

基于云模型的水闸泵站远程监测数据挖掘分析

李晓春¹, 张 生²

(1. 上海市青浦区河道水闸管理所, 上海 201799; 2. 上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 为提高水闸泵站监测的预测预警能力, 考虑到影响水闸泵站安全的多因素性和监测数据量庞大的特点, 提出了一种将云模型与改进 Apriori 算法相结合的水闸泵站监测数据关联规则挖掘方法。首先建立监测数据属性空间, 利用逆向云模型实现属性空间的软划分, 将监测数据离散化。为避免传统 Apriori 多次扫描数据库时间长的缺点, 采用二进制存储数据并通过“与运算”获取频繁项集; 考虑到数据的动态增加, 采用增量更新方法对规则进行更新。最后以某市水闸泵站的监测数据分析为例, 验证了所提方法的有效性。

关键词: 安全预警; 云模型; 数据挖掘; 二进制存储; Apriori 算法

中图分类号: TH 122 **文献标志码:** A

Data mining analysis of remote monitoring of sluice pumping station based on cloud model

LI Xiaochun¹, ZHANG Sheng²

(1. Shanghai Qingpu District River Sluice Management Institute, Shanghai 201799, China; 2. School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In order to improve the forecasting and warning ability of sluice pumping station monitoring, considering the multi-factor affecting the safety of sluice pumping station and the huge amount of monitoring data, a method for mining association rules of sluice pumping station monitoring data was proposed, which combines the cloud model with the improved Apriori algorithm. Firstly, the attribute space of monitoring data was established, and the soft partition of attribute space was realized by using the reverse cloud model, and the monitoring data were discretized. In order to avoid the disadvantage of long time taken for multi-scanning database with the traditional Apriori algorithm, the binary was used to storage data and frequent itemsets were got through ‘and operation’. Considering the dynamic increase of data, an incremental update method was proposed to update rules. Finally, the monitoring data analysis of a sluice pumping station in a city was taken as an example to verify the effectiveness of the proposed method.

Keywords: safety warning; cloud model; data mining; binary storage; Apriori algorithm

收稿日期: 2019-03-21

基金项目: 国家重点研发计划项目 (2018YFB1700902)

第一作者: 李晓春(1984-), 男, 工程师。研究方向: 水利信息化。E-mail: 304047616@qq.com

通信作者: 张 生(1968-), 男, 高级工程师。研究方向: 计算机应用。E-mail: zhangsheng@usst.edu.cn

水闸泵站远程监测系统主要用于对区域内河道水位、水质等参数以及水泵、水闸等设备的运行情况进行实时监测, 记录数据并形成历史数据库, 以确保水闸泵站的安全运行并准确预测可能发生的水灾^[1]。当前对于水灾以及水闸泵站安全预测预警方面的研究大多采用单测点模型, 例如文献[2]基于我国山洪灾害雨量预警的需求, 介绍了采用分布式水文模型分析临界雨量的基本思路与方法。文献[3]以南京三汊河闸门失稳风险为例, 采用故障树法对闸门失稳事故进行分析。文献[4]以某近海软基拦河闸变形监测资料为基础, 建立基于变形监测数据的近海软基水闸水平位移监控指标拟定方法。文献[5]结合风险分析对水闸安全监测项目的设置和测点布置等进行了研究。以上采用单测点模型的研究虽然可以在一定程度上对水灾或水闸泵站的安全进行预测预警, 但实际上导致水灾以及影响水闸泵站安全方面的因素往往不止一个, 水闸内外水位高低、雨量大小等都是影响水闸泵站安全的重要因素, 在多因素影响环境中只考虑单因素而得到的规律有一定的应用局限性。考虑多影响因素则需要更多的监测信息, 水闸泵站远程监测系统实时获取的数据是对水闸泵站安全预警的重要依据, 如何分析和使用这些数据是准确预警的关键。针对水闸泵站远程监测数据库中的大量数据, 关联规则挖掘方法能够提取指标属性间的关联规则, 是数据分析的有效方法。文献[6]提出了基于关联规则挖掘的电力生产安全事故事件关键诱因筛选分析方法。文献[7]提出改进的 FP-Growth 算法挖掘隧道交通事故关联规则。文献[8]使用时间序列规则挖掘算法建立交通拥堵传导规律模型, 提高交通效率。文献[9]提出了一种基于关联规则分析的电力变压器故障马尔科夫预测模型, 使用 Apriori 算法挖掘状态参量与状态之间的关联规则以及各状态之间的关联规则。以上研究算法均采用关联规则挖掘算法辅助决策, 其中 Apriori 算法是关联规则挖掘的常用方法。

