文章编号:1007-4619(2008)02-0370-08

基于认知的模糊地理要素建模

——以中关村为例

刘 瑜,袁一泓,张 毅

(北京大学 遥感与地理信息系统研究所,北京 100871)

摘 要: 地理空间中的地理要素往往具有模糊性,这种模糊源于人对现实世界的概念化过程,因此具有主观特性。基于模糊集理论,尽管有很多途径对模糊地理要素对应的隶属度函数进行探讨,但是基于认知实验的方法最直接的反映了人们对相应要素的概念化过程中的模糊性。以中关村地区为例,设计了基于地标的问卷调查,并计算每个地标属于"中关村地区"这一概念的隶属度,进而采用支持向量回归方法,得到该要素的隶属度函数。该方法具有实验实施简单,结果便于管理的特点。最后,我们分析了中关村的隶属度函数的一些空间分布特征及其原因。

关键词: 空间认知;模糊地理要素;支持向量回归;中关村

中图分类号: P208 **文献标识码:** A

1 引 言

在地理信息科学领域,研究地理空间认知的一个重要目的就是通过探讨人类对于现实世界的认知和概念化过程,建立更为符合人的空间知识表达和加工过程的应用。由于现实世界的无限复杂性,我们在数据建模中需进行必要的概括和抽象[1]。根据 OGC 规范,人类在进行抽象过程中,首先建立的是符合常识性地理空间认知、适合于自然语言表达的概念世界模型[2]。在概念世界中,由于相应的概念化过程,概念的模糊性普遍存在,这种模糊性既包括地理要素的模糊性[3,4],也包括空间关系的模糊性[5,6]。从人类地理空间知识表达的角度,人类通过为地理要素命名,形成一个个地名,并基于这些地名组织了表现为文本形式的空间知识。为了提取并且加工这些地理空间知识,我们需要建立数字地名辞典[7—9],记录每个地名对应的空间范围,模糊要素的建模无疑有助于对上述空间知识进行更为精确的处理。

根据 ISO19109[10]的定义,要素是指"在选定语境中一个有意义的对象"。文献[11]和[12]认为,地理要素是根据其关联的地理语义——即其本体,被人类从更像一个"场模型"的现实世界中识别出来。其本体的不同,决定了地理要素具有不同的模糊特性,例如青藏高原和长江三角洲地区,其模糊性的机制并不一致。要素的模糊特性通常体现在边界的渐变性上,文献[13]根据地理要素的边界,将地理对象分为 Fiat对象和 Bona Fide对象,前者是指其边界是认知的或者基于法令等规定的,后者则具有明确的物理边界。大部分自然要素在人们归纳推理直至产生概念模型的过程中会具有不确定的边界(如半干旱地区、亚热带地区),属于 Fiat对象中的认知边界类;而人造要素(如城市)虽然内部的构成成分多属于 Bona Fide对象,边缘相对清晰,但由于内部各结构之间存在空隙,也会产生边界的模糊性。Montello[14]认为:地物的模糊性不仅是由"错误"或"逻辑学上的不明确导致",而是真实的模糊,并将地物模糊性产生的原因归纳为以下 5类,这些原因是地理要素模糊性的直接来源:(1) 度量错误或不准确;(2) 多地区接合处;(3) 边界随时间变化;(4) 有争议的地物边界;(5) 地物基本概念的模糊性。

许多学者对具有模糊边界的地物进行了研究,其中模糊集[15]是最经常采用的数学工具。根据模

收稿日期:2007-08-01;**修订日期**:2007-10-30

基金项目:国家高科技研究发展计划(863计划)(编号:2007AA12Z216),国家自然科学基金(编号:40701134),香港、澳门青年学者合作研究基金(编号:40629001)资助项目。

作者简介:刘 瑜(1971—),男。副教授。主要研究方向为地理信息科学,在国内外相关学术刊物和会议上发表论文 70余篇,E-mail: liuyu@urban.pku.edu.cn。

