

空间拓扑在地理信息系统的应用与前景

武夏夏

(重庆大学计算机学院 重庆 400044)

摘 要 本文首先介绍了地理信息系统(GIS),并阐述了空间拓扑这一工具在地理信息系统中的应用,介绍了几种常用的空间拓扑关系表达方式。通过分析近年来空间拓扑在地理信息系统应用要求下的发展趋势,提出了今后在地理信息系统应用中的空间拓扑的发展方向与研究热点。

关键词 地理信息系统, GIS, 空间拓扑, RCC, N-交集, HBM

1 引言

以计算机为核心的信息处理系统技术是二次世界大战后科技革命的主要标志之一。在信息的诸多类型中与空间相关的信息是十分重要的一类。人类生存的地球这个三维空间中的万物无不与空间位置相关,如何利用计算机处理空间相关信息是地理信息系统(Geographic Information System,简称 GIS)产生和发展的原动力。

地理信息系统(GIS)由于其在生产生活中的适用性而具有广泛的应用空间,它不仅包括了对空间数据的存储、管理还包括了对空间数据的进一步处理:空间查询与分析,其中空间查询与分析是地理信息系统(GIS)的核心。

空间拓扑关系是指在拓扑变换(旋转、平移、缩放等)下保持不变的空间关系,即拓扑不变量。由于对象间空间关系的描述形式各不相同,对象间的空间关系是极为复杂多样的。拓扑关系是不考虑度量 and 方向的空间实体之间的空间关系,由于拓扑学是研究图形在拓扑变化下不变的性质,拓扑学已成为空间关系研究的基础,为空间对象间的包含、覆盖、相离和相接等空间关系的描述提供直接的理论依据^[1],是GIS中最重要的一类空间关系。在GIS中,地理空间中的点、线、面实体之间存在着各种各样的拓扑关系,表示拓扑关系的数据是空间数据的重要组成部分。

空间对象关系是空间查询与分析的基础。空间关系的描述与识别非常重要。一方面它为地理信息系统数据库的有效建立、空间查询、空间分析、辅助决策等提供了最基础的关系,另一方面是将空间关系理论应用于地理信息系统查询语言,形成一个标准的SQL空间查询语言,从而通过应用程序进行空间特征的存储、提取、查询、更新等。

综上所述,拓扑空间关系是空间关系研究的重点和难点。

2 空间拓扑与 GIS 结合

2.1 GIS 应用中的空间拓扑关系研究

近年来,在拓扑空间关系描述方面,人们已经作了许多工作,对确定性区域间拓扑关系模型的研究取得了很大进展,也为不确定性区域间拓扑关系的研究奠定了基础。目前主要有基于逻辑的公理化拓扑理论和传统的数学拓扑两大类方法,其中最具有代表性的是RCC形式化模型和n-交集模型,另外还有许多学者提出的改进方法^[8]。

1988年Guting以点集为基础,运用集合运算 $=$ 、 $\neq$ 、 $\cap$,给出了相等(equal),不相等(unequal),包含(inside),相离(outside)和相交(intersect)等拓扑空间关系的定义。Pullar(1988)将点集方法加以扩充,运用拓扑学理论中点集的边界(boundary)和内部(in-terrier)的概念,给出了覆盖(overlap)和相邻(neighbor)两个关系定义。

Wagner(1988)定义了相邻、相离、严格包含、相交四种拓扑空间关系。Egenhofer和Franzosa(1991)提出了一个四元组表达的拓扑空间关系描述框架。Egenhofer(1993)引进了点集的余,构造了一个由点集的边界、内部、余之间的交集组成的九元组,以此作为描述两个点集间拓扑空间关系的框架。Clementini等人(1993)在Egenhofer所定义的四元组基础上,运用维数扩展法(dimension extended method),即用两个空间目标内部与边界之间交集的维数,作为描述两个点集间拓扑空间关系的框架。进一步,他们给出了二维拓扑空间关系的最小集。这种方法形式化地描述了二维空间目标之间的拓扑空间关系。

