

文章编号: 1007-4619(2007)02-0152-07

关于线状地理特征空间关系的自然语言描述的形式化表达

许 璠

(中国科学院 地理科学与资源研究所, 北京 100101)

摘 要: 目前的地理信息系统都使用定量的方法存储和查询空间信息,不能反映人们的定性思维方式,因此它表达地理空间信息的能力受到限制。在关于空间关系的自然语言描述的调查的基础上,本文定义了一组能反映线状物体空间关系的度量指标,结合反映拓扑关系的定量指标,使用决策树的数据挖掘算法,对调查结果进行分析,形式化了描述空间关系的自然语言词汇,建立了模糊的自然语言表达和图形的几何特征之间的联系。自然语言形式化后得到的规则可用于空间关系的自然语言查询系统。

关键词: 线状地理特征; 拓扑关系; 度量指标; 自然语言; 形式化; 决策树

中图分类号: P208 文献标识码: A

Fomalizing the Natural-language Descriptions about the Spatial Relations between Linear Geographic Objects

XU Jun

(Institute of Geographic Science and Natural Resources Research, CAS, Beijing 100101, China)

Abstract People usually use qualitative terms to express spatial relations, while current geographic information systems (GIS) all use quantitative approaches to store spatial information. The abilities of current GIS to represent and query spatial information about geographic space are limited. If GIS are to be made easier to use, it must be made more intuitive, so that it will help people focus on the work they want to do without thinking about a computer program or application. In order to incorporate the concepts and methods people used to infer information about geographic space into GIS, research on the formal model of common sense geography becomes increasingly important. This paper worked on how to formalize the qualitative natural-language descriptions of spatial relations between linear geographic objects. Based on the result of a human subject test about natural-language description of spatial relations, this paper defines some quantitative topological and metric indices which are related with the natural-language spatial relation terms, and uses these indices to formalize the natural-language representation with a decision tree algorithm, thus bridging the gap between natural-language terms and the computational model of spatial relations. This method provided a way of completely understanding the relationship between the ambiguous natural-language representations and the geometric spatial relations of geographic objects. The rules extracted from the trees can be used in natural-language spatial query systems.

Key words linear geographic objects; topological relations; metric indices; natural language; formalization; decision tree

收稿日期: 2005-11-23 修订日期: 2006-04-20

作者简介: 许璠 (1972—), 女, 博士, 毕业于纽约州立大学布法罗分校。主要从事地理信息系统和地理认知的研究。Email: xujun@reis.ac.cn.

1 引言

空间关系在地理信息系统中占有重要地位,空间数据分析的一个基本问题就是如何正式理解空间物体的几何关系^[1]。当人们理解和描述空间关系的时候,通常采用定性的方法。目前的地理信息系统都采用定量的方法存储和查询空间信息,因而它表达地理空间信息的能力受到一定限制。要使 GIS 更容易使用, GIS 的操作必须变得更加直观化,使得用户以一种能反映他们的自然思维过程的方式来操作系统。将人们对于时空推理的概念和方法,人类的思维模式,尤其是人类怎样处理地理空间概念和物体融合在地理信息系统中,是未来地理信息系统应解决的问题^[2]。因此对于地理空间的普遍认识和空间关系正式模型的研究变得日益重要。

空间关系可分为三种类型:拓扑关系,度量关系及整体和局部的次序关系^[3]。各种类型的空间关系都已有量化的计算模型。在拓扑关系方面,根据研究本体的不同,有基于区域的 RCC 模型^[4-8]和基于点集的 n -交集模型^[1,9]。在表达方位的空间关系方面,有 Frank 的四方位和八方位模型^[10], Hernández 表达相对位置的八角模型^[11], Freksa 等人的格网模型^[12,13]。在远近距离方面,有各种基于模糊集的表达方式^[14-16]。关于地理空间的认知和语言学方面,人们也做了一定研究^[17-19]。此外,Mark 等人通过调查的方式研究了线状物体与面状物体之间空间关系的自然语言理解^[20,21],并且比较了英语和西班牙语对于空间关系理解的差异^[22]。但是对于如何建立地理认知与正式模型的关系,以及如何将人们对于空间关系的理解用于直观化的 GIS 设计则研究的非常少,只有 Shariff 等人作了研究。在 Mark 等人调查研究的基础上,Shariff 等定义了一组度量指标,用之与 9-交集拓扑关系模型相结合来形式化自然语言描述的线状地理特征与面状地理特征之间的空间关系,并实现了用类似于自然语言的 SQL 语句查询空间数据库^[23,24]。

