

基于数据挖掘技术的白水河滑坡多场信息关联准则分析

陈锐, 范小光, 吴益平

Analysis on association rules of multi-field information of Baishuihe landslide based on the data mining

CHEN Rui, FAN Xiaoguang, and WU Yiping

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.2021.06-01>

您可能感兴趣的其他文章

Articles you may be interested in

基于快速聚类-信息量模型的汶川及周边两县滑坡易发性评价

Landslide susceptibility assessment based on K-means cluster information model in Wenchuan and two neighboring counties, China

周天伦, 曾超, 范晨, 毕鸿基, 龚恩慧, 刘晓 中国地质灾害与防治学报. 2021, 32(5): 137-150

基于遥感影像多尺度分割与地质因子评价的滑坡易发性区划

Landslide susceptibility assessment based on multi-scale segmentation of remote sensing and geological factor evaluation

李文娟, 邵海 中国地质灾害与防治学报. 2021, 32(2): 94-99

基于MSBAS技术的金沙江上游色拉滑坡形变分析

Deformation analysis of Sela landslide in the upper reaches of Jinsha River based on MSBAS technology

熊国华, 杨成生, 朱赛楠, 董继红, 张勤 中国地质灾害与防治学报. 2021, 32(5): 1-9

国产GB-InSAR在特大型水库滑坡变形监测中的应用

Application of GB-InSAR in deformation monitoring of huge landslide in reservoir area

郭延辉, 杨溢, 杨志全, 高才坤, 田卫明, 何玉童 中国地质灾害与防治学报. 2021, 32(2): 66-72

中国地质灾害数据质量评价指标体系构建

Formulation of an evaluation index system of geological hazard data quality in China

尹春荣, 李媛, 曲雪妍, 张艳玲, 佟彬, 杨旭东, 房浩 中国地质灾害与防治学报. 2021, 32(4): 120-125

基于PFC^{3D}的鱼鳅坡滑坡运动过程分析

Analysis of Yuqiupo landslide motion process based on PFC^{3D}

张家勇, 邹银先, 杨大山 中国地质灾害与防治学报. 2021, 32(4): 33-39



关注微信公众号, 获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.2021.06-01

基于数据挖掘技术的白水河滑坡 多场信息关联准则分析

陈 锐¹, 范小光², 吴益平¹

- (1. 中国地质大学(武汉)工程学院, 湖北 武汉 430074;
2. 中国电建集团河南省电力勘测设计院有限公司, 河南 郑州 450007)

摘要:为探究滑坡多场监测数据间的关联准则,采用数据挖掘技术中的两步聚类法与Apriori算法,开展滑坡多场信息关联准则研究。以三峡库区白水河滑坡为例,分析ZG93监测点于2003年6月—2016年12月期间的监测数据,选取影响滑坡变形的主要诱发因子,采用两步聚类法对不同的影响因子进行预聚类和聚类,将数值型变量转化为离散型变量后,应用Apriori算法进行处理,生成满足最小置信度的关联准则,建立白水河滑坡多场耦合作用模式下的影响因子与滑坡位移变形关联准则判据。研究表明,关联准则对于滑坡灾害的变形分析具有重要的意义,数据挖掘技术可较好地应用于三峡库区地质灾害位移预测预报中。

关键词:水库滑坡;数据挖掘;两步聚类法;Apriori算法;关联准则

中图分类号: P642.22 **文献标志码:** A **文章编号:** 1003-8035(2021)06-0001-08

Analysis on association rules of multi-field information of Baishuihe landslide based on the data mining

CHEN Rui¹, FAN Xiaoguang², WU Yiping¹

- (1. Faculty of Engineering, China University of Geosciences, Wuhan, Hubei 430074, China;
2. Henan Electric Power Survey & Design Institute, Zhengzhou, Henan 450007, China)

Abstract: In order to explore the association criteria of landslide multi-field monitoring data, we have adopted the two-step clustering method and Apriori algorithm, which belong to the classical data mining method, and we have also proposed the process of landslide monitoring data mining. Based on the Baishuihe landslide in the Three Gorges Reservoir Area, we analyzed the monitoring data of ZG93 from June 2003 to June 2016. The main inducing factors of the landslide displacement were selected, and the two-step clustering method was used to pre-cluster and cluster the different influence factors. We used Apriori algorithm to deal with the classified variables to generate frequent item sets that satisfy the minimum support degree. The association rules between the precipitating factors and the landslide deformation are established under the multi-field coupling mode of Baishuihe landslide. The results show that the correlation criterion is of great significance to the deformation analysis of landslide hazards and the data mining technology can be applied to the displacement prediction of geological hazards in the Three Gorges Reservoir Area.

Keywords: reservoir landslide; data mining; two-step clustering method; Apriori algorithm; association rules

收稿日期: 2020-12-29; 修订日期: 2021-02-28

基金项目: 国家自然科学基金项目(41977244; 42007267)

第一作者: 陈 锐(1997-), 男, 硕士研究生, 主要从事地质灾害预测预报研究。E-mail: 1336104030@qq.com

通讯作者: 吴益平(1971-), 女, 教授, 博士生导师, 从事地质灾害预测预报与风险评估研究。E-mail: ypwu@cug.edu.cn

1 研究背景

三峡工程作为世界上最大的水利枢纽工程,在防洪、发电及航运等方面发挥着巨大的作用。自三峡工程建成以来,由库水位变化所引起的地质灾害愈来愈受到人们的重视。其中,滑坡是三峡库区最频发且影响最大的地质灾害,为确保能实时监测滑坡位移变化,滑坡多场监测技术被广泛应用。