Clementini和Felice基于拓扑分类不变量给出了空间拓扑关系的分类^[10]。在国内,廖士中等人(1997)针对基于4-交集模型难以推导出拓扑关系的完备集、概念邻域和复合表等不足,以拓扑学为基础,提出了n维实体间拓扑关系完备集的概念,建立

了拓扑关系的闭球模型,基于闭球模型可以直接推导出拓扑关系的完备集、概念邻域以及复合表,闭球模型比 4-交集模型更为简单、有效。

2.2 几个常用的空间拓扑模型

a) RCC 形式化模型

RCC 模型以区域为基元,区域可以是任意维,但在特定的形式化模型中,所有区域的维数是相同的,如在考虑二维模型时,区域边界和区域间的交点不被考虑进来。

RCC 模型假设一个原始的二元关系 $C(x, y)$ 表示区域 x 与 y 连接。关系 C 具有自反性和对称性,可以根据点出现在区域中来给出关系 C 的拓扑解释。 $C(x, y)$ 表示 x 和 y 的拓扑闭包共享至少一个

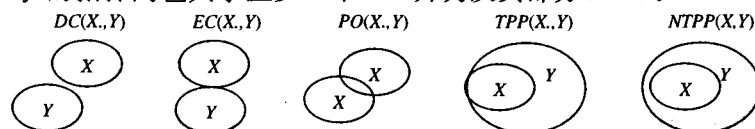


图 1 RCC-8

b) 9-交集模型

通过考虑点集边界和内部的交集,由点集运算定义的两个空间物体间的拓扑关系集,通过考虑点集边界和内部的交集给出了拓扑空间关系的描述,但是这些关系既不互斥也不完备。

对于 9-交集来说,是将两个对象 A, B , 分别按照边缘,内部点集,外部点集分为 $\partial A, A^\circ, A^-$ 与 $\partial B, B^\circ, B^-$, 由此, A, B 的拓扑关系就可以用它们边界、内点集和外部点集所组成的 +9 个交集的二元关系 $R(A, B)$ 来表示,分别为: $\partial A \cap \partial B, \partial A \cap B^\circ, \partial A \cap B^-, A^\circ \cap \partial B, A^\circ \cap B^\circ, B^\circ \cap B^-, A^- \cap \partial B, A^- \cap B^-$ 。

c) 基于层次的模型 HBM (Hierarchy-based Method)

人类认知过程具有由粗到细,由简到繁的层次化特征。将这一特征应用到拓扑关系的分类中,形成一个逐步细化,按层次展开的分类过程。这一过程得到一个树形结构,叶节点上的 5 种基本空间拓扑关系是 disjoint, touch, overlap, in 和 equal。

用三元组 $\langle \lambda_1, r, \lambda_2 \rangle$ 表示二维空间对象 λ_1 和 λ_2 间存在着关系 r 。在 IR^2 中定义 5 种拓扑关系如下:

(1) 分离 (disjoint)

$$\langle \lambda_1, \text{disjoint}, \lambda_2 \rangle \Leftrightarrow (\lambda_1 \cap \lambda_2 = \phi)$$

(2) 接触 (touch)

$$\langle \lambda_1, \text{touch}, \lambda_2 \rangle \Leftrightarrow (\lambda_1^\circ \cap \lambda_2^\circ = \phi) \wedge (\lambda_1 \cap \lambda_2 \neq \phi)$$

(3) 覆盖 (overlap)

$$\langle \lambda_1, \text{overlap}, \lambda_2 \rangle \Leftrightarrow (\lambda_1^\circ \cap \lambda_2^\circ \neq \phi) \wedge (\lambda_1^\circ \cap \lambda_2^- \neq \phi) \wedge$$

$$(\lambda_1^- \cap \lambda_2^\circ \neq \phi)$$

(4) 包含 (in)

$$\langle \lambda_1, \text{in}, \lambda_2 \rangle \Leftrightarrow (\lambda_1^\circ \cap \lambda_2^\circ \neq \phi) \wedge (\lambda_1 \cap \lambda_2 = \lambda_1) \wedge (\lambda_1$$

点,使用关系 C 可以定义 8 个基本关系。在 RCC 模型中,定义在区域上的关系通常被分组为关系集合,集合中的元素互不相交且联合完备 (jointly exhaustive and pairwise disjoint, 简称 JEPD), 即对于任何两个区域,有且仅有一个特定的 JEPD 关系被满足。

