

黔南森林火险因子数据仓库的设计

肖云丹, 纪 平

(中国林业科学研究院 资源信息研究所, 北京 100091)

摘 要:收集近 10 a 来贵州省黔南州的气象及火灾数据,建立贵州省黔南火险因子数据仓库。设计和建立黔南州火灾影响因子多维数据模式,确定多维数据的组织方式,进行数据仓库多维分析。对数据仓库内各气象因子与火灾进行关联分析,找出气候对黔南森林火灾的影响。

关键词:数据仓库;森林火灾;气象因子

中图分类号:S762.22 **文献标志码:**A **文章编号:**1001-7461(2012)02-0127-04

Design of Forest Fire Factor Data Warehouse for Qiannan of Guizhou

XIAO Yun-dan, JI Ping

(Institute of Forest Resources Information Techniques, CAF, Beijing 100091, China)

Abstract: This paper aimed to collecting meteorological information and fire data over past decade in Qiannan of Guizhou Province, establishing the forest fire factor data warehouse, and designing multi-dimensional data model of impact factors of forest fire, and identifying the organization structure of multidimensional data, as well as make multidimensional analysis. Association analysis was used to analyze the relationship between meteorological factors and fire occurring to find the climate effects on forest fire in Qiannan.

Key words: data warehouse; forest fire; meteorological factor

森林火灾的发生,绝不是偶然的,从气候角度来考察,它都有一个孕育的过程,森林火灾的发生往往具有明显的地域性和季节性的特点,并且随着气候的准周期变化,森林火灾的多发年和少发年也有准周期的振动^[1]。对于火灾这样一个复杂的大系统,各因素之间有着复杂的相互关系,这种关系在大量的历史数据记录中表现为火灾和气象报告中各指标数据项之间定性的关联关系。对这些数据的应用大多还停留在查询、检索阶段,数据库中隐含的具有丰富的决策意义知识和信息远远没有得到充分的发掘和利用。如何在不精确、难以建模、动态海量的火灾信息中发现有价值的信息,建立火险因子数据仓库,并在大量的历史数据记录中找出各指标数据项之间定性的关联关系,将这种关系量化“挖掘”出来并反映出火灾实际情况的定量关联规则。对于减少火灾对林业经济造成的损失,分析灾后恢复方案并积极推动灾后重建等都具有重大的现实意义。

本研究收集近 10 a 来贵州省黔南州的气象及火灾数据,旨在建立贵州省黔南火险因子数据仓库;以数据仓库为研究基础,对数据进行查询、联机分析处理及数据挖掘等,为下一步对数据挖掘工作打好基础,以期能对数据仓库内各气象因子与火灾进行关联分析,找出气候对黔南森林火灾的影响,建立关联规则,以及对随着时间空间变化呈现出明显的或潜在的变化特点的指标,揭示黔南州春季防火期主要气象要素的时空变化特征,探索该地区春季森林火灾发生规律及其影响因素。

1 数据仓库的概念

数据仓库是决策支持系统和联机分析应用数据源的结构化数据环境。数据仓库的特征在于面向主题、集成性、稳定性和时变性,它是一个过程而不是一个项目。数据仓库的出现,并不是要取代数据库。目前,大部分数据仓库还是用关系数据库管理系统来管

理的。可以说,数据库、数据仓库相辅相成、各有千秋。数据库是面向事务的设计,数据仓库是面向主题设计的。数据库一般存储在线交易数据,数据仓库存储的一般是历史数据。数据库设计是尽量避免冗余,一般采用符合范式的规则来设计,数据仓库在设计时有意引入冗余,采用反范式的方式来设计。数据库是为捕获数据而设计,数据仓库是为分析数据而设计,它的 2 个基本的元素是维表和事实表。