随着监测技术的进步与监测周期的增加,滑坡多场监测实时传输,取得了海量的监测数据。近些年数据挖掘技术在滑坡研究领域发展迅速,王树良等^[1]结合宝塔滑坡的监测数据,研究了滑坡监测数据挖掘的视角及其泛层次关系;张纯志^[2]以万源市太平镇滑坡为研究对象,应用线性回归方程处理滑坡数据,挖掘数据间的联系;徐峰等^[3]应用时间序列分析方法分析了三峡库区八字门滑坡,将滑坡位移分解,建立滑坡位移预报模型。但以上两种分析方法在实际的滑坡数据处理应用上还存在着一一定局限性,都只考虑单因子条件下的分析和预测,实际滑坡预测分析是一个复杂的系统,受多因子共同影响,同时随着滑坡数据的复杂多样化,常规数据处理方法已无法满足研究需要。近年来,许多新型数据挖掘技术在国内外兴起,并在滑坡领域应用广泛^[4-7]。马水山等^[8]采用两步聚类等方法,得到滑坡变形同监测数据间的关联规则;段功豪^[9]利用 Apriori 算法,以降雨量及库水位为主要影响因子,挖掘出树坪滑坡位移变形与影响因子间的关联准则。HUANG 等^[10]通过数据挖掘研究了三峡库区滑坡活动同库水位及降雨强度之间的关联准则。孙义杰^[11]采用数据挖掘方法中的 Apriori 算法,依据马家沟滑坡变形影响因子,确定滑坡变形位移同各因子间的关联准则。TSAI 等^[12]利用资料挖掘技术分析地形与植被因子,以验证区域性强降雨所诱发的滑坡,采用决策树和贝叶斯网络等数据挖掘算法,从海量数据间获取有效信息。马俊伟^[13]采用两步聚类法、Apriori 算法及决策树 C5.0 算法,对马家沟滑坡和朱家店滑坡的信息关联规则及信息阈值进行分析研究。

结合上述数据挖掘方法,本文选取白水河滑坡为例,结合滑坡资料从降雨与库水位角度选取影响因子^[14-16],分别为:月累计降雨量、日降雨量月度最大值、库水位月平均值、库水位波动速度、单月最大有效连续降雨、单月库水位日浮动最大值。依据数据挖掘流程,采用两步聚类法使滑坡变形演化定性化,将 6 种影响因子由数值型变量转化为离散型变量,随后应用 Apriori 算法挖掘出滑坡数据间的关联准则,输入影响因子作为关联准则前项,滑坡位移速度为关联准则后

项,生成影响因子与滑坡位移速度的关联准则,选取其中的有效规则,实现滑坡监测数据处理与关联准则挖掘。研究表明,关联准则对于滑坡灾害的变形分析具有重要的意义,数据挖掘技术可较好地应用于三峡库区地质灾害位移预测预报中。

2 白水河滑坡概况及数据分析

2.1 滑坡概况

白水河滑坡位于长江主干道南岸,属秭归县沙镇溪镇乐丰村。由滑坡勘察报告得到白水河工程地质平面图及剖面图(图 1、图 2)。南北长 500 m,白水河滑坡东西宽 430 m,面积 $21.5 \times 10^4 \text{ m}^2$,滑体平均厚度约 30 m,体积 $645 \times 10^4 \text{ m}^3$,主滑方向为 15° ,属大型顺层土质滑坡。2004 年 7 月,白水河滑坡出现明显变形,根据其变形特征划分出滑坡预警区。预警区东侧以黄土包凹槽为界,西侧以滑体西部山羊沟为界,后缘以高程约 297 m 为界,前缘剪出口在长江库水位 145 m 水位以下。

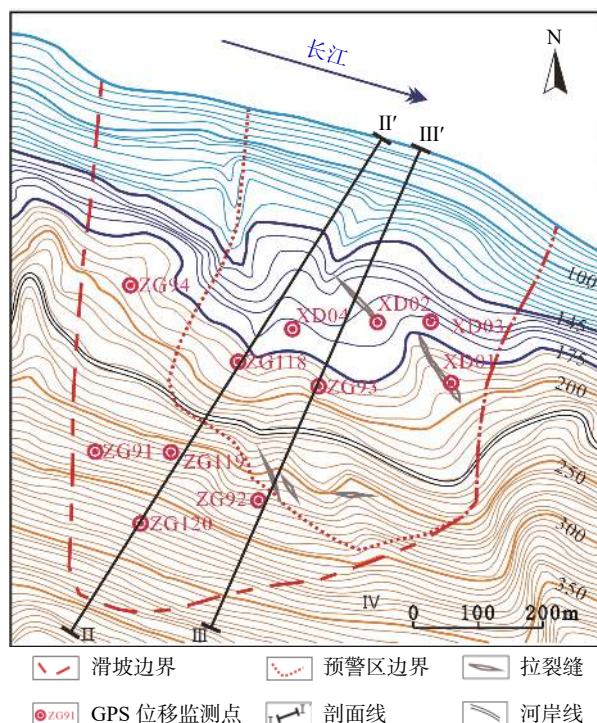


图 1 白水河滑坡工程地质平面图

Fig. 1 Engineering geological plan of Baishuihe Landslide

2.2 白水河滑坡变形特征分析

选取白水河滑坡监测点 ZG93 数据进行研究,通过对往年降雨强度及库水位变化数据归纳分析,获得白水河滑坡监测数据曲线图 3,将白水河滑坡位移变形特征分为三个阶段。