本文采用 Apriori 算法提取监测数据的关联规则, 为水闸泵站的安全预警提供依据。采用 Apriori 算法提取规则的问题之一是多次扫描庞大的数据库而造成的时间成本较高, 本文采用二进制的方式存储数据。基于加权二进制的 Apriori 算法既能节约存储空间, 又能通过二进制的“与运算”得到频繁项集, 降低时间成本。为解决数据挖掘过程中监测数据实时更新的问题, 采用基于增量更

新的 Apriori 算法提取潜在规则, 即利用水闸站中的闸内外水位、雨量和流量数据, 引入动态更新方法对原有规则进行更新, 通过对比规则对可能发生的水灾作出及时、准确的预测。

在实际情况中, 水闸泵站的监测数据往往具有连续属性, 数据挖掘算法无法直接处理这些数据, 需要将连续属性进行离散化处理。常见的连续数据离散化方法如等距离划分方法、布尔逻辑和粗糙集理论结合的离散化算法等主观性较强, 且在本质上是属性空间的硬划分。文献[10]将云理论引入数据挖掘领域, 提出了一种基于云理论的属性空间软划分方法, 并将其用于煤矿安全监测数据关联规则的挖掘。本文采用云模型对连续数据进行离散化处理, 该方法不仅克服了划分区间过硬的问题, 还考虑了划分过程的不确定性。

本文提出了云模型和改进 Apriori 算法相结合的水闸泵站远程监测数据挖掘分析方法。首先利用云模型将具有连续属性的监测数据进行离散化处理, 得到相应的定性概念语言值; 然后利用水闸泵站中的闸内外水位、雨量和流量数据, 通过采用改进的 Apriori 算法挖掘水闸泵站安全规则; 针对不断更新的动态数据库, 引入动态更新方法对原有规则进行更新, 从而为水闸泵站提供比较好的安全策略。

1 云理论相关知识

1.1 云模型及相关概念

云模型由李德毅院士首次提出, 是一种用语言值描述某定性概念与其数值表示间不确定转换的模型^[11]。

定义 1^[11] 云和云滴: 设 U 是一个普通集合。 $U=\{x_i, i=1, 2, \cdots, n\}$, 称为论域。设 T 为与集合 U 相联系的定性语言值, x_i 对于 T 所表达的定性概念的隶属度为 $\mu(x_i)$ 是一个具有稳定倾向的随机数, 其取值为 $[0, 1]$, 则称 x_i 在论域 U 上的分布为隶属云, 简称云。其中, 每一个确定值和它的隶属度组成的数据对 $(x_i, \mu(x_i))$ 称为一个云滴。

云模型的数字特征用期望值 Ex 、熵 En 和超熵 He 这 3 个数值来表征, 记作向量 $C (Ex, En, He)$, 称为云模型的特征向量, 它构成了定性与定量的相互映射, 把模糊性和随机性集成到了一起。

期望值 Ex : 即论域中定性概念的信息中心值。

熵 En : 即不确定性的度量, 反映在论域中能

够被该定性概念所接受的范围。熵越大，定性概念越模糊，随机性也越大。

超熵 He ：即 En 的熵，是熵的不确定性的度量，间接反映了云的“厚度”。超熵越大，云滴离散度越大，“厚度”也越大。

1.2 云发生器

云发生器 (cloud generator, CG) 即通过软件实现或固化硬件实现的云的生成算法。云可以根据不同条件和不同输入输出变量生成，于是存在不同的云发生器，本文运用逆向云发生器进行属性空间软划分，运用 X 条件云发生器进行数据语言值转换。

由定性概念值到定量数值的过程，即从数字特征到云滴的转换过程，称为正向云发生器；则逆向云发生器就是由定量表示到定性概念的过程，即从云滴转化到数字特征的过程，如图 1 所示。

