

度量参数与空间关系描述的研究

陈学工, 张弛伟, 张文艺, 张 厅

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

摘要:度量参数作为对9交集拓扑关系模型所识别的关系种类的细化,在判断一种几何结构是否匹配一个空间术语时,相比于单独使用拓扑,它能提供更精确的描述。文中分析了9交集模型在实际应用中的不足以及目前的度量参数不支持带孔面对象的缺点,定义了三对细化的面面拓扑关系,并提出了一组支持带孔面对象的面对象间空间关系度量参数。这些度量参数加强了面对象间空间关系的描述,可广泛应用于空间查询和空间分析中。

关键词:度量参数;9交集模型;拓扑;空间关系

中图分类号:TP391.72

文献标识码:A

文章编号:1673-629X(2007)05-0187-04

Research on Metric Parameters and Spatial Relations Description

CHEN Xue-gong, ZHANG Chi-wei, ZHANG Wen-yi, ZHANG Ting

(Sch. of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: The metric parameters are expressed as refinements of the categories identified by 9-intersection, a model for topological spatial relations, and provide a more precise measure than does topology alone as to whether a geometric configuration matches with a spatial term or not. Some shortages of the 9-intersection model on practical applications and the disadvantages of metric parameters at present that no support area object with holes are pointed out, 3 pairs of refined area-area topological relations are defined, and a set of metric parameters of area-area topological relations are proposed, the parameters can support area object with holes. The description of area-area spatial relations are enhanced by these metric parameters, and the parameters are widespread applied on spatial query and spatial analysis.

Key words: metric parameters; 9-intersection model; topology; spatial relations

0 引言

拓扑关系是地理信息系统领域中非常重要的一种空间关系。完备和形式化的拓扑关系描述和表达是设计空间查询语言的关键,是实现有效空间查询的前提,同时也是空间分析的基础^[1]。近年来,国内外学者在这个方面进行了大量研究,其中最著名的模型是 Egenhofer 提出的 9 交集模型,它基于点集拓扑理论来定性描述目标间的拓扑关系。许多学者对 9 交集模型进行了改进,使其在识别能力和对空间对象的支持性等方面都有所加强,但还是没有解决区分的拓扑关系有限这一问题,这将直接影响到空间查询和空间分析的准确性。为此,笔者定义了三对细化的面面拓扑关系,并提出了一组面对象间空间关系的度量参数,弥补了之前提出的度量参数不支持带孔面对象的缺点,并将度量参数应用在了拓扑查询中。

1 9 交集模型在实际应用中的不足

1.1 9 交集模型框架

Egenhofer 在 1991 年提出了 9 交集模型,将平面上的两个对象 A 和 B 分为了 A 的内部(A^+)、边界(∂A)和外部(A^-)与 B 的内部(B^+)、边界(∂B)和外部(B^-)。对象的这 6 个部分构成 9 交矩阵,它定义了一个拓扑关系,可以用下面的矩阵来表达这个关系:

$$R_9(A, B) = \begin{bmatrix} A^+ \cap B^+ & A^+ \cap \partial B & A^+ \cap B^- \\ \partial A \cap B^+ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^+ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix} \quad (1)$$

在式(1)中矩阵中每一项可以取值空(0)或非空(1),所以 9 交矩阵可以确定 $2^9 = 512$ 种拓扑关系。就面面关系来说,有 8 种关系是可以实现的,并且它们满足互斥性和完备性。这 8 种关系为:相离(disjoint)、相接(meet)、交叠(overlap)、相等(equal)、包含(contain)、在内部(inside)、覆盖(cover)和被覆盖(covered by)。

1.2 实际应用中的不足

为了让 9 交集模型能区分交集为点、线、面的情况和支持混合对象,相继发展了维数扩展法 9 交模型^[2]

收稿日期:2006-08-10

作者简介:陈学工(1965-),男,湖南长沙人,副教授,博士,研究方向为信息系统和数据库技术。

和集成性维数扩展法 9 交模型^[3]。由于 9 交模型本质上是一种定性的模型,改进后的模型还是没有解决区分的拓扑关系有限这一问题,人们通常感知到的一些拓扑关系间的区别通过 9 交集模型无法体现出来。例如,图 1 中 A 和 B 的关系,9 交集模型中都描述为相离,不过相比之下,图 1(a) 中 A 和 B 更趋近于相接。若在城市规划中,A 为居民区,B 为化工厂,显然图 1(a) 中化工厂对居民区的影响要严重得多。此外,定性描述的拓扑关系易受误差和不确定性的影响,也不适用于动态信息分析^[4],因此 9 交集模型还不能很好地满足实际应用的需要。

