification of maximum rainfall in three days			
类别	聚类中心/0.1mm	包含数目/个	贡献大小
L1	133.9	71	小
L2	431.7	44	中
L3	907.7	16	大

表 9 月降雨量分类

Table 9 Classification of monthly rainfall			
类别	聚类中心/0.1mm	包含数目/个	贡献大小
M1	358.6	64	小
M2	1227.5	46	中
M3	2911.3	21	大

表 10 库水位降速分类

Table 10 Classification of falling speed of reservoir water levels			
类别	聚类中心/0.1m	包含数目/个	贡献大小
N1	-11.41	22	小
N2	-0.09	70	中
N3	9.69	39	大

表 11 库水位分类

Table 11 Classification of reservoir water levels

类别	聚类中心/0.1m	包含数目/个	贡献大小
O1	144.31	27	小
O2	156.88	47	中
O3	170.92	57	大

(2) 地表位移数据聚类分析

滑坡的变形主要经历缓慢变形、累进变形和加速变形 3 个阶段<sup>[19]</sup>。本文通过对该区域内 25 个滑坡地表位移的监测数据进行清洗和整理后, 同样调取 K-means 聚类方法将每个滑坡各监测点的月位移分为 3 类, 如表 12 列举了忠县某滑坡位于坡体前缘的监测点 Z309 的分类情况。在监测过程中, 若监测发现有 3 个监测点以上处于加速变形阶段, 就可以认为该滑坡处于不稳定状况, 发布预警信息, 做好人员财产迁移准备; 若监测发现只有中前部有 2 个监测点处于加速变形阶段, 则可认为该滑坡属于牵拉式滑坡, 应对滑坡前缘采取防治措施; 若监测发现中后部有 2 个监测点处于加速变形阶段, 而其余 4 个监测点未进入该阶段, 则可认为是推动式滑坡, 应对坡体后缘采取治理; 若前缘后缘中部均有 1 个以上监测点位移处于加速变形阶段, 则该滑坡属于平移式滑坡, 需要对该滑坡采取全面防治措施。

表 12 忠县某滑坡监测点位移分类

Table 12 Classification of displacement at monitoring of landslides points in Zhongxian

监测点	类别	位移聚类中心/mm	所属阶段	包含数目	所属部位
Z309	O1	48.5422	缓慢变形阶段	82	前缘
	O2	370.3887	累进变形阶段	37	
	O3	889.5758	加速变形阶段	12	

本文采集了研究区内 25 个典型大中型涉水滑坡 11 年的监测数据, 共 19650 条监测数据, 经过前期数据清理, 滑坡依次编号为 V1~V25, 将滑坡发育的基础因子、诱导因子和滑坡位移监测等级划分数据按照时间顺序, 合并生成一个数据集, 存放在 HDFS 文件系统中, 待下一步关联规则挖掘时调用, 生成的数据集如表 13 所示。

3 滑坡危险性规则挖掘和分析

本文提出的 FRPFP 模型在 Spark 框架下, 采用 Scala 编程实现。该算法最终部署到实验室的分布式 Spark 集群中。该集群的硬件配置为: 3 台 CPU 主频为 3.6 GHz, 内存为 16G 的台式机。集群的环境配置为: 操作系统 Ubuntu16.04, Hadoop 版本 Hadoop2.7.2,

Spark 版本为 Spark1.6.1, 算法的编程语言为 scala2.10.4, JDK 版本为 1.7.0\_51。Master 节点 1 个, Slave 节点两个。

滑坡发育的基础因子和诱导因子与监测点位移大小之间有密切的关系, 通过本文提出的 FRPFP 模型应用到该研究区监测数据挖掘和知识发现, 针对用户感兴趣的区域性滑坡、局部滑坡群和单体滑坡, 合理设置前部为孕灾因子, 后部集合为监测点位移变形类别, 可以分别生成区域性滑坡危险性关联规则、典型滑坡危险性关联规则; 通过对换前后部集合, 可以生成滑坡发生情况下, 挖掘滑坡发生的原因分析规则。具体功能如图 3 所示。