(1) 第一阶段: 2003 年 6 月—2006 年 8 月,库水位

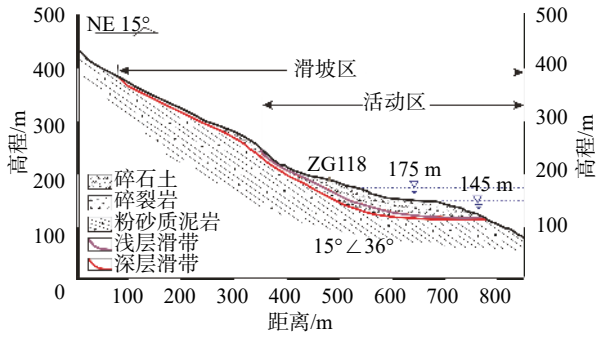


图2 白水河滑坡工程地质剖面图

Fig. 2 Engineering geological section of Baishuihe Landslide

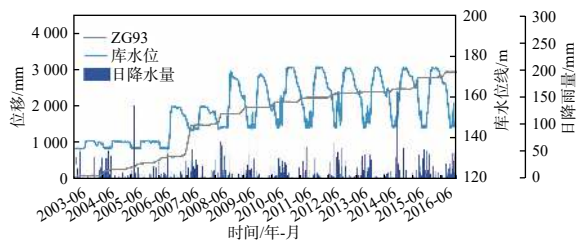


图3 白水河滑坡监测数据曲线

Fig. 3 Monitoring data curve of Baishuihe landslide

在 135 ~ 140 m, 这一阶段为低水位时期, 且库水位波动幅度相对较小, 滑坡前缘受库水位影响效果不明显, 此时滑坡位移变化特征主要与降雨强度相关, 主要表现为滑坡位移变形同降雨强度的增大而增大, 随降雨强度的降低而趋于稳定。

(2) 第二阶段: 2006年8月—2008年8月, 库水位抬升至 155 并稳定在 145 ~ 155 m, 相对第一阶段而言, 库水位抬升同时库水位波动幅度相对增大, 滑体内渗流场、应力场及岩土体结构发生改变。当库水位下降时, 随降雨强度的增大, 滑坡的位移变形同步大幅增大。此时滑坡位移变形特征主要受降雨强度影响, 库水位有一定影响。

(3) 第三阶段: 2008年8月—2016年12月, 库水位抬升至 175 m 并稳定在 145 ~ 175 m, 这一阶段库水位高度和波动幅度再次增加。在第二阶段中, 滑体内渗流场、应力场及岩土体结构发生改变调整后趋于稳定, 当库水位下降时, 随降雨强度的增大, 滑坡位移变形也会增大。相对第二阶段产生的大幅变化而言, 这一阶段变形量显著降低。随着库水位规律变化, 滑体内渗流场、应力场及岩土体结构进一步趋于稳定, 此时滑坡位移变形特征受降雨强度及库水位共同影响。

综上所述, 当降雨强度较低, 库水位抬升时, 滑坡体内的地下水抬升会滞后于库水位的抬升, 使得被库水浸没的滑体受到与滑面正交的静水压力作用, 此时滑坡整体相对稳定, 变形特征不明显。但当降雨强度增大同时

库水位下降时, 滑坡体内的地下水位下降会明显滞后于库水位的下降, 使得滑体在库水位下降后一段时间内仍受到一个与滑面平行且指向坡外的动水压力作用, 由于降雨强度大, 此时滑坡稳定性会大幅降低, 变形特征显著。因此, 高强度的降雨及库水位波动幅度是影响白水河滑坡位移变形的主要影响因素。

3 滑坡多场信息关联规则研究

3.1 数据挖掘流程

数据挖掘流程如图 4 所示, 主要由数据采集、数据预处理、数据离散化和关联准则挖掘四步组成。本文主要采用两步聚类算法及 Apriori 算法, 通过数据挖掘流程处理滑坡监测数据, 得到滑坡位移速度同影响因子间的关联准则。

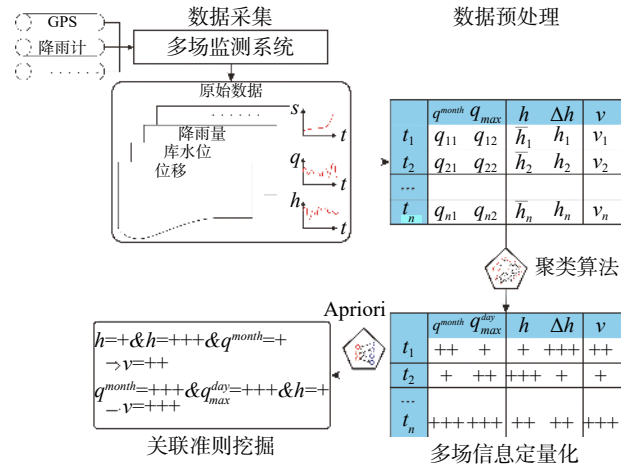


图4 滑坡多维信息时序关联数据数据挖掘流程图

Fig. 4 Data mining flow chart of multi-dimensional landslide information time series association criterion