要将人类关于空间关系的语言和认知用于 GIS 关键是要建立模糊的、定性的语言或认知与量化的计算机模型之间的联系,即解决自然语言的形式化问题。本文作者定义了一组可以用来表示线状物体空间拓扑和度量关系的定量指标。在社会调查了解英语自然语言对两个线状物体特征空间关系的描述的基础上,使用量化的度量指标以及拓扑

关系模型,应用数据挖掘的算法,形式化了描述线状物体空间关系的英语自然语言词汇,得到可以表达定性的语言描述的量化计算规则,可用于空间关系的自然语言查询。

2 量化指标

本文选取了一组几何指标,包括反映拓扑关系的指标和反映度量关系的指标,用于形式化线状物体空间关系的自然语言描述。

2.1 拓扑关系模型

本文用 9-交集模型来表达两条线状地理特征的拓扑关系。9-交集模型是一个表示二元的空间拓扑关系的模型^[1],它采用点作为原始的空间实体,地理目标被定义为点的集合^[9]。给定一个区域 Y ,与之相关的有 3 个点集:内部点集 (Y°)、外部点集 (Y^-)和边界点集 (∂Y)。两个物体 A 和 B 之间的二元拓扑关系 R 是基于 A 的 3 个点集和 B 的 3 个点集的有序交集,可表示为一个 3×3 的矩阵。

$$R(A, B) = \begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}$$

如果用 0 和 1 分别表示空集和非空集,理论上 9-交集模型可以区分 512 种不同的关系,而实际上两个物体之间的拓扑关系不可能有这么多种。由于两个线状物体的外部点集的交集总是非空的,因此本文只用了 9-交集模型中其余的 8 个值作为表示拓扑关系的定量指标,分别被记为 $\{R_{11}, R_{12}, R_{13}, R_{21}, R_{22}, R_{23}, R_{31}, R_{32}\}$ 。

2.2 度量指标

单纯的拓扑关系不足以反映所有的空间关系信息,在有些情况下,度量关系甚至比拓扑关系更能影响人们的自然语言描述^[25]。效仿 Shariff 等人采用的表示一个线状物体和一个面状物体的空间关系的度量指标^[25],本文定义了一些表达两条线状物体空间关系的度量指标。在以下度量指标的定义中, A 和 B 代表图中的两个线状物体。

2.2.1 角度

两个线状物体的相对走向可以用它们之间的角度来表示。两个不规则的线状物体之间的角度有两个,一个是局部角度 (LA),即在两个物体相交之处所形成的锐角(图 1(a)),另一个是全局角度 (GA),

即两条线状物体的大体走向之间的角度, 可以通过物体的最小外接矩形来获取。用最小外接矩形的长轴代表物体的走向, 两个物体之间的全局角度就是两个最小外接矩形的长轴相交的形成的锐角(图 1 (b))。局部角度和全局角度的范围都是 $0^{\circ}-90^{\circ}$ 。

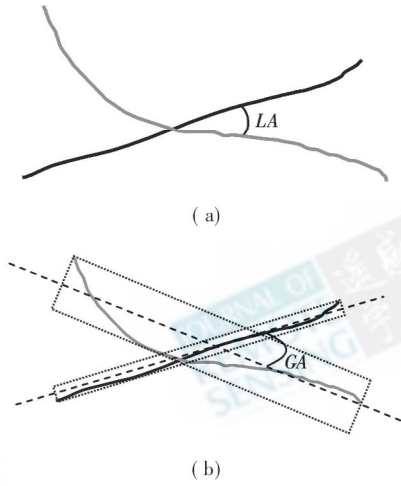


图 1 空间关系的度量指标: 角度
(a) 局部角度; (b) 全局角度

Fig 1 Metric indices of spatial relations - angles
(a) Local angle (b) Global angle

