维普资讯 http://www.cqvip.com

第35卷第2期 气 象 科 技 Vol. 35, No. 2
2007年4月 METEOROLOGICAL SCIENCE AND TECHNOLOGY Apr. 2007

应用KDD技术分析气象地质灾害与降水的关系

彭贵芬[1] 段旭[2] 舒康宁[1] 周毅[1]

(1 云南省气象台,昆明 650034; 2 云南省气象科学研究所,昆明 650034)

摘要 应用人工智能技术——数据库中知识发现(KDD)技术,分析研究发生在云南省境内造成重大损失的气象地质灾害的发生与当日及前期降水的关系。结果表明,绝大多数气象地质灾害都是由强降水诱发的,云南气象地质灾害与发生当日及前期降水的关系主要有3种类型,即暴雨诱发型、多日中大雨诱发型和连阴雨诱发型。所得到的云南气象地质灾害与当日及前期降水的定量关系,为云南地质灾害的气象监测、预报模式的建立提供了依据。

关键词 数据库 知识发现 气象地质灾害 降水

引言

云南的地质环境、人类不当的经济活动、雨季大雨、暴雨的集中出现,导致云南省存在发生地质灾害的隐患。云南全省每年都要发生上百次大型地质灾害,死亡人数都在100人左右,直接经济损失达10多亿元。为了做出有实际指导意义的气象地质灾害(MGD:Meteorological Geologic Disaster)危险等级预报,进而发出警报,关键要摸清MGD与诱发因子——降水的定量关系,建立MGD的中短期预报模式,通过监测和预报降水,开展MGD的气象监测和危险等级预报。

早期对泥石流滑坡灾害与降水关系的研究结果,主要是建立泥石流滑坡发生和降水临界值之间的经验公式。Onordera等(1974年)发现,在日本累积雨量大于等于150 mm,或每小时降雨强度大于等于20 mm时,将发生泥石流滑波;美国San Benito和Alameda把过程降水量累积大于等于180 mm定为泥石流滑坡发生的临界值;加拿大的临界累积雨量值大于等于250 mm;中国香港则大于等于350 mm,且日雨量大于100 mm,1 h雨量大于40 mm;巴西滑坡的暴雨强度临界值为250~300 mm。近年国内单九生等[1,2]、李晓[3]、马力等[4]、李军等[5]、钟荫乾[6]、李媛等[7]、林孝松[8]的研究指出:泥石流滑坡除24 h、小时临界降水是主要因素外,还与前期有效降水与连续降水的累积值有关。本研究对发生在云南境内、造成重大损失的514个灾害个例的研究证明,发生MGD灾害的当天出现50 mm以上暴雨的有47个,前1日出现暴雨的有25个,两日合计暴雨数仅占总数的14%,即514个个例的历史概括率为14%,这样低的概括率一方面说明仅用各县级气象站的降水量资料进行研究,观测分辨率不够,未能完全反映发灾地点的实际降水情况,另一方面也说明MGD的发生不仅与当日或前一日的暴雨有关,还与发生前一段时期的降水有关。如果不对MGD发生前期的降水情况进行分型,找出各型对应的临界值(多个临界值)的指标,仅用一两个临界值做预报,将出现大量的漏报现象。通过数据库中的知识发现技术对MGD发生当日及前期降水关系的分型,找出各型对应的临界值(一组临界值),可将514个个例的概括率提高到92%。

1 资料和方法

1.1 资料来源

本文研究所用MGD资料为1953~2002年发生在云南省境内造成重大损失的地质灾害特例共514个。资料有3个来源:①有气象记录以来(1953~1994年)各气象站气象大事记中有关地质灾害的记录;②中国地质科学院地质力学研究所网站上公布的地质灾害数据库(1955~2000年);③云南省国土资源厅信息处提

中国气象局新技术推广项目(CMATG2005M39)资助

作者简介:彭贵芬,女,1955年生,学士,高级工程师,主要从事决策气象服务工作,Email: gfpeng0725@163.com

收稿日期:2005年11月28日;定稿日期:2006年2月21日

维普资讯 http://www.cqvip.com

供的1995~2002年的地质灾害资料。将各种渠道收集到的地质灾害资料按发生日期、发生县、发生地点、灾情、死亡人数、受伤人数、失踪人数、经济损失、明细说明等属性建立地质灾害数据库。