其中 RCC-8 (见图 1) 包括不连接 (DC)、外部连接 (EC)、部分交叠 (PO)、正切真部分 (TPP)、非正切真部分 (NTPP)、相等 (EQ)、反正切真部分 (TP-PI) 和反非正切真部分 (NTPPI)。RCC-5 没有考虑区域的边界,即将 DC 和 EC 合并为分离 (DR), TPP 和 NTPP 合并为真部分 (PP), TPPI 和 NTPPI 合并为反真部分 (PPI)。

$$\cap \lambda_2 \neq \lambda_2)$$

(5) 相等 (equal)

$$\langle \lambda_1, \text{equal}, \lambda_2 \rangle \Leftrightarrow (\lambda_1 = \lambda_2)$$

3 前景展望

由于 GIS 的应用日益广泛,空间拓扑的发展也日益重要起来。目前来看,关于空间拓扑的发展方向主要有如下几种。

(1) 在空间推理方面的应用。空间推理是指利用空间理论和人工智能 AI (Artificial Intelligence) 技术对空间对象进行建模、描述和表示,并据此对空间对象间的空间关系进行定性或定量分析和处理的过程。空间推理的研究起源于 20 世纪 70 年代初,最初是以量空间为研究对象的,多维的并且不能通过单一的纯量充分表示的空间,只是在近年来才成为知识表示的一个重要领域。目前,空间推理被广泛应用于地理信息系统、机器人导航、高级视觉、自然语言理解、工程设计和物理位置的常识推理等方面,并且正在不断向其他领域渗透,其内涵非常广泛。空间推理的研究在人工智能中占有很重要的地位。是人工智能领域的一个研究热点^[1]。地理信息系统是空间推理应用得最为广泛的领域之一,也是空间推理的研究热点之一^[12]。

定性空间推理是处理常识性空间知识的一种人工智能方法^[12],已经成为 GIS 的关键技术。由于空间知识本质上是定性的,所以研究空间推理的重点就是研究定性空间推理。定性空间推理是空间推理的重要组成部分。定性空间推理研究的关键问题有两个。一个是对空间实体及其相互位置关系在拓扑

(下转第 248 页)

型都不能反映文献与提问的相关性的大小。只有概率模型才能反映出文献与提问的相关性的大小,从而对相关文献进行排队,所以想要提高查准率还要采用概率模型才能得到实现。本文就不在这方面再作详细的讨论了。

总结 本文首先从信息检索数学模型做了简单的介绍入手,提出了现有的三种有代表性的信息检索模型:集合模型,代数模型,概率模型,并对这三种检索模型的优缺点作了分析,在此基础上提出了一种实用的二次检索方法,并且详细说明了此方法的检索过程,指出了采用二次检索的目的是为了提高检索的效率。文章最后通过对一组具体数据的采集和实验,用查全率和查准率来作为检索效率的评价标准,可以知道采用基于二次检索方法的查全率和查准率高于仅用布尔模型和向量模型的查全率和查准率。

各种模型的混合使用是多变、复杂的,结合各种模型的优势,采用一种混合模型就能够很好地提高检索效率。由于集合模型的成熟研究,现在的研究多在代数模型和概率方面,比如以概率论和模糊数学为补充手段的文献向量加权上,对比各种加权方