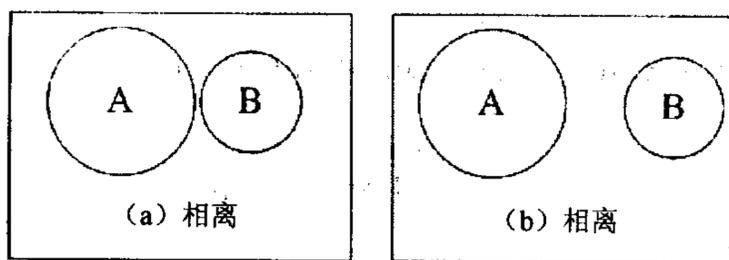


图 1 相离的两种情形

2 面面空间关系度量参数

2.1 目前度量参数的不足

为了使 9 交集模型更好地满足实际应用的需要,一些学者在面面空间关系中引入了度量参数^[4~7],增加了可区分的拓扑关系,但他们都没有考虑到带孔的面对象。在 OGC(Open GIS Consortium)制定的业界标准——OpenGIS 实现规范(OpenGIS Simple Features Specification for SQL)中,明确规定了平面面对象(即多边形)是由一个外部边界和零或多个内部边界所定义的^[8],每个内部边界定义面对象里的一个孔。由于孔的存在,面对象间的拓扑关系进一步细化了,相离分为内离和外离,相接分为内接和外接(见图 2),交叠分为内交叠和外交叠。给定实体 A 与 B,用 A_E 代表 A 的外部边界定义的实体, B_E 代表 B 的外部边界定义的实体,A 与 B 之间的以上 6 种细化拓扑关系定义如下:

- 1) 内离:A 与 B 相离,且 $A_E \cap B_E = A_E$ 或 $A_E \cap B_E = B_E$;
- 2) 外离:A 与 B 相离,且 A_E 与 B_E 也相离;
- 3) 内接:A 与 B 相接,且 $A_E \cap B_E = A_E$ 或 $A_E \cap B_E = B_E$;
- 4) 外接:A 与 B 相接,且 A_E 与 B_E 也相接;
- 5) 内交叠:A 与 B 交叠,且 $A_E \cap B_E = A_E$ 或 $A_E \cap B_E = B_E$;
- 6) 外交叠:A 与 B 交叠,且 A_E 与 B_E 也交叠。

以上三对细化的拓扑关系使得面对象间拓扑关系的描述变得更加准确了。例如,图 2 中实体 A 和 B 的拓扑关系分别为外接和内接,若在气象分析中,A 为

40mm 降雨带,A 中的孔为 60mm 降雨带,B 为居民区,显然图 2(b) 中居民区受降雨的影响更大。然而,由于目前的度量参数不支持带孔的面对象,所以不能区别三对细化的拓扑关系,因此要定义新的度量参数来解决这一问题。

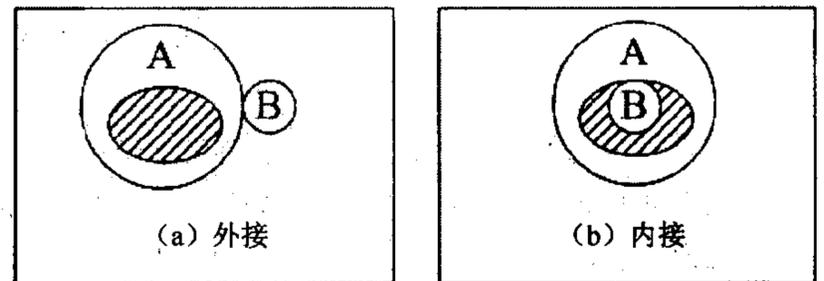


图 2 相接的两种情况

2.2 定义度量参数和取值范围

度量参数分为两类:分割度和临近度^[7]。分割度表示面对象的边界或内部被分割的程度,临近度表示两个面对象之间或一个面对象与另一个面对象的孔之间临近的程度。以 A 为参考对象,定义了面对象 A、B 之间空间关系的 10 个度量参数,其中名称以 S 结尾的表示分割度,以 C 结尾的表示临近度,以 H 开头的表示包含带孔的面对象。由于孔描述的是面对象的外部,而面对象的外部之间的关系没有实际的研究意义,因此没有定义孔之间的度量参数。