降水资料从云南省125个气象站有记录以来的逐日降水资料中选取。

1.2 技术方法

将514条MGD记录配上灾害发生当天及前39天当地的逐日降水量,构成了一个庞大的数据库。我们需要从大量的MGD数据和降水数据资源中挖掘出对建立预报模式有指导意义的知识,即通过对数据整体的全面而深刻的认识,对大量数据的高度浓缩和抽象,从而发现有用的知识。人工智能中的数据库中知识发现(KDD: Knowledge Discovery in Database)技术[9]就是能从地质灾害数据库中精确抽取大量琐碎数据中隐含的、预先未知的和潜在有用信息的方法和工具。知识发现的主要任务是找出事物之间的因果关系,所采用的方法有求同法、差异法、同异并存法、剩余法、共变法。前两种方法是基本法,后3种方法只是求同法和差异法的结合、引申和补充。KDD在本质上以逻辑归纳为主,采用从个别到一般,从感性到理性的知识抽象过程。KDD的主要实施对象是关系数据库,本文采用关系数据库存储资料,并实施知识的抽取。从关系数据库中可进行4类知识的抽取:依赖关系、分类知识、描述性知识、偏差型知识。抽取出来的知识必须以可以理解的方法表达,可以是一些产生式规则、函数解析式、自然语言文字报告等。

关系数据库中的数据是由若干关系表组织起来的,可以从逻辑上把他们虚拟地连接起来,构成一张单一的二维通用大表。通用大表的横向是属性,纵向是元组或记录,通过通用大表接口,外部应用可以将数据库看成是单一的平面文件,可根据用户提出的发现算法,利用SQL(Structured Query Language)的映射、投影等功能,将与发现任务有关的数据取出来。待处理的数据量过大时,先对数据进行抽样处理,最后形成与发现任务有关的初始数据集。

对于关系数据库系统的知识发现状态空间是一个三维空间,是发现系统实施多种发现算法的运作空间(图1)。通常从3个方面对知识模板进行操作。属性方向OA:对属性之间关系的认识和发现;宏元组方向OM:对各宏元组之间一致性和差异性的认识和发现;知识模板方向OT:是属性值从微观到宏观的操作,使知识模板上升到抽象级别高的知识模板。

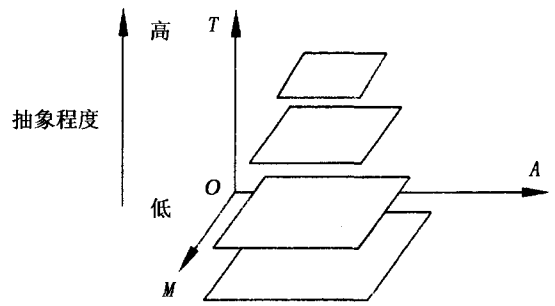


图1　知识发现状态空间示意

设二维大表由m个属性A_i和n个元素组成,每个属性各有V_i个不同的取值($1 \leqslant V_i \leqslant n$),这样就构成了大表的取值空间:$U = V_1 V_2 \cdots V_m$,每个元组都可以对应到$U$中的一个实点,$n$个实点的集合构成$R$。$U$中的数据分布模式反映出数据间的作用规律。设$X_{i,k}$为第$k$个元组的第$i$个属性的取值,则每个元组都可以用逻辑式表示:

$$\exists t(t \in R)[A_1(t) = X_{1,k} \cap A_2(t) = X_{2,k} \cap \ldots \cap A_n(t) = X_{n,k}]$$

R也可以用逻辑式表达:

$$\forall t(t \in R) \Leftrightarrow [A_1(t) = X_{1,1} \cap A_2(t) = X_{2,1} \cap \ldots \cap A_m(t) = X_{m,1}]$$
$$\cup [A_1(t) = X_{1,2} \cap A_2(t) = X_{2,2} \cap \ldots \cap A_m(t) = X_{m,2}]$$
$$\cup \ldots$$
$$\cup [A_1(t) = X_{1,n} \cap A_2(t) = X_{2,n} \cap \ldots \cap A_m(t) = X_{m,n}]$$

知识发现状态空间中的许多操作都是通过合并、投影等方式,简化关系表的取值描述形式。

2 MGD与降水关系的知识发现

将地质灾害发生日及发生县作为关键属性,从多年逐日降水资料库中以地质灾害发生日为起点,向前抽取40天的逐日降水量,横向以D_i($i = 0,1,2,\cdots,39$,代表MGD前发生前第i天)属性存储逐日降水量,纵向以元组(记录)构成用于知识发现的二维表。