糊集理论,变量隶属于一个模糊集合的程度可以用介于 0与 1之间的数值来表示,从而可以建立相应的隶属度函数 (membership function, MF)。而在模糊地理要素的表达中,隶属度函数刻画了空间 (通常是二维空间)上每一点隶属于模糊地理要素的程度,因而其隶属度函数是二维的,即它可以表示为 $z = f(x, y)$,其中 z表示了点 (x, y)处的隶属度数值。目前,模糊集方法在地理信息系统及相关领域中得到了广泛应用,并通常为了便于存储的原因表现为栅格形式。在确定一个模糊地物的隶属度函数时,有以下 3种常用方法。

1.1　认知实验法

该方法选择一组被试,以问卷等形式取得其对被研究地物模糊性的判断。如在 Montello等人的"确定 Santa Barbara城区范围"实验中[16],要求被试分别以 100%和 50%的确信度,在纸上画出"Santa Barbara城区"的范围,以此来拟合城区边界和模糊带。由于地理现象的模糊性主要来自人类认知中的概念模糊,认知实验方法将地理要素模糊研究与心理学结合起来,使实验数据很好的反映认知活动的结果,体现了模糊要素的主观性。但缺点是对每个地物都需要对多个被试进行调查,收集实验数据的成本较高,且样本人群的选择是否具有代表性、被试回复的可信度等因素皆对实验的最终准确度有所影响。

1.2　地理信息检索法

地理信息检索 (geographic information retrieval, GIR)是顾及地理语义的信息检索技术。由于 Web页中包含了丰富的、通常以地名为基础组织的地理信息,基于 GIR的方法正是根据 Web页中所包含的模糊地理概念,通过对网页所包含地理内容的分析,确定某模糊要素的范围。在 Clough等人的研究中[17],选取"地名 +概念类型"作为关键词利用 Google进行搜索,并抽取出搜索得到网页中的地理概念,根据地名匹配原则以点的方式标绘在地理坐标系中,进而计算每个实验点出现的次数,最终依据点密度绘制该模糊地物的隶属度函数。GIR方法的优势在于不需要进行大规模的问卷调查,当需要研究多个地理要素的隶属度函数时比较方便。但由于选取的搜索概念容易带有人口密度的特征或其他倾向性,实验结果容易产生偏差。

1.3　遥感图像分类法

在遥感图像分类中,由于光谱图像中各像元的类别隶属度不同,因此分类结果带有一定的不确定性,基于此结果进行区域划分便会出现边界的模糊现象。文献 [18]提出了不同的建模方法,将基于遥感图像分类结果的模糊地物分为 3类:外延与内部均模糊的地物、外延模糊内部明确的地物、外延明确内部模糊的地物,并针对 3类地物,在类别隶属度函数和地理区域隶属度函数之间建立不同的映射关系。基于遥感图像分类的方法建模过程简单清晰,但仅适用于能够通过遥感能够分辨的地理要素,并且其准确性直接受分类精度的制约,对于结构较为简单的地物类应用显示出良好的效果,但不适用于过于复杂或抽象的地理概念。此外,通过遥感途径确定的地物模糊性与认知模糊性之间的吻合程度,在很多情形下并不明确。

比较上述 3种方法,可以看出第 1种方法最直接体现了人们对于一个模糊地理要素的认知,第 2种次之,第 3种最弱。然而,在文献 [16]描述的认知实验中,如果被试对区域没有整体认知,就难以形成一个封闭的曲线,提高了实验的难度,并影响了结果的可靠性。在本研究中,考虑到地标在人类空间认知中占据重要的地位[19,20],在被研究要素的可能范围内,选择一系列地标,让被试评判它是否属于该模糊地理要素。根据文献 [21]的观点,如果 N个被试中,有 M个对一个地标属于该地理要素给以肯定回答,那么相应位置的隶属度可以认为是 M/N。在得到多个地标的隶属度之后,我们选择支持向量回归 (support vector regression, SVR)方法计算相应区域的隶属度分布。之所以采用 SVR,是因为它通过核函数对于空间分布能够达到比较准确的近似,而它支持解析形式的隶属度函数,也便于计算机管理和存储。采用插值方式生成的结果在数据库中就难以管理,因为难以采用确定的数学函数表达隶属度空间分布。

