

数据挖掘技术在水利工程管理中的应用研究

[摘要] 因能从海量数据中发现潜在的、有价值的知识,数据挖掘成为人们非常感兴趣的热点技术。水利工程管理面临的是海量的非空间数据和空间数据,对这些数据的挖掘,能够提炼出有价值的知识,从而提高水利工程管理的科学化水平和决策水平。在水利工程管理中利用空间数据挖掘技术,需要解决数据仓库建设、数据挖掘与GIS集成和数据挖掘系统模型等3个方面的关键问题。数据仓库是建设水利工程数据挖掘系统的基础。数据挖掘与GIS集成的方式有嵌入式、松散耦合式、紧密耦合式3种,紧密耦合式是两者集成的最好方式。水利工程系统的数据挖掘适合采用基于OLAP和OLAM的探查性数据挖掘模型。

[关键词] 空间数据挖掘; 水利工程; 数据挖掘模型; 地理信息系统; 数据仓库

doi: 10. 3969/j. issn. 1673 - 0194. 2010. 04. 027

[中图分类号] F270.7 [文献标识码] A [文章编号] 1673-0194(2010)04-0076-04

1 引言

我国长期以来兴建了一大批水利工程,初步形成了具有防洪、排涝、灌溉、供水、发电、养殖、种植、旅游等功能要素的水利工程体系,为国民经济的高速发展发挥了巨大的基础作用和支撑作用[1]。在水利工程建设取得辉煌成就的同时,人们逐渐意识到我们在水利工程的管理上还存在着手段比较落后,重建轻管、水利资源利用率低等突出问题,致使一大批水利工程不能发挥其价值,或者工程寿命大大缩短[1,2]。穆范楠等分别从制度管理、机制管理、人力资源管理等方面对水利工程管理中存在的问题进行了论述,并提出了不少可行性的解决措施[3]。不可否认,水利工程管理中出现的问题,不少是制度上的问题,但水利工程管理有其特殊性、复杂性,需要广博的知识和高超的技术,单纯靠“软管理”是



不能从根本上解决问题的，必须借助一些现代化的信息手段来辅助进行决策和管理，才能够更好、更科学地解决问题。

近年来，在水利工程信息化的过程中，我国建设了一大批水利工程管理信息系统，对于水利工程的建设和运行管理起到了很好的帮助作用。但是，这些系统所提供的功能大多是业务型的，很少面向管理决策。随着水利工程管理向现代化纵深发展，这些系统远远满足不了人们的需要。另一方面，水利工程管理信息系统在发展过程中积累了海量的数据，不少是空间类型的数据，而且这些数据还在不断地增长，而相比于数据的生产、运输和累积能力，人类对空间数据的分析能力还很落后 [4]。人们虽然深知这些海量数据中蕴含了很多有价值的知识，但是不知道如何利用它们，而依靠传统的信息系统是解决不了这些问题的。数据挖掘技术的出现为这些问题的解决带来了可能。所谓数据挖掘，就是从海量数据中发现潜在的、有价值的知识的过程。传统的数据挖掘技术和方法一般作用于非空间数据，而水利工程管理方面的数据不但有非空间数据，还有大量的空间数据。和非空间数据相比，空间数据除了具备非空间数据的特征外，还有拓扑、方位和距离等非空间特征，因此其挖掘技术的实现有其特殊性。在武汉大学李德仁院士首次提出空间数据挖掘这一概念后，国内外不少学者为此开展了广泛的研究 [5-7]。

2 空间数据挖掘在水利工程管理中应用需要解决的主要问题

水利工程管理信息系统中存在着大量的空间数据，因此需要采用空间数据挖掘技术。和一般的空间数据挖掘系统相比，对水利工程数据的挖掘需要考虑其历史发展因素和特殊性。首先，水利工程是一个系统工程，其有效管理往往需要多领域、多部门的专家相互协作，一项重要决策的做出往往需要对历史数据从各种维度进行分析，反复考虑各种因素，综合各个专家的意见才能形成，而不同的专家和决策者会从不同的角度来分析数据，因此对水利工程数据的挖掘需要交互探查或查询驱动的方法，在技术实现上需要采用数据仓库和数据立方体支持这种探查式的、快速的联机查询和分析。其次，在用的水利工程信息系统的主体是 GIS



