

基于关联规则的滑坡演化阶段判识指标

刘小珊, 罗文强, 李飞翱, 黄 丽

(中国地质大学数学与物理学院, 武汉 430074)

摘 要: 滑坡的演化一般要经历初始变形、等速变形和加速变形 3 个阶段, 准确地划分这 3 个阶段是进行滑坡预测预报的基础。依据大量滑坡从初始变形到失稳破坏的监测数据, 采用数据挖掘中的关联规则, 用经典的 Apriori 算法挖掘滑坡变形过程中累积位移、累积加速度等属性与演化阶段的关联关系, 发现滑坡演化规律与累积加速度变化特征关联性强, 在滑坡从初始变形进入等速变形, 以及从等速变形到加速变形阶段时, 都存在相应的阈值, 即累积加速度在滑坡进入等速和加速变形阶段前后呈现出不同的特点。针对此特点, 以湖北秭归县鸡鸣寺滑坡为例, 将累积加速度作为划分鸡鸣寺滑坡演化阶段的指标, 得出划分阈值, 并利用关联规则分析其相关性。最后将此指标用于其他滑坡, 均得到良好效果。因此根据该指标, 可望实现滑坡演化阶段的自动判识。

关键词: 滑坡演化; 关联规则; 累积加速度; 阶段判识

中图分类号: P642.22

文献标志码: A

文章编号: 1000-7849(2014)02-0160-05

大量的滑坡监测实例表明, 典型的、具有普遍意义的滑坡从出现变形到最终的失稳破坏, 其运动累积位移—时间曲线可划分为 3 个阶段, 即初始变形、等速变形和加速变形阶段^[1] (图 1)。研究表明, 滑坡进入加速变形阶段是滑坡发生的前提, 因此准确且定量地判别滑坡演化的各个阶段是进行滑坡预报的基础。

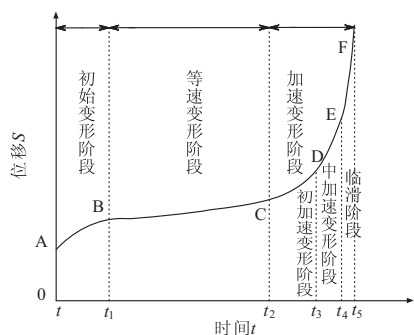


图 1 滑坡变形的 3 阶段演化图示^[1]

Fig.1 Evolution scheme of three deformation phases of landslide

目前, 国内外学者已提出了多种滑坡临滑预警指标, 将速度、加速度、切角等属性的变化作为滑坡是否发生急剧破坏的标准^[2-5], 但对滑坡演化过程中的 3 阶段的划分指标相对较少。另外, 判识指标大多都是通过分析总结得出, 其理论性不强。笔者首次引入基于关联规则的 Apriori 算法, 与传统预报时所关心的累积位移和变形速度相比, 挖掘滑坡演

化过程与累积加速度的相关性, 从而建立基于关联规则的滑坡演化阶段的判识指标——累积加速度, 为实现滑坡演化阶段的自动判识提供可能。

1 关联规则

1.1 基本概念

关联模型是数据挖掘中最先研究的问题之一, 也是数据挖掘的一个主要研究方向。关联规则就是挖掘数据库中不同项之间的联系。若设数据集为 D (一般为事务数据库), $I = \{i_1, i_2, \dots, i_m\}$ 为所有项目的集合, 事务 T 是一个项目子集 ($T \subseteq I$), 每一个事务具有唯一的标识 Tid 。设 X 是一个由项目构成的集合, 称为项集。事务 T 包含项集 X , 当且仅当 $X \subseteq T$ 。如果项集 X 中包含 k 个项目, 则称其为 k 项集^[6]。

关联规则是形如 $X \Rightarrow Y$ 的逻辑蕴含式, 其中 $X \subset I, Y \subset I$, 且 $X \cap Y = \emptyset$ 。规则 $X \Rightarrow Y$ 是事务数据库 D 的支持度, 即项集 X 在事务数据库中出现次数占 D 中总事务的百分比。正确率 (accuracy) 通常称为置信度或信任度 (confidence), 是将正确预测的实例数量表示为它在关联规则应用所涉及的全部实例中所占的比例^[7]。实际上, 支持度是一个概率值, 若项集 X 的支持度记为 $\text{support}(X)$, 那么规则的置信度为 $\text{support}(X \cup Y) / \text{support}(X)$ 。