(1) 数据采集: 数据采集是将野外监测数据(包括 GPS、降雨数据、位移数据)进行收集归纳, 获得数据基础。

(2) 数据预处理: 数据预处理主要指对已有数据进行审核、筛选、排序等处理措施。

(3) 数据离散化: 由于 Apriori 算法只能处理离散型变量, 故采用两步聚类法将监测数据离散化。

(4) 关联准则挖掘: 应用 Apriori 算法对离散后的变量进行关联规则集合选择, 生成有效的关联规则。

3.2 两步聚类法

两步聚类法算法是 CHIU 等^[17]在 2001 年在提出的一种能处理大规模类型数据的算法, 该算法是通过预聚类与聚类两步将数据划分整合, 进而完成数据分类, 其聚类过程详见图 5。

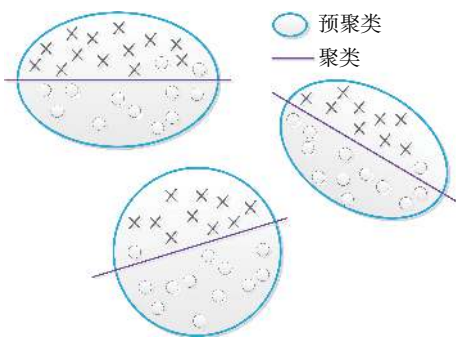


图 5 两步聚类法示意图

Fig. 5 Schematic diagram of two-step clustering method

对于样本数据既包括数值型变量又包括分类型变量, 两步聚类算法通常采用对数似然函数, 若聚成 j 类, 则其定义为:

$$l = \sum_{j=1}^J \sum_{i \in I_j} \ln p(X_i | \theta_j) = \sum_{j=1}^J l_j \quad (1)$$

式中: p ——似然函数;

I_j ——第 j 类的样本集合;

θ_j ——第 j 类的参数向量;

J ——聚类数目。

针对全部样本, 其对数似然聚类是各类对数似然聚类之和。

对于存在的第 i 类和第 j 类, 两者合并后的类记为 i, j , 则他们的距离定义为:

$$d(i, j) = \xi_i + \xi_j - \xi_{(i,j)} \quad (2)$$

式中: ξ_i ——第 i 类的对数似然距离;

ξ_j ——第 j 类的对数似然距离;

$\xi_{(i,j)}$ ——第 i 类和第 j 类合并后的对数似然距离。

ξ ——对数似然函数的具体形式, 定义为:

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \ln(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} E_{vk} \right) \quad (3)$$

其中:

$$E_{vk}^A = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \ln \frac{N_{vkl}}{N_v} \quad (4)$$

式中: K^A ——数值型变量的个数;

K^B ——分类型变量的个数;

$\hat{\sigma}_k^2$ ——第 k 个数值型变量的总方差;

$\hat{\sigma}_{vk}^2$ ——第 k 个数值型变量在第 v 类中的方差;

N_v ——第 v 类的样本量;

N_{vkl} ——在第 v 类中第 k 个分类型变量取第 l 个类别的样本量;

L_k ——第 k 个分类型变量的类别。

当第 i 类和第 j 类合并后, $-\xi_{(i,j)}$ 大于 $\xi_i + \xi_j$, 因此 $d(i,j)$ 小于 0。 $d(i,j)$ 越小, 说明第 i 类和第 j 类合并将不会引起类内部差异的显著增加。当小于阈值 C 时, 第 i 类和第 j 类可以合并; 当 $d(i,j)$ 大于阈值 C 时, 说明合并将会引起聚类簇内部的差异性显著增加, 第 i 类和第 j 类不能合并。

阈值 C 的定义为:

$$C = \ln V \quad (5)$$

其中:

$$V = \prod_k R_k \prod_m L_m \quad (6)$$

式中: R_k ——第 k 个数值型变量的取值范围;

L_m ——第 m 个分类型变量样本量。

3.3 Apriori 算法

通过两步聚类算法对滑坡监测数据类型进行转化, 将分类所得的变量采用 Apriori 算法计算, 挖掘监测数据的关联准则。

Apriori 算法由 AGRAWAL 等^[18]提出, 该算法首先生成高于最小支持度的频繁项目集, 在第一步产生的频繁项目集中生成高于最小可信度的关联准则。

频繁项目集是指对包含项目 a 的项集 T , 其支持度大于或等于用户指定的支持度阈值 (minsupp), 即:

$$\frac{|T(a)|}{|T|} \geq \text{minsupp} \quad (7)$$

包含 k 个项目的频繁项目集称为频繁 k 项集, 记为 L_k 。如图 6 上层的项目集 $ab, abc, abcd$, 当满足最小支持度时均为频繁 k 项集。

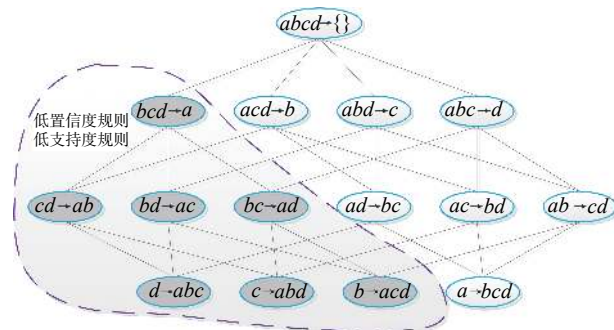


图 6 Apriori 算法示意图

Fig. 6 Schematic diagram of Apriori algorithm

Apriori 算法实现过程详见图 6。首先, 搜索出产生长度为 1 的频繁项集 L_1 , L_1 又用于产生长度为 2 的频繁项目集 L_2 , 如此循环, 搜索所有的频繁项目集。