本文研究区为北京市中关村地区,选择该地区是因为首先它具有模糊性,其次它蕴含了不同的地理语义。在当前的信息时代,中关村已经成为北京市高科技区的一个象征。因此从认知的角度,研究其地理分布及相应的模糊性,具有一定的实践意义。

2　基于支持向量回归的中关村范围认知实验

2.1　实验背景和概述

在现实中,"中关村"蕴含着历史、行政、商贸、教

育等多重涵义。由于它具有悠久的历史,因此对其覆盖区域的认知不可避免地受到历史因素的影响。根据历史地理学家的考证,在清代中叶,由于“三山五园”的修建,带动了海淀镇的繁华,中关村就坐落在海淀镇。在清末民初的地图上,始见今天中关村这块地域的标注,地理位置东西大约为今天的蓝旗营西侧至北大东门,南北从北大物理学院到北四环路。1961年,中关村街道办事处的成立标志着这一地理概念有了行政上的严格区域划分。近期,海淀区将原中关村街道与双榆树街道合并,形成了新的中关村行政区划(图 2)。从 20世纪 80年代开始,中关村地区成立了大批高科技领域的公司与产业,建立了中关村科技广场,电子一条街的模式基本形成。随着后来海外风险投资的注入,成为全国著名的科技园区。对于众多中国人来说,“中关村”是高科技园区的象征,科技商贸区的涵义对其范围的影响巨大。另外,该区里密集分布着北京大学、清华大学和中国科学研究院等中国一流学府和研究机构,教育和科研力量的雄厚也使得其既具有一定的文化底蕴,又不能忽略大学区的概念对其地域边界的作用。

众多学者认为人类对不同概念有特定的认知过程。因此,我们假设人类在对地理概念进行认知时,其判断和决策不仅受空间位置的影响,还受概念本身蕴涵的多重涵义的影响,并通过对认知实验数据的分析对这一假设进行验证。实验选择的地理概念为“中关村”,如前所述,该地名蕴含了丰富的历史、行政、商贸和教育等多重涵义,对中关村地理范围认知的探究,有助于对不同因素的影响进行讨论。认知实验一个重要因素是被试的选择,根据文献[22]的阐述,环境空间的尺度决定了人们无法在短时间内建立起对于一个环境的认识,是经过长时间、多层次、不同角度对地理环境的观察和了解才逐渐累积起来的。因此在本实验中,选择被试为长期居住于北京市区,对本地情况有多方面了解的人员,并且在回答问题时有中关村地区地图作为参考,这保证了收集数据的较高可信度,能够反映公众对于中关村这一地名的认知。

2.2 实验方法

尽管如前所述,不同视角的中关村蕴含了不同的地理范围,但是一个被公众广为接受的最大范围还是较为确定的,即北京市西北部,以北京大学、清华大学为中心的信息技术产业和高等教育密集区。为了保证选择区域能够覆盖潜在的中关村地区,我们对该范围进行扩展,并在该范围内选取有代表性的 30个地标,如图 1。

图 1 选取地标分布图(底图来自 http://map.baidu.com)

Fig. 1 Landmarks selected in the cognitive experiment (Background map is from http://map.baidu.com)

这 30个地标可根据类型的不同分为以下 4类:科技商贸类地标(Ⅰ);教育科研类地标(Ⅱ);生活类地标(Ⅲ)和自然及人文景点类地标(Ⅳ)(表 1)。以上地标名称中,字面上均未含有“中关村”字样,避免对被试的选择形成误导,影响实验结果的准确度。

我们以问卷的方式,对 20位在北京居住 10年以上的被试进行调查,分别对上述地标进行回答,判断其是否属于中关村范围内,并在地标后的选项中选择“是”、“否”或者“不确定”。根据被试的回答情况,我们用一个认同系数 d来表示每位被试对每个地标的回答。若选“是”,则 $d=1.0$;选“否”,$d=0.0$;选“不确定”,$d=0.5$。根据前面的讨论,对每个地标都可以计算一个认同程度,即地标对应位置的隶属度数值,公式如下:

$$\mu_i=\frac{1\times a_i+0\times b_i+0.5\times c_i}{N_i} \qquad (1)$$

式中,a_i为对第 i个地标做肯定判断的人数,b_i为否定判断的人数,c_i为选择“不能确定”的人数,N_i为被试总数。表 1详细给出了 30个地标的隶属度。

表 1　30个地标的类型和隶属度

Table 1　Types and membership degrees of 30 selected landmarks

序号	地标名称	类别	μ
1	中国科学院图书馆	Ⅱ	0.95
2	银科大厦	Ⅰ	0.925
3	未名湖	Ⅳ	0.85
4	海淀公园西门	Ⅲ	0.775
5	北京大学新政管楼	Ⅱ	0.825
6	资源大厦	Ⅰ	0.95
7	北京 101中学	Ⅱ	0.725
8	畅和堂	Ⅳ	0.225
9	海淀区法院	Ⅲ	0.475
10	郭林家常菜	Ⅲ	0.925
11	畅春园小区	Ⅲ	0.75
12	颐和园文昌阁	Ⅳ	0.2
13	清华附中	Ⅱ	0.375
14	鼎均大厦	Ⅰ	0.65
15	颐和园南如意门	Ⅳ	0.25
16	景帝陵	Ⅳ	0.125
17	三一八纪念碑	Ⅳ	0.225
18	清华同方科技广场	Ⅰ	0.825
19	清华侧门	Ⅱ	0.725
20	中国农大西区	Ⅱ	0.35
21	北京师范大学	Ⅱ	0.175
22	中发电子大厦	Ⅰ	0.875
23	文化大厦	Ⅰ	0.875
24	北坞村	Ⅲ	0.25
25	万泉河	Ⅳ	0.35
26	信息大厦	Ⅰ	0.40
27	胜利饭店	Ⅲ	0.30
28	世宁大厦	Ⅰ	0.50
29	云航大厦	Ⅰ	0.40
30	观音殿	Ⅳ	0.40

2.3　基于 SVR的数据处理

支持向量机(support vector machine, SVM)是在统计学习理论的基础上发展出来的一种通用学习方法[23]。它遵循了有严格理论基础的结构风险最小化原理,能较好地解决非线性、高维数据和时间序列预测等实际问题,另外 SVM的求解最后转化成二次规划问题的求解,因此,SVM的结果具有最优和惟一的双重优势。SVR则是基于支持向量机,将非线性回归转化为线性回归求解的方法。在本研究中,采用 SVR对 30个地标进行训练,建立地标坐标同隶属度的函数关系,从而得到中关村这一概念的二维隶属度函数。采用 LIBSVM软件包[24],对数据进行处理和训练。具体包括以下步骤:

2.3.1　数据预处理

以图 1所示研究区中心点为原点,建立直角坐标系,取得选定地标的坐标值,即完成地标点数字化的过程。将每个地标的横纵坐标及隶属度整理为 (x_i, y_i, μ_i)的格式,作为 SVM软件可接受的输入向量。

2.3.2　核函数选择

虽然在研究中已经证明,只要满足 Mercer条件的对称函数即可作为核函数[23]。对于一般 SVM应用,有以下 4类核函数,即线性函数、多项式函数、RBF(Radial Basis Function)函数和 Sigmoid函数。本实验中选择 RBF函数(公式(2))作为核函数,是由于 RBF可以将非线性映射的数据转化到高维空间中,使其具有线形映射的特征。同时,一般的线性核函数也可以看作是 RBF的一个特例[25],而 Sigmoid核函数性质与 RBF相似[26]。

$$K(x_i, x_j)=e^{-\gamma\|x_i-x_j\|^2},\ \gamma>0 \qquad (2)$$

2.3.3　确定 SVR的最优参数

在 SVR中 γ参数的选择对于结果的影响较大。具体到本应用,γ数值偏低则拟合精度低,并且空间分布比较规则,如近似于椭圆形;而 γ数值偏高则会出现过度训练的情形。对于一个特定的问题,事先无法知道 γ取何值时最优,因此有必要采用一定的方法事先进行参数搜索,找到最优的参数 γ。按照文献[24]的建议,我们采用交叉验证(cross-validation)方法[27]进行搜索。首先随机将数据分为 n个大小相同的子集,训练其中的 $n-1$个子集得到分类器,再使用所得到的分类器对剩下的一个子集进行预测,并评估训练效果;上述步骤循环进行 m(通常 $m>n$)次,这样可以保证整个

