

基于数据仓库技术的大坝资料分析与安全决策系统研究

宋小刚¹, 李德仁^{1,2}, 华锡生³, 黄红女³

(1. 武汉大学遥感信息工程学院, 湖北 武汉 430072; 2. 测绘遥感信息工程国家重点实验室, 湖北 武汉 430072;
3. 河海大学土木工程学院, 江苏 南京 210098)

摘要: 在现有基于数据库技术的大坝安全监测系统研究的基础上, 结合工程项目的具体情况以及大坝监测数据分析的特点, 设计并构建了大坝安全监测数据仓库系统, 为监测资料的分析查询、信息挖掘和综合评判提供了良好的数据环境. 利用联机分析处理(OLAP)与数据挖掘(DM)技术对监测数据仓库中的数据进行了分析和挖掘, 使安全监控资料的作用得到了进一步的发挥, 为安全监控评判和决策的深化提供新的途径.

关键词: 大坝安全监控; 数据仓库; 联机分析处理; 数据挖掘

中图分类号: TV698.1; TP274 文献标识码: A 文章编号: 1000-1980(2006)03-0280-05

目前, 对大坝的安全监控主要是通过建立各种监控系统实现的, 这种监控系统在大坝日常运行的安全评估中体现了重要作用. 但是, 这种基于数据库技术的监控系统在开发和运行的过程中存在的问题和局限性慢慢地显露出来: (a) 现有系统的数据组织方式限制了一些分析手段和分析方法的应用, 不利于数据的再加工处理; 同时存在数据一致性问题; (b) 分析手段不够灵活, 分析效率不是很高, 而且对于历史数据的利用不够充分, 缺乏数据的二次分析能力; (c) 评判准则和专家知识有限, 而且大坝工程有其自身的特点, 来自于其他坝体案例的专家知识不是很适合每个大坝, 再加上知识运用过程中一些条件的限制, 使得知识的利用率很低.

局限性的存在限制了这些系统工程效益的进一步发挥, 也限制了智能水平及其先进性的进一步提高. 需要以新的技术和方法来拓展原系统, 延长其生命周期. 大坝安全监测数据仓库的建立以及相应技术、方法、工具的应用能较理想地解决目前存在的问题, 为大坝提供一个良好的数据体系化环境. 用传统的关系数据库系统去管理操作型数据, 满足日常的事务和应用处理; 用数据仓库去集成和组织分析型数据, 支持分析处理和信息挖掘, 产生的结果重新服务于操作型数据, 或直接服务于大坝的安全评价和决策. 二者取长补短, 无疑是较理想的结合^[1].

1 大坝安全监测数据仓库的框架设计

大坝安全监测数据仓库是一种体系结构, 它不仅仅用来存储监测数据, 还要提供数据进级(从实时数据到导出数据)、分析服务、数据挖掘等各种工具^[2]. 根据数据仓库体系结构思想, 结合工程项目的具体特点以及大坝监测数据分析的特点, 本文设计了大坝安全监测数据仓库系统的框架结构, 如图1所示.

图1中, 原安全监控系统中的内容作为数据源, 用来建立大坝安全监测数据仓库, 包括各种关系数据库, 比如生成库、知识库、现实库、模型库、环境量库等. 大坝安全监测数据仓库系统主要包括3个部分: 数据集成与管理、数据仓库存储和分析工具. 数据集成与管理主要是自动将数据从数据源中提取出来, 转移到监测数据仓库中, 并给予维护. 分析工具包括数据发掘工具、数据分析模型、可视化工具等, 其中数据发掘工具相对于原监控系统来说是一种新的工具, 是用户根据监测数据的特点采用相应的数据挖掘算法编制的一种应用程序, 用于发现知识、发现规则、发现新的监测数据关系.