2 火险因子数据仓库的建立

数据仓库主要工作的对象是多维数据,因此又称为多维数据库。采用 MS SQL Server2000 设计和建立火险因子数据仓库。设计和建立黔南州火灾影响因子多维数据概念,确定多维数据的组织方式,进行数据仓库多维分析。数据仓库采用 Analysis Service 进行多维数据的设计,以多维数据的组织方式管理。以建立的数据仓库为数据源,通过 SQL Server 的 Analysis 管理系统建立数据立方体,提供 OLAP 在线查询分析服务。本研究数据仓库采用基于关系表的存储方式,采用星型架构为基础建立多维数据立方体,核心是事实表,围绕事实表的是维度表。按时间、地区、气象建立相应的维度表,每个维度划分成为多个层次。选取不同的维度表和事实表,对数据进行多维分析。

2.1 建立数据仓库的步骤

2.1.1 确定数据仓库的用户及主题 确定数据仓库的主题是建立数据仓库的最开始阶段,应针对应用的特点进行分析设计。以贵州省黔南州森林春季防火期主要气象要素和林火情况为研究对象,收集整理数据并建立贵州省黔南森林火灾影响因子数据仓库,试图找出气象因子对森林火灾的影响与关系。

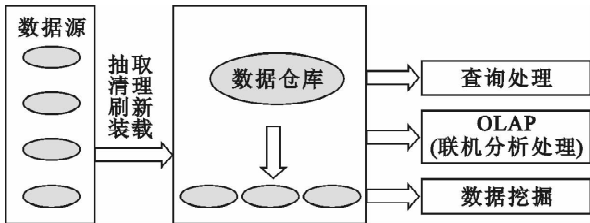


图 1 数据仓库结构图

Fig. 1 Constructional drawing of data warehouse

2.1.2 数据抽取及数据预处理 以现有的数据库为基础,同时考虑到数据仓库建立和运行中可能存在的问题,对数据库中数据进行完整性、一致性检查,对其中的噪声数据进行处理,对丢失的数据进行适当填补等数据预处理过程,针对项目主题确定最佳的可行方案,明确从每个数据源中所抽取的数据及定义数据如何转换、装载到主题的数据表中,对数

据进行事务处理、批处理等各种类型的数据处理工作。对数据进行清理、转换、并对数据进行重新组织和加工,装载到数据仓库的目标库中^[2]。

2.1.3 设计数据仓库

2.1.3.1 数据仓库的事实表的设计 事实表的主要特点是包含数字数据,而这些数字数据可以汇总,以提供有关单位动作的历史数据。每个事实表还包括一个或多个部分组成的索引,该索引包含作为外键的相关性维度表的主键^[3]。本研究的事实表包括火灾面积、火灾次数、其他灾害受灾面积、气象等基本观测数据。

2.1.3.2 数据仓库维度表的设计 维是人们观察数据的特定角度,是考虑问题时的一类属性。属性的集合构成一个维,如时间维、地区维、气象维等。维的一个具体取值是数据项在某维中位置的描述。同一维度还可以存在细节程度不同的各个描述方面,这些多个描述方面为维的层次(图 2)。

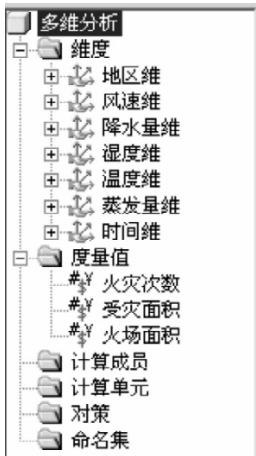


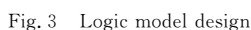
图 2 维度设计

Fig. 2 Multi-dimensional design

2.1.3.3 逻辑模型设计 逻辑建模是数据仓库实施中的重要环节,它能直接反映用户需求,同时对系统的物理实施有重要的指导作用^[4]。在构建多维数据集时常用的架构有星型架构、雪花状架构和星型雪花架构,本项目采用的是星形架构建模(图 3)。星型架构是以事实表为核心,其他的维度表围绕这个核心表星型分布。维度表只与事实表关联,维度表彼此之间没有任何联系,每个维度表中的主码都只能是单列的,同时该主码被放置在事实数据表中,作为事实数据表与维度表连接的外码。

2.1.4 数据仓库的粒度 粒度是指数据仓库的数据单位中保存数据的细化或综合程度的级别。细化程度越高,粒度级别就越小;细化程度越低,粒度级别就越大^[5]。黔南火险因子数据仓库中时间粒度包括月、季度。地区维的粒度包括州、县(市)。