(1)IAS:表示 B 的内部被 A 的内部分割的程度,定义为 $LAS = \text{area}(A \cap B^*) \setminus \text{area}(B)$,取值范围为 $[0,1]$;

(2)IBS:表示 B 的边界被 A 的内部分割的程度,定义为 $IBS = \text{length}(A^* \cap \partial B) \setminus \text{length}(\partial B)$,取值范围为 $[0,1]$;

(3)AS:表示 B 的边界被 A 的边界分割的程度,定义为 $AS = \text{length}(\partial A \cap \partial B) \setminus \text{length}(\partial B)$,取值范围为 $[0,1]$;

(4)IC:表示 B 与 A 在内部的临近程度(如图 3(a)所示),定义为 $IC = \text{area}(\Delta_I(A)) \setminus \text{area}(A)$,取值范围为 $[0,1]$;

(5)OC:表示 B 与 A 在外部的临近程度(如图 3(b)所示),定义为 $OC = (\text{area}(\Delta_O(A)) + \text{area}(A)) \setminus \text{area}(A)$,取值范围为 $[1, +\infty)$;

(6)HIAS:表示 B 的内部被 A 的孔的内部分割的程度,定义为 $HIAS = \text{area}(\text{hole}_A \cap B^*) \setminus \text{area}(B)$,取值范围为 $[0,1]$;

(7)HIBS:表示 B 的边界被 A 的孔的内部分割的程度,定义为 $HIBS = \text{length}(\text{hole}_A \cap \partial B) \setminus \text{length}(\partial B)$,取值范围为 $[0,1]$;

(8)HAS:表示 B 的边界被 A 的孔的边界分割的程度,定义为 $HAS = \text{length}(\partial \text{hole}_A \cap \partial B)$,取值范围为 $[0,1]$;

(9)HIC:表示B与A的孔在内部的临近程度,定义为 $HIC = \text{area}(\Delta_I(\text{hole}_A)) \setminus \text{area}(\text{hole}_A)$,取值范围为[0,1];

(10)HOC:在 $B \subset A$ 的条件下,表示B与A的孔在外部的临近程度,定义为 $HOC = (\text{area}(\Delta_O(\text{hole}_A)) + \text{area}(\text{hole}_A)) \setminus \text{area}(\text{hole}_A)$,取值范围为[1, +∞)。

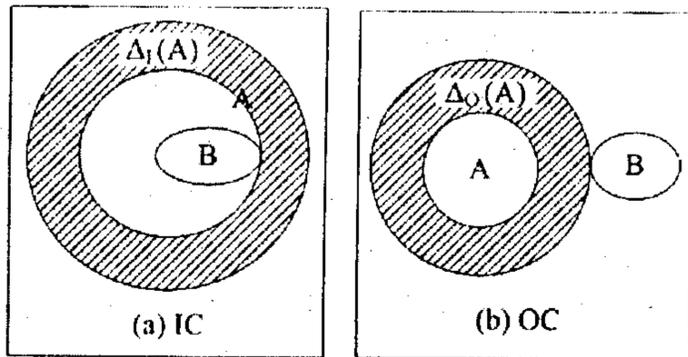


图3 IC和OC

3 度量参数在拓扑查询中的应用

3.1 两类拓扑查询中度量参数的应用

空间数据库中常见的拓扑查询有两类:一类是找出所有与给定对象存在拓扑关系R的对象,另一类是对象A和B之间存在什么样的拓扑关系^[9]。

对于第一类拓扑查询,拓扑关系是已知的,查询结果是一个对象或一个对象集。因此,要针对给定对象和目标对象的不同情况来计算不同的度量参数。例如,已知拓扑关系为相离,如果给定对象A和目标对象B都不含孔,那就只要计算OC,通过OC就能区别图1中的两种情形。为了使查询结果更接近用户的要求,可以通过度量参数对查询结果进行筛选和排序。例如,选出与对象A相离的对象中 $OC > 1.5$ 的对象并按升序排列。