训练数据集中的每个数据被预测了一次。基于交叉验证法,最终求得 γ最优系数为在 45—55之间。由于采用了随机方法,γ的精确数值并不能确定,我们设定其值为 50。

2.3.4 绘制中关村区域的隶属度函数

在利用 SVR训练并建立坐标和隶属度之间的数学联系后,我们可以计算空间上任意一点的隶属度数值,将该结果离散化以栅格数据的方式存储。基于该栅格数据可以计算中关村的质心位置,它可看作是空间认知意义上的中关村地理范围的中心。如图 2,质心所处位置在北京大学东侧,与历史地理学者认定的中关村历史位置较为吻合。基于隶属度函数,可绘制相应的等值线图 (图 2)。不同等值线包围的区域是相应隶属度数值为 α的模糊集截集 (cut set)。其中 α = 0.9截集范围可以认为是非常可信的中关村区域,0.5截集可以认为是广义中关村区域,而在 0.5截集外基本上可以认为不属于中关村地区。根据“鸡蛋/蛋黄 (Egg/Yolk)”模型[28],0.9截集范围可以认为是“蛋黄”,即核心区域,而 0.5截集范围则对应于“鸡蛋”。

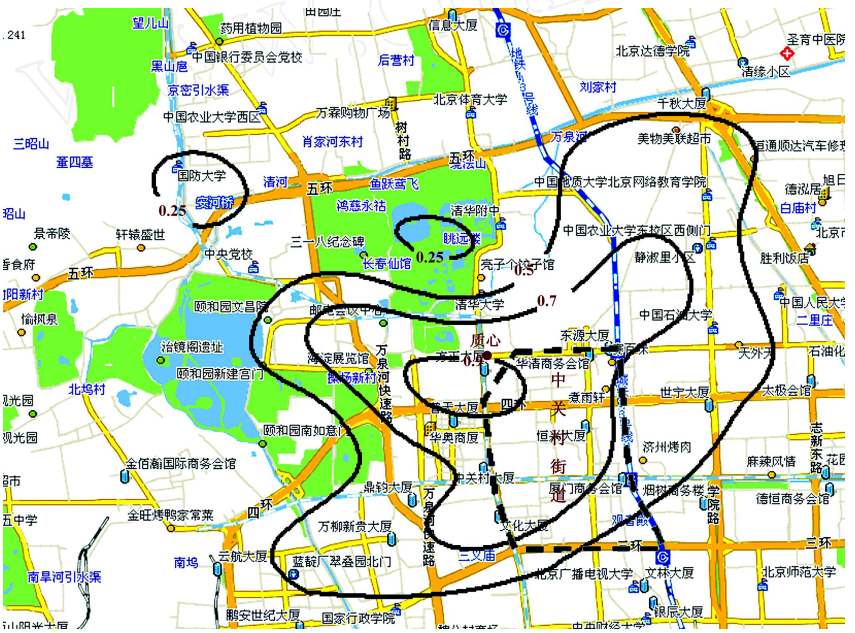

图 2 隶属度函数的等值线图

Fig. 2 Isoline map of the membership function associated with Zhongguancun

2.4 实验结果分析和讨论

根据表 1和图 2,我们可以对中关村这样一个模糊地理要素的空间分布得到以下结论。

(1) 中关村核心区域 (0.9截集)覆盖了传统的中关村范围,而其质心也与历史地理学家的考证吻合。这首先说明中关村作为一个历史地名,尽管经过多年沿革,但是其地理范围具有相对稳定性,并且影响了人们对于其地理范围的认知。核心区域除了包括传统中关村范围外,还向西扩展覆盖了北京大学校园的南部。这说明了北京大学对于人们在中关村概念形成中的影响,也反映了中关村所蕴含的教育职能。

(2) 中关村外围区域 (0.5截集)在东部和南部形成两个延伸。东部区域越过了学院路,而覆盖了清华大学和语言学院等高校,南部区域则覆盖了西北四环和西北三环之间这样一个信息技术产业活动密集区,包括了大量写字楼,其认同系数较高。以上充分说明了人们在形成中关村这一概念时,高科技和高等教育因素在其中起到了重要作用。