从频繁项目集中产生简单的关联准则, 按置信度大于置信度阈值的条件, 选择出有效规则集合。对每个频

繁项目集 L , 计算 L 所有非空子集 L' 的置信度, 如果 $C_{L' \rightarrow (L-L')}$ 大于用户指定的置信度阈值 (minconf), 即

$$C_{L' \rightarrow (L-L')} = \frac{|T(L)|}{|T(L')|} \geq \text{minconf} \quad (8)$$

则生成关联准则 $L' \Rightarrow (L-L')$

4 白水河滑坡关联挖掘分析

4.1 影响因子选取

结合以往相关研究的成果^[19-20], 从降雨和库水位两种角度共选取了 6 种影响因子作为关联准则的前项。

(1) 降雨: 降雨入渗是影响滑坡位移变形的主要影响因素, 针对白水河滑坡而言, 由于区域内两次降雨的时间间隔较长, 且降雨入渗与蒸发作用同时进行, 故不考虑降雨时长的作用, 将降雨强度划分为 3 个影响因子进行分析, 分别为: 月累计降雨量 ($\sum q^{\text{month}}$)、日降雨量月度最大值 ($q_{\text{max}}^{\text{day}}$) 及单月最大有效连续降雨量 ($q_{\text{continuous}}^{\text{effective}}$)。

(2) 库水位: 根据库水位高度及波动幅度对滑坡位移变形特征影响分析可以得知, 当库水位产生变动时, 滑体内地下水会出现较明显的滞后效应, 对于白水河滑坡而言, 滑坡的位移变形受库水位较大影响, 此次将库水位划分为 3 个影响因子来探究其对滑坡变形的影响, 分别为: 库水位月平均值 (\bar{h})、月库水位波动速度 (Δh)、单月库水位日浮动最大值 ($h_{\text{max}}^{\text{float}}$)。

4.2 聚类分析

按照上述介绍的滑坡数据挖掘流程, 针对白水河滑坡特定的影响因子进行滑坡信息关联规则挖掘。

月累计降雨量、日降雨量月度最大值、库水位月平均值、库水位波动速度、单月最大有效连续降雨、单月库水位日浮动最大值、月位移速度的定性化成果见表 1~表 7。

由表 1 中月累计降雨量的两步聚类结果可知: 月度累计降雨量被划分为 Heavy_Rainfall、Moderate_Rainfall、Light_Rainfall 三簇, 分布表征月累计降雨量为 183.5 ~ 517.6 mm、69.9 ~ 179.8 mm、3.1 ~ 66.1 mm。

表 1 白水河滑坡月累计降雨量定性化成果
Table 1 Qualitative results of monthly accumulated rainfall of Baishuihe landslide

月累计降雨量 $\sum q^{\text{month}}/\text{mm}$	定性化值
(183.5, 517.6)	Heavy_Rainfall
(69.9, 179.8)	Moderate_Rainfall
(3.1, 66.1)	Light_Rainfall

由表 2 日降雨量月度最大值两步聚类结果可知: 日降

雨量月度最大值按照强度分布被划分为 55.9 ~ 160.7 mm、26.5 ~ 55.2 mm、1.3 ~ 25.6 mm 被划分为 Heavy_Rain_Shower、Medium_Rain_Shower、Light_Rain_Shower 三簇。

表 2 白水河滑坡日降雨量月度最大值定性化成果
Table 2 Qualitative results of monthly maximum rainfall of Baishuihe landslide

日降雨量月度最大值 $q_{\text{max}}^{\text{day}}/\text{mm}$	定性化值
(55.9, 160.7)	Heavy_Rain_Shower
(26.5, 55.2)	Medium_Rain_Shower
(1.3, 25.6)	Light_Rain_Shower

由表 3 中库水位月度平均值两步聚类结果可知: 库水位月度平均值被划分为 High_Water_Level、Medium_Water_Level、Low_Water_Level 三簇, 分别表示月累计降雨量为 160.14 ~ 174.74 m、144.21 ~ 158.47 m、135.13 ~ 138.95 m。

表 3 白水河滑坡库水位月平均值定性化成果
Table 3 Qualitative results of monthly average water level of Baishuihe landslide reservoir

库水位月度平均值 \bar{h}/m	定性化值
(160.14, 174.74)	High_Water_Level
(144.21, 158.47)	Medium_Water_Level
(135.13, 138.95)	Low_Water_Level

由表 4 月库水位波动速度两步聚类结果可知: 月库水位波动速度按照 (13.26, 17.35)、(7.23, 11.36)、(1.57, 5.89)、(-1.56, 1.31)、(-7.09, -3.41)、(-13.02, -8.59) 被划分为 Sharply_Rise、Medium_Rise、Slowly_Rise、Smooth Fluctuation、Medium_Drop、Sharply_Drop 六簇。

表 4 白水河滑坡月库水位波动速度定性化成果
Table 4 Qualitative results of water level fluctuation rate of Baishuihe landslide monthly reservoir

月库水位波动速度 $\Delta h/\text{m}$	定性化值
(13.26, 17.35)	Sharply_Rise
(7.23, 11.36)	Medium_Rise
(1.57, 5.89)	Slowly_Rise
(-1.56, 1.31)	Smooth Fluctuation
(-7.09, -3.41)	Medium_Drop
(-13.02, -8.59)	Sharply_Drop

由表 5 中单月最大有效连续降雨量两步聚类结果可知: 单月最大有效连续降雨量被划分为 High_Effective_Rainfall、Medium_Effective_Rainfall、Low_Effective_Rainfall 三簇, 分布单月最大有效连续降雨量为 110.5 ~ 239.4 mm、36.6 ~ 109.8 mm、1.5 ~ 36.1 mm。