对于第二类拓扑查询,对象A、B是已知的,查询结果是它们之间的拓扑关系。因此,要针对不同的拓扑关系来计算不同的度量参数。例如,已知对象A有一个孔,对象B没有孔,如果它们的拓扑关系为相接,那么先计算HIC以区别图2中的两种情况,再计算AS和HAS以区别交集为点、线、面的情况以及交集的大小。通过计算度量参数可以区分上文定义的细化拓扑关系,使面对象间空间关系的描述更加准确。

3.2 应用举例

数据库采用 Oracle10g,空间数据通过 Oracle Spatial 进行存储和管理,使用 Oracle 提供的空间拓扑关系查询算子进行拓扑查询,再针对不同情况计算相应的度量参数。

3.2.1 例1

将一幅世界地图通过 Mapinfo 的工具 easyload 导入 Oracle 中,查询中国的邻国,SQL 语句如下:

```
Select b.country
```

```
from world a,world b
```

```
Where a.country = 'China'
```

```
and SDO_RELATE(a.geoloc,b.geoloc,'mask=touch querytype=WINDOW') = 'TRUE'
```

查询结果为俄罗斯、朝鲜等13个国家(实际上是14个,由于地图的误差,巴基斯坦不在其中)。该查询属于第一类拓扑查询,因为所有对象都不含孔,故只需计算AS这一个度量参数。AS越大,说明该国与中国之间的共有边界越多,结果如表1所示(按降序排列)。

表1 中国与其邻国之间的度量参数AS

国家	度量参数 AS
蒙古	0.185
俄罗斯	0.139
印度	0.127
哈萨克斯坦	0.053
缅甸	0.050
尼泊尔	0.035
朝鲜	0.030
越南	0.030
吉尔吉斯斯坦	0.029
不丹	0.018
塔吉克斯坦	0.015
老挝	0.011
阿富汗	0.003

3.2.2 例2

假设有一个居民区(用R表示),内有一座公园(用P表示),居民区周围有6片绿化带(用A~F表示),分别查询这6片绿化带与居民区的拓扑关系,SQL 语句如下:

```
Select SDO_GEOM.RELATE(mdsys.sdo_geometry(2003, NULL, NULL, mdsys.sdo_elem_info_array(1,1003,3,5,2003,3), mdsys.sdo_ordinate_array(20,275,145,195,47,258,117,212)), 'DETERMINE', a.geoloc,0.005) from green_area a where a.name = 'A'
```

该查询属于第二类拓扑查询,所以要根据查询得到的拓扑关系计算相应的度量参数,结果如表2所示。

表2 居民区与绿化带间的拓扑关系和度量参数

绿化带代号	拓扑关系	度量参数			
		AS	OC	HAS	HIC
A	DISJOINT	0	/	0	0.497
B	DISJOINT	0	/	0	0.913
C	TOUCH	0.133	/	0.133	0
D	DISJOINT	0	1.215	0	/
E	DISJOINT	0	1.705	0	/
F	TOUCH	0.385	0	0	/

通过表2可以发现,有4片绿化带与居民区相离,

2 片绿化带与居民区相接,利用度量参数可以区分这些相近的拓扑关系。 $HIC_B > HIC_A > 0$,说明 A 和 B 为内离,且 B 比 A 更远离 P 的边界; $OC_E > OC_D > 1$,说明 D 和 E 为外离,且 E 比 D 更远离 R; $HIC_C = OC_F = 0$ 且 $AS_F > AS_C$,说明 C 为内接 F 为外接且 F 与 R 之间的共有边界更多。绿化带的实际分布如图 4 所示(阴影部分为居民区)。

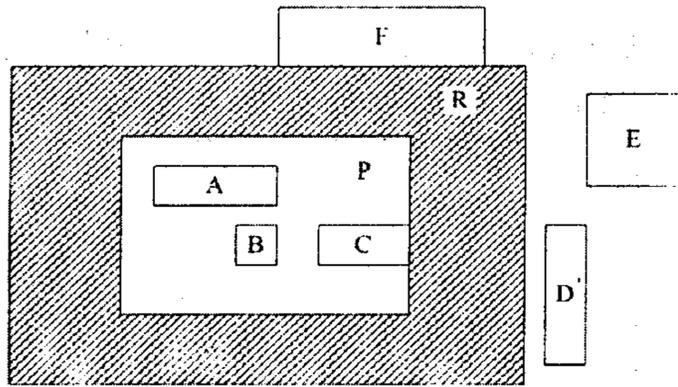


图 4 居民区与绿化带

4 结束语

通过分析 9 交集模型在描述面对象间拓扑关系上的不足以及目前度量参数的不足,定义了三对细化的面面拓扑关系,提出了支持带孔面对象的度量参数,增强了面对象间空间关系的描述,并把度量参数应用在了拓扑查询中。基于文中的思想,还可以对线面对象间空间关系的度量参数进行改进。