(Geographical Information System, 地理信息系统), 大部分的空间数据是由 GIS 系统生成的, 空间数据的查询、计算、分析和可视化显示是一种复杂的技术, 因此如何利用原有的 GIS 系统中的数据, 数据挖掘如何和 GIS 集成以进行复杂的空间数据处理成为一个需要解决的重要问题 [8, 9]。最后, 要实现水利工程的数据挖掘, 需要建立一个数据挖掘系统模型, 模型在系统工程的研究、设计和实现中是一个非常重要的问题, 一个好的模型对了解系统本质特征、揭示系统的规律起到非常重要的作用, 建模也是实现一个工程系统的重要一步。因此, 要想实现空间数据挖掘技术在水利工程管理中的应用, 这 3 个问题是不可回避的、必须研究的核心问题。

3 空间数据仓库

水利工程信息化的过程中产生了海量的数据, 而数据仓库是处理海量数据的关键技术, 它可以将不同来源的数据统一到语义上一致的环境下。在水利工程信息系统中除了有丰富的非空间数据外, 还有大量的空间数据, 如地图、预处理过的遥感图像、视频等。空间数据与非空间数据相比, 除了具备传统数据库数据的特征外, 还携带了空间特征, 如拓扑、方位、距离等。“空间数据仓库是面向主题的、集成的、时变的和非易失性的非空间数据和空间数据的集合”, 用于支持空间数据挖掘和与空间数据相关的决策过程。建立空间数据仓库是一个具有挑战性的工作, 需要解决两个方面的问题: 集成来自异构数据源和系统的空间数据; 如何在空间数据仓库中实现快速而灵活的联机分析处理 [10]。

影响水利工程建设和管理决策的数据来源是丰富多样的, 如气象数据库、蓄滞洪区空间分布式社会经济数据库、雨情和水情数据库、水旱灾情数据库等, 它们往往存在于异构的环境中, 可能来自于不同的系统, 数据格式多种多样。数据格式不仅与特定的结构有关, 如光栅格式和矢量格式, 而且与特定的厂家有关。为了能够进行空间数据的分析和处理, 需要首先对这些异构的数据进行清洗、变换和集成, 以清晰一致的格式存放在数据仓库中, 然后可以调用相应的数据挖掘算法获取有用的知识。空间数据仓库已成为联机数据分析处理和数据



挖掘必不可缺的平台。利用空间数据仓库技术,可以对异构的各类信息进行过滤、集中和综合,完成水情信息采集、工情信息采集、防汛抗旱信息等水利工程信息的自动接收、处理等功能,在此基础上可以进行汛情分析、暴雨洪水预报、调度、灾情评估以及旱情预测等知识发现功能 [11]。

空间数据仓库、OLAP (On-Line Analytic Process, 联机分析处理) 和 OLAM (On-Line Analytic Mining, 联机分析挖掘) 的实现基于多维数据模型,这种模型围绕中心主题组织数据,将数据看作数据立方体的形式。数据立方体允许从多维对数据建模和观察,它由维和事实来定义。数据仓库有星型模式、雪花型模式或事实星座型模式。在这 3 种结构中,星型模式提供了简洁而有组织的仓库结构,便于进行 OLAP 和 OLAM 操作,所以是空间数据仓库建模的好选择。相比于传统的数据立方体,空间数据立方体中存在 3 种类型的维:非空间维、空间到非空间维和空间到空间维;有两种不同的度量:数值度量和空间度量 [10]。

4 水利工程 GIS 系统与数据挖掘系统结合的方式

水利工程的建设和管理与其所在地的地形、地质、社会、经济以及河流的水文等空间要素有关,而 GIS 善于处理和分析空间信息,因此大多水利工程在信息系统中采用了 GIS 技术 [12]。GIS 是空间数据库发展的主体 [13]。GIS 中含有大量的空间和属性数据,有着比一般关系数据库和事务数据库更加丰富和复杂的语义信息,隐藏着丰富的知识。空间数据挖掘和知识发现技术,一方面可使 GIS 查询和分析技术提高到发现知识的新阶段,另一方面从中发现的知识可构成知识库用于建立智能化的 GIS 系统,同时也将促进 3S