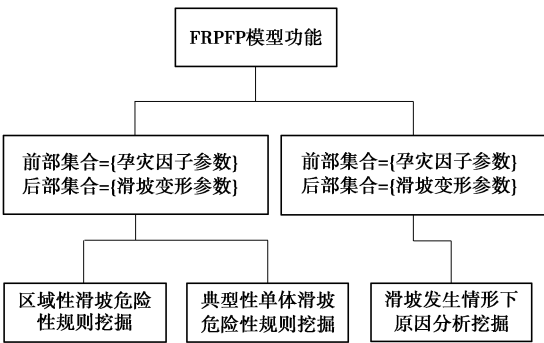


图 3 FRPFP 模型功能图

Fig. 3 Function of FRPFP model

3.1 区域性滑坡危险性规则挖掘

由于该研究区在地层岩性、地貌、气候等具有共同特征<sup>[20]</sup>, 基于此, 对该区域进行关联规则知识发现, 得到区域滑坡危险性规则, 具有研究意义和使用价值。本文提出的 FRPFP 关联规则模型应用于研究区内滑坡的 11 个孕灾因子与 6 个监测点变形阶段知识发现, 需要提前设置前后部集合。由于挖掘前未知滑坡诱因, 将 11 个孕灾因子 33 个等级划分参数均输入为前部集合, 设置前部集合为  $F=\{B1, B2, B3, D1, D2, D3, E1, E2, E3, F1, F2, F3, G1, G2, G3, I1, I2, I3, J1, J2, J3, L1, L2, L3, M1, M2, M3, N1, N2, N3, O1, O2, O3\}$ ; 设置后部集合时, 感兴趣的是监测变形量处于加速变形的阶段, 只需取 6 个监测点处于加速变形阶段对应的等级划分参数, 后部集合为  $R=\{P3, Q3, R3, S3, T3, U3\}$ ; 设置  $min\_sup=0.10$ ,  $min\_conf=0.6$ 。

在 Hadoop 集群环境下, 启动 Hadoop 集群, 再启动 Spark 集群, 使用 Spark-submit 提交任务, 执行 FRPFP 模型算法, 并将挖掘结果以 txt 文件形式存储, 得到三峡库区奉节县白帝城至江津库段区域性滑坡关联规则如表 14 所示, 运行时间为 297349 ms。



表 13 待挖掘数据集  
Table 13 Dataset to be mined

监测日期	地貌	地层岩性	坡度	高差	体积	地下水	地质模型	月最大降雨量	月降雨量	水位降速	库岸月水位	前缘监测点一	前缘监测点二	中部监测点一	中部监测点二	后缘监测点一	后缘监测点二	滑坡编号
2006-01-20	B3	D3	E1	F2	G1	I2	J3	L1	M1	N2	O1	P1	Q2	R1	S1	T1	U1	V1
2006-02-20	B3	D3	E1	F2	G1	I2	J3	L1	M1	N2	O1	P1	Q3	R1	S1	T1	U1	V1
2006-03-20	B3	D3	E1	F2	G1	I2	J3	L1	M1	N2	O1	P1	Q2	R1	S1	T1	U1	V1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

表 14 区域性滑坡关联规则 Table 14 Association rules of regional landslide			
编号	前项	后项	置信度
1	B1,D3,E3,F3,G2,I3,J2,O3,M3	T4,S4,Q4,P4	76.56%
2	B1,D3,F3,G2,I3, J2,O3,M3	T4,S4,Q4,P4	76.36%
3	B1,D3,E3,G2,I3,J2,O3,M3	T4,S4,Q4,P4	73.47%
4	B1,D3,E3,G2,J2,F3,O3,M3	T4,S4,Q4,P4	73.47%
5	B1,D3,E3,F3,I3,G2,O3,M3	T4,S4,Q4,P4	73.47%
6	D1,G2,L2, M3,N2	R4,Q4,P4	94.73%
7	D1,G2,I3,J1,L2,M3,N3	R4,Q4,P4	92.30%
8	D1,G2,L2,J1,M3,N3	R4,Q4,P4	83.72%
9	D1,G2,I3,L2,M3,N3	R4,Q4,P4	92.31%
10	D1,G2,J1,L2.M2,N3	R4,Q4	97.98%
11	B2,D1,E1,G2,I2,J2, L2,M2,N3,O1	P4,Q4	100%
13	B1,D2,E2,F2,I3,J2, L3,M2,N1,O2	R4,S4	74.03%
14	B1,D2,F2,G2,I3,J2, M2,N1,O3	R4,S4	100%
15	B2,E2,F2,G1,J2,L2,M2,N1,O3	U4,T4	100%
16	D1,E3,F1, G1, J2, L1,M1,N1,O2	U4,T4	100%