收稿日期: 2013-06-01 编辑: 杨 勇

基金项目: 国家重点基础研究发展计划“973”计划(2011CB710605); 国家自然科学基金项目(40372120)

作者简介: 刘小珊(1989—), 女, 现正攻读应用数学专业硕士学位, 主要从事工程概率及地质灾害预测方面的研究工作。E-mail: cuglxs@126.com

1.2 Apriori 算法

挖掘关联问题就是寻找支持度和置信度分别满足用户给定的最小支持度和最小可信度的关联规则。其中,Apriori 算法在关联规则领域中具有很大的影响力,命名源于算法使用了频繁项集性质的先验(prior)知识。Apriori 算法是一种最具影响力的挖掘布尔关联规则的频繁项集算法,其核心是基于两阶段频集思想的递推算法。

一般来说,由 m 个项目形成的不同项集的数目可以达到 2^m-1 个,尤其在海量数据库 D 中,这是一个 NP 难度问题。为了避免计算所有项集的支持度(实际上频繁集只占很少一部分),Apriori 算法引入潜在频繁项集的概念^[6-8]。若潜在频繁 k 项集的集合记为 C_k ,频繁 k 项集的集合记为 L_k , m 个项目构成的 k 项集的集合为 C_m^k ,则三者之间满足关系 $L_k \subseteq C_k \subseteq C_m^k$ 。构成潜在频繁项集所遵循的原则是“频繁项集的子集必须为频繁项集”^[8-9]。在具体实现时,Apriori 算法将发现关联规则的过程分为 2 个步骤:第 1 步通过迭代,检验出事务数据库中所有频繁项集,即支持度不低于用户设定的阈值项集;第 2 步利用频繁项集构造出满足用户最小信任度的规则。其中,挖掘或识别出所有频繁项集是该算法的核心,占整个计算量的大部分。在使用第 1 步找到的频繁项集产生期望的规则时,会产生只包含集合项的所有规则,而且其中每一条规则的右部只有一项,这里采用的是中规则的定义。一旦这些规则被

生成,那么只有那些大于用户给定的最小可信度的规则才被保留下来^[10]。

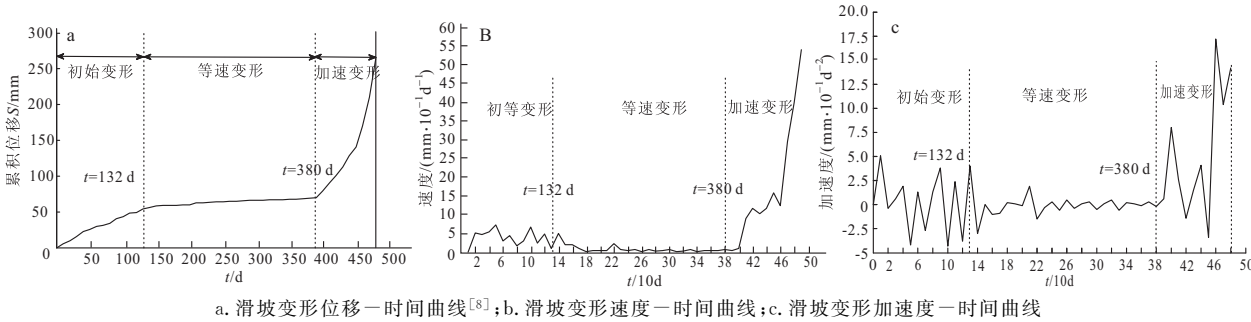
2 滑坡实例分析

选取湖北秭归县鸡鸣寺滑坡为例,介绍使用关联规则中的 Apriori 算法对滑坡演化过程中累积加速度分别与初始变形、等速变形、加速变形这 3 个阶段的关联关系。

2.1 数据准备

鸡鸣寺滑坡为一人为活动诱发产生的基岩滑坡,自 1990 年 3 月 5 日在秭归县水泥厂采石场上方发现裂缝开始进入初始变形阶段;1990 年雨季过后滑坡的变形有所减缓,呈缓慢增长状态,即等速变形阶段,直到 1991 年 3 月滑坡滑面开始贯通;同年 5 月滑坡变形速度开始加快,即进入了加速变形阶段,到 1991 年 6 月 29 日滑坡发生^[11],演化经历 16 个月。图 2-a 为鸡鸣寺滑坡变形位移—时间曲线。

