

【水利水电工程】

水利工程管理数据挖掘模型研究

韩红旗

(华北水利水电学院 管理与经济学院, 河南 郑州 450011)

摘要:水利工程管理数据挖掘模型构建要考虑到数据仓库建设及与原有GIS集成问题。数据仓库可以将异构的数据源统一到语义一致的模式下,是建设数据挖掘系统的基础。数据挖掘系统与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。对水利工程管理数据的挖掘宜采用基于OLAP和OLAM的探查性数据挖掘模型。

关键词:水利工程管理;空间数据挖掘;数据挖掘模型;地理信息系统;数据仓库

中图分类号:TV697 **文献标识码:**A **doi:**10.3969/j.issn.1000-1379.2010.01.035

在水利工程建设和管理中积累的海量数据,大部分是空间数据,其中蕴藏了丰富的知识,对它们的理解有助于提高管理能力和决策水平^[1]。近年来兴起的数据挖掘技术,为处理和分析海量数据提供了强大的支撑。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仅包括非空间数据,而且包括大量的空间数据。空间数据除了具备非空间数据的特征外,还有拓扑、方位和距离等特征,因此其挖掘技术的实现存在特殊性。李德仁院士首次提出空间数据挖掘这一概念后,国内外不少学者对此开展了广泛的研究^[2-4]。

在水利工程管理中运用数据挖掘技术,要考虑如何建立一个适合水利工程管理特点的模式。首先,水利工程是一个系统工程,为实现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采用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多维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,目前水利工程业务管理系统的主体是GIS,大部分空间数据由GIS系统生成,对空间数据的查询、计算、分析和可视化显示是一种复杂的技术,如何将原有的GIS系统中的数据与数据挖掘系统集成以进行复杂的空间数据处理,成为需要解决的重要问题^[5-6]。

1 空间数据仓库和多维数据立方体

水利工程管理所面对的海量数据来自于异构的环境,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。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

据仓库需要解决两个方面的问题:如何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^[1]。

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。为了能够进行空间数据的分析和处理,首先需要对这些异构数据进行清洗、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知识。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、防汛抗旱信息接收、处理等,在此基础上还可以进行汛情分析、暴雨洪水预报、灾情评估以及旱情预测等^[7]。

交互探查或查询驱动的数据挖掘主要基于空间数据仓库、OLAP(On-Line Analytic Process, 联机分析处理)和OLAM(On-Line Analytic Mining, 联机分析挖掘)技术,这些技术基于多维数据模型。在多维数据模型中,往往围绕某些中心主题组织数据,将数据视为立方体形式。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,因此是空间数据仓库建模的较好选择。相比于传统的数据立方体,空间数据立方体中存在非空间维、空间到非空间维和空间到空间维3种类型的维,有数值度量 and 空间度量2种不同的度量^[1]。巨大的计算和存储开销,需要在联机计算代价和空间聚集计算及存储代价之间进行平衡。

收稿日期:2009-01-20

基金项目:华北水利水电学院青年基金资助项目(HSQJ2006026);河南省科技攻关计划项目(0624220107)。

作者简介:韩红旗(1971—),男,河南洛阳人,讲师,硕士,主要从事数据挖掘、信息系统研究工作。

E-mail:bithq@163.com

2 数据挖掘系统与GIS系统集成的方式

水利工程建设和管理与其所在地的地形、地质、社会、经济情况以及河流的水文特征等空间要素有关,而GIS善于处理和分析空间信息,因此水利工程业务信息系统中广泛采用了GIS技术^[8]。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可使GIS查询、分析技术提高到发现知识的新阶段,从中可以发现构建智能化GIS系统的知识,促进“3S”技术的智能化集成^[9]。不管是从保护用户以前的投资考虑,还是从降低数据挖掘系统中空间数据处理的难度考虑,都很有必要探讨GIS系统与数据挖掘系统的集成方式。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时,其与GIS系统的集成可以采用不耦合、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,但缺点很多^[1]。雷宝龙^[5]和李春梅^[10]等提出了GIS与空间数据挖掘集成的3种模式,即松散耦合式、嵌入式和混合型空间模型法,在此基础上提出嵌入式、松散耦合式和紧密耦合式3种集成模式。

