

空间规则的可视化解释^{*})

郭平 范丽 叶莲

(重庆大学计算机学院 重庆400044)

摘要 随着地理信息系统(GIS)软件的成熟,其空间信息分析、处理和展示能力日趋完善。利用空间数据挖掘等空间信息分析抽取手段进行空间规则发现并对这些规则进行规范化与形式化描述进而利用GIS进行可视化展现已成为一种趋势。本文研究了空间关联规则、分布规则、分类规则的形式化描述,并且给出了这些规则的可视化解释过程。

关键词 地理信息系统(GIS),空间规则,空间规则形式化描述,空间规则可视化

The Visual Explanation of the Spatial Rules

GUO Ping FAN Li YE Lian

(College of Computer Science, Chongqing University, Chongqing 400044)

Abstract As the improvement of the Geographic Information System (GIS) software, the capability in spatial information analysis, handling and display is perfected gradually. It has become one kind of tendency that utilizes the way of analyzing and extracting spatial information, such as spatial data Mining, to discover spatial rules and describe these rules regularly and formally, furthermore utilizes GIS to visual expression. This paper has a research on formal description of spatial association rule, spatial distribution rule and spatial classification rule, and then elaborates the visual explanation process of the spatial rules.

Keywords Geographic Information System (GIS), Spatial rule, Spatial rule formal description, Spatial rule visual expression

1 引言

地理信息系统(GIS)以地理空间数据库为基础,通过可视化方式提供多种空间的和动态的地理信息,为管理和决策提供服务与支持。通俗地讲, GIS 就是整个地球或部分区域的资源、环境在计算机中的缩影。

在GIS中,空间数据可视化更重要的是为人们提供一种空间认知的工具,让人完成最终的知识发现。可视化数据分析技术拓宽了传统的图表功能,使用户对数据的剖析更清楚。当显示空间数据挖掘和知识发现(SDMKD)的结果时,将地图同时显示作为背景,一方面能够显示其知识特征的分布规律,另一方面也可对挖掘出的结果进行可视化解释,从而达到更好的分析效果。

本文主要研究分析空间数据挖掘中获得的空间关联规则、空间分布规则和分类规则的形式化描述及可视化解释。

2 空间规则的描述

GIS的核心是空间数据。目前,绝大多数GIS对空间数据的利用主要是查询、空间分析和简单的统计,人们更迫切需要的是从海量空间数据中挖掘出对决策具有指导意义的知识,为决策支持提供重要依据。因此,空间数据挖掘和知识发现应运而生,它是从空间数据库中提取隐含的、用户感兴趣的空间和非空间的模式和知识的过程。

2.1 从空间数据中可发现的知识类型

从GIS及其应用的角度上看,能从空间数据库中发现的知识(规则)主要有以下几类^[1]:普遍的几何知识,空间关联规

则,空间分布规则,空间分类(聚类)规则,空间特征规则,空间区分规则,空间演变规则,面向对象知识。本文主要探讨空间关联规则、空间分布规则和空间分类规则。

(1)空间关联规则是指空间对象之间的关联性和相关性方面的规则,主要包括相邻、相连、共生、包含等规则。例如,村落与道路相连,道路与河流的交叉处是桥梁等。

(2)空间分布规则是指对象在地理空间的分布规律,包括在垂直方向(高程坐标)、水平方向(地理坐标)的分布规律以及垂直向和水平向的混合分布规律。

(3)空间分类规则是根据对象的空间或非空间特征将对象划分为不同的类别的规则。例如根据人均收入,可把不同的地区分为发达地区、发展地区、贫困地区。

2.2 基本空间关系与空间谓词

空间关系是人们认知和表示空间形态的途径,它也是进行空间规则描述的基础。主要的空间关系包括拓扑关系、方位关系和度量关系三种基本类型。

(1)空间拓扑关系及拓扑关系谓词 拓扑关系是基本的空间关系,1992年Randell, Cui和Cohn基于连通性建立了空间区域的RCC(region connection calculus)理论^[2],得到了8种基本的空间拓扑关系如图1所示。

这8种拓扑关系可以用谓词描述: $disjoint(x, y)$ —— x 和 y 相离, $touch(x, y)$ —— x 和 y 相接, $overlap(x, y)$ —— x 和 y 交叠, $cover(x, y)$ —— x 覆盖 y , $contain(x, y)$ —— x 包含 y , $coveredby(x, y)$ —— x 被 y 覆盖, $in(x, y)$ —— x 包含于 y , $equal(x, y)$ —— x 和 y 相等。