(3) 不论是核心区域还是外围区域,与中关村街道办事处的管辖范围存在较大差异。这里有两个可能的原因:首先中关村街道办事处成立较晚;其次,这说明在人们形成城市空间认知的过程中,街道办事处作为一级行政区,在人们的空间认知中概念较为模糊。

(4) 中关村外围区域在北部和西部,明显存在两个缺口。分别对应于圆明园遗址公园和颐和园南部。而这两个地区属于旅游景点。这说明人们在对中关村进行认知时,排斥了其旅游景点语义。从表 1可以看出,自然及人文景点类地标,如颐和园文昌阁及三一八纪念碑,认同系数偏低,这从反

面也表明了高科技和教育在中关村概念形成中的作用。

(5) 未名湖作为自然及人文景点类地标,认同系数 0.85,远远高于其余同类地标,与北京大学新政管楼的系数 0.825十分接近。由于未名湖在北京大学的范围之内这个事实广为人知,因此人们对它的认知时通常会和对北大的认知结合;而新政管楼和未名湖两者同处于北京大学范围内。由于未名湖和中关村并没有直接的包含关系,因此被试倾向于用更大范围的地理概念 (即北京大学)的性质取代地标自身的性质,这使得处于同一熟知的地物范围内的地标容易被赋予相近的认同系数。与此类似,同处于颐和园范围内的畅和堂、南如意门和文昌阁 3个地标,认同系数亦十分接近。该结论在文献[29]等对层次空间推理的研究中也可以得到印证,也说明了中关村同未名湖等概念在空间知识分层体系中至少相差两层。

3　结　论

模糊地物是人们认知地理世界过程中普遍存在的现象,本文利用认知实验结合支持向量回归的方法对模糊地物的空间分布进行研究和讨论,表明 SVR具有良好的拟合能力,便于在研究中发掘人们的认知规律。本研究得到有意义的结论有如下两点。

首先,我们通过研究被试对不同类别地标的判断倾向,可以发现人们在对模糊地物进行认知时,经常会受到概念本身蕴涵的地理语义的影响,认知实验的方法体现了各种影响的综合。在以"中关村"地域为研究对象的认知实验中,被试普遍显示出对科技类和教育类地标较高的隶属度,反映了中关村这一概念所蕴含的科技和教育意义;然而与行政区划范围则存在较大差异。这说明在人们的空间信息表达中,中关村这一概念所代表的社会经济含义远远大于行政管理含义。当然,前者也不能无限夸大,在本次实验设计中,同时选取了位于"丰台中关村科技园区"和"上地中关村软件园"的建筑物进行调查,结果得到了非常低的认同度。该事实以及隶属度分布质心同历史上中关村位置的吻合,说明了地名的历史沿革对其位置认知的作用。另外值得指出的是,不同尺度、不同类型的地理要素的模糊性机制存在较大差异。因此上述试验过程不仅可以用于确定模糊地物和隶属度函数,还可通过对不同类型地标的选取来判断相应地理要素的认知分类,如中关村由于科技类地标及教育类地标认同系数较高,可被划分为"科技及教育类"地物。

其次,在 GIS应用中,如何管理具有渐变边界的模糊地物一直是个难题。采用栅格方法数据量偏大,不易存放于空间数据库中,并且受到分辨率影响;采用"鸡蛋/蛋黄"模型在使用时需要内插;其他简化为几何形状的方法又过于粗略,难以反映地物的实际空间分布。采用 SVR方法,隶属度函数可以用一个数学公式表达,并且可以通过调整得到合适的 γ数值,以达到最好的拟合效果,从而在可管理性和准确描述空间分布之间取得一种均衡。

值得指出的是,本文提出的方法仍然需要大量的问卷调查以确定一个地物的隶属度函数,仍存在工作量较大的不足。因此在未来的研究中,拟采用基于 Web的 GIR方法计算得到中关村概念的隶属度函数,实验结果进行相互对照,从而更为深入地探讨中关村这一地物的模糊特性。

参考文献 (References)

[1] Longley P A, Goodchild M F, Maguire D J, *et al*. Geographic Information Systems and Science [M]. Second Edition, New York: Wiley, 2005.