2.1 嵌入式

嵌入式是将数据挖掘系统和GIS融为一体,优点是可以充分利用GIS系统的空间数据处理功能来开发数据挖掘系统,减少开发工作量,降低开发难度;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,另外这种方式会因考虑到一种系统的用户需求而限制另一系统的部分用户需求,使系统功能的开发受到限制。嵌入式模型见图1。

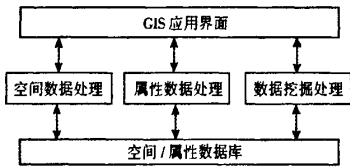


图1 嵌入式模型

2.2 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。其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处理,系统开发的难度很高。松散耦合式的模型见图2。

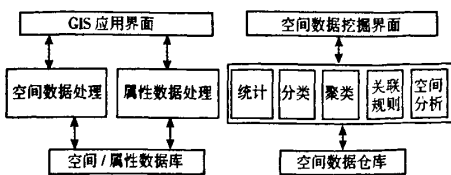


图2 松散耦合式模型

2.3 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系统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。其缺点是开发的数据挖掘系统往往依赖于GIS系统。紧密耦合式的模型见图3。

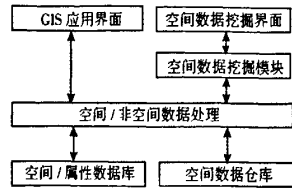


图3 紧密耦合式

紧密耦合式比其他两种集成方式有着明显的优点,是建立水利工程数据挖掘系统应优先考虑的方式。

3 水利工程数据挖掘系统模型

水利工程管理决策大多是复杂的非结构化决策,因此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多次人机对话,然后结合专家的隐性知识,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的数据挖掘。针对水利工程数据挖掘系统的特点,提出图4所示的水利工程数据挖掘系统模型。

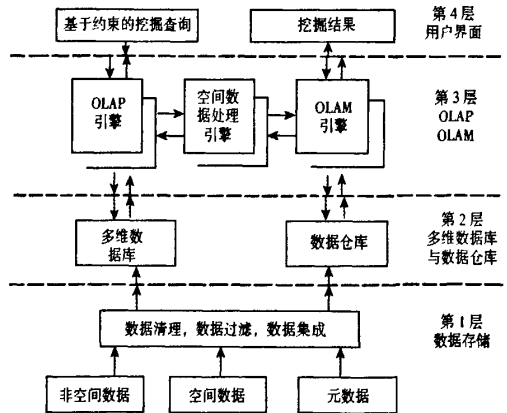


图4 水利工程数据挖掘系统模型

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。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。第1层数据经过清洗、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3层OLAP/OLAM分析的重要数据源。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。为了实现空间数据的查询、聚集、分析、显示等功能,第3层要设置一个空间数据处理引擎,可以利用GIS系统中的空间数据处理功能。第4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。(下转第76页)

2 公式误差分析及应用实例

通过设定立方抛物线形断面不同的特征水深,由式(6)计算特征水深 H_0 ,误差 $\Delta = (H_0' - H_0) / H_0 \times 100\%$,结果见表1。可见,最大误差约为0.60%,精度完全满足工程需要。

已知某立方抛物线形断面渠道 $y = 0.4|x|^{3/2}$,渠道底坡

$i = 0.52 \times 10^{-3}$,粗糙系数 $n = 0.025$,流量 $Q = 20 \text{ m}^3/\text{s}$ 。计算得无量纲参数 $M = 0.1654328$,根据式(6)得 $H_0 = 0.65842 \text{ m}$,则正常水深 $h = 3.339 \text{ m}$ (精确解 $h = 3.341 \text{ m}$),相对误差为 -0.057% 。无量纲参数 $K = 40775$,由式(3)求得临界水深 $h_k = 1.836$ 。可见, $h > h_k$,此时水流状态为缓流。