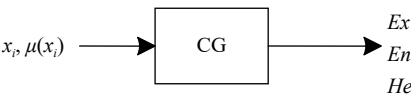


图 1 逆向云发生器示意图

Fig.1 Schematic diagram of a backward cloud generator

在给定论域中特定数值 x 的条件下，云发生器称为 X 条件云发生器^[12]，如果给定云的 3 个数字特征 (Ex, En, He) 和论域上的特定值 x ，利用 X 条件云发生器则可以得到与特征向量个数相对应的 n 个隶属度数值，如图 2 所示。

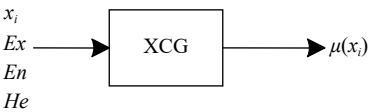


图 2 X 条件发生器示意图

Fig.2 Schematic diagram of a X condition generator

图中， $\mu(x_i) = e^{-\frac{(x_i - Ex)^2}{2En'^2}}$ ， $En'_i \sim N(En, He)$ 。

云模型和云发生器能够实现定量数据与定性数据之间的相互转换，通过定量到定性的转换过程能够将原始的监测数据离散化，从而将离散化后的数据用于后续的 Apriori 算法，达到挖掘关联规则的目的。

2 基于云模型的属性空间软划分

本文挖掘关联规则所使用的 Apriori 算法，其

适用的数据类型为离散型数据，但水闸泵站的监测数据为连续型，因此需要对原始监测数据进行处理。通过分析水闸泵站的监测数据和指标，可以建立属性空间。本文采用基于云模型的属性空间软划分方法^[12]，通过对数据归纳分析，将数字特征值与定性语言值对应起来，即可从数据中产生概念，定量的数据就可以表达成定性的语言概念以实现连续监测数据的离散化。

2.1 监测数据属性空间软划分

设水闸泵站监测数据的关联规则由 m 个条件属性和 1 个决策属性组成，则空间矢量数据 $P = \{p_1, p_2, \dots, p_m, p_{m+1}\}$ 就是一条包含 $(m+1)$ 个属性的历史监测数据， n 条监测数据记录可以表示为矢量域 $P = \{p_1, p_2, \dots, p_n\}$ 。

若集合 $\{C_1, C_2, \dots, C_m, C_{m+1}\}$ 为 P 的非空子集， $C_i = \{C_{i1}, C_{i2}, \dots, C_{iz}\}$ 为任一属性的数字特征集合，其中任意一个数字特征可能对应 $z (1 \leq z \leq n-1)$ 个该属性的基础数据，且 $\forall C_i, C_k (i \neq k)$ ，都有 $C_i \cap C_k = \emptyset$ ，若 $C_1 \cup C_2 \cup \dots \cup C_m \cup C_{m+1} = P$ ，则 $C_1, C_2, \dots, C_m, C_{m+1}$ 为 P 的一个划分。

本文以生成 3 个云 ($l=3$) 为例，选取 {好，中，差} 3 个语言值，对水闸泵站监测数据属性空间进行软划分，方法如下：

根据水闸泵站的历史检测数据，利用 1.3 中的逆向云发生器算法得到属性的数字特征 $C(Ex, En, He)$ 。由于本文选取云的数量为 3，设中间云的特征值为 $C_2(Ex_2, En_2, He_2)$ ，则两侧云的特征值为 $C_1(Ex_1, En_1, He_1)$ ， $C_3(Ex_3, En_3, He_3)$ ，其计算公式如下：

属性

$$\begin{cases} \text{好, } Ex_1 = Ex_2 - vEn_2 \\ \text{中, } Ex_2 = Ex, En_2 = En, He_2 = He \\ \text{差, } Ex_3 = Ex_2 + vEn_2 \end{cases} \quad (1)$$

式中， v 为根据实际情况选取的权值。

2.2 隶属度计算及语言值转换

对监测数据属性空间进行划分后，还需要将数据映射到对应的定性概念中。本文中任意一个属性对应的概念集为 $C_i = \{C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2), C_3(Ex_3, En_3, He_3)\}$ ，与之对应的语言值集合为 {好，中，差}。利用 1.3 中的 X 条件云发生器可以计算属性值 x 对概念集 C_i 中概念的隶属度值，记为 $\mu_i = \{\mu_1, \mu_2, \mu_3\}$ 。比较隶属度大小，选取最大的隶属度对应的语言值作为属性值 x 的最终表现值。例如，若 $\mu_2 > \mu_1 > \mu_3$ ，则