[2] OpenGIS Consortium. The OpenGIS™ Abstract Specification [M]. Topic 5: Features, 1999.

[3] Burrough P A. Natural Objects with Indeterminate Boundaries [A]. Geographic Objects with Indeterminate Boundaries [C], Burrough P A, Frank A U. London: Taylor & Francis Ltd, 1996.

[4] Bittner T, Stell J G. Vagueness and Rough Location [J]. *GeoInformatica*, 2002, **6**(2): 99—121.

[5] Dutta S. Qualitative Spatial Reasoning: a Semi-Quantitative Approach Using Fuzzy Logic [A]. Proceedings of the First Symposium on Design and Implementation of Large Spatial Databases [C]. 1990.

[6] Bloch I, Fuzzy Spatial Relationships for Image Processing and Interpretation: a Review [J]. *Image and Vision Computing*, 2005, **23**: 89—110.

[7] ADL. Alexandria Digital Library [DBOL]. http://www.alexandria.ucsb.edu/, 2002.

[8] Chaves M S, Silva M J, Martins B. GKB-Geographic Knowledge Base [R]. Departamento de Informatica, Faculdade de Ciencias da Universidade de Lisboa, Campo Grande, Lisboa, Portugal, 2005.

[9] Liu Y, Zhang Y, Tian Y, *et al*. On Generalized Place Names and Associated Ontologies [J]. *Geography and Geo-Information Science, (in press)*. [刘瑜,张毅,田原等. 广义地名及其本体研究,地理与地理信息科学 (出版中).]

[10] ISO19109, http://www.seegrid.csiro.au/twiki/pub/Xmml/

Feature Model / 19109 DIS2002. pdf.

[11] Goodchild M F, Yuan M, Cova T J. Towards a General Theory of Geographic Representation in GIS [J]. *International Journal of Geographic Information Science*, 2007, **21**(3): 239—260.

[12] Liu Y, Goodchild M F, Guo Q, *et al*. Towards a General Field Model and its Order in GIS [J]. *International Journal of Geographical Information Science*, *(in press)*.

[13] Smith B. On Drawing Lines on a Map [A]. Proceedings of COSIT 1995 [C]. 1995.

[14] Montello D R. Regions in Geography: Process and Content [A]. Duckham M, Goodchild M F, Worboys M F. Foundations of Geographic Information Science [C]. London: Taylor & Francis, 2003: 173—189.

[15] Zadeh L A. Fuzzy Sets [J]. *Information and Control*, 1965, **8**: 338—353.

[16] Montello D R, Goodchild M F, Gottsegen J, *et al*. Where's Downtown? Behavioral Methods for Determining Referents of Vague Spatial Queries [J]. *Spatial Cognition and Computation*, 2003, **3**: 185—204.

[17] Clough P, Joho H, Jones C B, *et al*. Modelling Vague Places with Knowledge from the Web [A]. CIKM'05 [C]. Germany: Bremen, 2005.

[18] Cheng T, Molenaar M, Lin H. Formalizing Fuzzy Objects from Uncertain Classification Results [J]. *International Journal of Geographical Information Science*, 2001, **15**(1): 27—42.

[19] Lynch K. The Image of the City [M]. MIT Press, 1960.

[20] Montello D. Spatial cognition [A]. Smelse N J, Baltes P B. International Encyclopedia of the Social & Behavioral Science [C]. Oxford: Pergamon Press, 2001.

[21] Dubois T, Prade H. Fuzzy Sets and Probability: Misunderstandings, Bridges and Gaps [A]. Proceedings of 2nd IEEE Conf on Fuzzy Systems [C]. San Francisco, CA, 1993.

[22] Montello D R. Scale and Multiple Psychologies of Space [A]. Spatial Information Theory: A Theoretical Basis for GIS [C]. Lecture Notes in Computer Science 716, Berlin: Springer-Verlag, 1993.

