DOI:10.3880/j.issn.1006-7647.2019.05.014

基于关联规则的土石坝渗流推理预测方法及应用

李 俊[1],苏怀智[1,2],周仁练[1]

(1.河海大学水利水电学院,江苏 南京 210098;
2.河海大学水文水资源与水利工程科学国家重点实验室,江苏 南京 210098)

摘要:将关联规则算法引入到土石坝渗流的预测中,并建立预测模型。首先将渗流量和上游水位、降雨量、下游水位等组成事务组,利用 K-means 算法对实测数据进行预处理,然后通过 Apriori 算法进行频繁项集的挖掘,产生强关联规则,最后将强关联规则用于渗流量的预测。本模型可以较好的解释影响渗流变化的内在原因,将本预测模型应用于某土石坝工程中,结果表明该模型具有较好的精度。

关键词:土石坝;渗流预测;关联规则;Apriori 算法;K-means 算法

中图分类号:TV314 **文献标志码**:A **文章编号**:1006-7647(2019)05-0089-06

Seepage prediction method based on association rules and its application to earth-rock dams //LI Jun[1], SU Huaizhi[1,2], ZHOU Renlian[1] (1. *College of Water Conservancy and Hydropower Engineering*, *Hohai University*, *Nanjing* 210098, *China*; 2. *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering*, *Hohai University*, *Nanjing* 210098, *China*)

Abstract: The association rule algorithm was introduced into the prediction of seepage behavior of earth-rock dams, and a prediction model was established. Firstly, the seepage flow, upstream water level, rainfall, downstream water level, etc. were formed into a transaction group. Secondly, the K-means algorithm was used to preprocess the measured data, and then the Apriori algorithm was used to mine the frequent item sets, resulting in strong association rules. Finally, strong association rules were used for the prediction of the seepage flow. This model can better explain the intrinsic reasons of the seepage variation. The prediction model was applied to a certain earth-rock dam project, and the results show that the model has good accuracy.

Key words: earth-rock dam; seepage prediction; association rule; Apriori algorithm; K-means algorithm

我国已建大坝 9.8 万余座,其中绝大多数为土石坝。渗流是影响土石坝安全的重要因素之一,土石坝的各种破坏形式都直接或间接与渗流有关,所以准确预测渗流量的大小对实现土石坝的安全监控有着重要意义[1-2]。根据已建立的渗流量统计模型,土石坝渗流量主要受上下游水位、降雨入渗以及坝前淤积和防渗体的时变过程等影响[3-4]。但土石坝渗流性态受多种因素的影响,预测难度很大[5]。

目前,界内对于渗流预测的研究主要集中在神经网络、时间序列、回归分析等方面,这些方法或多或少存在一定的缺陷。神经网络存在收敛速度慢、泛化能力弱等缺点;时间序列预测方法在处理模糊数据时可能会出现模型过度拟合的危险;回归分析需要建立复杂的数学模型,通过求解线性方程组得到各个回归系数,当线性方程组出现病态或奇异时,回归方法是无法处理的[6-12]。1993 年,Agrawal 等首先提出关联规则的概念。关联规则是形如 $X \rightarrow Y$ 的蕴涵式,即由事物 X 推导出事物 Y。其中,X 和 Y 分别称为关联规则的先导和后继,并且关联规则 $X \rightarrow Y$ 存在支持度和信任度。Apriori 是一种经典的关联规则算法,用于挖掘大数据中潜在的事物联系。该算法采用自底向上的遍历思想逐级挖掘,以确保关联规则的准确性。针对经典 Apriori 算法执行效率低的缺点,周发超等[13]引进 TID 标识码,使得算法效率大大提高;林郎碟等[14]将 Apriori 算法应用于图书推荐服务中,结合“分割-整合”的思想,为建设智能图书推荐系统打下理论基础;Guo 等[15]将温度、湿度、气压和风速打包成事务组,进行短期风速

基金项目:国家重点研发计划(2018YFC0407101,2017YFC0804607);广西重点研发计划(桂科 AB17195074)