规则 1~5 中, 规则前项都包含 M3 (月降雨量大) 和 O3 (库岸水位=高), 说明三峡库区奉节至江津段整体对月降雨量和库岸水位响应较强, 在月降雨量大和库岸水位高的时, 出现 4 个监测点位移处于加速变形阶段几率较大, 坡体易发生滑移。这是由于该研究区强降雨和库岸水位提高, 使得地下水位急剧抬升, 孔隙水压力和浮托力骤然增加, 潜在滑移面的土层被软化, 内聚力和摩擦角都减小, 使坡体处于不稳定状态。

规则 6~10 中, 规则前项中较多含有 D1 (地层岩性=J<sub>2s</sub>)、J1 (地质模型=顺层)、N3 (水位降速=最大), 说明 J<sub>2s</sub> 这种地层的顺层滑坡, 在水位降速较大时, 滑坡中前部 3 个监测点处于加速变形阶段, 滑坡处于欠稳定状态。这是由于水位快速下降时, 坡体渗透性

低导致地下水的排出过缓, 使坡体浸润线滞后而产生孔隙水压力, 同时地下水位高于库水位, 产生了水力梯度, 这两方面增大了滑体的下滑力, 导致滑坡处于欠稳定状态。

规则 11~15 中, 规则前项包含 B1、B2 (地貌=丘陵、低山), 坡度属性为 E1、E2 (边坡坡度=0° ~ 30°), 高差属性为 F1、F2 (高差=0~200 m), 规则后项产生了 2 个加速变形的监测点, 变形的规则与第二节的滑坡属性贡献率规律一致, 说明通过 FRPFP 模型得到的规则和统计学的规律相印证, 具有可靠性。

### 3.2 典型滑坡危险性规则挖掘

本文选取研究区内丰都县某古滑坡为例, 地处三峡工程腹心。该滑坡是土质顺向坡, 由两个边坡组成, 高差分布为 128~277 m, 滑坡总面积为 43.7×10<sup>4</sup> m<sup>2</sup>, 体积 448×104 m<sup>3</sup>, 大致主滑方向约 295°。滑体物质组成为碎石土, 滑床为砂岩、砂质页岩及泥岩。目前稳定性差, 滑坡前缘位于蓄水位 175 m 以下。

将 11 个孕灾因子 33 个等级划分参数均输入为前部集合, 设置前部集合为  $F=\{B1, B2, B3, D1, D2, D3, E1, E2, E3, F1, F2, F3, G1, G2, G3, I1, I2, I3, J1, J2, J3, L1, L2, L3, M1, M2, M3, N1, N2, N3, O1, O2, O3\}$ ; 不同之处在于设置后部集合时, 该滑坡编号为 V4, 将该参数加入到后部集合中, 后部集合为  $R=\{P3, Q3, R3, S3, T3, U3, V4\}$ ; 设置 min\_sup=0.05, min\_conf=0.6。得到该单体滑坡的关联规则如表 15 所示, 运行时间为 153791 ms。

规则 1~4 在月降雨量大、三日降雨量大时滑坡体前缘、中部和后缘均有监测点处于加速变形阶段, 这是由于该滑坡物质大部分为碎屑岩, 高强度的降雨通过孔隙进入滑坡体内, 使得滑坡体地下水位上升, 在一定程度上促进了该滑坡的整体变形; 规则 5, 6, 10 说明在水位降速最大时, 滑坡中部、后缘会处于加速变形阶段, 这是由于碎屑岩的主要成分主要位长石和石英, 长石亲水性较强, 接触到水同时产生较厚的液