^{*})基金项目:国家自然科学基金项目(编号:50378903)。郭平 副教授,主研方向:AI、GIS;范丽 硕士研究生,主研方向:GIS。

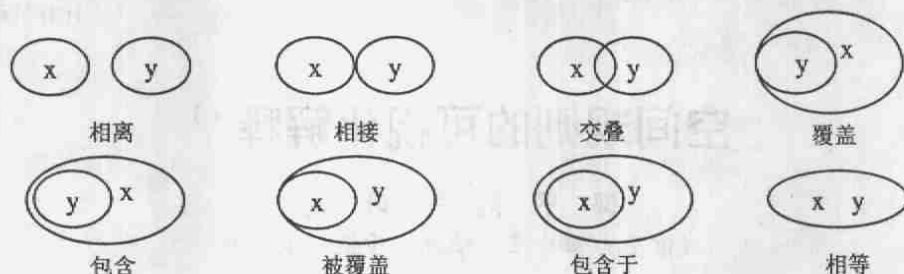


图1 RCC-8中的基本拓扑关系

(2) 空间方位关系及方位关系谓词 空间对象方位关系常用带中心区域的二维坐标系来描述^[3],如图2所示,它将整个空间划分为正东、正南、正西、正北、西北、东北、东南、西南、中心区9大块,可以用如下谓词描述:RE:East-of(x, y)—— x 位于 y 的正东方;RS:South-of(x, y)—— x 位于 y 的正南方;RW:West-of(x, y)—— x 位于 y 的正西方;RN:North-of(x, y)—— x 位于 y 的正北方;NW:NorthWest-of(x, y)—— x 位于 y 的西北方;NE:NorthEast-of(x, y)—— x 位于 y 的东北方;SE:SouthEast-of(x, y)—— x 位于 y 的东南方;SW:SouthWest-of(x, y)—— x 位于 y 的西南方;OC:Is-same(x, y)—— x 与 y 同位。

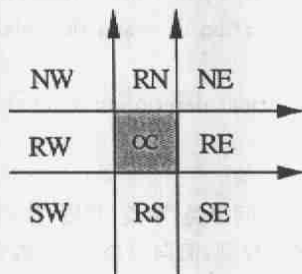


图2 带中心区域的二维坐标系空间

(3) 空间度量关系及距离谓词 空间度量关系是在空间表示中嵌入度量体系来表达对空间对象及它们间关系的度量。距离关系是一种常用的度量关系,本文仅考虑距离这种度量关系。

距离关系由距离系统描述^[6],图3是一个各向同性距离关系结构的例子。假设 y 为参考对象,其距离谓词为:VC:Very-Close(x, y)—— x 离 y 很近;CL:Close-To(x, y)—— x 临近 y ;CM:Commensurate(x, y)—— x 离 y 不远;FR:Far-Away(x, y)—— x 远离 y ;VF:Very-Far(x, y)—— x 离 y 很远。

可以根据不同的需要构造出不同的距离关系结构。

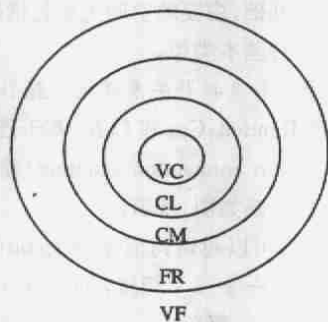


图3 距离结构

2.3 空间规则描述

空间规则是包含空间关系谓词的规则,它由空间谓词、常

量及逻辑运算符($\vee, \wedge, \neg, \rightarrow, \leftrightarrow$)等构成。

(1) 空间关联规则 用于描述空间对象相互之间的关联。根据关联规则的形式描述,空间关联规则一般可表示为^[5]:

$$A \rightarrow B [S\%, C\%]$$

其中, A 和 B 是由空间和非空间谓词构成的表达式。 $S\%$ 表示规则的支持度, $C\%$ 表示规则的置信度,它们分别表示了规则的有效性和确定性。例如:

$$\text{is-a}(X, \text{'school'}) \wedge \text{close-to}(X, \text{'sports-center'}) \rightarrow \text{close-to}(X, \text{'park'}) [0.5\%, 80\%]$$