作者简介:李俊(1995—),男,硕士研究生,主要从事水工结构与大坝安全监测研究。E-mail: 1921336117@qq.com

通信作者:苏怀智(1973—),男,教授,博士,主要从事涉水工程安全防控与提能延寿研究。E-mail: su_huaizhi@hhu.edu.cn

预测,将预测结果用于校正由混沌时间序列所预测的结果,取得较好的预测精度。本文将 Apriori 算法用于挖掘环境量(包含上下游水位、降雨量等)与渗流量之间的内在关系,建立推导法则,从而为大坝渗流监控和预报提供一种新途径。

本文的主要思路是利用某土石坝实测数据建立预测模型,然后检验该模型的预测精度并给予评价。在对实测数据进行预处理时,首先分析环境量和渗流量的相关关系,用 K-means 聚类算法对环境量和渗流量进行离散化处理,然后将处理过的数据转换成 Apriori 算法能够识别的布尔型矩阵进行关联规则的挖掘,最后将关联规则用于渗流量的预测。通过实例演示和精度分析,证实了该预测模型具有较好的预测精度。

1 土石坝渗流监测数据预处理

1.1 渗流量与环境量相关关系及滞后性分析

本文仅考虑渗流量受上游水位、降雨量和下游水位的影响[3]。当环境量(上游水位、降雨量和下游水位)发生变化时,渗流量不会立刻响应,所以在进行关联规则挖掘时,在同一组数据内,环境量和渗流量不应是同一时间,环境量应较渗流量有所提前,此性质恰好可以用于建立预测模型[16]。需要注意的是,环境量相对渗流量提前的时间也各不相同,应根据环境量的性质分别计算。

1.1.1 上游水位

上游水位是影响渗流量大小的重要因素,显然,在其他环境量保持稳定的前提下,渗流量与上游水位呈正相关关系。有两种方法估计上游水位的渗流响应滞后时间:

a. 利用大坝泄流。水库一般会在雨季来临之前或者需要放水抗旱时大量泄流以降低库水位,根据工程经验,一段时间后,坝体渗流量也会随之下降。所以,选取某一次泄洪前后且无降雨的时间段内一组数据,直接观察渗流量变化滞后的时间,即为上游水位的渗流响应滞后时间。

b. 利用稳定数据。如果没有合适的泄洪时间段,那就只能选用多段稳定数据来估算。当水库不泄洪且无降雨时,水库水量损失主要是因为蒸发和渗流,水库水位会稳步下降,渗流量也会相应变化。可以选取多段无降雨时段,观察渗流量的滞后时间,取均值(以天为单位,取整)作为上游水位的渗流响应滞后时间。

1.1.2 降雨量

降雨量对土石坝渗流量的影响较为明显,也比较容易判断渗流响应滞后时间。选取前后一段时间均无降雨,且上下游水位变化幅度较小或变化稳定的短时降雨,观察渗流量响应时间,作为降雨量的渗流响应滞后时间。

1.1.3 下游水位

通常情况下,下游水位变化不大,所以该模型不考虑下游水位的渗流响应滞后时间。为了方便建模,下游水位取与上游水位对应的同一天的值。

1.2 实测数据离散化

关联规则算法不具备处理连续数据的能力,所以在进行关联规则的挖掘前,应先将实测数据进行离散化处理。所谓离散化,即将连续型环境量(或渗流量)所处的区间,利用 K-means 算法划分成多个短区间(簇),所处同一短区间(簇)的数据归为一类,用区间编号代替。K-means 是一种基于形心的划分算法,使用簇 C_i 的形心代表该簇。K-means 算法把簇的形心定义为簇内点的均值[14]。

以渗流量为例介绍该算法的处理流程。假设 m 天的渗流量实测数据需要划分为 k 个簇,随机选取 k 个数据,每个数据作为一个簇的初始形心,对剩下的每个数据,根据其与各个簇中心的距离,把他分配到最相似的簇。然后,对于每个簇,使用上次迭代分配到该簇的对象,计算新的均值。然后,使用更新后的均值作为新的簇中心,重新分配所有对象。继续迭代,直到分配稳定,即本轮形成的簇与前一轮形成的簇相同[17]。