以原有的MGD与降水关系的经验背景知识为基础[10],将关系数据库的SQL语句用于数据表中

维普资讯 http://www.cqvip.com

元组的映射、投影、归纳。主要应用 SQL 的 3 个语句:查询语句(Select),新增、插入语句(Insert),数据更新语句(Update)进行知识发现。将文献[1]的发现方法与本研究的内容、资料和数据库技术、手段相结合,设计知识发现步骤,并用 VFP(Visual FoxPro)编制发现程序,进行 MGD 与降水关系的知识发现。具体步骤如下:

(1)将地质灾害资料和逐日降水资料通过数据库接口将数据汇聚到数据库(可手工录入或以格式化文件的方式成批录入),然后对雨量资料采样、过滤,抽取 MGD 发生当日及前 39 日的逐日降水量,初始化(清空)模板表。

(2)进行属性方向(*OA*)的操作。经过第一步采样以后的降水资料是发生 MGD 的降水集,按各属性的概括率(70%)作为控制信度进行操作,各属性均未通过信度检验,说明 MGD 与单日降水关系不明显。为了解决这个问题,应用背景知识基,分段求取累积降水量、最大降水量、暴雨日、大雨日、中雨日、雨日数据,以新属性存放。

(3)进行宏元组方向(*OM*)的操作。应用背景知识基,对数据进行宏元组的操作(归纳为不同的型):用"rmax1"代表 MGD 发生前 0~5 日的最大日降水量,则用"rmax1≥50 mm、rmax1<50 mm and rmax1≥10 mm、rmax1<10 mm"进行宏元组的初分和信度检验,初步分为暴雨组、多日中大雨组、连阴雨组、白噪声组。用 t 检验进行分组信度控制。对各宏元组组间一致性用属性的概括率作为信度控制。将诱发 MGD 的前期降水类型初步归纳为 3 种类型:暴雨型、多日中大雨型和连阴雨型。

(4)进行知识模板方向(*OT*)的操作。对各宏元组进行组内(微观)操作,按主属性、次主属性的次序,以概括率作为信度控制进行分类统计,提取、生成产生式规则和自然语言文字报告知识。

(5)对产生式规则进行回报模拟,进行模式评价和知识验证。

(6)输出可以理解和应用的知识。

3 MGD 与前期降水的关系

根据 2005 年 1 月中国地质环境监测院地质灾害气象预报预警技术要求规定,将 MGD 危险等级(可能性)划分为 5 级,即 1 级:可能性很小(<20%);2 级:可能性较小(20%~40%);3 级:可能性较大(40%~60%);4 级:可能性大(60%~80%);5 级:可能性很大(>80%)。其中 3 级是防范级,4、5 级是警报级。

降水与 MGD 的关系模型中暴雨诱发型、多日中大雨诱发型的降雨关键时段是前 0~5 天和前 6~11 天,连阴雨诱发型的降雨关键时段是前 0~30 天。根据 MGD 与降水的关系模式和关键时段,用"rmax1"和"rmax2"分别代表前 0~5 天、6~11 天内的最大日降水量,"trdays1"和"trdays2"分别代表前 0~5 天、6~11 天内的暴雨日数,"hrdays1"和"hrdays2"分别代表前 0~5 天、6~11 天内的大雨日数,"mrdays1"和"mrdays2"分别代表前 0~5 天、6~11 天内的中雨日数,"thmrdays"代表前 0~11 天内的中雨以上强降水日数,"rdays"代表前 0~30 天内的雨日数,来表征发现结果。

3.1 暴雨诱发型

暴雨诱发型组内记录的产生式知识规则、自然语言文字报告、信度(概括率)详见表 1。

由表 1 可见,暴雨型组内记录的一致性为 MGD 发生前 0~5 天内,均出现了至少 1 个暴雨日。而组内记录的差异性则有几个特点:①MGD 发生前 0~11 天内仅有 1 个暴雨日的记录非常少(仅占 5%);②有 1 个暴雨日且仅有 1 个中、大雨日的记录也很少(仅为 6%);③前 0~11 天内有 2 个以上暴雨日或仅有 1 个暴雨日但还有 2 个以上中、大雨日的记录很多,达 90%。