2 大坝安全监测数据仓库的构建

大坝安全监测数据仓库的构建是基于现有监测数据库基础上的一个不断循环、反馈而使系统不断增长与完善的过程, 大体上有以下几个步骤:

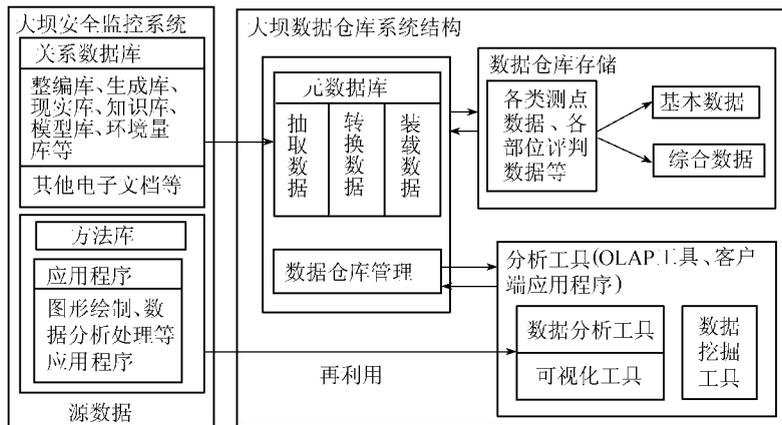


图1 大坝安全监测数据仓库系统体系结构

Fig.1 Frame of data warehouse system for dam safety monitoring

a. 确定主要的主题域及其内容. 以小浪底大坝安全监控系统为例, 该运行系统中, 数据都是围绕着测点、观测项目、工程部位或整个大坝来组合. 因此, 测点和观测项目为该系统的两个主题. 测点是大坝日常运行的基础, 大坝或某一工程部位运行的正常与否都是通过测点的监测数据来反映的. 观测项目是测点和大坝或工程部位之间的桥梁, 也就是说测点通过监测项目监测来反映大坝的运行情况.

b. 逻辑模型设计. 在确定主题之后, 分析主题域, 着重于某个主题, 设计多维数据结构的事实表和维表, 并采用星型模式来表达事实表和维表之间的逻辑关系^[3]. 在本监测数据仓库中, 有3个维表(时间维、位置维、环境量维)和一个事实表, 这些表之间依靠主题的公共码键联系在一起, 形成一个完整的主题. 另外, 每个维都要根据大坝安全评判的实时性和评判、报警的级别性进行粒度层次划分.

c. 数据集成. 通过利用 DTS(Data Transformation Services)把原系统数据库中的监测数据和一些文本形式的结果文件转换和合并到监测数据仓库中. 在转换和合并过程中, 对于简单数据转换或数据移动, 可以使用 DTS 导入\导出向导, 而对于复杂数据析取过程, 可以使用 DTS 设计器创建一个包含复杂工作流的 DTS 包或编写 DTS 包的应用程序^[4].

经过以上步骤, 建立了一个比较完善的大坝安全监测数据仓库系统, 为大坝检测部门提供了一个灵活、有效、实时、智能、相对准确的大坝资料分析和信息挖掘平台.

3 基于大坝安全监测数据仓库的资料分析和信息挖掘

大坝安全监测数据仓库是资料分析和信息挖掘的基础, 它的建立为用户提供了一个统一的、一致的分析环境. 它能够提供对数据的简单统计、查询, 但不能从多角度交互地、快速地观察监测数据, 不能从现有的数据中挖掘潜在的模式, 这就产生了对 OLAP 和数据挖掘的需求.

3.1 利用 OLAP 技术对大坝资料进行多维分析

基于 MS Analysis Services 系统, 利用系统自带的 OLAP 工具软件和自行开发的客户端应用程序, 可以从不同的应用角度对监测数据做丰富、动态的多维查询和分析^[5]. 在对大坝监测数据仓库中的数据进行多维查询和分析时, 首先要利用分析管理器(Analysis Manager)建立 OLAP 分析服务器, 然后根据不同的应用分析要求建立不同的多维数据集. 有关大坝的几个多维数据集建好后, 就可以利用 OLAP 工具软件和自行开发的应用程序对建好的各个多维数据集进行各种操作. 主要实现的功能有: 数据浏览和查询、数据报表生成和数据分析.

3.1.1 浏览和查询

OLAP 提供了建立在多维数据模型基础上的高水平查询, 可以从不同的角度(维)通过不同的度量值来观察、分析所关心的事实数据, 使监测部门的工作人员逐步摆脱了对固定报表的依赖, 通过丰富、动态的多维查询, 主要是对维的拖动和粒度的选择, 实现从不同角度、不同层次对监测数据的快速的查看, 了解测点、部位及整个大坝的运行情况. OLAP 也可以通过来回的动态比较, 验证监测部门工作人员根据对大坝的实际了解所作出的各种假设. 分析人员还可以浅化以看到较为概括的数据, 也可以深化以看到数据的细节. 例如, 在

查看历史异常数据时,对于异常次数多的测点或者是最近异常的测点,可以通过钻取操作查看该测点历史异常数据的详细信息.