[23] Vapnik V. The Nature of Statistical Learning Theory [M]. Berlin: Springer-Verlag, 1995.

[24] Chang C C, Lin C J. LIBSVM: a Library for Support Vector Machines [R]. Software Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[25] Keerthin S S, Lin C J. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel [J]. *Neural Computation*, 2003, **15**(7): 1667—1689.

[26] Lin H T, Lin C J. A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-Type Methods [R]. Department of Computer Science and Information Engineering Taiwan University, 2003.

[27] Hsu C W, Lin C J. A Simple Decomposition Method for Support Vector Machines [J]. *Machine Learning*, 2004, **46**: 291—314.

[28] Cohn A G, Gotts N M. The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries [A]. Burrough P A, Frank A U. Geographic Objects with Indeterminate Boundaries [C]. London: Taylor & Francis Ltd, 1996.

[29] Timpf S, Frank A U. Using Hierarchical Spatial Data Structures for Hierarchical Spatial Reasoning [A]. Hirtle S C, Frank A U. COSIT'97 [C]. Berlin: Springer-Verlag, 1997.

A Cognitive Approach to Modeling Vague Geographical Features: A Case Study of Zhongguancun

LIU Yu, YUAN Yi-hong, ZHANG Yi

(Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, *China)*

Abstract: In practice, vagueness is a common phenomenon of geographical features. The vagueness of a feature often comes from human conceptualization of the real world. Modeling of vague features will undoubtedly contribute to more precisely handling spatial knowledge. In recent studies, a number of theoretical methods have been employed to model vague features, where fuzzy set theory is in common use. Following that theory, the degree that an element belongs to a fuzzy set can be expressed by a number between 0 and 1. We can thus establish a corresponding membership function (MF) for a fuzzy set. Recently, much literature focuss on vague features and proposes approaches to establish corresponding MFs. They include approaches based on cognitive experiment, remotely sensed data and GIR (Geographical Information Retrieval). Due to the subjectivity of vagueness, spatial cognitive experiments provide a direct way to represent the vagueness of individuals' conceptualization of corresponding features. However, previous methods based on cognition cost highly and are somewhat hard to control the result, since subjects in such experiments are asked to delineate boundaries of vague areal objects. Landmarks play an important role in individuals' development of spatial cognition. It is thus relatively easy for individuals to perceive a landmark and decide whether it is within a given region or not. In this research, we took Zhongguancun in Beijing city as an example, since it is complex with different meanings, such as educational, political,

and historical meanings. A questionnaire is designed to collect membership degrees of 30 landmarks which are in the region of Zhongguancun. These 30 landmarks, which are selected from the maximum potential region corresponding to Zhongguancun, can be abstracted to point features. They belong to different types, such as office building, hotel, school, recreation place, and natural feature. For each landmark, the subjects are asked whether it is within Zhongguancun, for which three optional answers are provided: YES, NO, and NOT SURE. By collecting all answer sheets, we can compute a score of each landmark. Such a value can be viewed as the membership degree that the corresponding position belongs to the concept of "Zhongguancun". However, since Zhongguancun is a two-dimensional vague object, it should be represented by a membership function (MF) like $z = f(x, y)$. We thus need to find out an appropriate interpolation method to obtain the MF. In this research, support vector regression (SVR) is adopted to compute the MF. Compared with conventional interpolation methods, such as IDW (Inverse Distance Weighted) method, the proposed approach is easy to implement and the results are convenient to be managed. Additionally, SVR provides a mechanism to obtain a trade-off between goodness of fit and generality by adjusting some parameters, such as γ for radial basis kernel functions. Based on the result of membership function, we also investigate spatial distribution properties of Zhongguancun and find out some interesting points. Since Zhongguancun has been viewed as "the silicon valley in China", it is closely related with such concepts as high-tech industry, university, and so on. Consequently, the landmarks associated with these concepts always have higher membership degrees. On the contrary, lower scores are assigned to some recreation places and natural features. In summary, from a point of view of behavior, the current concept of Zhongguancun is far beyond the scope as an administrative unit, both spatially and functionally, since many factors influence its internal representation of individuals.

Key words: spatial cognition; vague geographical features; support vector regression; Zhongguancun