暴雨诱发型的前期降水特征:MGD 发生的当日及前 5 日中(共 6 日),出现 2 日或以上日降水量大于等于 50 mm 的暴雨;或出现 1 日降水量大于等于 50 mm 的暴雨,且前 11 日内,出现过 2 日或以上降水量在 10~50 mm 的中、大雨为暴雨诱发型。

从此型 MGD 的前期降水情况可见,MGD 不但与当日降水量有关,还与前期降水有直接关系。单日大暴雨只会造成洪灾,而不一定能诱发 MGD。因为 MGD 的发生,需要前期土壤有一定含水量,即需要一定的降水累积。此型诱发 MGD 的降水类型又可根据灾害发生前期的降水情况分为 3 类,即暴-暴(两次暴雨间隔在 10 天内)、连暴(连续 2 日暴雨或以上)、多日中大雨-暴类。此型 MGD 有 107 个,占总数的 21%(107/514),其诱发的 MGD 以泥石流灾害为主。

维普资讯 http://www.cqvip.com

表1　暴雨型知识规则表

产生式规则	自然语言文字报告	信度(概括率)
IF rmax1≥500 and trdays1+trdays2=1 and hrdays1+hrdays2=0 and mrdays1+mrdays2=0 THEN MGD=1,2	前0～5天内有1个暴雨日,但前0～12天内无中雨以上的降水,则发生MGD的可能性很小	0.047(5/107)
IF rmax1≥500 and trdays1+trdays2≥2 THEN MGD=5	前0～12天内有2个或以上暴雨日,发生MGD的可能性很大	0.308(33/107)
IF rmax1≥500 and trdays1+trdays2=1 and hmrdays≥2 THEN MGD=5	前0～5天内仅有1个暴雨日,但前0～12天内有2个或以上中、大雨日,则发生MGD的可能性很大	0.589(63/107)
IF rmax1≥500 and trdays1+trdays2=1 and hmrdays=1 THEN MGD=2	前0～5天内仅有1个暴雨日,且仅有1个中、大雨日,则发生MGD的可能性较小	0.056(6/107)
IF rmax1≥500 and (trdays1+trdays2≥2 or trdays1+trdays2=1 and hmrdays≥2) THEN MGD=4,5	前0～5天内出现了暴雨,且前0～11天内至少还有1个暴雨日或者2个或以上中、大雨日,则发生MGD的可能性大	0.897(99/107)

3.2　多日中、大雨诱发型

多日中、大雨诱发型组内记录的一致性和差异性知识、组内记录的产生式知识规则、自然语言文字报告、信度(概括率)详见表2。由表2可见,多日中、大雨型组内记录的一致性为MGD发生前0～5天内,均未出现暴雨,但出现了10～50 mm的中、大雨。而组内记录的差异性有几个特点:①MGD发生前0～11天内仅有1个中雨以上降水日的记录非常少;②前0～11天内有4个以上中雨以上降水日的记录占此型记录的54%;③前0～11天内有2～3个中雨以上降水日、且前25天内有13天或以上雨日的记录占此型记录的32%;④前0～11天内中雨以上强降水日小于等于3天、且前2天内有大于等于40 mm以上的降水的记录占此型记录的25%。

多日中、大雨诱发型前期降水特征:MGD发生的当日及前5日中,未出现50 mm以上的暴雨日,但发生当日及前11天中出现了多日(4日或以上)10 mm以上中雨以上强降水;或前11天中有2～3日中雨以上强降水,且前25天内有13个以上雨日或前6～11天内出现过至少1日40 mm以上强降水为多日中、大雨诱发型。

在本文所研究的514个个例中,有330个个例前期降水情况可归纳为此型(包括12日内有暴雨出现的情况),占总数的64%。在10多天中,多日出现中、大雨,降水被土壤充分吸收,土壤被降水充分浸透,在陡峻的坡地上形成了过饱合的土壤,只要再有中雨出现,就可能诱发MGD。这种类型的前期降水,既有一定的强度,也有一定的持续性,易造成大范围的滑坡灾害,对生命财产威胁很大。

3.3　连阴雨诱发型

连阴雨诱发型组内记录的一致性和差异性知识、组内记录的产生式知识规则、自然语言文字报告、信度(概括率)详见表3。由表3可见,此型MGD与持续不断的连绵阴雨关系很大。