值得提出的是,到目前为止所有的度量参数都只支持简单几何对象,支持复杂几何对象的度量参数将是下一步的研究方向。此外,如何利用度量参数定义自然语言的拓扑关系谓词,使拓扑查询变的人性化,也

(上接第 186 页)

就可以很方便地实现对企业信息资源的快速检索,出于安全考虑,可以对文档设置访问权限。

Lucene 并没有规定数据源的格式,而只提供了一个通用的结构(Document 对象)来接受索引的输入,因此输入的数据源可以是数据库,Word 文档,PDF 文档,HTML 文档……只要能够设计相应的解析转换器将数据源构造成 Document 对象即可进行索引,该实例默认支持 *.txt, *.html, *.htm 类型的电子文档的索引和检索,如因实际需要,还可以再开发设计 Word,PDF 等相应文档的解析转换器,将数据源构造成 Document 对象,就可以实现对 Word,PDF 等文档的检索支持。

5 结束语

文中提出了一种解决全文检索的方法,可以应用到搜索引擎、中小企业网站站内检索、个人用户桌面搜

是一个研究热点。

参考文献:

- [1] EGENHOFER M, KUHN W. Interacting with Geographical Information System[C]//In Geographical Information Systems: Principle, Techniques, Management and Applications. London: Taylor & Francis, 1998.
- [2] EGENHOFER M, FRANZOSA R. Point - set Topological Spatial Relations[J]. International Journal of Geographical Information Systems, 1991, 5(2): 161 - 174.
- [3] 钟志农,唐征武,张帆. 一种统一的拓扑关系判断模型[J]. 国防科技大学学报, 2004, 26(5): 57 - 62.
- [4] 邓敏,李成名,刘包文. 利用拓扑和度量相结合的方法描述面目标间的空间关系[J]. 测绘学报, 2002, 32(2): 164 - 169.
- [5] EGENHOFER M, NEDAS K. Splitting Radios: Metric Details of Topological Line - Line Relations[C]//17th International FLAIRS Conference. Miami Beach, FL: [s. n.], 2004.
- [6] EGENHOFER M. Query Processing in Spatial - Query - by - Sketch[J]. Journal of Visual Languages and Computing, 1997, 8(4): 403 - 424.
- [7] EGENHOFER M, RASHID A, SHARIFF B M. Metric Details for Natural - Language Spatial Relations [J]. ACM Transaction on Information Systems, 1998, 16 (4): 295 - 321.
- [8] OpenGIS Consortium Inc. OpenGIS Simple Features Specification for SQL[S/OL]. Revision 1. 1. 1999 - 05 - 05. Http://www.opengis.org.
- [9] SHEKHAR S, CHAWLA S. 空间数据库[M]. 谢昆青, 马修军, 杨冬青译. 北京: 机械工业出版社, 2004: 34 - 36.

索引建立、特定文档检索数据库建立等,从而实现目标文档方便的检索管理,提高检索效率。

参考文献:

- [1] 苏新宁. 信息检索理论与技术[M]. 北京: 科学技术文献出版社, 2004.
- [2] 徐宝文, 张卫丰. 搜索引擎与信息获取技术[M]. 北京: 清华大学出版社, 2003.
- [3] 肖创柏. 基于全文检索技术的商业信函处理系统的设计与实现[J]. 计算机应用研究, 2004(1): 150 - 152.
- [4] 曹元大, 贺海军. 全文检索字索引技术的研究与实现[J]. 计算机工程, 2004, 28 (6): 21 - 23.
- [5] Bookstein A, Swanson D R. A Decision Theoretic Foundation for Indexing[J]. Journal of the American Society for Information Science, 1975(26): 76 - 77.
- [6] Crossman D A, Frieder O. Information Retrieval: Algorithms and Heuristics [M]. Boston: Kluwer Academic Publishers, 1998: 49 - 51.