大量滑坡实例表明,滑坡从开始出现变形到最后失稳破坏经历的时间较长,除加速变形阶段中的临滑阶段可能在短短几天内变形外,3 个阶段一般都会持续稍长的演化时间。因此,笔者选 10 d 为计量单位,计算各个滑坡的属性值,既方便实际中的测量,也能更好地反映滑坡演化过程中 3 阶段的属性特征。鸡鸣寺滑坡变形每 10 d 的速度、加速度变化如图 2-b,c 所示。



a. 滑坡变形位移—时间曲线^[8]; b. 滑坡变形速度—时间曲线; c. 滑坡变形加速度—时间曲线

图 2 湖北秭归县鸡鸣寺滑坡变形曲线图

Fig. 2 Deformation curves of Jimingsi landslide in Zigui, Hubei Province

2.2 实验及结果分析

如果待挖掘的数据属性皆为布尔属性,那么挖掘出来的规则都是布尔联系规则。现在的问题是滑坡监测数据中累积加速度是数值型,则必须对该数量属性进行必要的处理,使得数量联系转化为布尔联系规则的挖掘^[12]。我们采取的方法是对连续的数据取两阈值,将滑坡演化的整个阶段的数据划分为 3 个布尔属性。对于鸡鸣寺滑坡而言,累积加速度如图 3 所示,我们取初始变形阶段累积加速度的最大值 7.5 和最小值 1 作为划分阈值,试验数据如表 1 所示,

其中指标为累积加速度 $\eta: \eta = \tilde{a}_i = \sum_{i=1}^t a_i$ 。

表 1 湖北秭归县鸡鸣寺滑坡试验数据

Table 1 Experimental data of Jimingsi landslide in Zigui, Hubei Province

演化阶段	累积加速度 $\eta/(mm \cdot 10^{-1} d^{-2})$
初始变形	(1, 7.5]
等速变形	[0, 1]
加速变形	(7.5, +∞)

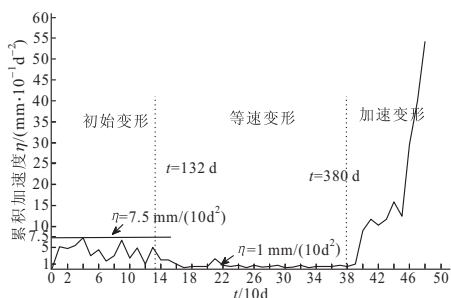


图3 湖北秭归县鸡鸣寺滑坡累积加速度曲线图

Fig. 3 Accumulative acceleration curves of Jimingsi landslide in Zigui, Hubei Province

在算法的具体实现过程中,当鸡鸣寺滑坡的累积加速度 $\eta \in (1, 7.5]$ 时,我们记为 A; 当 $\eta \in [0, 1]$ 时,记为 B; 当 $\eta \in (7.5, +\infty)$ 时记为 C。

用 Apriori 算法挖掘出的关联规则如下:

规则 1: indicator = C ==> stage = accelerated
conf: (1)

(规则 1 说明当累积加速度在 C 内取值,即: $\eta \in (7.5, +\infty)$ 时,滑坡演化阶段为加速变形阶段的置信度为 100%,规则 2~6 可类似理解)

规则 2: stage = accelerated ==> indicator = C
conf: (1)

规则 3: stage = constant ==> indicator = B
conf: (0.96)

规则 4: indicator = A ==> stage = initial
conf: (0.93)

规则 5: indicator = B ==> stage = constant
conf: (0.92)

规则 6: stage = initial ==> indicator = A
conf: (0.88)

分析以上规则可以看出:

(1) 当累积加速度 $\eta \in A$ 时,该阶段处于等速变形阶段的置信度为 93%,根据规则 4 基本可以确定鸡鸣寺滑坡处于初始变形阶段。

(2) 当累积加速度 $\eta \in B$ 时,该阶段处于等速变形阶段的置信度为 92%,根据规则 5 也可初步判断鸡鸣寺滑坡正处于等速变形阶段。

(3) 当累积加速度超过阈值 7.5,即 $\eta \in C$ 时,该阶段处于加速变形阶段的置信度为 100%,说明滑坡已经到了加速变形阶段。而当滑坡处于加速变形阶段也可判断出累积加速度 $\eta \in (7.5, +\infty)$,说明所取指标 η 与第 3 阶段的相关性非常大(根据规则 1 和 2)。