表示80%靠近体育中心的学校同时也靠近公园,并且在讨论的数据集中有0.5%的数据符合这一规则。

(2) 空间分布规则 用于描述属性与空间对象的关联,即反映所感兴趣的对象与空间位置或高程的关联。空间分布规则本质上也是一种空间关联知识。例如:

$$\text{is-a}(X, \text{'field'}) \wedge \text{wheat-yield}(X, \text{'large'}) \rightarrow \text{south-of}(X, \text{'greatwall'}) \wedge \text{north-of}(X, \text{'qing-huai'})$$

是描述“中国的大部分小麦种植区位于长城以南,秦岭和淮河以北”这条农作物空间分布规则。

(3) 空间分类规则 用于分析空间对象并导出与空间特征有关的分类模式。例如若需要根据家庭平均收入把地区按贫富分类,则要找出决定分类的重要空间和非空间因素。分类标准是基于描述分类对象和位于分类对象临近位置的对象间空间关系的属性和谓词来定的。最简单的形式是通过 IF—THEN 语句表示,形如:

$$\text{IF } X \text{ THEN } Y$$

其中, X 是属性——值对形成规则前件(IF部分)的一个合取项, Y 包含类预测,形成规则后件(THEN部分)。

例如:我们对重庆沙坪坝的3个对象 $O_i (i=1, 2, 3)$ 进行分类。

$$O_1: \text{close-to}(x, \text{重庆大学})$$

$$O_2: \text{close-to}(x, \text{三峡广场})$$

$$O_3: \text{close-to}(x, \text{重庆大学}) \wedge \text{close-to}(x, \text{三峡广场})$$

我们利用缓冲区 A, B, C 来分别表示 O_1, O_2, O_3 邻近对象的属性。缓冲区代表对分类对象的类标签属性有影响的区域^[7]。它既包含用于从专题地图中查找聚合信息的缓冲区,也包含用于决定空间谓词的缓冲区。

这个分类可以用 IF—THEN 规则表示为:

$$\text{IF in}(x, A\text{-buffer}) \text{ THEN close-to}(x, \text{重庆大学})$$

$$\text{IF in}(x, B\text{-buffer}) \text{ THEN close-to}(x, \text{三峡广场})$$

$$\text{IF in}(x, C\text{-buffer}) \text{ THEN close-to}(x, \text{重庆大学}) \wedge \text{close-to}(x, \text{三峡广场})$$

以上规则是说如果对象 x 在缓冲区 A 中,则离重庆大学很近;如果在缓冲区 B 中,则离三峡广场很近;如果在缓冲区 C 中,则离重庆大学和三峡广场都很近。

3 空间规则可视化解释

利用地理信息系统(GIS)的空间信息处理能力实现信息管理系统可视化已成为一种趋势,这里我们将基于GIS给出空间关联规则、分布规则和分类规则的可视化解释过程,并通过实例说明具体的步骤。下面的例子中用到的空间数据库 world.sdb 和属性数据库 world.sdd 均是 SuperMap II 自带的数据库。

3.1 空间关联规则解释

对空间关联规则的解释可以利用GIS软件的图层风格设置和叠置(或叠加)分析功能来完成。其处理过程是:

(1)根据规则前件所含的属性谓词和空间谓词查询空间数据库,并将所得的记录生成一个数据集。根据规则后件所含的属性谓词和空间谓词查询空间数据库,将所得记录生成第二个数据集。

(2)对两个数据集所表示的图层分别进行风格设置(着以不同的颜色或晕线)。

(3)将两个数据集进行叠置分析(Clip, Erase, Union, Intersect, Identity)或将两个图层进行叠加(overlay)^[8],便可获得规则前件和后件的关联性。

例如,我们从空间数据库 world.sdb 中发现有这样一条空间关联规则:

$is_a(X, 'country') \wedge catching_fish(X, 'true') \rightarrow close_to_sea(X, 'true') (19.4\%, 56\%)$