2.2.2 分裂度

分裂度表示 2 条线状物体如何相互切割。本文定义了 3 种分裂度:

- 交叉分裂度 (IS)描述两条线状物体的相交点是如何分割两个物体的。每个物体有一个交叉分裂度, 因此共有两个交叉分裂度指标。由于本文不考虑线状物体的方向性, 交叉分裂度指标总是被分割成的较短的一段物体的长度占物体总长度的比例, 因此它的范围是 $0-50\%$ (图 2 (a))。如果两个物体没有相交, 或者交点多于 1 个, 则没有交叉分裂度值, 用 -9999 表示。

- 重叠分裂度 (AS)是两条线状物重合的部分占物体总长度的比例, 共有两个重叠分裂度指标 (图 2 (b)), 它们的范围都是 $0-100\%$ 。如果两个物体只相交于一点, 重叠分裂度指标为 0 而如果两个物体不相交, 则没有重叠分裂度值, 记为 -9999。

- 内部穿越分裂度 (ITS)描述两个以上的交点如何分割两条线状物体, 是两个最靠近端点的相交点之间的物体长度占物体总长度的比例, 一共有两个内部穿越分裂度指标, 它们的范围都是 $0-100\%$ (图 2 (c))。如果两个物体只相交于一点, 内部穿越分裂度指标为 0 如果两个物体不相交, 则没有内

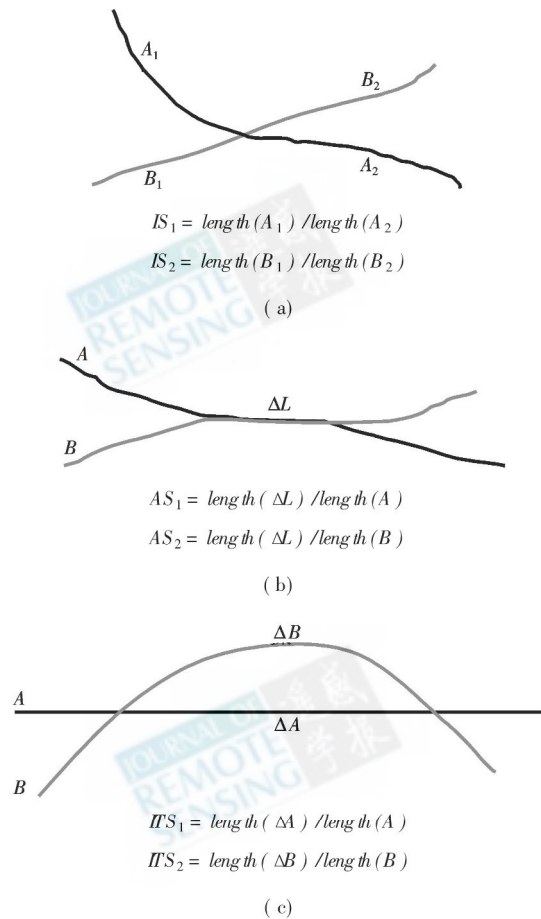


图 2 空间关系的度量指标: 分裂度
(a) 交叉分裂度; (b) 重叠分裂度; (c) 内部穿越分裂度
Fig 2 Metric indices of spatial relations - splitting ratios
(a) Intersect splitting (b) A longness splitting
(c) Interior traversal splitting

部穿越分裂度值, 记为 -9999。

2.2.3 距离

在定义与距离有关的度量指标之前, 必须定义两条线状物体之间的最小距离和最大距离。如图 3 所示, 每条线状物体的任何一点都有到另外一个物体的最短距离, 两条线之间的最小距离 (SD) 就是所有最短距离中最小的一个。每条线状物体的所有点

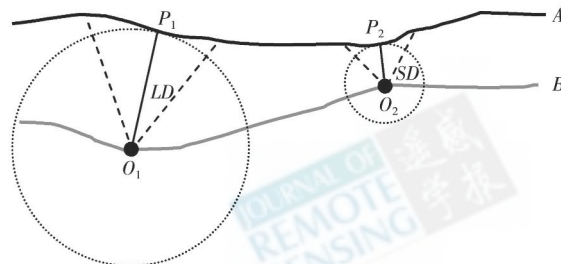


图 3 空间关系的度量指标: 距离
Fig 3 Metric indices of spatial relations - distance measures

到对方的最短距离中最大的一个是此物体到对方的最大距离, 而两条线之间的最大距离 (LD) 是两个最大距离中较小的一个。

基于最小距离和最大距离的定义, 关于距离的度量指标有:

• 最小距离-长度比值 (SDLR) 是两个物体间的最小距离与物体总长度的比值。最小距离-长度比值有两个:

$$SDLR1 = SD / length(A)$$

$$SDLR2 = SD / length(B)$$

• 最大距离-长度比值 (LDLR) 是两个物体间的最大距离与物体总长度的比值。最大距离-长度比值也有两个:

$$LDLR1 = LD / length(A)$$

$$LDLR2 = LD / length(B)$$

• 最小距离-最大距离比值 (SLDR) 是最小距离和最大距离的比值:

$$SLDR = SD / LD$$

2.2.4 重合度

两个线状物体的最小外接矩形的重叠部分的大小, 不仅反映两个物体的临近程度, 也反映了物体的弯曲程度。在图 4 中, 两个最小外接矩形的重叠部分用阴影表示, 重合度指标 (OR) 是阴影部分面积与最小外接矩形的总面积的比值。重合度指标也有两个:

$$OR1 = area(C) / area(A)$$

$$OR2 = area(C) / area(B)$$

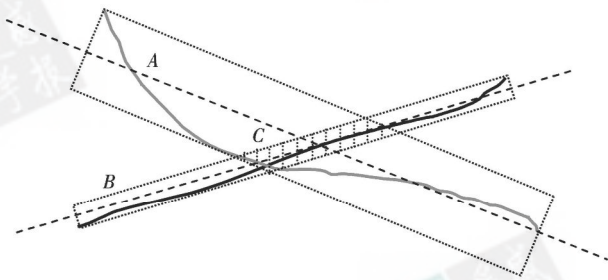


图 4 空间关系的度量指标: 重合度

Fig 4 Metric indices of spatial relations: overlap ratios

3 线状地理特征空间关系自然语言理解的调查

本文采用一个人机交互的调查问卷来获取几何形态如何影响人们对两个线状地理特征之间的空间关系的英语自然语言描述。一系列地图, 每幅地图显示两条线状地理特征, 依次显示在计算机屏幕上。同时, 一句描述这两个地理特征的空间关系的话也显示在屏幕上。被调查者要判断这句话表达的空间关系是否正确, 然后从 7 个选项中选出答案, 它们是: 坚决同意, 同意, 大致同意, 中立, 大致不同意, 不同意和坚决不同意。描述语句的句法结构是“主体 + 空间关系谓词 + 谓体”的形式, 其中主体和谓体分别为图中的两个线状物体, 空间关系谓词是表达它们之间空间关系的一个英语自然语言词汇^[25], 例如“ I-29 crosses I-290 ”“ Colorado River goes along I-15 ”。调查中共使用了 10 个空间关系谓词, 每个空间关系谓词被用来描述 114 幅地图, 其中每幅地图中的线状地物在长度、距离、角度等几何特征上各有不同, 所表示的空间关系, 有的是典型的, 可以用某个空间关系谓词来表示的; 有的却不那么典型; 而有的则不适合用这个空间关系谓词, 因此每个问题所得到的被调查者的同意程度也不同。例如, 同样用“ cross”这个词来描述空间关系, 图 5(a) 中的例子得到大部分被调查者的完全同意, 图 5(b) 得到被调查者的同意程度大部分介于大致同意和大致不同意之间, 而图 5(c) 得到大部分被调查者的答案为完全不同意。所有的问题被分为 3 组, 这样基本保证被调查者能在一个小时之内完成一组问题。

Egenhofer 和 Herring 用 9-交集模型描述了两个点、线和面之间的拓扑关系, 两条线状地物之间的空间关系最多有 57 种, 其中的 33 种可存在于两个简单的无分支的线状地物之间, 另外 24 种只能存在于复杂的有分支的线状地物之间^[11]。即使是两条简