(4) 鸡鸣寺滑坡演化过程中的初始变形阶段和等速变形阶段会有少许点表现出其他阶段指标的特性,这是由于滑坡的监测数据是很难严格地符合演化图 1 的规律^[13-14],较常见的一种演化方式是从滑

坡孕育开始便直接进入等速变形阶段,跳过了初始变形阶段,或者说初始变形和等速变形合为了一个阶段。这类滑坡的演化并非严格地按照“三段式”的演化,比如天荒坪开关站滑坡、智力露采边坡、黄茨滑坡等^[11,15-16]。

在收集了大量滑坡的监测数据后,做出各个滑坡的累积加速度曲线如图 4,利用关联规则挖掘演化阶段与累积加速度的关联关系,得到如表 2 所示的试验结果。其中,斋藤试验是由日本学者斋藤于 20 世纪 60 年代在实验室采用黏土滑坡模型进行观测研究得到的典型的滑坡预报经验公式的曲线^[11,16],从图 4-a~c 可以看出,斋藤试验曲线、宝成铁路滑坡以及大冶铁矿滑坡表现出与鸡鸣寺滑坡类似的演化特征,累积加速度的变化特征与 3 个阶段的关联性大,特别是在加速变形阶段,除了宝成铁路滑坡与累积加速度的关联性是 88%,其他 3 个滑坡的关联性达到 100%。另外,对于天荒坪开关站滑坡等直接进入等速变形阶段的几个滑坡,如图 4-d~g 所示。在等速变形阶段,累积加速度相对比较平稳,在某一个固定值上下波动,但是一旦呈现上升趋势,便进入加速变形阶段。再如金川露天矿滑坡,将第一阶段的均值点作为阈值来划分演化阶段,前后两阶段表现出完全不同的演化特征,类似的还有黄茨滑坡、智力露采边坡等。

由此可以看出,虽然不同滑坡划分的阈值不一样,但划分的指标累积加速度都存在这样的规律:第一阶段累积加速度在两阈值之间波动,第二阶段累积加速度总体都小于较小的一个阈值,第三个阶段累积加速度基本在阈值点以上,且呈上升趋势。

表 2 试验结果

Table 2 Experimental results

滑坡	演化阶段	加速度/(mm·10 ⁻¹ d ⁻²)	置信度
鸡鸣寺滑坡	初始变形	(1, 7.5]	93%
	等速变形	[0, 1]	92%
	加速变形	(7.5, +∞)	100%
斋藤试验	初始变形	(2.5, 6]	100%
	等速变形	[0, 2.5]	93%
	加速变形	(6, +∞)	100%
宝成铁路滑坡	初始变形	(6, 10]	58%
	等速变形	[0, 6]	86%
	加速变形	(10, +∞)	88%
天荒坪开关站滑坡	等速变形	[0, 1]	73%
	加速变形	(1, +∞)	96%
金川露天矿滑坡	等速变形	[0, 10]	91%
	加速变形	(10, +∞)	68%
黄茨滑坡	等速变形	[0, 8]	79%
	加速变形	(8, +∞)	88%
大冶铁矿滑坡	初始变形	[0, 3]	67%
	等速变形	[3, 5]	74%
	加速变形	(5, +∞)	100%
智力露采边坡	等速变形	[0, 5]	89%
	加速变形	(5, +∞)	100%