3.1.2 数据报表

数据报表的生成不依赖于数据分析、数据查询等功能,用户可以自由选择报表的行、列及数据区的内容,包括综合数据和细节数据,并能临时生成计算数据.该报表类似于 Excel 的风格,并能保存、打印、生成相应图表等.

3.1.3 数据分析

OLAP 工具(包括自行开发软件)根据用户提出的分析任务在特定的数据立方体上进行动态多维分析,其中包括跨维、在不同层次之间跨成员地计算和建模^[6],在时间序列上的对比分析、描述性分析、趋势和预测分析,下钻、上卷、切片和切块.可以从数据立方体的不同数据透视图中,在每一个粒度级和所有维度上实现从宏观到微观的交互查询和数据分析,并将分析结果通过多种形式,如直方图、饼图、分布图等返回给用户.

a. 对比分析.用 OLAP 对监测数据进行对比分析,就是在维和层次上分析数据,时间维、位置维、组合维都是 OLAP 分析的条件,这些条件根据决策的分析要求,通过 OLAP 切片、切块功能实现维和层次的取舍,通过 OLAP 的钻取功能实现从维的一层数据到另一层数据的转换.

b. 描述式数据分析.描述式数据分析是以简洁概要的方式描述监测信息,并提供数据的一般性质.根据建立不同应用的多维数据集,采用 OLAP 方式对监测数据在不同维度中进行下钻、上卷、旋转、切块和切片等分析操作,确定各指标在不同维度和粒度中的总和、最大值、最小值、均值、方差及其分布等反映数据一般性质的信息,以图形的方式直观地显示描述式数据分析的结果.如:对于历史正常数据,在时间维和位置维之上进行分析,可以确定各测点在不同时间段内测值的最大值、最小值、相同环境量测值、均值、方差等,分析者可根据分析结果清楚地掌握各种指标信息和特征数据.

c. 趋势和预测分析.趋势和预测分析是根据历史数据和模型来预测新数据集的行为.可以依赖趋势和预测数据分析确定下一个测值的范围以及判断测值的正常与否,确定各因素对测点测值的影响及程度等,让决策者根据结果提前提出相应的对策.

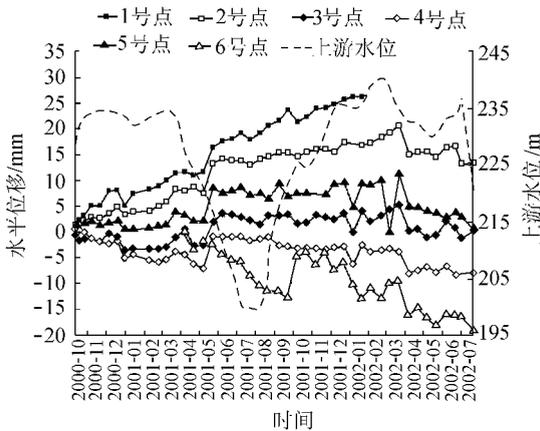


图2 多测点水平位移测值过程线

Fig.2 Curve for displacement at monitoring points

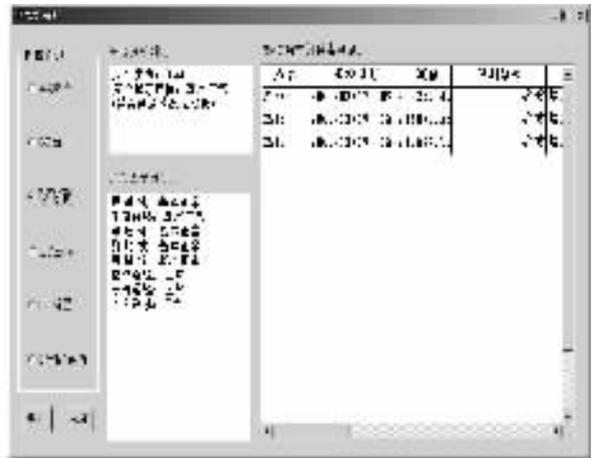


图3 大坝关键部位综合评判界面图

Fig.3 Interface for dam safety evaluation

d. 图形分析.图形是一种最直观明了的信息表示方式,各种分析结果或对象之间的隐含关系经过图形可视化,便会让人一目了然.本文对小浪底大坝监测数据进行了各种图形分析,包括坝体水平、垂直位移分布图、过程线、坝体浸润线、测值环境量关系曲线、包络线、等值线等,图2为相互关联的多个测点的测值过程线.

