

文章编号:1671-8860(2008)01-0046-05

文献标志码:A

基于实例的 GIS 数据库模式匹配方法

王育红^{1,2} 陈军²

(1 河南理工大学测绘与国土信息工程学院,焦作市世纪大道 2001 号,454003)

(2 国家基础地理信息中心,北京市海淀区紫竹院百胜村 1 号,100044)

摘要:针对空间数据库的特点,提出了一种改进的基于实例的层次式模式匹配方法,并解决了实体对应关系的自动建立、要素类相似性计算、属性相似性计算等问题。最后,通过实验验证了该方法的有效性。

关键词:模式匹配;相似性;GIS 数据库;更新;共同实体识别

中图法分类号:P208

随着空间信息技术的日益发展和成熟,许多部门纷纷建立了面向特定需求的 GIS 应用系统。为节省数据采集成本,加速系统构建进程,这些部门一般要从空间数据生产单位获取已有空间数据库的副本,通过一系列的加工处理,将其加载到自己的数据库中。为叙述方便并强调两者之间的派生关系,本文将生产单位所建的 GIS 数据库简称为主数据库,应用部门在主数据库副本基础之上所建的 GIS 数据库称为客户数据库。

在经过一系列处理之后,客户数据库和主数据库之间在拥有共性的同时,在要素类抽象层次、属性命名、数据内容等方面也存在一系列的差异。由于这些差异的存在,当利用新版主数据库对旧版客户数据库进行更新时,首先应该在主数据库和客户数据库之间执行模式匹配操作^[1,2],发现其间语义相同或相互的模式元素,并显式地声明彼此之间的映射关系,以确定新版主数据库中哪些类型的数据可以用于客户数据库的更新,据此选择最佳的更新集成策略。

本文在分析总结相关研究^[3-6]的基础上,针对空间数据库的特点和应用背景,提出一种基于实例的模式匹配方法,详细讨论了其中几个关键问题的解决途径,并通过实验验证了该方法的有效性。

1 模式匹配的实施框架

作为与传统数据库的显著区别之一, GIS 数

据库不仅存储空间实体的常规属性信息,而且包含几何定位信息。即使在实体相同 ID 不存在或者未知的情况下,也可利用其几何信息来确定两个要素类之间的共同实体^[7,8]。显然,如果两个要素类中存在共同的实体,这两个要素类就可以认为是相关的,然后通过对比分析共同实体的属性值,可进一步确定相关的属性^[5,6]。基于以上认识,本文提出了一种基于实例的模式匹配方法,其基本实施框架如图 1 所示。

GIS 数据库往往规定一个要素类中只能存储相同几何类型的空间实体,据此,将要素类划分为点、线、面三种基本要素类型。假设主数据库和客户数据库中分别存在 M 个要素类(包括 M_1 个点要素类、 M_2 个线要素类、 M_3 个面要素类)和 N 个要素类(包括 N_1 个点要素类、 N_2 个线要素类、 N_3 个面要素类)。如果采用两两直接比较的方式来判断其间是否存在共同实体,不仅需要较多的比较对数($M \times N$ 对),而且也很难判断不同几何类型的空间实体是否对应于同一现实地物。因此,首先对要素类进行分组,只在几何类型相同的要素类之间执行共同实体识别,以减少比较对数(现在的比较对数为 $M_1 \times N_1 + M_2 \times N_2 + M_3 \times N_2$)、降低系统处理时间。

为进一步提高执行效率,本文采用抽样统计方式发现两个要素类之间的共同实体,其基本流程如图 2 所示。首先,在客户数据库要素类中随

利用数据库中的现有实例数据,不需要其他辅助信息和工具的支持;②能够识别模式元素间的同物异名和同名异物现象,可避免模式匹配中的错配或漏配现象;③所采取的属性相似性计算方法不要求实例值严格相同,具有一定的容错性。

致谢:感谢国家地理信息中心蒋捷教授、香港理工大学李志林教授在本文写作时给予的指导和帮助。

参 考 文 献

- [1] Sperly L. A Framework for Update Process in GIS [C]. The Third International Conference on GeoComputation, University of Bristol, Bristol, 1998
- [2] Arnaud B. From the Schema Matching to the Integration of Updating Information into User Geographic Database [C]. The 12th International Conference on Geoinformatics, Geospatial Information Research: Bridging the Pacific and Atlantic, University of Gvle, Sweden, 2004
- [3] Rahm E, Bernstein P A. A Survey of Approaches to Automatic Schema Matching [J]. The VLDB Journal, 2001(10): 334-350
- [4] 关佶红,虞为,安扬, GML 模式匹配算法 [J]. 武汉大学学报·信息科学版, 2004, 29(2): 169-174
- [5] Fan Weiguo, Lu Hongjun, Madnick S E, et al. Discovering and Reconciling Value Conflicts for Numerical Data integration [J]. Information Systems, 2001, 26: 635-656
- [6] Chua C E H, Chiang R H L, Lim E P. Instance Based Attribute Identification in Database Integration [J]. VLDB Journal, 2003, 12(3): 228-243
- [7] Yuan Shuxin, Tao Chuang. Development of Conflation Components [C]. Geoinformatics'99, International conference on Geoinformatics and Socioinformatics, Ann Arbor, Michigan, USA, 1999
- [8] Volker W, Dieter F. Matching Spatial Data Sets: a Statistical Approach [J]. International Journal of Geographical Information Science, 1999, 13 (5): 445-473
- [9] Naphtali R, Rukshan A, Yuan Jun, et al. Semantic Relations: the Key to Integrating and Query Processing in Heterogeneous Databases [C]. The World Multiconference on Systemics, Cybernetics and Informatics, Computer Science and Engineering: Part I, USA, 2000

第一作者简介:王育红,博士生,现从事空间数据库更新、异构空间数据集成、GIS 应用软件系统开发等研究。

E-mail:wyh3003@tom.com

An Instance-Based Approach for Schema Matching Between GIS Databases

WANG Yuhong^{1, 2} CHEN Jun²

(1) School of Surveying and Land Information Engineering, Henan Polytechnic University, 2001 Shiji Road, Jiaozuo 454003, China)

(2) National Geomatics Center of China, 1 Baishengcun, Haidian District, Beijing 10004, China)

Abstract: According to the characteristic of GIS database, an improved instance-based hierarchical approach to schema matching is proposed, some key issues within which are also be resolved such as automatic identification of shared entities, similarity computation between feature classes, attribute similarity computation, etc. Finally, experimental result is present to validate the efficiency of this approach.

Key words: schema matching; similarity; GIS database; updating; shared entities identification

About the first author: WANG Yuhong, Ph.D candidate, majors in GIS database, etc.

E-mail: wyh3003@tom.com