MOLAP^[6]。联机分析处理的目的是多角度分析历年相关数据,对数据仓库内数据进行关联分析,找出气象因子对黔南森林火灾的影响。



上卷操作即通过概念分层向上攀升或者通过维约,在数据方上进行聚集。如在时间维上,月 \leq 季 \leq 年,上卷操作可由“月”层向“季”层聚集数据。当用维归约进行上卷时,一个或多个维由给定的数据方删除。如在考虑只包含两维“地区”和“时间”的数据方。上卷可以删除“时间”维,导致火灾面积是按地区而不是按时间和地区聚集,以此分析时间维或地区维上的气象因子和火灾的变化特征,进而探讨两者之间的关系。

下钻是上卷的逆操作,它由不太详细的数据到更详细的数据,能够综合分析各维度上气象因子和火灾的变化情况。下钻可以通过沿维的概念分层向下或引入新的维来实现。如沿着县(市) \leftarrow 州 \leftarrow 省 \leftarrow 国家定义的地区维的概念向下,在中心数据方执行下钻操作的结果。这里,下钻由地区维的分层向下,由州层到更详细的县(市)层。结果数据方详细地列出每个县(市)火灾面积的大小,而不是按州或者省的火灾面积求和(图4)。

[illegible]

Fig. 4 Drilling

本研究基于数据挖掘技术以贵州省黔南为研究对象,在前期获得贵州省历年林业基础数据的基础上,调查贵州省火灾后的相关林业数据,对数据库中数据进行完整性、一致性检查,对其中的噪声数据进

行处理,对丢失的数据进行适当填补,建立贵州省黔南火险因子数据仓库;在数据仓库的基础上能对数据进行查询、联机分析处理及数据挖掘等,探索该地区春季森林火灾发生规律及其影响因素,以期为预防和控制森林火灾提供基础数据和理论支持。

通过研究表明,都匀地区发生森林火灾次数最多,其次是平塘和罗甸;而总火场面积最大的是独山地区,其次是荔波和罗甸。数据说明都匀、平塘等地区易发生森林火灾,而独山、荔波地区森林火灾更易发生蔓延。将都匀、平塘和罗甸 3 地连成分界线,那么分界线东南 6 个县市(包括上述 3 个地区)发生森林火灾次数及火场面积均大于西北 6 个县市,分别占总数的 60.4%和 67.7%。

研究中还发现,2008 年黔南州森林春季防火期内发生森林火灾较为频繁,高达 847 次,是往年发生频率的 2~9 倍,这是由于雪凝灾害的发生,使得可燃物成倍增加,且火环境有所改变,使得火灾发生频率有所增加。气候干燥度西部地区大于东北地区;最大风速易出现在惠水、都匀等中部地区。都匀、荔波、平塘、独山、龙里县森林火情发生率最高,其次为罗甸、

贵定、长顺、惠水、三都,再次为福泉、瓮安,即每 1 万公顷森林面积发生森林火灾次数较少。发生森林火灾快速反应能力及控制能力,以瓮安、都匀、福泉、长顺为最快。龙里、罗甸、惠水、贵定其次,三都、荔波、独山、平塘扑救森林火灾快速反应能力相对较弱。森林火灾以荔波、独山、平塘、都匀为重,其次是罗甸、龙里、惠水、贵定、长顺、三都,较轻的是瓮安、福泉。

参考文献:

[1] 张艳平. 黑龙江大兴安岭地区气候变化对森林火灾影响的研究[D]. 东北林业大学,2008:2-4.

[2] 刘伟杰. 基于数据仓库理论的林业信息数据模型的设计[J]. 吉林林业科技,2006,11(3): 29-31.

[3] 杨卫民,梁红,谭骏珊. 数据仓库在森林资源决策支持系统中的应用[J]. 计算机技术与发展,2006, 16(4):187-189.

[4] 李柯. 基于数据仓库和 OLAP 的民航灾害预警决策支持系统[J]. 西南交通大学学报,2007,42(4):511-515.