法的优劣,选择出一种或几种好的加权方法也是现在的研究热点。总之以代数模型和概率模型为基础的综合性的研究将成为将来信息检索模型的研究重点。

参考文献

- 1 Miyamoto S. Information Retrieval Based on Fuzzy Association [J]. Fuzzy Sets and System, 1990, 38(2): 191~205
- 2 Lee C, Lee G G. Probabilistic Information Retrieval Model for Dependency Structured Indexing System. Information Processing and Management [M]. SIGIR2002, August 2002
- 3 Jones K S, Walker S, Robertson S E. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments Part 2 [J]. Information Processing and Management, 2000, 36(12): 779~808
- 4 沈一东,邢永康. 一种新的知识表达模型——信度网[J]. 计算机科学, 2000, 9(27): 40~43
- 5 邓路华. 信息检索系统数学模型的理论及其评价——谨以此文献给信息检索的先驱杰拉德·索顿先生[J]. 大学图书馆学报, 2002, 1(3): 6~4
- 6 邢永康, 马少平. 信息检索的概率模型[J]. 计算机科学, 2003, 30(8): 13~17
- 7 康耀红. 现代情报检索理论[M]. 北京: 科技文献出版社, 1990
- 8 王继成, 邹涛, 等. 基于 Internet 的信息资源发现技术与实现. 计算机研究与发展[J], 1999, 36(11): 1369~1374
- 9 何静, 刘海燕. 信息检索与过滤中的信息需求表示方法[J]. 计算机工程与设计, 2003, 24(8): 3~8

(上接第 238 页)

空间关系描述发展的基础上,可将其研究结果应用于空间推理,对空间推理的发展产生了正向的推动。由于空间推理的研究对象的转变,极大地扩展了空间推理的应用领域,使空间推理的理论和应用研究近年来有了长足的进展。在国外,近年来成立了许多专门从事空间推理方面研究的协会和联盟,如 NCGIA(National Center for Geographic and Analysis), USGS(U. S. Geological Survey), 欧洲定性空间推理网 SPACENET 以及匹兹堡大学的空间信息研究组和慕尼黑大学空间推理研究组等等。国际知名期刊 Artificial Intelligence 近年来发表了许多篇空间推理方面的文章,而且呈逐年增长的趋势。

(2) 空间拓扑本身方法的创新。由于拓扑空间关系表示是空间关系理论的重要组成部分,也是空间数据库设计的重要基础,其研究将有助于设计有效的空间查询和有效的数据处理方式。而由于对空间关系表示的侧重点不同,也产生了很多的空间拓扑关系表示方法,因此基于不同空间拓扑关系表示方法的空间推理方法一般来说是不可以通用的,这就导致了同一问题研究的重复和浪费。

因此,在空间拓扑领域的研究中,一方面应该研究出更符合地理信息系统本身特点的表达方式,另一方面应该找到将现有各种表示方法转化为一个统

一的较优的空间拓扑关系表达方式,使原来在不同空间拓扑关系条件下的空间推理方法得到新的应用。

参考文献

- 1 刘亚彬, 刘大有. 空间推理与地理信息系统综述. 软件学报, 2000
- 2 Renz J, Nebel B. On the complexity of qualitative spatial reasoning: a maximal tractable fragment of the region connection calculus. Artificial Intelligence, 1999
- 3 U. S. Geologic Survey URL. 1998. <http://nsdi.usgs.gov/nsdi>
- 4 肖乐斌, 等. 三维 GIS 的基本问题探讨. 见: 地理信息系统论坛 (GIS Forum)—学术论文, 2002
- 5 王康弘, 刘利, 钟耳顺. 地籍空间实体的空间拓扑关系和变更类型分析. 2001 中国 GIS 年会论文集, 2001. 3
- 6 肖乐斌, 钟耳顺, 等. GIS 空间概念模型的研究. 见: 地理信息系统论坛 (GIS Forum)—学术论文, 2002
- 7 GuoPing Tao Huang-Fu, Research on the Relationship Between 4-intersection and Classifying Invariant Base on the Simple Regions. In: International Conference on Machine Learning and Cybernetics (ICMLC 2003), 2003. 11
- 8 虞强源, 刘大有, 等. 空间区域拓扑关系分析方法综述. 软件学报, 2003
- 9 Egenhofer M J, Franzosa R D. On the Equivalence of Topological Relationships. Int. Jour. Of Geographical Information Systems, 1995(9): 133~152.
- 10 Clemintini E, Felice P Di. Topological invariants for lines. IEEE Transactions on Knowledge and Data Engineering, 1998, 10(1): 38~54
- 11 郭薇, 陈军. 基于点集拓扑学的三维拓扑空间关系形式化描述. 测绘学报, 1997, 26(2): 122~127
- 12 廖士中, 石纯一. 定性空间推理的研究与进展. 计算机科学, 1998, 25(4): 11~13