表 15 丰都某滑坡关联规则  
Table 15 Association rules of a landslide in Fengdu

编号	前项	前项规则意义	后项	变形部位	置信度
1	L3,M3,N1,O1	三日最大降雨量=76.1 mm (大), 月降雨量=242.92 mm (大), 水位降速=-11.41 m (小), 库水位=144.31(低)	U3,S3,Q3,V4	前中后	100%
2	L3,M2,N1,O2	三日最大降雨量=76.1 mm (大), 月降雨量=110.37 mm (中), 水位降速=-11.41 m (小), 库水位=156.88 m (中)	R3,S3,V4	中	100%
3	L3,M3,N1,O1	三日最大降雨量=76.1mm (大), 月降雨量=242.92 mm (大), 水位降速=-11.41 m (小), 库水位=144.31 (低)	S3,Q3,V4	前中	100%
4	L3,M3,N1,O1	三日最大降雨量=76.1 mm (大), 月降雨量=242.92 mm (大), 水位降速=-11.41 m (小), 库水位=144.31 (低)	U3,Q3,V4	前后	100%
5	L2,M1,N3,O3	三日最大降雨量=33.49 mm (中), 月降雨量=35.578 mm (小), 水位降速=9.69 m (大), 库水位=170.92 m (高)	U3,S3,V4	中后	100%
6	M1,N3,L2	三日最大降雨量=33.49 mm (中), 月降雨量=35.58 mm (小), 水位降速=9.69 m (快)	U3,S3,V4	中后	100%
7	M2,N1,O3	月降雨量=110.37 mm (中), 水位降速=-11.41 m (慢), 库水位=170.92m (高)	R3,S3,V4	中	100%
8	L2,M2,N1,O3	三日最大降雨量=33.49 mm (中), 月降雨量=110.37 mm (中), 水位降速=-11.41 m (小), 库水位=170.92 m (高)	R3,S3,V4	中	100%
9	L1,M2,N2,O3	三日最大降雨量=11.12 mm (小), 月降雨量=110.37 mm (中), 水位降速=-0.09 m (中), 库水位=170.92m (高)	R3, P3,V4	前中	100%
10	L2,M1,N3,O3	三日最大降雨量=33.49 mm (中), 月降雨量=35.58 mm (小), 水位降速=9.69 m (大), 库水位=170.92 m (高)	U3,S3,V4	中后	100%
11	M1,N3,L2	三日最大降雨量=33.49 mm, 月降雨量=35.58 mm, 水位降速=-0.09 m	U3,S3,V4	中后	100%
12	M2,N1,O3	月降雨量=110.37 mm (小), 水位降速=-11.41 m (小), 库水位=170.92 m (高)	R3,S3,V4	中	66.67%
13	L2,M2,N1,O3	三日最大降雨量=33.49 mm (中), 月降雨量=110.37 mm (中), 水位降速=-11.41 m (小), 库水位=170.92 m (高)	R3,S3,V4	中	66.67%

体膜, 使得坡体渗透系数变小, 另长石常常次生吸水性较强的高岭土和绢云母, 这些物质吸水膨胀堵塞原来岩土中的孔隙, 使得滑坡体渗透性差, 在库水快速下降时, 由于滑坡体渗透性差, 坡体的地下水水位会高于库水位, 使坡体稳定性降低, 这与研究区内滑坡危险性规则类似。由于滑坡体前缘被库水淹没, 库水快速下降时主要是滑坡体中部、后缘产生变形。规则 7, 8, 9, 11, 12, 13 是滑坡处于蓄水阶段, 处于库水缓慢上升、三日最大降雨量中、月降雨量小阶段, 坡体发生不稳定变形。这是由于此阶段是每年的 9 月底左右开始蓄水, 逐渐使三峡水位蓄至 175 m 的非汛期水位, 蓄水阶段会使江水倒灌入滑坡体, 使得坡体地下水增多, 浸软层力学性能变差, 并对前缘坡体部分产生托浮力, 滑坡体产生失稳现象。但是由于该滑坡体对降雨的响应更为明显, 库岸水位缓慢上升对滑坡体只引起短暂失稳, 蓄水稳定后, 该滑坡体又处于