(GIS/RS/GPS) 的智能化集成 [13],因此很有必要探讨 GIS 系统与数据挖掘系统的结合方式。当数据挖掘系统工作在一个需要与其他信息系统成分通信的环境下,可以采用不耦合、松散耦合、半紧密耦合和紧密耦合 4 种方案。不耦合方案虽然简单,但缺点不少,是一种非常糟糕的设计 [10]。雷宝龙和李春梅提出了 GIS 与空间数据挖掘集成的 3 种模式:松



散耦合式、嵌入式和混合型空间模型法 [8, 14]。在此基础上对上述 3 种模式进行了改进，以适合于水利工程 GIS 系统和空间数据挖掘系统的集成。

4.1 嵌入式

嵌入式是将数据挖掘系统融入到 GIS 中，也就是说系统既是一个 GIS 系统，又是一个数据挖掘系统。嵌入式的优点是可以充分利用 GIS 系统所提供的空间数据处理和分析功能来开发数据挖掘系统，减少了开发的工作量，降低了开发的难度；其缺点是数据挖掘功能被限制在特定的 GIS 系统中，难以移植到其他的 GIS 系统上，而且这种方式会因为考虑到一种用户的需求，而限制另一部分用户的需求，从而使系统功能的开发受到限制。图 1 是嵌入式模型的示意图。

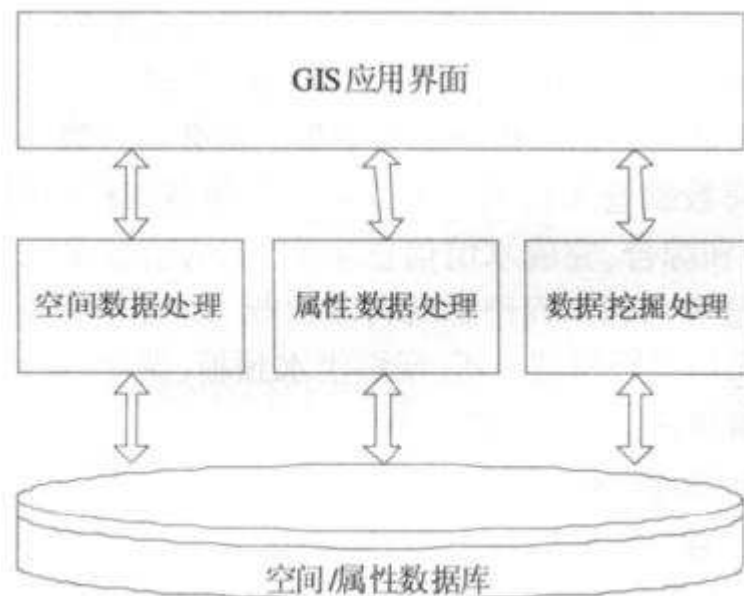


图 1 嵌入式模型

4.2 松散耦合式

在松散耦合式下，数据挖掘系统和 GIS 系统实际上是两个独立的系统，数据挖掘系统从 GIS 中获取空间数据和属性数据，经过清洗、过滤和变换后存入自身的数据库或数据仓



库中，数据挖掘所进行的其他工作与 GIS 系统没有任何联系。这种模式的优点是数据挖掘系统不依赖于特殊的 GIS 系统，可以开发出独立的、相对通用的空间数据挖掘系统；缺点是在数据挖掘系统中要融入复杂的空间数据的处理，系统开发的难度很高。图 2 是松散耦合式模型的示意图。

4.3 紧密耦合式

紧密耦合式克服了嵌入式和松散耦合式的缺点，既充分利用了原有 GIS 的处理空间数据的强大功能，降低了开发的难度，又不受制于原有 GIS 系统的用户需求的制约，具有较大的灵活性，提供了相对独立的数据挖掘功能。其缺点是和原来系统联系密切，开发的数据挖掘系统往往依赖于 GIS 系统。图 3 是紧密耦合式模型的示意图。