表2　多日中、大雨型知识规则表

产生式规则	自然语言文字报告	信度(概括率)
IF rmax1≥10 and max1<50 and thmrdays≥2 THEN MGD=2～5	前0～5天内无暴雨,但前0～11天内有2个以上中雨以上的降水,则有发生MGD的可能性	0.909(300/330)
1IF rmax1≥10 and max1<50 and thmrdays≥4 THEN MGD=3～5	前0～5天内无暴雨,但前0～11天内有4个以上中雨以上的降水,则发生MGD的可能性较大	0.5455(180/330)

维普资讯 http://www.cqvip.com

（续表2）

产生式规则	自然语言文字报告	信度(概括率)
IF rmax1≥10 and max1<50 and thmrdays≥5 THEN MGD=4	前0～5天内无暴雨,但前0～11天内有5个以上中雨以上的降水,则发生MGD的可能性大	0.3545(117/330)
IF rmax1≥10 and max1<50 and thmrdays≥6 THEN MGD=5	前0～5天内无暴雨,但前0～11天内有6个以上中雨以上的降水,则发生MGD的可能性很大	0.1970(65/330)
IF rmax1≥10 and max1<50 and thmrdays≥2 or thmrdays≥3 THEN MGD≥3	前0～5天内无暴雨,但前0～11天内有2～3个中雨以上强降水,则发生MGD的可能性较大	0.3636(120/330)
IF rmax1≥10 and max1<50 and thmrdays≥2 or thmrdays≥3 and yr25≥13 THEN MGD≥3	为前0～5天内无暴雨,但前0～11天内有2～3个中雨以上强降水,且前25天内有13个或以上雨日,则发生MGD的可能性较大	0.8917(107/120)
IF rmax1≥10 and max1<50 and thmrdays=3 and yr25≥13 THEN MGD=4	为前0～5天内无暴雨,但前0～11天内有3个中雨以上强降水,且前25天内有13个或以上雨日,则发生MGD的可能性大	0.45(54/120)
IF rmax1≥10 and max1<50 and thmrdays=2 and yr25≥17 THEN MGD=4	为前0～5天内无暴雨,但前0～11天内有2个中雨以上强降水,且前25天内有17日或以上雨日,则发生MGD的可能性大	0.30(36/120)
IF rmax1≥10 and max1<50 and thmrdays≤3 and rmax1≥400 and rmax2≥400 THEN MGD=3	前0～5天内无暴雨,且前12天内中雨以上强降水少于3个,但前12天内有40 mm以上的降水,则发生MGD的可能性较大	0.2459(30/120)

表3 连阴雨日的知识规则表

产生式规则	自然语言文字报告	信度(概括率)
IF rmax1<10mm and rdays31≥15 THEN MGD=3～5	前0～5天内日最大降水量小于10 mm,但前0～30天内有15个以上雨日出现,则有发生MGD的可能性	0.6640(83/125)
IF rmax1<10mm and rdays31≥18 THEN MGD=4～5	前0～5天内日最大降水量小于10 mm,但前0～30天内有18个以上雨日出现,则发生MGD的可能性大	0.4800(60/125)
IF rmax1<10mm and rdays31≥18 THEN MGD=4～5	前0～5天内日最大降水量小于10 mm,但前0～30天内有21个以上雨日出现,则发生MGD的可能性很大	0.2400(30/125)

连阴雨诱发型的前期降水特征:MGD发生当日及前5天内无中雨以上强降水出现,但30日内出现15个以上雨日(日降水量≥0.1 mm,下同),即有一半以上日子中均有降水,其间有3个左右相对集中的降水时段,每个降水时段间可有1～3个降水中断日,其间可有多个中雨以上强降水日。如果雨日在18个以上,则发生MGD的危险性较大;雨日达21个以上,则发生MGD的危险性很大。MGD发生当日及前30日内一直维持连绵阴雨(其间可中断1～3天),由于降水强度不大,降水被土壤完全吸收。降水的持续,使土壤含水量达到过饱合状态,原来稳定的土壤结构被破坏后,产生了MGD。

诱发MGD的此类型降水个例有83个,占总数的16%,出现的概率相对较小,但多诱发大面积滑坡灾害,造成的经济损失很大。

4 白噪声记录

所研究的514个MGD个例中,有42个当日和前期当地气象站没有强降水,似乎MGD的发生与降水没有关系,但没有强降水是不会发生MGD的。MGD发生前没有强降水记录的原因如下:

(1)降水量观测空间密度不够。云南具有"一山分四季、十里不同天"的气候特征,小尺度单点性强降水经常发生,由于离测站较远,没有观测记录到强降水,但这种强降水却造成了严重的灾害。