1.3 布尔型转换

将实测数据聚类以后,为满足 Apriori 算法对输入数据的要求,还需要将聚类后的实测数据转换成布尔型数据,即只用 0 和 1 表示。

以渗流量和上游水位为例,简要说明转换过程。假设有 m 组数据,矩阵表中每一行表示同一天的监测值,每行数据均包含渗流量、上游水位、降雨量和下游水位 4 个变量,构成 $m\times4$ 矩阵 $\boldsymbol{A}\in\mathbf{R}^{m\times4}$。利用 K-means 算法将渗流量和上游水位分别划分为 k_1、k_2 类,对于矩阵中的所有实测数据,均用该实测数据所处的簇编号代替。

比如渗流量被聚类成 k_1 类,转换成布尔型矩阵时,渗流量对应布尔型矩阵的前 k_1 列,且每个实测数据的簇编号对应的布尔型矩阵中渗流量列编号,用 1 标注该列,其余列均为 0。上游水位、降水量、下游水位都可用相同方法转换。

2 基于 Apriori 算法的土石坝渗流关联规则

利用 Apriori 算法挖掘关联规则可分为两步,首先挖掘实测数据中的频繁项集,然后将频繁项集推理成关联规则。支持度是指项集在数据库中出现的

次数或频率,置信度是指关联规则在数据库中的准确性。算法中,所有支持度大于最小支持度的项集称为频繁项集,所有置信度大于最小置信度的关联规则称为强关联规则[18]。最小支持度和最小置信度的取值应保证算法所挖掘出来的强关联规则的数目合理,且将强关联规则用于工程预测时,预测精度要满足工程要求[19]。

利用 Apriori 算法产生频繁项集的过程主要分为连接和剪枝两步:首先,将包含 i 个变量的项集称为 i -项集,将包含 i 个变量的频繁项集称为 i -频繁项集,扫描所有实测资料,产生候选 1 -项集,根据最小支持度,产生 1 -频繁项集。然后由 1 -频繁项集自连接产生 2 -项集,对 2 -项集剪枝处理,即剔除 2 -项集中有非空子集是非频繁项集的项,再根据最小支持度,产生 2 -频繁项集。重复以上步骤,直到得出包含上游水位、降雨量、下游水位和渗流量 4 个变量的 4 -频繁项集[20]。

然后由频繁项集产生强关联规则,例如 2 -频繁项集($\boldsymbol{U}$, $\boldsymbol{S}$),其中 $\boldsymbol{U}$ 为取值在区间[a,b]内的上游水位;S 为取值在区间[c,d]内的渗流量。

则关联规则由 $\boldsymbol{U}$ 推导出 $\boldsymbol{S}$ 的支持度 S_S 和置信度 C_C 的计算公式分别为:

$$S_S(\boldsymbol{U} \rightarrow \boldsymbol{S}) = P(\boldsymbol{U} \cup \boldsymbol{S}) \tag{1}$$

$$C_C(\boldsymbol{U} \rightarrow \boldsymbol{S}) = P(\boldsymbol{U} \mid \boldsymbol{S}) \tag{2}$$

该关联规则可以表述为:当上游水位在区间[a,b]内时,渗流量有 $P(\boldsymbol{U} \mid \boldsymbol{S})$ 的可能性出现在区间[c,d]内,且这种情况出现的频率为 $P(\boldsymbol{U} \cup \boldsymbol{S})$。

3 基于关联规则推理的土石坝渗流预测实现过程

综合应用 K-means 算法和 Apriori 算法,基于土石坝渗流关联规则的推理,实现土石坝渗流预测的具体流程如图 1 所示,主要步骤如下。

步骤 1:渗流量与环境量相关关系及滞后性分析。利用前文提到的方法分别计算环境量的渗流响应滞后时间,根据滞后时间,将理论上具有因果关系的渗流量与环境量调整为同一组数据,构成新的数据组。