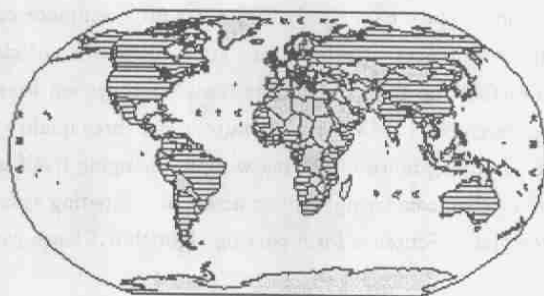


图4 “捕鱼国家”图层显示

(3)我们把“临海的国家”图层以黑色垂直晕线(■)显示,并叠加在“捕鱼国家”和“World”之上,可以看出,在图5中水平晕线区域没有了,因为其完全被垂直晕线覆盖。从这个显示可以肯定我们的结论:捕鱼的国家都是临海的国家。

通过以上步骤,我们将规则以可视的方式表示在“World”之上,并且验证了存在临海但不捕鱼的国家。

3.2 空间分布规则解释

对空间分布规则的可视化解释主要是通过制作各种专题图来完成的。制作专题地图是渲染GIS图层的一种方式,但是它更能表现属性数据与地理空间的关系,并以不同的颜色、面积、图形等特性直观地表现出来,通过风格渲染来达到可视、直观的效果。

根据空间分布规则来制作专题图可分为以下几个步骤:

(1)根据规则前件的属性谓词确定专题图类型(单值专题图、范围分段图、点密度图、统计图、等级符号图、标签专题图等)^[9]。

(2)从空间目标的属性数据库中选取与规则前件相关联的属性字段。

(3)将属性的值划分为不同的取值区间,并确定不同取值

对它进行可视化解释和分析计算的过程是:

(1)根据规则前件在 world.sdb 中查询满足 $catching_fish=true$ 条件的国家,生成一个名为“捕鱼的国家”的数据集。根据规则后件在 world.sdb 中查询满足 $close_to_sea=true$ 条件的国家,生成一个名为“临海的国家”的数据集。图3以高亮方式在地图窗口显示了生成的数据集,并可从数据集浏览窗口查看记录数。



图3 浏览数据集窗口

(2)我们通过图例的风格设置把“捕鱼的国家”图层以黑色水平晕线(■)显示并放在 World 图层之上(如图4),可以看出这些国家基本上是临海的国家。



图5 关联规则分析参考图

区间对应的符号、图形、标签和颜色。

(4)根据专题图显示解释空间目标属性的地理空间分布。

例如,我们从空间数据库中查询1994年世界人口在空间位置上的分布规则:

$is_a(X, 'country') \wedge population(X, '>6 \times 10^8') \rightarrow in(X, 'Asia') \wedge South_of(X, 'Asia')$



图6 分布规则参考图(点密度专题图)

(下转第186页)

实验中使用了尺寸各不相同的24位 RGB 格式图像,图像类别涉及植物、动物、实物。结果表明,使用传统的单一特征进行检索,不能兼顾图像的整体内容,检索效果不够理想。本文提到的方法不仅解决了这一问题,而且在形状特征提取时,由于利用了特征点信息,从而不仅克服了边界点信息的局限性而且也减少了计算量。对于采用的图像分块策略,使得此方法应用于图像的对象或区域查询中也得到了很好的效果。图3是实验得到的一个检索结果,因为综合了图像的颜色和形状信息,所以与单一的特征信息(颜色相交直方图)检索相比较,具有较高的准确度。

总结 本文根据传统单一图像特征检索的不足,结合颜色特征与形状特征的优越性,提出了一种新的基于分块颜色和形状信息的图像检索方法。主要以图像的颜色平均直方图作为颜色信息的描述,以特征点为依据利用边缘类型直方图作为形状信息的描述。实验结果表明,该方法具有明显的优越性,在应用范围上,不仅适用于一般的图像查询,而且也适用于基于对象或区域的图像查询;在检索效率上,与常规的方法