表 5 白水河滑坡单月最大有效连续降雨量定性化成果
Table 5 Qualitative results of maximum effective continuous rainfall in a single month of Baishuihe landslide

单月最大有效连续降雨量 $q_{\text{continuous}}^{\text{effective}}/\text{mm}$	定性化值
(110.5, 239.4)	High_Effective Rainfall
(36.6, 109.8)	Medium_Effective Rainfall
(1.5, 36.1)	Low_Effective Rainfall

由表 6 单月库水位日浮动最大值聚类结果可知: 单月库水位日浮动最大值按 (1.66, 3.223)、(0.744, 1.513)、(0.063, 0.63)、(-0.414, 0)、(-1.697, -0.49) 划分为 Sharply_Rise_Water、Medium_Rise_Water、Slowly_Rise_Water、Slowly_Drop_Water、Medium_Drop_Water 五簇。

表 6 白水河滑坡单月库水位日浮动最大值定性化成果
Table 6 Qualitative results of the maximum daily fluctuation of the water level in a single month of Baishuihe landslide

单月库水位日浮动最大值 $h_{\text{max}}^{\text{float}}/\text{m}$	定性化值
(1.66, 3.223)	Sharply_Rise_Water
(0.744, 1.513)	Medium_Rise_Water
(0.063, 0.63)	Slowly_Rise_Water
(-0.414, 0)	Slowly_Drop_Water
(-1.697, -0.49)	Medium_Drop_Water

白水河滑坡为较典型的阶跃型滑坡, 采用两步聚类法将白水河滑坡的变形演化划分为三个阶段(表 7), 主要包括: 变形起始阶段(I)、变形稳定阶段(II)和变形加速阶段(III)。

白水河滑坡月位移速度定性化结果详见表 7。由

表 7 白水河滑坡月位移速度定性化成果
Table 7 Qualitative results of monthly displacement rate of Baishuihe landslide

月位移速度 $v/(\text{mm}\cdot\text{mon}^{-1})$	定性化值
(1.042, 10.669)	III
(0.092, 0.939)	II
(-0.195, 0.078)	I

表可知: 变形起始阶段(I)表示坡面监测点以 -0.195 ~ 0.078 mm/mon 的速度变形, 此集合占有所有集合的 42.3%; 变形稳定阶段(II)表示坡面监测点以 0.092 ~ 0.939 mm/mon 的速度变形, 此集合占有所有集合的 40.5%; 变形加速阶段(III)表示坡面监测点以 1.042 ~ 10.669 mm/mon 速度变形, 此集合占有所有集合的 17.2%。

4.3 关联准则挖掘

采用 Apriori 算法将影响因子作为关联前项, 滑坡位移速度为关联后项。考虑实际数据的有效性与实用性, 设定支持度阈值为 5%, 置信度阈值为 80%, 生成白水河滑坡位移同影响因子的关联准则。共生成关联准则 87 项, 结果显示大多数规则后项处于变形第 I、II 阶段——变形起始阶段及变形稳定阶段。这两阶段坡面位移变化较小, 仅为 -0.195 ~ 0.939 mm/mon。考虑实际滑坡变形时主要考虑第 III 阶段——变形加速阶段的位移变化。从众多关联结果中选取提升度大于 1 的准则, 提升度大于 1 时表明关联前项对关联后项的出现有促进作用, 则此准则具有较好的参考与实际意义。经过筛选后的关联准则成果详见表 8。

表 8 白水河滑坡多场信息关联准则

Table 8 Multi field information association criterion of Baishuihe landslide

规则 ID	规则	支持度 /%	置信度 /%	提升度
1	$\bar{h} = \text{High_Water_Level} \ \& \ q_{\text{continuous}}^{\text{effective}} = \text{Low_Effective Rainfall} \Rightarrow \text{I}$	25.77	85.71	2.02
2	$\bar{h} = \text{High_Water_Level} \ \& \ \sum q^{\text{month}} = \text{Light_Rainfall} \ \& \ q_{\text{max}}^{\text{day}} = \text{Light_Rain_Shower} \Rightarrow \text{I}$	25.15	85.37	2.02
3	$\bar{h} = \text{High_Water_Level} \ \& \ q_{\text{max}}^{\text{day}} = \text{Light_Rain_Shower} \Rightarrow \text{I}$	28.22	86.96	2.05
4	$q_{\text{continuous}}^{\text{effective}} = \text{Low_Effective Rainfall} \ \& \ h_{\text{max}}^{\text{float}} = \text{Slowly_Rise_Water} \ \& \ \bar{h} = \text{Low_Water_Level} \Rightarrow \text{II}$	8.42	100.00	4.13
5	$q_{\text{continuous}}^{\text{effective}} = \text{Low_Effective Rainfall} \ \& \ \sum q^{\text{month}} = \text{Moderate_Rainfall} \ \& \ \Delta h = \text{Slowly_Rise} \ \& \ h_{\text{max}}^{\text{float}} = \text{Slowly_Rise_Water} \Rightarrow \text{II}$	7.56	100.00	4.58
6	$q_{\text{continuous}}^{\text{effective}} = \text{Medium_Effective Rainfall} \ \& \ \bar{h} = \text{Low_Water_Level} \ \& \ \Delta h = \text{Slowly_Rise} \Rightarrow \text{II}$	7.56	100.00	4.58
7	$q_{\text{continuous}}^{\text{effective}} = \text{High_Effective Rainfall} \ \& \ \bar{h} = \text{Medium_Water_Level} \Rightarrow \text{III}$	6.13	90.00	5.24
8	$q_{\text{continuous}}^{\text{effective}} = \text{High_Effective Rainfall} \ \& \ \sum q^{\text{month}} = \text{Heavy_Rainfall} \ \& \ \bar{h} = \text{Medium_Water_Level} \Rightarrow \text{III}$	5.52	88.89	5.17
9	$q_{\text{continuous}}^{\text{effective}} = \text{High_Effective Rainfall} \ \& \ q_{\text{max}}^{\text{day}} = \text{Heavy_Rain_Shower} \ \& \ \sum q^{\text{month}} = \text{Heavy_Rainfall} \ \& \ \bar{h} = \text{Medium_Water_Level} \Rightarrow \text{III}$	5.52	88.89	5.17
10	$q_{\text{max}}^{\text{day}} = \text{Heavy_Rain_Shower} \ \& \ \sum q^{\text{month}} = \text{Heavy_Rainfall} \ \& \ \bar{h} = \text{Medium_Water_Level} \Rightarrow \text{III}$	6.13	80.00	4.66