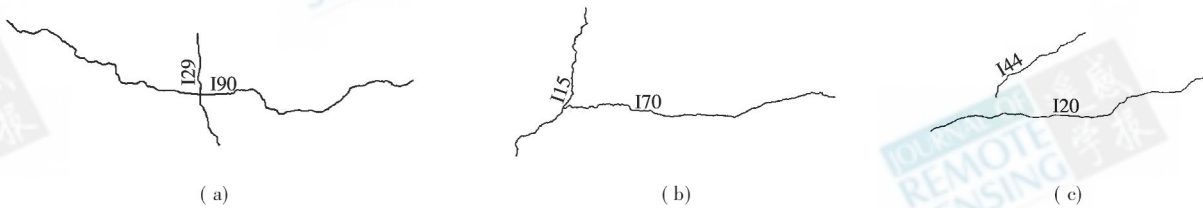


图 5 用于调查的 3 个实例

Fig 5 Three cases used in the human-subject test

单的线状地物之间的空间关系,也不是全部都能用一个空间关系谓词就能描述清楚的,有的必须用两个以上的词汇才能表达它们之间的空间关系。在本研究的调查中只选取了那些用一个空间关系谓词就可以表达空间关系的简单的线状地理特征。

本次调查在美国纽约州立大学布法罗分校完成,所有调查对象从在校的学生和工作人员中征集。在校园中发出广告,采取自愿参加的原则。共有 108 人参加了本次调查,每组 36 人,即对于每个问题可以得到 36 个回答。参加者全部为校内的教职工和学生,分别来自 31 个科系,男性 55 人,女性 53 人,以英语为母语的被测者最多,共有 44 人,其他参加者分别来自另外 18 种语言的国家和地区。其中 20 岁以下的有 18 人,20—25 岁年龄段的有 40 人,25—35 岁年龄段的有 36 人,35 岁以上的有 14 人。

对调查结果的分析表明,物体间的拓扑关系和度量关系影响了人们对空间关系描述词汇的选择,而且对于不同的词汇,拓扑关系和度量关系对其影响程度也不相同,有的词汇受拓扑关系影响较大,有的词汇受度量关系影响较大^[25]。受样本数量的限制,不同人群的比较结果除在个别情况下,未发现显著差异。究竟不同人群对于空间关系的理解是否有显著差异,需要进一步研究。受调查范围的限制,本调查结果在一定程度上只代表一定范围内、受过高等教育的人群对于空间关系的理解。

4 用定量指标形式化关于空间关系的自然语言描述

对于调查中使用的每对线状地理特征,都可以通过计算得出以上定义的 15 个度量指标和 8 个表示拓扑关系的指标,然而只需要其中的一部分指标就可以判断物体之间的空间关系是否属于某一个自然语言空间关系谓词描述的范畴,而且判断不同的空间关系谓词所需的指标是不同的。本文使用数据挖掘中的决策树算法来提取判断空间关系谓词所需的定量指标及它们的取值范围。

4.1 ID3 决策树算法

决策树算法最初由 Quinlan 提出^[26],是一种广泛使用的数据挖掘算法和机器学习的算法。这种算法通过自上而下地划分训练数据,将习得的训练集函数表示成树结构。最初将所有训练数据赋予根结点,然后选择一个属性,根据可能得到的属性值将根

分枝,这样将训练集分为子集,每个分支包含一个子集。对每个分支重复这个过程,直到一个结点下的所有训练数据都属于同一类或者没有剩余属性可划分。如果没有属性可分,则将结点中大多数数据所属的类赋给这个结点^[27, 28]。问题是如何选择分裂每个结点的属性。决策树中的 ID3 算法选择分割后信息增益最大的属性。一个结点的信息量可以用信息熵表示:

$$Entropy(S) \equiv \sum_{i=1}^c - p_i \log_2 p_i$$

式中, S 是训练数据集, p_i 是 S 中属于类别 i 的数据的比例。根据某个属性分割数据后所获得的信息增益为:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

式中, $value(A)$ 是属性 A 的所有可能的值, S_v 是 S 中属性 A 的值为 v 的子集^[27, 28]。

4.2 连续数据的离散化

ID3 算法只适用于所有属性都是基于种类字段的情况,而本文所用的度量指标都是连续的数值型的。在 ID3 算法基础上改进的 C4.5 算法,可以通过自动离散化的方式处理连续数据,并且可以处理缺失数据、噪声和修剪决策树^[29]。