步骤 2:聚类离散化。利用 K-means 算法对新的数据组进行聚类分析,并用聚类出来的簇编号代替实测数据,构成离散型实测数据矩阵。

步骤 3:布尔型转换。将离散型实测数据矩阵转换成 Apriori 算法能识别的布尔型矩阵。

步骤 4:关联规则的挖掘。将得到的布尔型矩阵作为 Apriori 算法的输入矩阵 $\boldsymbol{D}$,同时输入最小支持度和最小置信度进行关联规则的挖掘。

步骤 5:关联规则的筛选。经过上述关联规则挖掘后,会产生多个 4 -频繁项集,并由此产生多种强关联规则。但是并不是所有的强关联规则都能用于预测,只有形如 $\boldsymbol{U}$、$\boldsymbol{R}$、$\boldsymbol{D} \rightarrow \boldsymbol{S}$(其中 $\boldsymbol{U}$、$\boldsymbol{R}$、$\boldsymbol{D}$ 分别为上游水位、降雨量和下游水位,$\boldsymbol{S}$ 为渗流量)的强关联规则才是有用的[19]。所以需要对所产生的强关联规则进行筛选。

步骤 6:渗流量的预测。对于所筛选出的强关联规则 $\boldsymbol{U}$、$\boldsymbol{R}$、$\boldsymbol{D} \rightarrow \boldsymbol{S}$,环境量均提前渗流量一段时间,设环境量提前时间分别为 m_U、m_R、m_D,若要预测某一天的渗流量,环境量应分别取 m_U、m_R、m_D 前的值。显然,所产生的渗流量预测结果是一个聚类区间,而不是一个准确的值。考虑到聚类所产生的区间大小不一,不适合作为预测结果,本文采用区间上下边界的均值作为预测结果。

步骤 7:预测结果评价。本文选取平均绝对误差、平均相对误差以及均方根误差作为评价指标[21]。

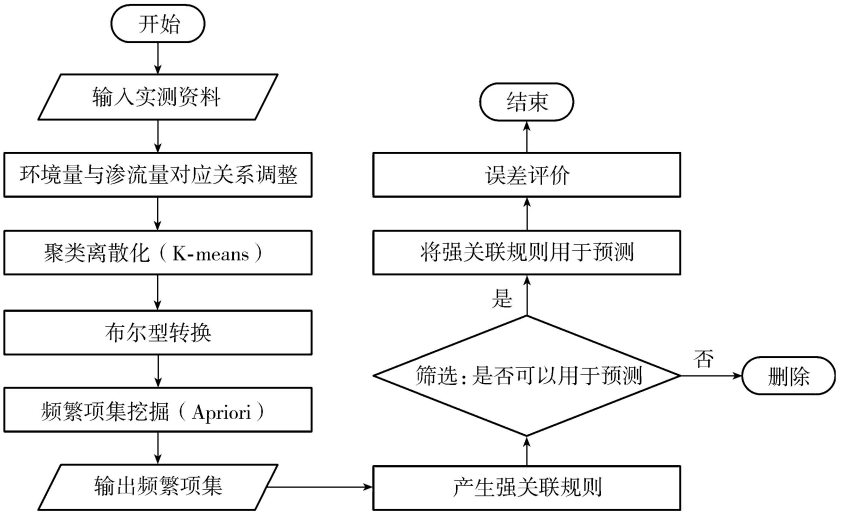
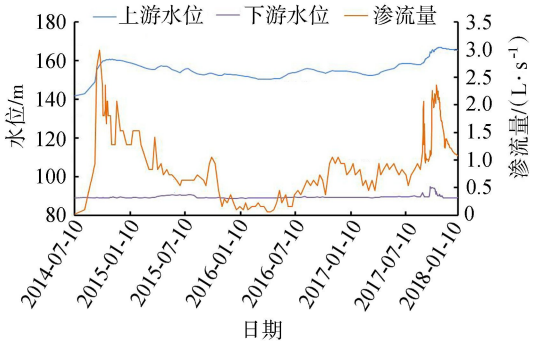