结果显示: 规则 1 为关于坡面监测点处于变形起始阶段且包含库水位月平均值(High_Water_Leve)的关联准则。可解读为: 当库水位月均值处于高水位线时($160.14 \text{ m} \leq \bar{h} \leq 174.74 \text{ m}$), 坡面监测点进入变形起始阶段。规则 2~3 为关于坡面监测点处于变形起始阶段且包含高强度日降雨量月最大值的关联准则。可解读为: 当滑坡区域日降雨量月度最大值为高值时(Light_Rain_Shower), 坡面监测点进入变形起始阶段。

规则 4~6 为坡面监测点处于变形稳定阶段的关联准则。规则中包含的水文因子主要有中低值的最大有效连续降雨量(Medium_Effective Rainfall、Low_Effective Rainfall)及中低等速度的库水位变化过程(Slowly_Rise、Smooth Fluctuation)。由关联结果可知, 中低程度的最大有效连续降雨及中低程度的库水位变化对坡面监测点位移变形不会造成太大影响。

规则 7~10 为坡面监测点处于变形加速阶段的关联准则。规则中包含的水文诱发因子主要由强降雨(Heavy_Rainfall)和高强度的最大有效连续降雨(High_Effective Rainfall)组成。由关联结果可知, 强降雨和高强度最大有效连续降雨是引起坡面监测点的显著变形的主要因素。

通过关联准则挖掘得到的结论与白水河滑坡实际位移变形影响因子结果进行对比得知, 数据挖掘流程得到的关联准则能较好的用于滑坡位移监测预报。

5 结论

通过对完整的数据挖掘流程与关联准则结论进行分析, 得到主要的成果与结论如下:

(1) 对白水河滑坡 ZG93 监测点数据进行分析研究, 在 6~9 月汛期来临之时, 白水河滑坡变形位移会受强降雨的影响。库水位对滑坡位移的影响主要表现在 2007 年 6 月第一次蓄水引起的较大变形及水库开始蓄水后, 当库水位下降时, 随降雨强度的增大, 滑坡的位移变形同步大幅增大。

(2) 分别对六种影响因子进行两步聚类并得到对应因子的定性化成果: 月累计降雨量($\sum q^{\text{month}}$)定性化成果、日降雨量月度最大值($q_{\text{max}}^{\text{day}}$)定性化成果、单月最大有效连续降雨量($q_{\text{continuous}}^{\text{effective}}$)定性化成果、库水位月度平均值(\bar{h})定性化成果、月库水位波动速度(Δh)定性化成果、单月库水位日浮动最大值($h_{\text{max}}^{\text{float}}$)定性化成果。

(3) 结合数据挖掘流程, 建立了白水河滑坡多场耦合作用模式下的影响因子与滑坡位移变形关联准则判据。共生成白水河滑坡多场信息关联准则 87 项, 选取

其中关联度强的 10 条规则进行分析, 关联准则成果显示: 强降雨和高强度最大有效连续降雨是引起坡面监测点的显著变形的主要因素。

(4) 以白水河滑坡为例, 按照数据挖掘流程得到多场信息关联规则, 对于滑坡灾害的变形分析具有重要的意义, 数据挖掘技术可较好地应用于三峡库区地质灾害位移预测预报中。

参考文献 (References):