C4.5 算法通过定义一个新的动态不连续属性,将连续的属性值离散化为不连续的间隔序列。它是对连续数据的二向分裂。假设属性 A 的值为连续数据,算法产生一个动态的布尔属性 A_c 。如果 $A < c$ 则属性 $A_c = 1$ 否则属性 $A_c = 0$ 。而这个选择的临界值 c 必须能产生最大的信息增益。将训练样本根据属性 A 的值排序,任何属于不同类别的相邻样本的属性 A 值的均值都可能是临界值, c 值是其中能产生最大信息增益的值。

4.3 用决策树算法提取自然语言词汇的定量表示规则

本文使用数据挖掘软件包 WEKA (Waikato Environment for Knowledge Analysis)^[28] 中的决策树算法 J4.8 来提取影响空间关系的自然语言描述的几何因子。J4.8 是 WEKA 软件包中相当于 C4.5 的决策树算法,能够处理连续数值型的属性。

如前所述,每一个自然语言的空间关系谓词都被用来描述 114 幅地图中的空间关系,并且每个描述都得到 36 个被调查者的评价,所以每个词汇可以

有 4104 个训练数据。由于有些被调查者对于有些问题没有给出答案, 实际训练数据略少于这个数。在调查中, 被调查者可以从 7 个选项中选择答案, 而实际应用中, 只需查询某种情况是否可以用某个空间谓词来描述, 因此只需“是”与“否”两种类型。调查的结果被重划分为两种类型, 其中“坚决同意”和“同意”被分为“是”这一类, 其他结果被分为“否”这一类。

用 WEKA 中的 J4.8 算法对训练数据分类, 得到能划分某个空间关系谓词的最佳属性组合, 形成这个空间关系谓词的树结构, 并且可以从中提取判断某种情况是否可以用这个空间谓词描述的规则。分类的结果用 10 倍交叉验证的方法来检验。即每次用 9/10 的数据来分类, 用剩下的 1/10 的数据来检验分类精度, 将这个过程重复 10 次, 使得每个数据都恰好有一次被用于检验, 这样, 将 10 次分类结果的精度平均, 得到总体精度。

以下列举一些英语中常用空间关系谓词的分类结果和精度评价。

(1) *Cross*

以下三条是判断空间关系谓词“*cross*”的规则:

$$\bullet R_{11} = 1 \wedge S_2 < 6.35 \wedge S_2 \leq 12 \wedge GA < 39.35 \wedge ITS_1 \leq 0 \Rightarrow cross$$

$$\bullet R_{11} = 1 \wedge S_2 < 6.35 \wedge S_2 \leq 12 \wedge GA > 56.9 \Rightarrow cross$$

$$\bullet R_{11} = 1 \wedge S_2 < 6.35 \wedge S_2 > 12 \Rightarrow cross$$

这些规则表明拓扑关系是决定两条线状物体是否相交的重要因素。两条线状物体有内部点的交集是它们相交的必要条件, 在此前提下, 分裂度指标和角度也是判断它们是否相交的因子。关于“*cross*”的 4078 个样本中, 有 3288 个被正确分类, 分类的总体精度为 80.63%。

(2) *is parallel to*

空间关系谓词“*is parallel to*”的判断规则比较简单, 只有一条。它的总体分类精度为 78.96%。

$$\bullet S_{LDR} > 25.410617 \Rightarrow is\ parallel\ to$$

如果两条线状物体之间的最小距离和最大距离的比值超过 25%, 则两个物体相互平行。受到调查中所用的图形的限制, 这条规则并不完善。调查中所选的都是在整个图层中比较接近的两条线状地物, 此时这条规则是适用的。然而当两条线状物体之间的距离非常大时, 则不论它们的形状和角度如何, 这个条件总是能够满足的。因此在实际应用中, 必须给这条规则加上距离因子的限制。在今后设计类似的调查时, 应考虑到从全图范围内选择目标。