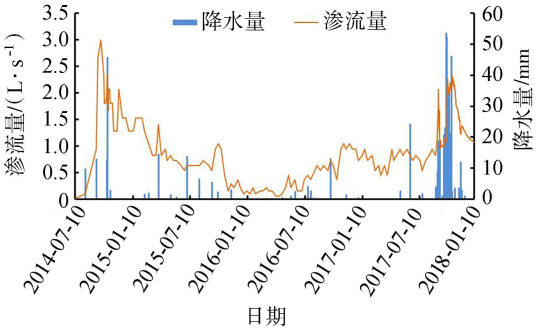
图 1 预测流程

4 工程实例分析

以某水库左岸土石坝段渗流量为预测对象。该坝段坝顶高程 177.60 m,设计蓄水位 170.0 m,共布设 3 个渗流监测点,分别监测先锋沟、左下挡和结合面部位。对先锋沟观测点在 2014 年 7 月 10 日至 2017 年 12 月 29 日期间的渗流量(L/s)进行 220 次观测,其中前 200 组数据作为训练集,后 20 组数据作为检验集。具体渗流量及相关环境量过程线见图 2。



(a)渗流量及上下游水位过程线



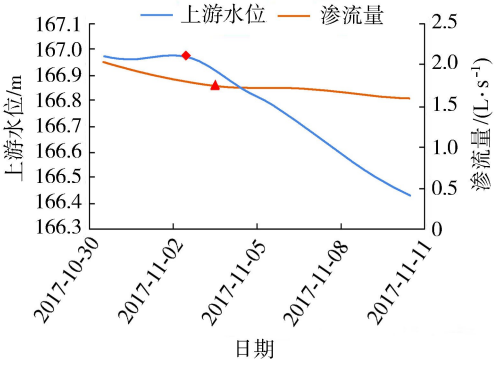
(b)渗流量及降水量过程线

图 2 某水库左岸土石坝段渗流量及相关环境量过程线

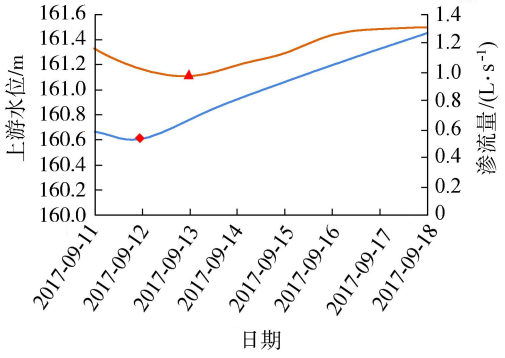
4.1 渗流量与环境量相关关系及滞后性分析

选取两组典型的上游水位与渗流量变化趋势,绘制成折线图,观察两组曲线的关键折点,可确定上游水位渗流量响应滞后时间为 1 d,同时下游水位渗流量响应滞后时间也为 1 d。