- [1] 王树良, 王新洲, 曾旭平, 等. 滑坡监测数据挖掘视角 [J]. 武汉大学学报(信息科学版), 2004, 29(7): 608-610. [WANG Shuliang, WANG Xinzhou, ZENG Xuping, et al. View angle of landslide-monitoring data mining [J]. Geomatics and Information Science of Wuhan University, 2004, 29(7): 608-610. (in Chinese with English abstract)]
- [2] 张纯志. 滑坡监测数据处理回归模型分析及探讨 [J]. 测绘与空间地理信息, 2016, 39(8): 214-217. [ZHANG Chunzhi. Landslide monitoring data processing regression model analysis and discussion [J]. Geomatics & Spatial Information Technology, 2016, 39(8): 214-217. (in Chinese with English abstract)]
- [3] 徐峰, 汪洋, 杜娟, 等. 基于时间序列分析的滑坡位移预测模型研究 [J]. 岩石力学与工程学报, 2011, 30(4): 746-751. [XU Feng, WANG Yang, DU Juan, et al. Study of displacement prediction model of landslide based on time series analysis [J]. Chinese Journal of Rock Mechanics and Engineering, 2011, 30(4): 746-751. (in Chinese with English abstract)]
- [4] MA J W, TANG H M, LIU X, et al. Establishment of a deformation forecasting model for a step-like landslide based on decision tree C5.0 and two-step cluster algorithms: a case study in the Three Gorges Reservoir area, China [J]. Landslides, 2017, 14(3): 1275-1281.
- [5] MA J W, TANG H M, HU X L, et al. Identification of causal factors for the Majiagou landslide using modern data mining methods [J]. Landslides, 2017, 14(1): 311-322.
- [6] 元呈明, 崔守梅, 陈辉, 等. 滑坡成因决策树挖掘 [J]. 中国地质灾害与防治学报, 2006, 17(1): 73-76. [QI Chengming, CUI Shoumei, CHEN Hui, et al. Decision-making tree for analysis of landslide causes [J]. The Chinese Journal of Geological Hazard and Control, 2006, 17(1): 73-76. (in Chinese with English abstract)]
- [7] 何少其, 刘元雪, 梁叶, 等. “阶跃式”滑坡突变预测与核心因子提取的平衡集成树模型 [J]. 中国地质灾害与防治学报, 2019, 30(5): 27-36. [HE Shaoqi, LIU Yuanxue, LIANG Ye, et al. Balanced decision tree ensemble model for catastrophe prediction and key factors' extraction of step-like

- landslides [J] . The Chinese Journal of Geological Hazard and Control, 2019, 30(5): 27 - 36. (in Chinese with English abstract)]
- [8] 马水山, 王志旺, 张漫. 基于关联规则挖掘的滑坡监测资料分析 [J] . 长江科学院院报, 2004, 21(5): 48 - 51. [MA Shuishan, WANG Zhiwang, ZHANG Man. Analysis on monitoring data of landslide based on associative data mining [J] . Journal of Yangtze River Scientific Research Institute, 2004, 21(5): 48 - 51. (in Chinese with English abstract)]
- [9] 段功豪. 三峡库区堆积层滑坡预报判据的动态挖掘研究 [D] . 武汉: 中国地质大学, 2013. [DUAN Gonghao. Study on prediction criteria of accumulated layer landslide in Three Gorges reservoir using dynamic data mining [D] . Wuhan: China University of Geosciences, 2013. (in Chinese with English abstract)]
- [10] HUANG H F, YI W, LU S Q, et al. Use of monitoring data to interpret active landslide movements and hydrological triggers in Three Gorges reservoir [J] . Journal of Performance of Constructed Facilities, 2016, 30(1): C4014005.
- [11] 孙义杰. 库岸边坡多场光纤监测技术与稳定性评价研究 [D] . 南京: 南京大学, 2015. [SUN Yijie. Bank slope multi-fields monitoring based on fiber optic sensing technologies and stability evaluation study [D] . Nanjing: Nanjing University, 2015. (in Chinese with English abstract)]
- [12] TSAI F, LAI J S, CHEN W W, et al. Analysis of topographic and vegetative factors with data mining for landslide verification [J] . Ecological Engineering, 2013, 61: 669 - 677.
- [13] 马俊伟. 渐进式滑坡多场信息演化特征与数据挖掘研究 [D] . 武汉: 中国地质大学, 2016. [MA Junwei. Evolution characteristics of multiple physics-based parameters and data mining technology for progressive landslides [D] . Wuhan: China University of Geosciences, 2016. (in Chinese with English abstract)]
- [14] SUN G H, ZHENG H, HUANG Y Y, et al. Parameter inversion and deformation mechanism of Sanmendong landslide in the Three Gorges Reservoir region under the combined effect of reservoir water level fluctuation and rainfall [J] . Engineering Geology, 2016, 205: 133 - 145.
- [15] 李麟玮, 吴益平, 苗发盛, 等. 考虑变形状态动态切换的阶跃型滑坡位移区间预测方法 [J] . 岩石力学与工程学报, 2019, 38(11): 2272 - 2287. [LI Linwei, WU Yiping, MIAO Fasheng, et al. Displacement interval prediction method for step-like landslides considering deformation state dynamic switching [J] . Chinese Journal of Rock Mechanics and Engineering, 2019, 38(11): 2272 - 2287. (in Chinese with English abstract)]
- [16] 易庆林, 张明玉, 文凯, 等. 三峡库区白水河滑坡变形特征及影响因素的阶段分析 [J] . 三峡大学学报(自然科学版), 2017, 39(1): 38 - 42. [YI Qinglin, ZHANG Mingyu, WEN Kai, et al. Periodic analysis of deformation characteristics and influential factors of Baishuihe landslide in Three Gorges reservoir area [J] . Journal of China Three Gorges University (Natural Sciences), 2017, 39(1): 38 - 42. (in Chinese with English abstract)]
- [17] CHIU T, FANG D P, CHEN J, et al. A robust and scalable clustering algorithm for mixed type attributes in large database environment [C] // Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01. August 26-29, 2001. San Francisco, California. New York: ACM Press, 2001: 263-268.
- [18] AGRAWAL R, IMIELŃSKI T, SWAMI A. Mining association rules between sets of items in large databases [C] // Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. May 25-28, 1993. Washington, D. C., USA. New York: ACM Press, 1993: 207-216.
- [19] WU X L, BENJAMIN ZHAN F, ZHANG K X, 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): 1 - 16.
- [20] LI D Y, SUN Y Q, YIN K L, et al. Displacement characteristics and prediction of Baishuihe landslide in the Three Gorges Reservoir [J] . Journal of Mountain Science, 2019, 16(9): 2203 - 2214.