相比有了明显的提高。本文接下来要做的工作是对算法进一步改进,并利用存储结构更好地提高检索效率。

参考文献

- 1 Rui Y, Huang T S, Chang S-F. Image retrieval: Past present and future. *Journal of Visual Communication and Image Representation*, 1999, 10: 1~23
- 2 章毓晋. 基于内容的视觉信息检索. 北京: 科学出版社, 2003
- 3 Gouet V, Boujemaa N. Object-based queries using color points of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL/CVPR 2001)*, Kauai, Hawaii, USA, 2001. 30~36
- 4 曹莉花. 用颜色特征进行基于形状的检索. *计算机工程与应用*, 1999, 12
- 5 Montesinos P, Gouet V, Deriche R, Pelé D. Matching color uncalibrated images using differential invariants. *Revue Image and Vision Computing*, 2000, 18(9): 659~671
- 6 Prasad B G, Gupta S K, Biswas K K. Color and Shape Index for Region-Based Image Retrieval, *IWVF4, LNCS 2059*, 2001. 716~725
- 7 章毓晋. 图像工程 上册: 图像处理和分析. 北京: 清华大学出版社, 1999
- 8 贾云得. 机器视觉. 北京: 科学出版社, 2000

(上接第171页)

首先可以确定用点密度专题图(点密度专题图是用不同颜色的点来表示目标对象的分布范围、数量特征和分布密度)比较合适。然后我们在属性数据库 world.sdd 中找到与谓词 Population 相关的属性字段 Pop_1994(1994年各国人口数), 并确定每个点代表的人口数为6亿人, 再制作基于 Pop_1994 字段的点密度专题图, 如图6。图6的显示与实际情况是相符的。

3.3 空间分类规则解释

空间分类规则可以通过规则前件的分类属性制作单值专题图(将具有不同属性值的空间对象以不同的颜色或晕线显示)或分段专题图(将属性值处于不同数值范围的空间对象以不同颜色或晕线显示)来对规则进行可视化解释。

例如, 以国家为单位给出人均年收入增长率(pop-grw-rt)分类规则如下:

```
if pop-grw-rt <="1%" then develop-speed = "no"
if pop-grw-rt > "1%" and pop-grw-rt <="2%" then develop-speed = "slow"
if pop-grw-rt > "2%" and pop-grw-rt <="4%" then develop-speed = "middle"
if pop-grw-rt > "4%" and pop-grw-rt <="5%" then develop-speed = "fast"
if pop-grw-rt > "5%" then develop-speed = "super"
```



图7 分类规则参考图(分段专题图)

图7通过对每一个分类赋予不同的填充晕线(左斜线、右

斜线、点、垂直线、水平线), 展示了1994年世界各国人均收入增长率的一个可视化分类(数据来自 world.sdb 和 world.sdd)。图的左边是不同填充晕线表示的分类标准。

结束语 本文以 GIS 为背景, 以提高 GIS 数据分析的智能化和可视化为目标, 对空间数据挖掘和知识发现理论中的部分规则进行了形式化的抽象, 并基于 GIS 软件 Supermap II 进行了可视化的分析。

基于一定 GIS 软件进行空间规则分析虽然很方便, 但不是完全智能的, 其应用还主要停留在数据库、空间叠加分析上, 缺乏知识处理能力和推断能力。智能 GIS 的发展, 需要大家共同努力: 通过各种空间分析要求, 促进软件智能化的集成; 通过更加智能化的软件, 增强空间分析能力。这实际上是基于知识的专家系统在 GIS 中的应用。

参考文献

- 1 邱凯昌著. 空间数据挖掘和知识发现. 武汉大学出版社, 2001. 16~17, 38~39
- 2 Randell D A, Cui Z, Cohn A G. A Spatial Logic Based on Regions and Connection. In: *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, 1992. 165~176
- 3 刘亚彬, 刘大有, 等. 定性空间表示与定性空间的研究与发展. *计算机科学*, 2003, 30(3): 65~67
- 4 Ester M, Kriegel H-P, Sander J. Knowledge Discovery in Spatial Databases. *Invited Paper at 23rd German Conf. on Artificial Intelligence (KI '99)*, Bonn, Germany, 1999
- 5 Han Jiawei, Kamber M 著, 范明, 孟小峰, 等译. 数据挖掘概念与技术. 北京机械工业出版社, 2001
- 6 Hern'andez D, Clementini E, Felice P D. Qualitative distances. In: Kuhn W, Frank A, eds. *Spatial Information Theory: a theoretical basis for GIS*, number 988 in LNCS, Berlin, Springer Verlag, 1995. 45~58
- 7 石云, 孙玉芳, 左春. 基于 Rough Set 的空间数据分类方法. *软件学报*, 2000, 11(5): 673~678
- 8 滕寿威, 陈俊华. SuperMap II 的特点和应用. <http://www.supermap.com>