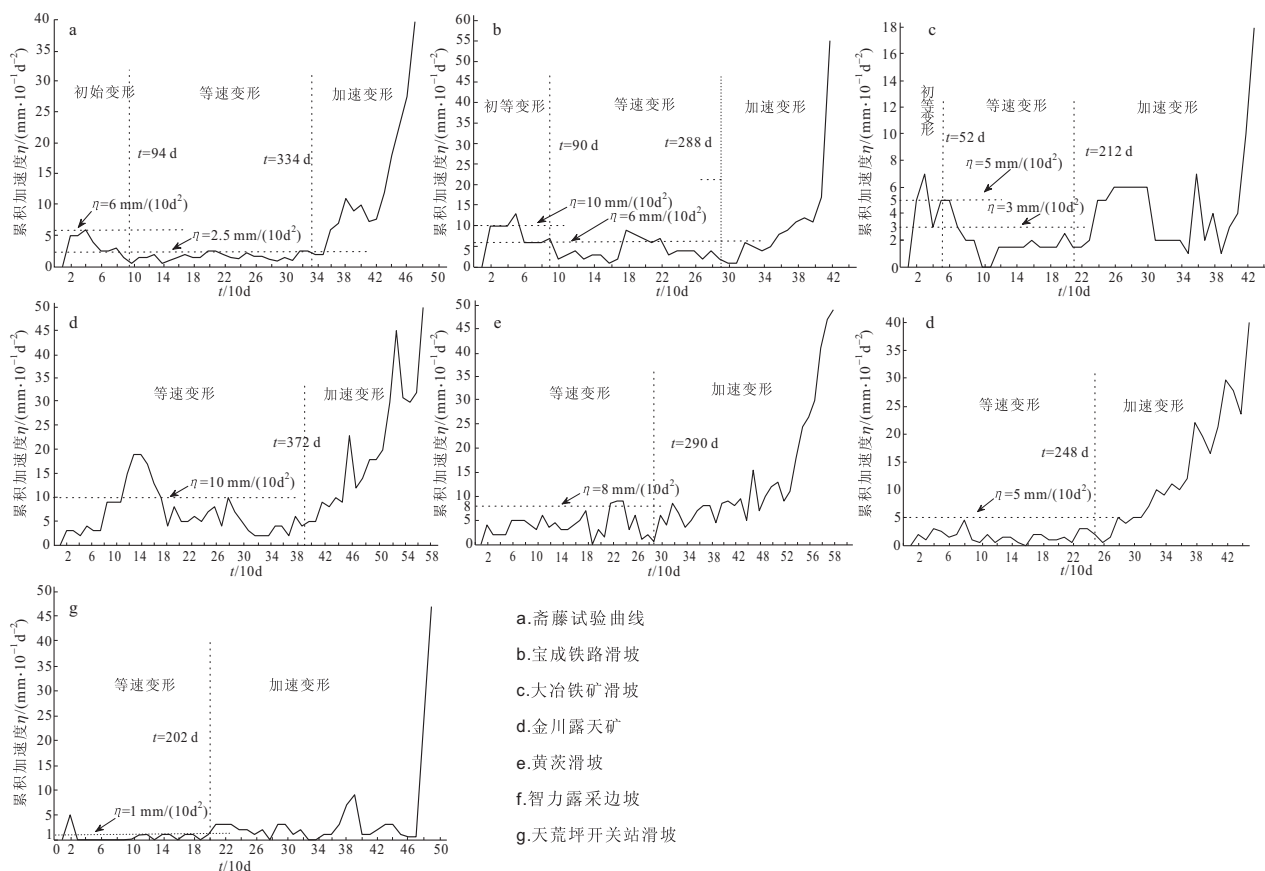


图4 各个滑坡的累积加速度变形曲线

Fig. 4 Accumulative acceleration curves of separate landslide

从试验结果可以看出:滑坡演化越到后期,累积加速度的变化趋势越明显。因此,如果要将加速变形阶段继续细分为初加速、中加速和临滑阶段,则可采用同样的方法,即利用关联规则对加速变形阶段的累积加速度进行挖掘,进而得出临滑阶段的阈值。鸡鸣寺滑坡临滑阶段的划分见图5。

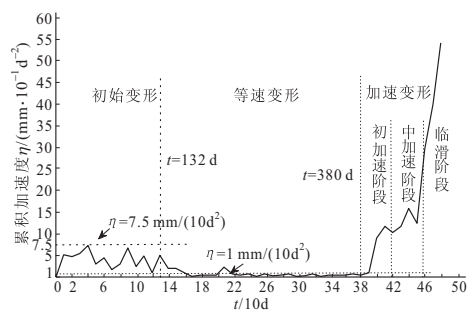


图5 湖北秭归县鸡鸣寺滑坡临滑阶段划分示例

Fig. 5 Critical sliding stage division of Jimingsi landslide in Zigui, Hubei Province