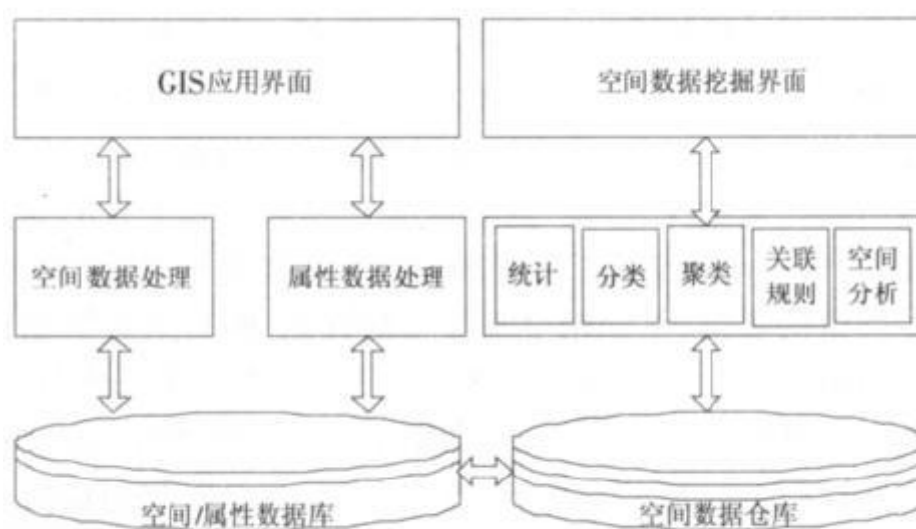


图 2 松散耦合式模型

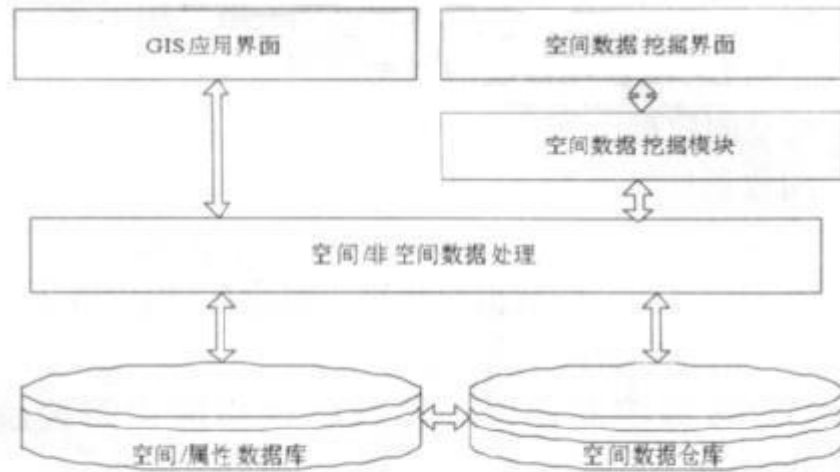


图3 紧密耦合式模型

在这3种结合方式中，紧密耦合式有着明显的优点，是建立水利工程数据挖掘系统优先考虑的方式。

5 水利工程数据挖掘系统模型

文献[4]介绍了国外几个相对比较成熟的空间数据挖掘系统：GeoMiner、MultiMediaMiner、SKICAT等，然后提出了作者领导的空间数据挖掘团队研究和开发的两种空间数据挖掘原形系统GISDBMiner和RSImageMiner，并提出了GIS空间数据挖掘系统的体系结构。文献[8]介绍了现有的数据挖掘模型：OLAM模型和影响域模型，以及GeoMiner原形系统的体系结构，最后提出了一个基于空间立方体的数据挖掘模型。文献[6]提到了Han提出的通用数据采掘原型DBLEARN/DBMINER、Holzheimer等人提出的并行体系结构，以及Matheus等人提出的多组件体系结构，并重点介绍了Matheus等人的多组件体系结构。