(3) *is perpendicular to*

空间关系谓词“*is perpendicular to*”的总体分类精度是 72.77%。它的判断规则如下:

$$\bullet GA > 53.32 \vee R_1 > 0 \wedge S_{DLR_1} \leq 1.15 \wedge GA < 78.28 \wedge S_{DLR_1} \leq 77.54 \wedge ITS_1 \leq 18.66 \Rightarrow is\ perpendicular\ to$$

$$\bullet GA > 78.28 \wedge R_1 > 0 \wedge S_{DLR_1} \leq 1.15 \Rightarrow is\ perpendicular\ to$$

从规则可以看出, 在两条线状物体之间的角度足够大并且它们之间的距离比较接近的情况下, 人们认为它们相互垂直。

(4) *meet*

空间关系谓词“*meet*”的总体分类精度是 80%。它的判断规则如下:

$$\bullet S_{DLR_2} \leq 0.139365 \Rightarrow meet$$

$$\bullet S_{DLR_2} > 0.139365 \Rightarrow meet$$

两条线状物体不一定要有内部或边界点的交集, 只要在它们非常接近的情况下, 就可以认为他们相遇。

5 结 论

在通过调查发现影响人们描述空间关系的几何因子的基础上, 本文定义了几种反映线状物体空间关系的度量指标, 并结合 9-交集拓扑关系模型, 用决策树算法从这些定量的指标中挖掘出有用的信息, 用以形式化描述空间关系的自然语言词汇。本方法与 Shariff 等人^[23, 24]使用的方法的不同之处, 在于他们只对度量指标进行聚类分析, 以拓扑关系为主要判别因素, 度量指标作为对拓扑关系的补充。实际应用中, 首先使用一个查找表找到某一词汇对应的各种拓扑关系, 然后根据不同拓扑关系所对应的度量指标判别条件分类。本文是使用数据挖掘的算法, 将拓扑关系和度量关系指标都用来训练, 得到包含两种关系指标的规则, 建立了模糊的自然语言表达和定量的几何特征指标之间的桥梁, 使之便于将规则直接应用于空间关系的自然语言查询系统。

参 考 文 献 (References)

- [1] Egenhofer M J, Herring J. Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases[A]. Technical Report[C]. Department of Surveying Engineering, University of Maine, 1994.
- [2] Frank A U, Mark D M. Language Issues for GIS[A]. Maguire D J, Goodchild M F, Rhind D W. Geographical Information