[5] 陈昌鹏,吴保国. 林业数据仓库的设计[J]. 农业网络信息:信息资源建设,2004(4): 30-32.

[6] 王红霞,朱喜林,马季兰,等. 气象数据仓库建立及数据统计与挖掘[J]. 太原理工大学学报, 2006, 37(Supp. 1): 101-103.

(上接第 78 页)

[3] 李联地,张恩生,王忠民,等. 坝上野生金莲花调查初报[J]. 河北林业科技,2003(5):19.

[4] 李勇,丁万隆. 我国部分金莲花种质资源遗传多样性的 RAPD 研究[J]. 植物遗传资源学报,2009,10(4): 535-539.

LI Y, DING W L. Genetic diversity analysis of partial Chinese *Trollius* populations based on RAPD markers[J]. Journal of Plant Genetic Resources, 2009,10(4): 535-539. (in Chinese)

[5] 任正博,张振学,周伟利,等. 金莲花化学成分研究[J]. 中国现代中药,2006,8(2):14-15.

REN Z B,ZHANG Z X,ZHOU W L, *et al.* Studies on the bio-activity consituents *Trollius ledebourii* [J]. Modern Chinese Medicine,2006,8(2):14-15. (in Chinese)

[6] 宋冬梅,孙启时. 金莲花属植物研究进展[J]. 沈阳药科大学学报,2005,22(3):231-234.

SONG D M, SUN Q S. Advances on *Trollius*[J]. Journal of Shenyang Pharmaceutical University, 2005, 22(3): 231-234. (in Chinese)

[7] 吴利霞,黄金祥,崔同. 金莲花植物成分及其药剂的研究进展[J]. 河北林果研究,2004,19(4):392-396.

WU L X,HUANG J X,CUI T. A review of the studies on elements and preparations of flos *Trollius*[J]. Hebei Journal of Forestry and Orchard Research, 2004, 19(4): 392-396. (in Chinese)

[8] 朱殿龙,丁万隆,陈士林. 金莲花属植物的研究进展[J]. 世界科学技术—中医药现代化,2006,8(4):26-33.

ZHU D L, DING W L, CHEN S L. Advances on *Trollius* [J]. World Science and Technology/Modernization of Traditional Chinese medicine and Materia Medica,2006,8(4):26-

33. (in Chinese)

[9] 赵二劳,王爱玲. 金莲花中总黄酮提取工艺的优化[J]. 精细化工,2009,26(12):1185-1187.

ZHAO E L,WANG A L. Optimization for extracting total flavonoids from *Trollius chinensis* Bunge[J]. Fine Chemicals, 2009,26(12):1185-1187. (in Chinese)

[10] 杨玉芳,王玄,赵红霞,等. 金莲花组织培养和快繁体系建立的研究[J]. 中国农学通报,2011,27(8):136-139.

YANG Y F,WANG X,ZHAO H X,*et al.* Research on optimization system on tissue culture and rapid propagation of *Trollius chinensis* [J]. Chinese Agricultural Science Bulletin, 2011,27(8):136-139. (in Chinese)

[11] 周应群,陈士林,张本刚,等. 中药资源调查方法研究[J]. 世界科学技术—中医药现代化,2005,7(6):130-136.

ZHOU Y Q,CHEN S L,ZHANG B G, *et al.* Study on methods of traditional Chinese medicinal resources survey [J]. World Science and Technology/Modernization of Traditional Chinese medicine and Materia Medica,2005,7(6):130-136. (in Chinese)

[12] 丁万隆,杨春清,张泽印,等. 金莲花生产标准操作规程(SOP) [J]. 现代中药研究与实践,2006,20(5):12-15.

[13] 陈宏伟,殷鸣放,刘娜,等. 辽东山区林下人参生长与林内温度和光照因子关系初步研究[J]. 西北林学院学报,2007, 22(2): 20-23.

CHEN H W,YIN M F,LIU N, *et al.* A preliminary study on the relationship between growth of ginseng and temperature and illumination in forest[J]. Journal of Northwest Forestry University, 2007, 22(2):20-23. (in Chinese)