e. 综合评判.除了以上4种主要的分析外,用户还可以根据需求编制其他算法程序对监测数据从不同侧面进行分析和评价(包括使用原有系统编制好的算法程序).例如,通过神经网络算法对各类测点的当前状态进行分析,从而对各工程部位运行状态作出评价.图3即为小浪底大坝几个关键断面的综合评判界面图,图中所示为大坝 C-C 断面的评判结果,左边为断面各个评判因子和综合评判结果,右边为断面当前状态

下的异常测点,可以看出,虽然有部分异常测点,但总的评判结果基本正常。

3.2 大坝安全监控的数据挖掘

小浪底大坝运行所积累下来的监测数据和评判资料数量非常巨大,利用数据挖掘技术的优势,从巨大的监测资料中发现隐含在数据背后潜在的有用信息,然后用来服务于大坝的评判,这将能更好、更准确地预测各类测点以及各个工程部位所出现的异常变化,并及时采取措施,确保大坝的安全与正常运行。数据挖掘技术所用到的算法有多种,本文仅采用了关联规则和决策树算法对小浪底大坝的监测数据进行了挖掘。

3.2.1 关联规则挖掘

通过关联规则的挖掘主要想发现测点与环境量、测点与测点、观测项目与测点以及观测项目之间的关系。本文通过关联规则算法对大坝监测资料进行信息挖掘时,所涉及的几张数据表(相应于不同的挖掘目标由数据仓库系统生成)为(a)测点与环境量关系表(点名,时间,测值,异常类型,上游水位,工程部位,下游水位,气温等)(b)测点间关系表(时间,异常测点 1,异常测点 2,……异常测点 n)(c)观测项目与测点关系表(部位名,时间,异常观测项目,异常测点 1,异常测点 2,……异常测点 n)(d)观测项目间关系表(部位名,评判时间,异常观测项目 1,异常观测项目 2,……异常观测项目 n)。

通过初步的挖掘发现,小浪底大坝安全监控中,环境量对测点测值影响的顺序为:上游水位、位置、时间、下游水位、气温等。异常测值中异常类型为 3(监控指标或模型预报值检验异常)的测值有 81% 发生在上游水位超过 220m,有 43% 的测值是属于 F1 断层部位,有 30% 的测值发生在 6、7、8 月份。测点之间的关系:渗流测点 P33 与 P34 及 P35, P42 与 P43 及 P44, P81 与 P82 等,它们之间的相关性很大,在前者出现异常的情况下,后者有 70% 以上的可能出现异常。观测项目与测点的关系:F1 断层的异常往往是由于渗流测点 P36, P37, P45, P46, P47, P48 出现了异常。观测项目之间的关系为:渗流异常时位移也往往异常。经过验证和人工巡视检查发现,这些挖掘出来的规则、知识是合理的,比如 F1 断层的异常测值较多是由于 F1 断层出现了渗流异常,有 30% 的异常测值是发生在 6、7、8 月份,这是由于 6、7、8 月份雨量比较充足。挖掘出来的这些规则可以用来辅助测点和部位的评判,也可以帮助分析人员进行各种分析。

3.2.2 决策树挖掘

本文利用 Analysis Manager 的挖掘模型向导,创建一个数据挖掘模型以训练各测点历史正常数据,并使用“Microsoft 决策树”算法在历史数据中找出各个测点运行的内在模式。通过挖掘,对历史数据进行了分类,从小浪底监测数据仓库的挖掘中,发现各测点在不同上游水位范围内相应的正常测值范围。部分测点挖掘结果如表 1 所示。

表 1 决策树挖掘结果

Table 1 Results of data mining by decision-making tree

测点	正常测值范围				
	130 ~ 150	190 ~ 210	210 ~ 230	230 ~ 250	250 ~ 270
P25	137 ~ 144	174 ~ 199	178 ~ 217	199 ~ 223	200 ~ 225
P26	134 ~ 144	173 ~ 199	177 ~ 216	200 ~ 222	201 ~ 223
P28	141 ~ 150	189 ~ 218	194 ~ 241	206 ~ 245	223 ~ 246
P31	142 ~ 148	190 ~ 211	207 ~ 226	216 ~ 232	233 ~ 242
P32	133 ~ 141	154 ~ 179	166 ~ 189	177 ~ 194	183 ~ 198
P36	130 ~ 138	145 ~ 164	148 ~ 170	153 ~ 171	154 ~ 175
P37	134 ~ 142	153 ~ 177	167 ~ 189	174 ~ 198	187 ~ 199