- Systems Principles and Applications [C]. London: Longmans Publishers, 1991, 147—163
- [3] Egenhofer M J, Franzosa R D. Point-set Topological Spatial Relations [J]. *International Journal of Geographical Information Systems*, 1991, 5(2): 161—174
- [4] Randell D A, Cui Z, Cohn A G. A Spatial Logic Based on Regions and Connection [A]. 3rd International Conference on Knowledge Representation and Reasoning [C]. San Francisco: Morgan Kaufmann, 1992
- [5] Cohn A G, Bennett B, Goodday J *et al*. Qualitative Spatial Representation and Reasoning with Region Connection Calculus [J]. *Geoinformatica*, 1997, 1(3): 215—316
- [6] Cohn A G, Bennett B, Goodday J *et al*. Representing and Reasoning with Qualitative Spatial Relations about Regions [A]. Stock O, Spatial and Temporal Reasoning [C]. Kluwer Academic Publisher, 1997
- [7] Cohn A G. Calculi for Qualitative Spatial Reasoning [A]. Cahnet J, Campbell J A, Pfalzgraf J. Proceedings of the International Conference AISM-3 on Artificial Intelligence and Symbolic Mathematical Computation. Lecture Notes In Computer Science [C], Springer-Verlag, 1996, 1138, 124—143
- [8] Cohn A G. Qualitative Spatial Representation and Reasoning Techniques [A]. Brewka G, Habel C, Nebel B. KI-97: Advances in Artificial Intelligence [C]. Berlin: Springer, 1997
- [9] Egenhofer M J, Franzosa R D. Point-set Topological Spatial Relations [J]. *International Journal of Geographical Information Systems*, 1991, 5(2): 161—174
- [10] Frank A U. Qualitative Spatial Reasoning about Cardinal Directions [A]. Proceedings of AutoCarto 10, Baltimore [C]. MD, 1991
- [11] Hernández D. Relative Representation of Spatial Knowledge: the 2-D Case [A]. Mark D, Frank A. Cognitive and Linguistic Aspects of Geographic Space. Kluwer Academic Publisher, 1991
- [12] Freksa C, Zimmernann K. On the Utilization of Spatial Structures for Cognitively Plausible and Efficient Reasoning [A]. Proceedings of the IEEE International Conference on Systems Man, and Cybernetics [C]. Chicago, October, 1992
- [13] Freksa C. Using Orientation Information for Qualitative Spatial Reasoning [A]. Frank A U, Campari I, Fomenini U. International Conference GIS-Frm Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space [C]. Pisa, Italy: Springer-Verlag, 1992
- [14] Gahegan M. Proximity Operators for Qualitative Spatial Reasoning [A]. Frank A, Kuhn W. COSIT' 95 Proceedings: Spatial Information Theory: A Theoretical Basis for GIS [C]. Berlin: Springer, 1995
- [15] Hernández D, Clementini E, Felice P D. Qualitative Distance [A]. Frank A, Kuhn W. COSIT' 95 Proceedings: Spatial Information Theory: A Theoretical Basis for GIS [C]. Berlin: Springer, 1995
- [16] Yao X. Qualitative Georeferencing and Proximity Modeling for Geospatial Information System [D]. Ph D Dissertation, Department of Geography, State University of New York at Buffalo, 2002
- [17] Montello D R. A New Framework for Understanding the Acquisition of Spatial Knowledge in Large-scale Environments [A]. Egenhofer M J, Gollidge R G. Spatial and Temporal Reasoning in Geographic Information Systems [C]. New York: Oxford University Press, 1998
- [18] Mark D M, Freska C, Hirtle S C, *et al*. Cognitive Models of Geographical Space [J]. *International Journal of Information Science*, 1999, 13(8): 747—774
- [19] Liu Y, Liu X, Wang X, *et al*. Qualitative Spatial Reasoning about Internal Cardinal Direction Relations [A]. Proceedings of International Symposium on Spatio-temporal Modeling: Spatial Reasoning, Spatial Analysis, Data Mining and Data Fusion [C]. August 27—29, Beijing, China, 2005
- [20] Mark D M, Comas D, Egenhofer M J *et al*. Evaluating and Refining Computational Models of Spatial Relations through Cross-linguistic Human-subjects Testing [A]. Frank A, Kuhn W. Spatial Information Theory—A theoretical Basis for GIS [C]. International Conference COSIT', Summering Austria, Lecture Notes in Computer Science, vol 988, Berlin: Springer-Verlag, 1995
- [21] Mark D M, Egenhofer M J. Calibrating the Meanings of Spatial Predicates from Natural Language: Line-region Relations [A]. Proceedings: Spatial Data Handling [C]. 1994
- [22] Mark D M, Egenhofer M J. Topology of Prototypical Spatial Relations between Lines and Regions in English and Spanish [A]. Proceedings: Auto Carto 12 [C]. Charlotte, North Carolina, 1995
- [23] Shariff R, Egenhofer M J, Mark D M. Natural Language Spatial Relations between Linear and Areal Objects: the Topology and Metric of English-language Terms [J]. *International Journal of Geographical Information Science*, 1998, 12(3): 215—246
- [24] Egenhofer M J, Shariff R. Metric Details for Natural Language Spatial Relations [J]. *ACM Transactions on Information Systems*, 1998, 16(4): 295—321
- [25] Xu J. Natural Language Representation and Query of Linear Geographic Objects in GIS [D]. Ph D Dissertation, Department of Geography, State University of New York at Buffalo, 2005
- [26] Quinlan J R. Induction of Decision Trees [J]. *Machine Learning*, 1986, 1: 81—106
- [27] Mitchell T M. Machine Learning [M]. New York: McGraw-Hill, 1997
- [28] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations [M]. San Francisco: Morgan Kaufmann, 2000
- [29] Quinlan J R. C4.5: Programming for Machine Learning [M]. San Mateo, CA: Morgan Kaufmann, 1993