水利工程管理决策大多是复杂的非结构化决策，需要进行探查性或查询驱动型的数据挖掘，以方便不同的决策者和专家从不同的领域或角度进行数据探查和分析。一般情况下，在挖掘过程中需要进行人机的多次对话，然后结合人类专家的隐性知识，才能够发现有价值的知识。因此自动化的挖掘方法不适合于水利工程数据挖掘。针对水利工程数据挖掘系统的特

点，本文在对上述各类体系结构研究的基础上，提出如图 4 所示的水利工程数据挖掘系统模型。

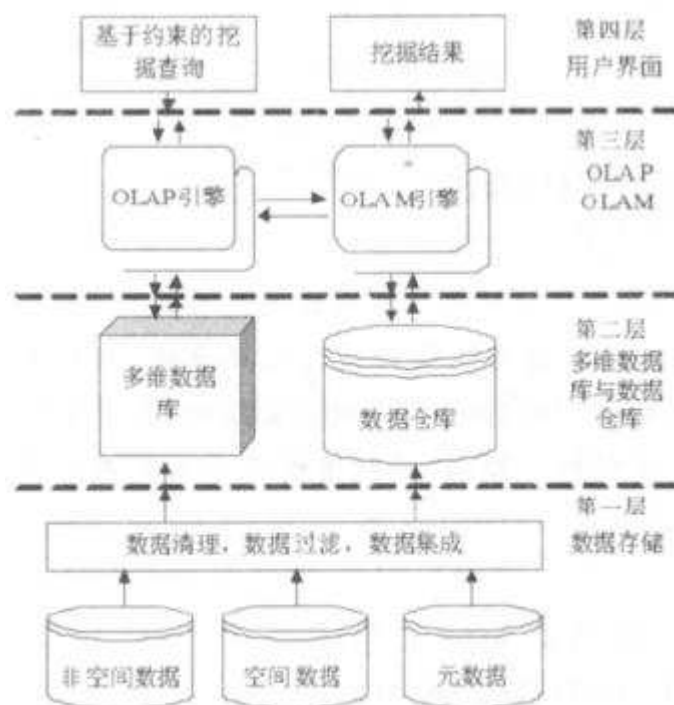


图 4 水利工程数据挖掘系统模型

模型分为 4 层，分别为数据存储层、多维数据库与数据仓库层、OLAP/OLAM 层、用户界面层。第一层数据存储层的数据主要来源于水利工程数据库和相关的异构数据库，元数据用于指导数据的清理、过滤和集成，是构建水利工程数据仓库重要的技术手段。第一层的数据经过变换和集成后，存储到数据仓库和多维数据库中，它们是实现第三层 OLAP/OLAM 分析所需要的重要的数据源。该模型的核心是 OLAP/OLAM，它们是支持探查性知识发现的核心技术。第四层是用户界面层，用来帮助用户实现基于约束的挖掘查询，并将挖掘结果显示给用户。

6 空间挖掘可以采用的方法与发现的知识类型



数据挖掘在水利工程管理上的应用,不仅可以建设智能型的 GIS 系统,促进遥感技术和 GIS 技术的深入应用,还可以从数据中发现潜在的、有价值的知识或规则,用于指导水利工程的建设和管理。一般来说,传统的数据挖掘方法如统计、分类、聚类等都可用于空间数据挖掘,但我们不能简单地把这些方法直接应用在空间数据的挖掘上 [7, 15]。一方面,因为空间数据除了具备一般非空间数据的特征外,还具备拓扑、方位、距离等空间特征;另一方面,传统的数据挖掘算法一般假定数据对象统计不相关、相邻的数据对象是独立产生的,而空间数据的相邻对象间存在着关联和相互影响,因此需要对原有的方法进行改进,使得数据挖掘方法适合于地理空间数据的挖掘。在空间数据挖掘与知识发现中可采用的方法主要有:统计方法、归纳方法、聚类方法、空间分析方法、探测性的数据分析、Rough 集方法、云理论、图像分析和模式识别等 [5, 13]。能发现的知识类型有:(1) 普遍的几何知识,如计算和统计出空间目标几何特征量的最小值、最大值、均值、方差、众数等;(2) 空间分布规律,如机井、水库的分布规律。能发现的规则有:(1) 空间关联规则,如地下水与降雨量的关系,河水质量与污染企业分布的关系;(2) 空间的聚类规则;(3) 空间演变规则,如水库泥沙淤积的演变规律,河道周围生态的演变规律 [13, 14, 16]。需要注意的是,为了便于理解空间数据、发现空间联系、发现空间数据与非空间数据之间的关系,应重视可视化的方法在水利工程数据挖掘过程和挖掘结果的使用 [10, 17]。