x 对应的概念为 $C_2(Ex_2, En_2, He_2)$, 则语言表现为“中”。

通过上述步骤, 将水闸泵站监测数据属性空间进行软划分后, 将历史监测数据变成了对应语言值, 进而可以利用对应的语言值得到最终的关联规则。

3 考虑增量更新的关联规则挖掘方法

为了确保水闸泵站安全经济地运行, 本文采用 Apriori 关联规则方法对水闸、水泵等设备的运行数据进行挖掘, 即从运行数据中提取关联规则, 从而对可能发生的水灾作出及时、准确的预测。水闸泵站监测数据库中数据量庞大, 且采用 Apriori 算法提取关联规则需要多次扫描数据库, 时间成本高。本文采用二进制的存储方式来存储数据, 该方法能节省空间, 且其采用矩阵乘积形式筛选频繁项集可以降低扫描数据集的次数, 从而降低时间成本。考虑到数据是在动态更新的, 采用传统 Apriori 算法需要对更新后的所有数据进行挖掘, 耗时较多。为此本文采用基于增量更新的改进 Apriori 算法挖掘监测数据库中的潜在安全规则, 从而为水闸泵站可能发生的危害做好预防措施。现将关联规则的相关定义进行介绍。

3.1 基于加权二进制的 Apriori 算法

关联规则挖掘作为数据挖掘中的重要方法, 通过提取指标属性间的关联规则对数据进行分析。Apriori 算法是关联规则挖掘的常用方法, 主要通过筛选频繁项集和候选项集得到支持度和置信度, 从而作为水闸安全规则的提取依据。 P 表示整个水闸泵的数据库, 其中 $Tid=\{t_1, t_2, t_3, \cdots, t_q\}$ 为监测记录的集合, $t_i(i=1, 2, 3, \cdots, q)$ 表示 P 中的每条行记录, 且 $t_i \subset P$ 。传统 Apriori 算法的主要概念如下。

定义 1 频繁项集

给定一个项集 P , 通过项集间的多次筛选获得候选项集 L_k , 若项集 L_k 的支持度大于或等于最小支持度, 记作 $minsup$, 则称项集 L_k 为频繁项集, 也称为频繁模式。

定义 2 Apriori-gen 函数

通过 Apriori-gen 函数生成频繁 k 项集的操作过程包含两个步骤: 连接和剪枝。将 $(k-1)$ 个项集

两两连接, 得到候选项集 C_k ; 对候选项集进行删减, 除去一个不属于 L_{k-1} 的 $(k-1)$ 个子集。

定义 3 强关联规则

设最小置信度为 $minconf$, 当频繁 k 项集的支持度 $sup(P) \geq minsup$, 置信度 $conf(P) \geq minconf$, 即频繁 k 项集为强关联规则。

当提取了关联规则后, 会存在规则冗余或者冲突, 需要进行处理。如果规则库中两个强关联规则 r_1 和 r_2 中的元素相互包含, 若 $r_2 \subseteq r_1$, 则 r_2 是一个冗余规则, 需要删除。如果 r_1 和 r_2 中的元素包含性以及规则优先性相互矛盾, 则规则 r_1 和 r_2 冲突, 需要删除优先度低的规则。

针对 Apriori 算法执行过程中, 需要多次扫描项目集、时间成本高且效率低的问题, 本文提出采用加权二进制的关联规则挖掘方法。针对数据集的冗余性比较高, 为降低频繁项集的效率, 本文给出权重定义, 即重复数据只显示一次, 将重复次数与全部记录数的比值作为权重, 记作 w , 如式(2)所示。

$$w = \frac{count(t_i)}{count(Tid)}$$

(2)

式中: $count(t_i)$ 为每个行记录的重复次数; $count(Tid)$ 为所有记录的总和; w 为记录的比重。

本文所提基于加权二进制的 Apriori 算法步骤如下:

步骤 1 对数据库 P 进行扫描, 项目在数据库中的显示与否设置为“1”或者“0”, 将重复出现的行记录只存储一次, 重复次数的比值权重设置为 w , 并将行数与 w 的乘积结果存储到最后一列, 得到候选集 C_1 , 依据原先设定的最小阈值, 筛选出频繁项集 L_1 。