3 结论

(1) 滑坡演化规律与累积加速度变化特征关联

性强。对于正在监测的滑坡而言,可将累积加速度作为划分滑坡演化阶段的指标。当该指标呈现出不同于现阶段的属性特征时,则说明进入了下一个演化阶段。特别是在加速变形阶段,累积加速度的变化呈现出非常明显的变化特征。

(2) 每个滑坡除了具有共性外(例如演化的3阶段),还有明显的个性特征。因此,将累积加速度作为划分指标其阈值并不一致,即每一个滑坡因其演化特征不同而具有其特定的阈值。

(3) 通过对滑坡监测资料进行关联规则的数据挖掘,所获取的知识是可靠且符合事实规则的,证明该方法用在滑坡预警预报中是可行的。

参考文献:

- [1] 许强,黄润秋,李秀珍. 滑坡时间预测预报研究进展[J]. 地球科学进展, 2004, 19(3): 478-483.
- [2] 杨杰. 用加速运动预报高速滑坡和山崩[J]. 水文地质工程地质, 1995(4): 11-12.
- [3] 高华喜,殷坤龙. 降雨与滑坡灾害相关性分析及预警预报阈值之探讨[J]. 岩土力学, 2007, 28(5): 1055-1060.
- [4] 梅其岳. 天荒坪开关站滑坡的形成条件和滑动机制[J]. 岩石力学与工程学报, 2001, 20(1): 25-28.
- [5] 周志斌. 大冶铁矿东采场边坡破坏特征及滑坡时间预报[J]. 中国矿业, 2000(增刊2): 79-82.

- [6] 李雄飞, 懂元方, 李军, 等. 数据挖掘与知识发现[M]. 北京: 高等教育出版社, 2010.
- [7] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases[J]. Proceedings of the ACM SIGMOD Conference on Management of Data, 1993, 22(2): 207-216.
- [8] Ian H, Frank W E. 数据挖掘实用机器学习技术[M]. 董琳, 译. 北京: 机械工业出版社, 2007.
- [9] 吴益平, 殷坤龙, 陈丽霞. 滑坡空间预测数学模型的对比及其应用[J]. 地质科技情报, 2007, 26(6): 95-100.
- [10] 潘俊辉, 王辉. 一种基于改进的遗传算法的关联规则挖掘及应用[J]. 齐齐哈尔大学学报, 2011, 27(2): 11-14.
- [11] 曾裕平. 重大突发性滑坡灾害预测预报研究[D]. 成都: 成都理工大学, 2009.
- [12] 马水山, 王志旺, 张漫, 等. 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.
- [13] 吴益平, 唐辉明. 滑坡灾害空间预测研究[J]. 地质科技情报, 2001, 20(2): 87-90.
- [14] William B O. Landslides associated with the May 12, 2008 Wenchuan earthquake: Implications for the erosion and tectonic evolution of the Longmen Shan[J]. Tectonophysics, 2010, 491(1): 244-252.
- [15] McPeck M A. Biochemical evolution associated with antipredator adaptation in damselflies[J]. Evolution, 1999, 53(6): 1835-1845.
- [16] 殷坤龙. 滑坡灾害预测预报研究[M]. 武汉: 中国地质大学出版社, 2004.

Identification Index of Landslide Evolution Stage Based on Association Rule

Liu Xiaoshan, Luo Wenqiang, Li Fei'ao, Huang Li

(School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, China)

Abstract: Landslide generally experiences three phases including initial deformation, uniform deformation and accelerated deformation. Precise classification of these three phases is the base for landslide forecast. Based on a large amount of monitoring data of landslide evolution from initial deformation to instability failure phase, using association rules of data mining, correlation of cumulative displacement, cumulative acceleration and evolution phases of landslide deformation are revealed by classical Apriori algorithm. And it is found that there is a high correlation between the rule of landslide evolution and the variation characteristics of accumulative acceleration. Once the landslide changes from initial deformation into uniform deformation phase, or from uniform deformation into accelerated deformation phase, there is a corresponding threshold. That is to say, the accumulative acceleration shows completely different characteristics before and after the landslide getting into other phases. Taking Jimingsi landslide as an example, accumulative acceleration is taken as a stage division index for landslide evolution and division threshold is obtained. Then the correlation between them is analyzed by association rule. This index is applied well as well to other cases of landslides. Therefore according to this indicator, automatic stage division of landslide evolution is expected to achieve for disaster prevention.

Key words: landslide evolution; association rule; accumulative acceleration; stage division