表1 计算结果精度评价

H_0/m	$\Delta/\%$	H_0/m	$\Delta/\%$	H_0/m	$\Delta/\%$	H_0/m	$\Delta/\%$
0.01	0.603203	0.51	0.223185	1.01	-0.489800	1.51	-0.487840
0.05	-0.367240	0.55	0.151828	1.05	-0.519470	1.55	-0.451630
0.10	0.152426	0.60	0.063082	1.10	-0.549540	1.60	-0.398700
0.15	0.411149	0.65	-0.023430	1.15	-0.571660	1.65	-0.337260
0.20	0.522795	0.70	-0.106300	1.20	-0.585710	1.70	-0.267270
0.25	0.551313	0.75	-0.184450	1.25	-0.591600	1.75	-0.188730
0.30	0.530268	0.80	-0.257090	1.30	-0.589230	1.80	-0.101640
0.35	0.478841	0.85	-0.323560	1.35	-0.578530	1.85	-0.005970
0.40	0.408802	0.90	-0.383390	1.40	-0.559470	1.90	0.098264
0.45	0.327777	0.95	-0.436180	1.45	-0.531980	1.95	0.211082
0.50	0.240919	1.00	-0.481620	1.50	-0.496050	2.00	0.332482

误差分析及实例计算表明,该正常水深计算公式具有适用范围广、形式较简洁、计算精度高的特点。

参考文献:

[1] 文辉,李风玲,黄寿生.圆管明渠均匀流的新近似计算公式[J].人民黄河,2006,28(2):67-68.
 [2] 文辉,李风玲,彭波,等.圆管明渠临界水深的直接近似计算公式[J].人民黄河,2007,29(4):67-68.
 [3] 李风玲,文辉,黄寿生.窄深式U形渠道正常水深的近似计算公式[J].人民黄河,2006,28(12):75-76.

[4] 李风玲,文辉,耿军利,等.宽浅式U形渠道正常水深的近似计算公式[J].人民长江,2007,38(8):170-171.
 [5] 文辉,李风玲.抛物线形断面渠道收缩水深的解析解[J].长江科学院院报,2009,26(9):32-34.
 [6] 文辉,李风玲.立方抛物线断面渠道收缩水深的直接计算方法[J].人民长江,2009,40(13):38-39.
 [7] 魏文礼,杨国丽.立方抛物线形渠道水力最优断面的计算[J].武汉大学学报:工学版,2006,39(3):49-51.
 [8] 吴持恭.水力学[M].北京:高等教育出版社,1982.

【责任编辑 张华岩】

(上接第74页)

4 结 语

空间数据挖掘技术是解决水利工程管理中数据丰富、知识贫乏问题的重要途径。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,能够发现潜在的有价值的知识,利用这些知识,能够降低水利工程管理的成本,为水利工程的管理决策提供支持。空间数据挖掘模型的建立是构建水利工程管理数据挖掘系统的重要一步。根据水利工程管理数据挖掘的特点,构建模型重点要考虑数据仓库和多维数据立方体建设、空间数据挖掘系统与原有GIS集成两个方面的重要问题。数据仓库技术能够将异构的数据源集成到语义一致的环境下,是水利工程数据挖掘的基础。通过3种空间数据挖掘系统与GIS集成方式的对比分析,认为宜采用紧密耦合式结构。水利工程数据挖掘系统宜采用交互探查式或查询驱动的方式发现知识,其模型应采用基于OLAP和OLAM的4层框架。

参考文献:

[1] Jiawei Han, Micheline Kamber. 数据挖掘概念与技术[M].北京:机械工业出

版社,2007.
 [2] 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,2005,21(4):6-10.
 [3] 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,1999,36(11):1301-1309.
 [4] 毕硕本,耿焱同,周国年.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(1):21-27.
 [5] 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.
 [6] 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,2004,27(1):14-17.
 [7] 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.
 [8] 刘立显.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.
 [9] 邱凯昌,李德仁,李德毅.空间数据挖掘和知识发现的框架[J].武汉测绘科技大学学报,1997,22(4):328-332.
 [10] 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,2005,20(3):6-10.

【责任编辑 张华岩 潘乐】