步骤 2 将 L_1 看做向量 L_1 进行自乘, 通过“与运算”连接, 得出 C_2 候选项集, 扫描 P 得出 L_2 , 以此类推迭代。

步骤 3 扫描数据库 k 次后, 且 $L_k \otimes L_k = C_{k+1}$, 即 C_{k+1} 为自乘的结果, 当 $C_{k+1} = \varnothing$, 停止数据库扫描。

步骤 4 获得最大频繁集 L_k 。

3.2 加权二进制 Apriori 算法的增量更新过程

Apriori 算法虽然能够帮助解决规则提取的问题, 但是, 随着数据的不断增长, 水闸泵站中的数据不断更新。采用 Apriori 算法挖掘更新后的数据耗时多、效率低, 因此, 本文采用 Apriori 算法的增量更新方式来获取频繁项集, 无需对所有数

据进行更新，可以降低时间成本，具体的方法描述如下。

本文水闸泵站的数据集设为 P ，新增数据集为 p ，则新数据集为 $(P+p)$ 。采用 Apriori 算法可以获取原始数据的频繁项集 $L(P)$ ，针对更新后的数据集，本文采用改进的加权二进制 Apriori 算法，相关定义如下：

a. 对于原始数据集形成频繁项集为 $L(P)$ ，新增数据集的频繁项集为 $L(p)$ ，依据某项集 d 在 P 和 p 中为频繁项集，在 $(P+p)$ 中也为频繁项集，找出在 P 和 p 中相同的频繁项集，放入 $(P+p)$ 的频繁项集 $L(P+p)$ 中。

b. 当项集 $d \in L(p)$ 且 $d \notin L(P)$ 时，扫描数据集 P 得到支持度 $support_P$ ，再在 p 中求出 $support_p$ ，以及求出 d 在 $(P+p)$ 中的 $support_{P+p}$ ，如式 (3) 所示，若 $support_{P+p} \geq minsup$ ，则将 d 作为频繁项集 $L(P+p)$ 中的一部分，反之不是频繁项集。

$$support_{P+p} = \frac{support_P|P| + support_p|p|}{|P| + |p|} \quad (3)$$

c. 当项集 $d \in L(P)$ 且 $d \notin L(p)$ 时，扫描数据集 p 得到支持度 $support_p$ ，再在 P 中求出 $support_P$ 以及求出 d 在 $(P+p)$ 中的 $support_{P+p}$ ，如式 (4) 所示，若 $support_{P+p} \geq minsup$ ，则将 d 作为频繁项集 $L(P+p)$ 中的一部分，反之不是频繁项集。

$$support_{P+p} = \frac{support_P|P| + support_p|p|}{|P| + |p|} \quad (4)$$

d. 若项集 d 在 P 中为非频繁项集，在 p 中也为非频繁项集，那项目集在 $(P+p)$ 中也为非频繁项集。

基于以上相关概念，本文所提 Apriori 算法的增量更新步骤如下。

输入水闸站监测原始数据集 P ，新增数据集 p ，最小阈值设定为 $minsup$ 。

步骤 1 对于某项集 d 而言， $d \in L(P)$ ，如果 $\{d\} = L(P+p)$ ，删除与 $L(p)$ 相同的频繁项集，一直循环，直到得到既属于 P 也属于 p 的频繁项集 d 。

步骤 2 当 $d \in L(P)$ ，如果有事务集 $s \in p$ 且 $d \subseteq s$ ，将 d 出现的次数累加为 $count$ 。

步骤 3 当 $count \geq minsup$ ，将数据集 d 添加到频繁项集 $L(P+p)$ 当中，重复步骤 2，直到扫描结束，得到最终的频繁项集。

步骤 4 根据频繁项集提出数据的安全规则的指标，即支持度和置信度，来验证水闸站的安全性。

4 案例分析

某市地势低平，水力比较平缓，同时河道中有大量聚集沉淀的污染物，当大暴雨突然袭击时，由于城市里多余的水量很难及时排出而常常形成水涝。水涝危害市内居民安全且会造成大量财产损失，因此需要对当地水闸泵站的远程监测数据进行分析，从而对水灾进行准确的预测预警。在预测预警过程中，仅仅依靠系统监测到的闸内外水位、雨量、流量和水闸位数据并不能满足要求，还需要根据已有的数据对潜在的安全规则进行挖掘，以便对水灾进行准确的预测预警。结合某市水务局的水务监测数据和该市某水闸泵站的水闸位等记录数据，利用本文的方法对安全规则进行挖掘。其中水闸泵站监测的部分监测记录基础数据如表 1 所示。