选取前后一段时间内均无降水的一次短时降水后的渗流量变化趋势图,如图 4 所示,观察得到降水量渗流响应滞后时间为 0。在预测时,降水量的取值应根据天气预报来确定。



(a)趋势一



(b)趋势二

图 3 典型上游水位与渗流量变化趋势

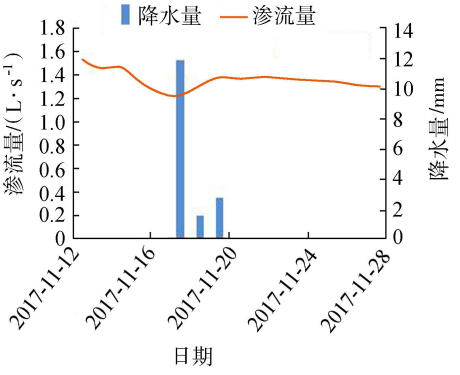


图 4 典型降水量与渗流量变化趋势

经过上述计算,应先将当天的渗流量、降水量与前一天的上、下游水位调整为同一组数据,然后进行后续处理。

4.2 聚类离散化及布尔型转换

对渗流量及环境量用 K-means 聚类算法离散化,根据区间长度分别取 $k_1=5$,$k_2=5$,$k_3=4$,$k_4=3$,结果如表 1 所示。

表 1 K-means 聚类算法对各变量聚类结果

聚类簇	渗流量			上游水位			降水量			下游水位		
	区间/(L·s⁻¹)	均值/(L·s⁻¹)	占比/%	区间/m	均值/m	占比/%	区间/mm	均值/mm	占比/%	区间/m	均值/m	占比/%
1	[0.02,0.47]	0.245	16.3	[141.60,145.07]	143.335	7.8	[0,0.1]	0.05	63.2	[88.72,89.91]	89.315	28.5
2	[0.48,0.92]	0.700	17.4	[145.42,149.46]	147.440	12.6	[0.1,9.7]	4.90	14.8	[89.93,91.34]	90.635	44.3
3	[0.94,1.37]	1.155	25.3	[149.72,155.63]	152.675	25.4	[9.9,18.6]	14.25	10.4	[91.35,94.70]	93.025	27.2
4	[1.38,2.12]	1.750	21.2	[156.02,161.33]	158.675	28.7	[18.7,53.5]	36.10	11.6			
5	[2.15,2.99]	2.570	19.8	[161.35,166.98]	164.165	25.5						

根据上述结果,可将 200×4 原始数据矩阵转换成 200×17 布尔型矩阵,用于关联规则挖掘。

4.3 关联规则的挖掘与筛选

将处理好的 200×17 布尔型矩阵作为 Apriori 算法的输入数据,由于最小支持度和最小置信度的取值尚无经验可循,为得到合理数量的强关联规则并达到预期的精度,经反复尝试,最终确定取最小支持度为 3,最小置信度为 0.6,进行关联规则的挖掘,得到 18 个 4 -频繁项集,由该频繁项集,可推导出 64 组强关联规则。筛选出可以用于预测的关联规则,整理结果如表 2 所示。表中下标表示该物理量所在的聚类簇,如表中第一行 U_3、R_1、$D_1 \rightarrow S_1$ 表示当上游水位在区间 3,降水量在区间 1,下游水位在区间 1 时,降水量有 75% 的可能出现在区间 1,且这种情况出现了 27 次。

表 2 筛选出的可以用于预测的强关联规则

序号	关联规则	支持度	置信度
1	U_3、R_1、$D_1 \rightarrow S_1$	27	0.75
2	U_3、R_1、$D_2 \rightarrow S_2$	3	1
3	U_3、R_2、$D_1 \rightarrow S_1$	4	0.80
4	U_4、R_1、$D_1 \rightarrow S_2$	17	0.71
5	U_4、R_1、$D_2 \rightarrow S_2$	6	0.67
6	U_5、R_1、$D_2 \rightarrow S_5$	10	1
7	U_5、R_2、$D_1 \rightarrow S_4$	6	1
8	U_3、R_1、$D_3 \rightarrow S_5$	3	0.75
9	U_5、R_4、$D_3 \rightarrow S_4$	4	0.80

4.4 利用强关联规则预测及评价

由于用于挖掘的数据较少,并不能挖掘出所有环境量区间的任意组合(共有 120 种组合)的关联规则。所以,并不是检验集中的所有数据都能用于预测检验。只有环境量符合挖掘出的关联规则的组合,才可以预测。这里只对符合所产生关联规则的 6 组数据做预测,结果如表 3 所示。

表 3 渗流量预测结果误差

序号	真实值/($L \cdot s^{-1}$)	预测值/($L \cdot s^{-1}$)	绝对误差/($L \cdot s^{-1}$)	相对误差/%
1	1.41	1.750	0.340	19.4
2	1.48	1.750	0.270	15.4
3	1.38	1.750	0.370	21.1
4	1.38	1.750	0.370	21.1
5	0.97	1.155	0.185	16.0
6	1.29	1.155	-0.135	-11.7