稳定状态。  
滑坡发生情形下原因分析的挖掘方面, 根据发生滑坡的部位为前缘、后缘、中部或者整体, 设置前部集合为: {前缘、中部或后缘监测点处于加速变形的参数}, 后部集合设置为: {滑坡的孕灾因子}, 设置合理的最小支持度和最小置信度, 则可以得出可靠度高的原因分析规则。

4 结 论

本文采用 FRFPF 模型对三峡库区奉节至江津段滑坡的监测数据进行了关联分析, 得出如下结论:  
(1) 由于该模型采用了前后部项约束, 与传统的关联算法相比较, 该模型得到的规则数量较少, 冗余度低, 有效比高, 且得到的规则能够反映孕灾因子与滑坡变形的逻辑关系, 能够全面反映滑坡受力学特征、诱发因素、外界环境变化等因素的影响, 与现场监测



结果和专家经验规律相符, 具有更强的实用意义;

(2) 通过实例分析进一步表明, 该模型在挖掘复杂影响因素的区域滑坡危险性判据、典型滑坡危险性判据, 考虑了 11 类 33 种孕灾因子的影响, 在监测数据巨量情况下, 得到的规则能够真实反映滑坡变形机理, 危险性判据具有很高的可靠性, 为研究滑坡变形特征和其预报判据提供了有效途径;

(3) 研究过程中进行区域性滑坡危险性规则挖掘时, FRPFP 模型在 Spark 单机模式 (Local) 运行时间为 897993.98 ms, 在 Spark 集群框架 (Cluster) 模式下运行时间为 297349 ms, 说明该模型由于采用并行化设计, 运行实时性更强, 效率更高。