表 1 水闸泵站监测基础数据

Tab.1 Monitoring data of the sluice pumping station

监测记录	内河水位/ m	外河水位/ m	雨量/ (mm · d ⁻¹)	流量/ (m ³ · s ⁻¹)	水闸位/ m
S ₁	3.02	1.72	1.22	0.98	3.22
S ₂	2.60	3.50	1.22	1.23	2.22
S ₃	2.34	2.29	1.12	1.13	3.22
S ₄	2.55	2.08	1.13	1.09	3.22
S ₅	2.58	1.78	1.10	1.16	2.22
S ₆	3.05	2.56	1.13	1.15	2.22
S ₇	2.57	1.86	1.14	1.13	2.22
S ₈	2.86	2.75	1.21	1.22	2.22
S ₉	3.07	1.90	1.23	0.87	2.22
S ₁₀	3.10	2.25	0	1.23	2.52
S ₁₁	3.11	1.98	1.22	1.11	2.22
S ₁₂	2.88	3.01	1.25	1.32	3.22
S ₁₃	3.01	1.86	1.24	1.08	2.22
S ₁₄	3.03	1.95	1.22	0.96	2.22
S ₁₅	2.57	2.38	1.11	1.06	3.22
S ₁₆	2.67	3.02	1.21	0.89	2.22
S ₁₇	2.85	2.21	1.12	1.01	2.22
S ₁₈	3.00	2.67	1.22	1.11	2.22
⋮	⋮	⋮	⋮	⋮	⋮

利用基于云模型的属性空间软划分方法对表 1 中的监测数据集进行属性空间软划分。将监测的各项属性均看作一维语言变量, 对于雨量和流量属性, 定义小、中、大 3 个语言值; 对于内外河水位及水闸位, 定义低、中、高 3 个语言值。

以内河水位为例, 对表 1 中数据调用逆向云算法, 计算得到反映内河水位的数字特征为 $Ex=2.696$, $En=0.264$, $He=0.118$, 对应语言值为“中”。取权重 $v=1$, 则与语言值“高”对应的数字特征为 $Ex_3=2.959$, $En_3=0.377$, $He_3=0.349$, 与“低”对应的数字特征为 $Ex_1=2.433$, $En_1=0.359$, $He_1=0.330$ 。

对表 1 中的初始数据进行软划分后, 使用 X 条件云发生器计算隶属度并根据隶属度进行排序, 得到如表 2 所示的数据。以内河数据中“3.02”为例, 计算得到隶属于 3 个概念的隶属度

表 2 经过软划分的数据表
Tab.2 Soft partitioned data

监测记录	内河水位	外河水位	雨量	流量	水闸位
S_1	高	低	大	小	高
S_2	中	高	大	大	低
S_3	低	中	中	中	高
S_4	低	低	中	中	高
S_5	低	低	中	中	低
S_6	高	中	大	中	低
S_7	低	低	中	中	低
S_8	中	中	大	大	低
S_9	高	低	大	小	低
S_{10}	高	中	小	大	中
S_{11}	高	低	大	小	低
S_{12}	中	高	大	大	高
S_{13}	高	低	大	中	低
S_{14}	高	低	大	小	低
S_{15}	低	中	中	中	高
S_{16}	低	高	大	小	低
S_{17}	中	低	中	中	低
S_{18}	高	中	大	中	低
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

分别为: 0.19, 0.03, 0.95。可以得到结论: 该值隶属于“内河水位高”这一概念的程度最大。将该结论记录到结果数据库中, 该条记录在数据库中应为“内河水位=高”。

针对表 2 中的数据, 采用改进的 Apriori 算法挖掘潜在关联规则。设置支持度为 10%, 置信度为 50%, 得出关联规则如表 3 所示。

表 3 数据库集中提取的关联规则
Tab.3 Association rules extracted from the database set

序号	IF	THEN	支持度/%	置信度/%
1	内河水位=高 雨量=大 流量=小	外河水 位=低 水闸位=低	16.70	42.86
2	内河水位=高 雨量=大 流量=中	外河水 位=高 水闸位=低	16.70	67.70
3	内河水位=高 雨量=大 流量=小	外河水 位=低 水闸位=高	5.60	25.00
4	内河水位=低 雨量=中 流量=中	外河水 位=低 水闸位=低	11.11	67.70
5	内河水位=低 雨量=中 流量=中	外河水 位=中 水闸位=高	11.11	50.00
6	内河水位=中 雨量=大 流量=大	外河水 位=高 水闸位=高	5.60	33.40
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