7 结语

利用空间数据挖掘技术,对具有空间特征的水利工程数据进行分析,能够发现潜在有价值的知识,利用这些知识,能够降低工程管理的成本,有效利用建设和维护资金,更好地发挥水利工程的效益,为水利工程的管理决策提供依据。要实现数据挖掘技术在水利工程中的应用,必须研究和解决数据仓库和数据立方体的应用、数据挖掘与 GIS 集成和水利工程数据挖掘系统模型 3 个核心问题。本文对这 3 个问题进行了探讨,认为数据仓库是水利工程数



据挖掘的基础,宜采用紧密耦合式结构与 GIS 系统进行集成,在挖掘模型上可以采用基于 OLAP 和 OLAM 的 4 层框架。

主要参考文献

- [1] 张敬光. 实施水利工程管理体制改革的思考 [J] 四川水利, 2004 (2) : 38-42.
- [2] 陈彩妮. 水利工程管理中存在的问题及对策 [J] . 中国水运, 2007, 5 (4) : 74-75.
- [3] 穆范楠, 赵玉红. 综合类大型水利工程管理体制研究 [J] . 人民黄河, 2007, 29 (8) : 6-10.
- [4] 李德仁, 王树良, 李德毅. 空间数据挖掘理论与应用 [M] . 北京: 科学出版社, 2006.
- [5] 王海起, 王劲峰. 空间数据挖掘技术研究进展 [J] . 地理与地理信息科学, 2005 (7) : 6-10.
- [6] 石云, 孙玉方, 左春. 空间数据采掘的研究与发展 [J] . 计算机研究与发展, 1999 (11) : 1301-1309.
- [7] 毕硕本, 耿焕同, 闫国年. 国内空间数据挖掘研究进展与技术体系探讨 [J] . 地理信息世界, 2008, 2 (1) : 21-27.
- [8] 雷宝龙, 刘艳, 邹瑜. 基于 GIS 的空间数据挖掘模型研究 [J]. 软件导刊, 2007 (3) : 7-9.
- [9] 毛克彪, 覃志豪, 李昕, 等. 空间数据挖掘与 GIS 集成及应用研究 [J] . 测绘与空间地理信息, 2004 (2) : 14-17.
- [10] [加] Jiawei Han, Micheline Kamber. 数据挖掘: 概念与技术 [M] . 北京: 机械工业出版社, 2007.



- [11]李琼. 数据仓库与数据挖掘技术在水利信息化中的应用[J]. 前沿, 2005(12):59-61.
- [12]刘立昱. 地理信息系统在水利工程中的应用综述[J]. 甘肃水利水电技术, 1999(2):63-65.
- [13]邸凯昌,李德仁,李德毅. 空间数据发掘和知识发现的框架[J]. 武汉测绘科技大学学报, 1997(12):328-332.
- [14]李春梅,范全润. 空间数据挖掘及其在地理信息系统中的应用[J]. 楚雄师范学院学报, 2005(6):6-10.
- [15]毛克彪,田庆久. 空间数据挖掘技术方法及应用[J]. 遥感技术与应用, 2002(8):198-204.
- [16]蓝荣钦,林丽霞,陈良友,等. 空间数据挖掘和知识发现的现状与发展[J]. 地理空间信息, 2004(6):19-21.
- [17]贾泽露,刘耀林,张彤. 可视化交互空间数据挖掘技术的探讨[J]. 测绘科学, 2004(10):34-37.
- [作者简介] 韩红旗(1971-), 河南伊川人, 华北水利水电学院管理与经济学院讲师, 主要研究方向: 数据挖掘, 信息系统。