(2)降水量分析研究时间密度不够。本文用来研究MGD与降水关系的降水资料为日降水量(24 h降水量),还未对更小时间尺度(6 h、3 h、1 h)降水量与MGD关系进行研究。

在知识发现中将这样的记录标记为白噪声,即

维普资讯 http://www.cqvip.com

不真实记录。今后需采取增设自动站、用雷达回波追踪等办法,从空间上和时间上对降水量加密监测、收集降水资料,以便开展进一步的研究工作。

5 实际应用结果

用数据库中知识发现技术建立的泥石流滑坡灾害危险等级预报系统于2003年起在云南省气象台投入业务运行,在云南省气象宽带网上发布逐日分县指导预报。云南省2003年发生了11次造成重大人员、财产损失的MGD,系统报准了7次,1次基本准确,报错3次(报错的3次,是由于测站监测密度不够,发灾地点的强降水未被观测到,无法做出预报,从灾后收集到的最临近雨量点的降水量来看,当天当地都出现了暴雨)。2004云南发生了7次特大型MGD,均预报准确。

2004年7月19日至20日,德宏州盈江县发生特大型泥石流灾害,49 502人受灾,因灾死亡12人,失踪48人,重伤4人,直接经济损失1.47亿元。对于本次MGD,业务系统7月18日、19日两日均报危险等级为5级(发生MGD可能性很大),提前两天做出了准确的预报,德宏州气象局根据预报及时向州政府发了警报,减少了因灾造成的损失。

6 结论

(1)对于MGD这种比暴雨发生概率小的小概率事件,用相关分析等统计方法无法得出其与降水的定量关系,用人工智能——数据库中知识发现技术能较好地得到有实际意义的结果。

(2)强降水是云南发生MGD的诱因。由于MGD发生前期均有一段降水积累过程,通过对前期降水情况的跟踪和监测,结合短期预报,可以对MGD发生的危险性做出有实际指导意义的预报。

(3)诱发云南MGD的降水类型为3型:暴雨诱发型;多日中、大雨诱发型;连续阴雨诱发型。

参考文献

[1] 单九生,魏丽,刘修奋,等.诱发江西滑坡的降水特征分析[J].气象,2004,30(1):13-15.

[2] 单九生,魏丽,刘修奋,等.诱发江西2002年重大地质灾害的气象条件分析[J].大气科学研究与应用,2004,26(1):26-35.

[3] 李晓.重庆地区的强降雨过程与地质灾害的关系[J].中国地质灾害与防治学报,1995,6(3):24-32.

[4] 马力,廖代强.重庆市山体滑坡气象条件等级预报业务系统[J].应用气象学报.2003,14(1):(122-124).

[5] 李军,周成虎.香港地区滑坡体积与前期降水关系分析[J].自然灾害学报,2002,11(2):36-44.

[6] 钟荫乾.滑坡与降雨关系及其预报[J].中国地质灾害与防治学报,1998,9(4):24-32.

[7] 李媛,朱晓冬,杨冰.滑坡泥石流与降雨关系研究动态[G]//典型地区突发性地质灾害时空预警示范区建设进展材料.北京:中国地质环境监测院,2002:16-28.

[8] 林孝松.滑坡与降雨研究[J].地质灾害与环境保护,2001,12(3):2-7.

[9] 邵军力,张景,魏长华.人工智能基础[M].北京:电子工业出版社,2000:193-202.

[10] 中国气象局预测减灾司.地质灾害气象预报预警技术文集[G].北京:气象出版社,2004:172-176.

Using KDD Technology to Study Relationship between MGD and Precipitation

Peng Guifeng[1] Duan Xu[2] Shu Kanglin[1] Zhou Yi[1]

(1 Yunnan Provincial Meteorological Office, 2 Yunnan Provincial Institute of Meteorology, Kunming 650034)

Abstract: The Knowledge Discovery in Database (KDD) technology is used to analyze and investigate the relationship between precipitation and the damage-inducing meteorological geologic disasters in Yunnan Province. The findings conclude that most meteorological geologic disasters are induced by strong precipitation; there mainly are three types of relationship: induced by heavy rainfall, induced by multi-day moderate-to-heavy rainfall, and induced by consecutive rainfall. The quantitative relationship provides the evidence not only for the meteorological surveillance and forecast of meteorological geologic disasters, but also for the prediction model development in the Yunnan Province.

Key words: knowledge discovery in database, meteorological geologic disaster, precipitation