从表 3 中规则可以得知, 当内河水位高, 雨量大且流量小, 则得出外河水位低且水闸位低的支持度和置信度比较大, 而外河水位低以及水闸位高的支持度和置信度较低。由此得出, 当内外河水位差较大, 雨量大且水流慢时, 若水闸位较低, 可以降低水灾的危险性。当内河水位低, 雨量和流量中等时, 得出外河水位高低以及水闸位高低的支持度和置信度较大, 可以推出水闸位与内外河水位, 雨量流量密切相关, 此时雨量和流量大小直接影响水位、水闸位高度大小, 与水库安全性有着直接关系。当内河水位较低, 雨量和流量处于中等程度时, 水闸位较低的置信度较高, 水闸位高的置信度较低, 说明水闸泵站的闸高度未拉开太大, 才能预防水闸可能产生隐患的危险性。

5 结束语

本文针对水闸泵站监测预警的影响因素多、数据庞大的特点，提出了基于云模型的水闸泵站远程监测数据挖掘分析方法，特点如下：

- a. 基于云模型的属性空间软划分方法处理水闸泵站监测数据，将数值数据转化为数据挖掘算法可用的定性概念语言值。
- b. 利用二进制存储监测数据值，并通过“与运算”获取频繁项集。考虑到数据是不断更新的，采用基于增量更新的 Apriori 算法挖掘水闸泵站指标的规则，分析规则后的隐含意义，从而为水闸泵站可能的隐患给出预测和预警。

通过对某市水闸泵站的监测数据进行挖掘分析后得出关联规则，验证了所提方法的可行性和有效性。后续研究考虑将基于云模型的属性空间软划分方法与水闸泵站智能监测系统结合起来对实时监测数据进行转换，以呈现更直观具体的水闸泵站安全情况。

参考文献：

[1] 张大庆. 区域水闸泵站远程监测系统[J]. 仪器仪表与分析监测, 2013(2): 16–18.

[2] 李昌志, 郭良, 刘昌军, 等. 基于分布式水文模型的山洪预警临界雨量分析——以涔水南支小流域为例[J]. 中国

防汛抗旱, 2015, 25(1): 70–76.

[3] 臧英平, 沃玉报, 徐影. 基于故障树法的水闸工程安全预警体系研究[J]. 水资源与水工程学报, 2015, 26(4): 163–168.

[4] 付传雄, 张君禄, 廖文来, 等. 近海软基水闸水平位移监控指标研究[J]. 水利与建筑工程学报, 2017, 15(6): 173–176.

[5] 方卫华. 水闸安全监测及可靠性评价研究[J]. 贵州水力发电, 2006, 20(2): 61–65.

[6] 陈碧云, 丁晋, 陈绍南. 基于关联规则挖掘的电力生产安全事件事件关键诱因筛选[J]. 电力自动化设备, 2018, 38(4): 68–74.

[7] 刘翔翔, 韩贝. 基于改进 FP 算法的隧道交通事故关联分析[J]. 现代电子技术, 2018, 41(17): 141–144.

[8] 周辉宇. 基于大数据规则挖掘的交通拥堵治理研究[J]. 统计与信息论坛, 2017, 32(5): 96–101.

[9] 王有元, 周立玮, 梁玄鸿, 等. 基于关联规则分析的电力变压器故障马尔科夫预测模型[J]. 高电压技术, 2018, 44(4): 1051–1058.

[10] 孟凡荣, 郑中珮, 陈佩佩. 一种基于云理论的空间关联规则提取算法[J]. 微计算机信息, 2008, 24(27): 135–137.

[11] 李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(6): 15–20.

[12] 孟凡荣, 宋春景, 郑中珮, 等. 基于云理论的煤矿安全监测数据关联规则挖掘[J]. 小型微型计算机系统, 2008, 29(9): 1622–1626.

（编辑：丁红艺）