由表 3 可知,6 组数据的平均绝对误差为 0.278 L/s,相对误差为 17.45%,均方差为 0.293 L/s,根据强关联规则所预测的结果均有较好的精度,相对误差控制在 20% 左右,预测精度良好。这表明,在进行关联规则的挖掘时,取置信度为 0.6 是足够的,若为了提高预测的准确率而过度的提高置信度,会使得可用的强关联规则数量减少,这显然是不合理的。

5 结 语

基于 Apriori 算法对关联规则的挖掘,本文建立了一种新的渗流量预测模型,该模型的特征总结如下:

a. 内部逻辑关系明确,易于通过各种编程软件实现。这为继续深入研究改进提供了很大的便利。

b. 预测精度高。渗流量的预测不同于坝体变形等方面的预测,渗流量的影响因素多,各影响因素之间关系复杂且变化多样,常规算法很难对渗流量进行准确预测。本文所构建的预测模型,将预测精度控制在 20% 左右,基本实现预期目标。

c. 通过对比相似的关联规则,可初步判断环境量与渗流量间的变化规律。例如,在表 4 中,对比规则 5、6,可知上游水位对渗流量的影响非常显著;同样的,对比规则 1、3,可知当降水量较小时,对渗流量的影响也较小。

d. 显然,增大置信度或支持度会提高模型的预测精度,但也会导致产生的强关联规则较少,即模型的可预测范围会缩小;反之,减小置信度或支持度会大幅增加强关联规则的数量,使模型预测范围扩大,但同时预测精度也会降低。所以,本模型的后续研究应该着重于置信度及支持度的大小(预测精度)和关联规则数目(预测范围)的权衡问题。

参考文献:

[1] 孙立昌. 土石坝渗流预测研究[D]. 南宁:广西大学, 2012.

[2] 沈振中,邱莉婷,周华雷. 深厚覆盖层上土石坝防渗技术研究进展[J]. 水利水电科技进展, 2015,35(5):27-35. (SHEN Zhenzhong, QIU Liting, ZHOU Hualei. Review of seepage control of earth-rockfill dams on thick overburden layer [J]. Advances in Science and Technology of Water Resources, 2015,35(5):27-35. (in Chinese))

[3] 吴中如. 水工建筑物安全监控理论及其应用[M]. 北京:高等教育出版社,2003.

[4] WANG Shaowei, XU Yingli, GU Chongshi, et al. Monitoring models for base flow effect and daily variation of dam seepage elements considering time lag effect[J]. Water Science and Engineering, 2018, 11(4): 344-354.

[5] XIANG Yan, FU Shuyan, ZHU Kai, et al. Seepage safety monitoring model for an earth rock dam under influence of high-impact typhoons based on particle swarm optimization algorithm [J]. Water Science and Engineering, 2017, 10(1): 70-77.

[6] CHEN Shoukai, HE Qidong, CAO Jigang. Seepage simulation of high concrete-faced rockfill dams based on

generalized equivalent continuum model [J]. Water Science and Engineering, 2018, 11(3): 250-257.

[7] 陈端,曹阳,夏辉,等. GRNN 神经网络在坝基渗流预测中的应用[J]. 人民黄河,2012,34(10):118-119,123. (CHEN Duan, CAO Yang, XIA Hui, et al. Application of generalized regression neural network in prediction of dam foundation seepage [J]. Yellow River, 2012, 34 (10): 118-119,123. (in Chinese))

[8] 钱镜林,陈闲忠,张晔. 基于相空间重构理论的神经网络在大坝渗流预测中的应用[J]. 水利与建筑工程学报, 2006, 4 (1): 21-23. (QIAN Jinglin, CHEN Xianzhong, ZHANG Hao. Application of ANN based on phase-space reconstruction in prediction of seepage in dam [J]. Journal of Water Resources and Architectural Engineering,2006,4(1):21-23. (in Chinese))