### 参考文献:

- [1] 李 炯, 牛瑞卿. 滑坡稳定性与土地利用关联规则挖掘及应用——以三峡库区秭归到巴东段为例[J]. 计算机工程与应用, 2013, **49**(6): 234 - 239. (LI Jiong, NIU Rui-qing. Association rule mining and using between stability and land use of landslides—taking area from Ziguito Badong in Three Gorges Reservoir Area as an example[J]. Computer Engineering and Applications, 2013, **49**(6): 234 - 239. (in Chinese))
- [2] 揭 奇, 施 斌, 罗文强, 等. 基于 DFOS 的边坡多场信息关联规则分析[J]. 工程地质学报, 2015, **23**(6): 1146 - 1152. (JIE Qi, SHI Bin, LUO Wen-qiang, et al. Analysis of DFOS-based association rules in multi-fields data monitoring of slopes[J]. Journal of Engineering Geology, 2015, **23**(6): 1146 - 1152. (in Chinese))
- [3] 胡 畅. 三峡库区秭归—巴东段典型顺层滑坡预测判据研究[D]. 武汉: 中国地质大学, 2013. (HU Chang. Study on the prediction criterion of bedding landslide in the area from Zigui to Badong segment of the Three Gorges Reservoir area[D]. Wuhan: China University of Geosciences, 2013. (in Chinese))
- [4] 付 杰, 牛瑞卿. 三峡库区老蛇窝滑坡变形规律数据挖掘[J]. 人民长江, 2012, **43**(23): 6 - 9, 13. (FU Jie, NIU Rui-qing. Data mining of the deformation law of the tiger landslide in the Three Gorges Reservoir Area[J]. Yangtze River, 2012, **43**(23): 6 - 9, 13. (in Chinese))
- [5] 刘小珊, 罗文强, 李飞翱, 等. 基于关联规则的滑坡演化阶段判识指标[J]. 地质科技情报, 2014, **33**(2): 160 - 164. (LIU Xiao-shan, LUO Wen-qiang, LI Fei-ao, et al. Identification index of landslide evolution phase based on association rules[J]. Geological Science and Technology Information, 2014, **33**(2): 160 - 164. (in Chinese))
- [6] SETHI K K, RAMESH D. HFIM: a Spark-based hybrid frequent itemset mining algorithm for big data processing[J]. J Supercomput, 2017, **73**(8): 3652 - 3668.
- [7] TAN J. Different types of association rules mining review[J]. Applied Mechanics & Materials, 2013, **241-244**: 1589 - 1592.
- [8] 章志刚, 吉根林. 一种基于 FP-growth 的频繁项目集并行挖掘算法[J]. 计算机工程与应用, 2014, **50**(2): 103 - 106. (ZHANG Zhi-gang, JI Gen-lin. Parallel algorithm for mining frequent item sets based on FP-growth[J]. Computer Engineering and Applications, 2014, **50**(2): 103 - 106. (in Chinese))
- [9] 刘传正, 刘艳辉, 温铭生, 等. 长江三峡库区地质灾害成因与评价研究[M]. 北京: 地质出版社, 2007: 51 - 86. (LIU Chuang-zheng, LIU Yan-hui, WEN Ming-sheng, et al. Study on the causes and evaluation of geological disasters in the Three Gorges Reservoir area of the Yangtze River[M]. Beijing: Dizhi Press, 2007: 51 - 86. (in Chinese))
- [10] 王尚庆, 徐进军, 罗 勉. 三峡库区白水河滑坡险情预警方法研究[J]. 武汉大学学报(信息科学版), 2009, **34**(10): 1218 - 1221. (WANG Shang-qing, XU Jin-jun, LUO Mian. Study on the risk warning method of Baishuihe landslide in Three Gorges Reservoir Area[J]. Geomatics and Information Science of Wuhan University (Information Science Edition), 2009, **34**(10): 1218 - 1221. (in Chinese))
- [11] 祁生文, 伍法权, 常中华, 等. 三峡地区奉节县城缓倾层状岸坡变形破坏模式及成因机制[J]. 岩土工程学报, 2006, **28**(1): 88 - 91. (QI Sheng-wen, WU Fa-quan, CHANG Zhong-hua, et al. Mechanism and model for deformation of bank slope with slightly inclined soil layers in Fengjie County of Three Gorges[J]. Chinese Journal of Geotechnical Engineering, 2006, **28**(1): 88 - 91. (in Chinese))
- [12] 黄 达, 顾东明, 陈智强, 等. 三峡库区塔坪 H2 古滑坡台阶状复活变形的库水—降雨耦合作用机制[J]. 岩土工程学报, 2017, **39**(12): 2203 - 2211. (HUANG Da, GU Dong-ming, CHEN Zhi-qiang, et al. Hybrid effects of rainfall and reservoir level fluctuation on old Taping H2 landslide in Wushan County in Three Gorges Reservoir area[J]. Chinese Journal of Geotechnical Engineering, 2017, **39**(12): 2203 - 2211. (in Chinese))
- [13] 张 俊, 殷坤龙, 王佳佳, 等. 三峡库区万州区滑坡灾害易发性评价研究[J]. 岩石力学与工程学报, 2016, **35**(2): 284 - 296. (ZHANG Jun, YIN Kun-long, WANG Jia-jia, et al.

- Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir[J]. Chinese Journal of Rock Mechanics and Engineering, 2016, **35**(2): 284 - 296. (in Chinese))
- [14] 乔建平, 吴彩燕, 田宏岭. 三峡库区云阳—巫山段地层因素对滑坡发育的贡献率研究[J]. 岩石力学与工程学报, 2004, **23**(17): 2920 - 2924. (QIAO Jian-ping, WU Cai-yan, TIAN Hong-ling. Contribution rate research of stratum to landslide growth of Yunyang-Wushan segment in Three Gorges Reservoir region[J]. Chinese Journal of Rock Mechanics and Engineering, 2004, **23**(17): 2920 - 2924. (in Chinese))
- [15] 许星宇, 朱鸿鹄, 张巍, 等. 基于光纤监测的边坡应变场可视化系统研究[J]. 岩土工程学报, 2017, **39**(增刊 1): 96 - 100. (XU Xing-yu, ZHU Hong-hu, ZHANG Wei, et al. Development of fiber optic monitoring-based visualization system for strain fields of slopes[J]. Chinese Journal of Geotechnical Engineering, 2017, **39**(S1): 96 - 100. (in Chinese))
- [16] HUANG S, LI C, LUO L. Landslide monitoring network establishment within unified datum and stability analysis in the Three Gorges Reservoir area[J]. Journal of Sensors, 2017(9):1 - 13.
- [17] PAL A, AGRAWAL S. An experimental approach towards big data for analyzing memory utilization on a Hadoop cluster using HDFS and MapReduce[C]// First International Conference on Networks & Soft Computing. IEEE, 2014, **64**(18): 442 - 447.
- [18] ZHAO N, HU B, YI Q, et al. The coupling effect of rainfall and reservoir water level decline on the Baijiabao Landslide in the Three Gorges Reservoir area, China[J]. Geofluids, 2017(12): 1 - 12.
- [19] WU X, ZHAN F B, ZHANG K, et al. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China[J]. Environmental Earth Sciences, 2016, **75**(2): 146.
- [20] 丁仁杰, 汪成民, 胡正华, 等. 试论长江三峡库区地震、地质灾害及其监测、预报与防治[J]. 自然灾害学报, 2001, **10**(3): 71 - 78. (DING Ren-jie, WANG Cheng-min, HU Zheng-hua, et al. Seismic and geological disasters in Three Gorges Reservoir area and their prevention and control[J]. Journal of Natural Disasters, 2001, **10**(3): 71 - 78. (in Chinese))

## “我与《岩土工程学报》”征文通知

《岩土工程学报》由我国六个一级学会主办, 自1979年创刊至今已经走过了40个春秋。40年来, 在岩土工程界同仁及广大读者的关心支持和全体编者的共同努力下, 学报已经成为我国岩土工程领域中具有重要影响的学术期刊, 是岩土工程理论和实践的重要论坛, 是我国水利水电、交通建筑和油矿资源等领域中从事岩土工程勘测、设计、施工、科学研究和教学等工作人员发表学术观点、交流实践经验的重要园地。《岩土工程学报》为我国培养了一大批岩土工程学科的带头人, 为我国的基础性工程建设事业, 特别是水利工程建设事业做出了重要贡献。学报连续多年被评为“百种中国杰出学术期刊”, 连续获得中国科协精品科技期刊工程项目资助, 并连续多年入选“中国最具国际影响力学术期刊”。

回顾过去的历程, 总结过去的经验, 可以使我们更好地扬帆前行。为了学报更加美好的明天, 为使学报取得更好的发展, 值此创刊40周年之际, 我们决定开展“我与《岩土工程学报》”为主题的征文活动, 并计划出版“《岩土工程学报》创刊40周年纪念文集”。

征文对象: 《岩土工程学报》的读者、作者、审稿人和

编者等。

征文内容(包括但不限于): ①与期刊之间的故事; ②从期刊获得的启示、帮助; ③投稿经历和感受; ④审稿中的付出和辛劳; ⑤主编印象; ⑥参与期刊学术活动的经历和感受; ⑦对期刊今后发展的建议和期望。

征文方式: 征文采用书面稿或电子稿形式, 文内可含插图和照片, 篇幅以2000字左右为宜, 可长可短。书面稿请邮寄至南京虎踞关34号《岩土工程学报》编辑部, 邮编210024, 信封右上角请注明“我与《岩土工程学报》征文”字样; 电子稿件可发至邮箱ge@nhri.cn, 邮件主题请注明“我与《岩土工程学报》征文”。

征文时间: 即日起至2019年6月30日。

优秀论文评选: 编辑部将评选优秀论文, 并对所有优秀论文予以奖励。2019年3月5日之前收到的优秀论文, 将在创刊40周年纪念文集中刊出; 2019年3月6日至6月30日收到的稿件将择优在学报网站或期刊刊出。

竭诚欢迎广大读者踊跃撰文投稿, 其他未尽事宜请向编辑部垂询, 垂询电话: 025-85829534, 85829556。

《岩土工程学报》编辑部