如果想更准确一些,可以把上游水位区间划分更窄一点,重新进行分类。笔者把挖掘到的这些各测点的内在模式用于在线分析中,作为另外一项评判指标去评判新测数据,或者在离线分析中作为一项辅助评判准则,从而能更准确地发现或分析异常测值。

此外,还可以开发 Analysis Services 的客户端应用程序,利用其他数据挖掘算法例如以粗集方法、神经网络方法、遗传算法、统计分析方法、模糊论^[7]等多种方法去发现大坝安全监测数据仓库或所建立的 OLAP 多维数据集中的有用信息,更好地为大坝评判和资料分析服务。

4 结 束 语

数据仓库技术的引入,弥补了目前基于关系数据库技术的大坝安全监控专家系统的不足。

大坝安全监测数据仓库的建立能满足资料分析、信息挖掘和综合评判、安全监控等需求,为安全监控评判和决策的深化提供了新的途径。一些相应技术、方法、工具的应用将在一定程度上解决原系统存在的问题。原系统与监测数据仓库系统的结合为大坝的安全监控提供了一个良好的数据体系化环境,使安全监控资料的作用得到进一步发挥。

由大坝安全监控与数据仓库两者的特点可知,在现有大坝安全监控系统中引入数据仓库技术有以下优

点 (a)提高系统分析与查询的效率和灵活性 (b)提升对历史数据二次分析的能力 ,发现大坝自身的运行规律 ,服务于大坝安全评判 ,既做到“取之于坝 ,用之于坝” ,又解决了原系统“知识贫乏”的困境 (c)可以解决大坝安全监控部门多源数据的集成和数据的不一致性问题。

参考文献 :

- [1] 宋小刚 . 数据仓库技术在大坝资料分析与安全决策中的应用研究 [D] . 南京 : 河海大学 , 2004 .
- [2] 舒云星 , 李晶 , 贺贵明 . 三峡工程安全监测信息系统决策支持数据仓库技术 [J] . 计算机工程 , 2002 , 28(11) : 109-112 .
- [3] LOU A . 数据仓库技术指南 [M] . 北京 : 人民邮电出版社 , 2000 .
- [4] 陈晓云 , 郭朝珍 . 数据仓库数据析取工具的设计与实现 [J] . 福州大学学报 : 自然科学版 , 2002 , 30(3) : 35-39 .
- [5] Microsoft 公司 . Microsoft SQL Server 2000 资源大全 [M] . 北京 : 机械工业出版社 , 2002 .
- [6] 张升文 , 张少仲 . 数据仓库、数据挖掘和 OLAP 分析及其应用实例 [J] . 微机发展 , 2002(1) : 10-15 .
- [7] 邵峰晶 , 于忠清 . 数据挖掘原理与算法 [M] . 北京 : 中国水利水电出版社 , 2003 .

Data warehouse-based system for data analysis and safety decision-making of dam projects

SONG Xiao-gang¹ , LI De-ren^{1, 2} , HUA Xi-sheng³ , HUANG Hong-ni³

(1. School of Remote Sensing Information Engineering , Wuhan University , Wuhan 430072 , China ;

2. State Key Laboratory of Information Engineering in Surveying , Mapping and Remote Sensing , Wuhan 430072 , China ;

3. College of Civil Engineering , Hohai University , Nanjing 210098 , China)

Abstract : In combination with the reality and monitoring data of dam projects , a data warehouse (DW) for dam safety monitoring was designed and constructed , providing a good environment for monitoring data analysis and query , data mining (DM) and comprehensive evaluation . By use of the on-line analytical processing (OLAP) tool and DM technology , dam monitoring data in data warehouse were analyzed and mined , and the function of the monitoring data was further utilized . The study provides a new approach for evaluation of result of safety monitoring and decision making for dam projects .

Key words : dam safety monitoring ; data warehouse ; on-line analytical processing ; data mining