[9] 李鹏犇,苏亮渊,贾亚杰,等. 基于多因素影响的 BP-RBF 神经网络渗流预测模型[J]. 人民黄河,2018,40(4):132-135. (LI Pengben, SU Liangyuan, JIA Yajie, et al. BP-RBF neural network seepage model under the influence of various factors [J]. Yellow River, 2018, 40 (4):132-135. (in Chinese))

[10] SU Huaizhi, OU Bin, YANG Lifu, et al. Distributed optical fiber-based monitoring approach of spatial seepage behavior in dike engineering [J]. Optics and Laser Technology. 2018, 103: 346-353.

[11] SU Huaizhi, FU Zhaoqing, GAO Ang, et al. Particle flow code method-based meso-scale identification for seepage failure of soil levee. Transport in Porous Media [J], 2017, 119(2): 311-336.

[12] 王仁钟,李雷,彭雪辉. 透水覆盖层上土石坝渗流量观测设计方案综合评价方法[J]. 水利水电科技进展, 2011,31(1):42-44. (WANG Renzhong, LI Lei, PENG Xuehui. Comprehensive evaluation method for seepage observation design plans of earth and rock-fill dams on permeable overburden layer[J]. Advances in Science and Technology of Water Resources, 2011,31(1):42-44. (in Chinese))

[13] 周发超,王志坚,叶枫,等. 关联规则挖掘算法 Apriori 的研究改进[J]. 计算机科学与探索,2015,9(9):1075-1083(ZHOU Fachao, WANG Zhijian, YE Feng, et al. Research and improvement of Apriori algorithm for mining association rules[J]. Computer Science and Exploration, 2015, 9(9): 1075-1083. (in Chinese))

[14] 林郎碟,王灿辉. Apriori 算法在图书推荐服务中的应用与研究[J]. 计算机技术与发展,2011,21(5):22-24. (LIN Langdie, WANG Canhui. Application and research of Apriori algorithm in library's book recommendation service [J]. Computer Technology and Development, 2011, 21(5):22-24. (in Chinese))

[15] GUO Zhenhai, CHI Dezhong, WU Jie, et al. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriorialgorithm[J]. Energy Conversion and Management,2014,84:140-151.

[16] 胡江,郑鹏翔. 基于滞后效应和防洪调度的大坝渗流预测模型研究[J]. 三峡大学学报(自然科学版),2008, 30 (6): 16-19. (HU Jiang, ZHENG Pengxiang. Forecasting model of dam seepage flow based on lagging effect and flood control operation [J]. Journal of China Three Gorges University (Natural Sciences), 2008, 30 (6):16-19. (in Chinese))

[17] 王国胤,刘群,于洪,曾宪华. 大数据挖掘及应用[M]. 北京:清华大学出版社,2017.

[18] 郭涛,张代远. 基于关联规则数据挖掘 Apriori 算法的研究与应用[J]. 计算机技术与发展,2011,21(6):101-103. (GUO Tao, ZHANG Daiyuan. Research and application on association rules based on Apriori algorithm [J]. Computer Technology and Development, 2011, 21 (6): 101-103. (in Chinese))

[19] RAJENDRA KR, SARANSH V, ASHU K, et al. A novel modified Apriori approach for web document clustering [M]. New Delhi:Springer India,2015.

[20] ZHANG Wenjing, MA Donglai, YAO Wei. Medical diagnosis data mining based on improved Apriori algorithm [J]. Journal of Networks,2014,9(5):1339-1345.

[21] 张豪,许四法. 基于经验模态分解和遗传支持向量机的多尺度大坝变形预测[J]. 岩石力学与工程学报,2011, 30(增刊2):3681-3688. (ZHANG Hao, XU Sifa. Multi-scale dam deformation prediction based on empirical mode decomposition and genetic algorithm for support vector machine (GA-SVM) [J]. Chinese Journal of Rock Mechanics and Engineering,2011,30(Sup2):3681-3688. (in Chinese))

(收稿日期:2018-08-01　编辑:郑孝宇)