

宽泛地理注记的投放模型

李佳田¹ 康 顺¹ 李晓娟¹ 张 蓝¹ 罗富丽¹ 白建军²

1 昆明理工大学国土资源工程学院, 云南 昆明, 650093

2 陕西师范大学旅游与环境学院, 陕西 西安, 710062

摘 要:针对网络地图服务环境中的地理注记评价问题,对传统地理注记内容进行非结构化文本形式扩展,定义宽泛地理注记,用投放过程来模型化标注行为。首先,依据词频分类方法确定地理注记文本内容分类,并根据 Voronoi k 阶邻近关系建立投放邻域的空间相关性定量收敛描述,进而结合地理注记文本与地理注记存在邻域构建基于地理注记类型与类型转化的投放模型。实验结果表明,在已知两种类型地理注记集的实际情况下,投放模型能够有效地对新增地理注记进行合理性评价。

关键词:地理注记;非结构化文本;词频分类;Voronoi k 阶邻近;空间自相关性

中图法分类号:P208;P283.1 **文献标志码:**A

地理注记是地图符号的重要组成类型之一。地图学中对地理注记的普通解释是,作为一种地图语言说明地图上所表示的地物名称、位置、范围、高低、等级(主次)等^[1]。就注记内容分析,其主要形式是结构化数据的单词或数值,然而,地图服务环境赋予地理注记更多的内涵,通常以自然语言文字描述为主要形式来表达地理现象。地理注记是空间位置与内容的复合体。学术界关于地理注记的研究主要集中在三个方面:①以地理实体命名目标的注记体系和标注规范研究^[2-4];②以解决注记内容叠置或优化布置的制图综合方法研究^[5-6];③栅格扫描地图中的注记识别方法研究^[7]。

现阶段,多数网络地图环境提供具有典型众包模式特征的地理注记服务。在这种应用模式中,受众不再单纯地作为地图服务使用者,更为重要的是,其同时是数据或信息的提供者。随着地理标注服务的深入应用,受众提供的数据量必将呈几何级数增长。维护地理注记数据准确性,人工干预效率低下,迫切需要一个有效地对地理注记适宜性进行评价的方法。目前鲜见这方面的研究工作。本文通过将传统地理注记扩展为宽泛地理注记,将网络环境中的标记行为理解为投放过程,分解为投放的是什么、投放到哪里及将其投放到这里是否合适三个步骤,形式化投放过程并建

立地理注记投放计算模型,用以解决地理注记适宜性评价问题。

1 宽泛地理注记

地理注记分为名称注记、数字注记与说明注记三种类型,其构成元素包括字体、字级、字色与字距^[1]。提取地理注记的位置与内容两个主要性质,忽略构成元素,其可以被形式化为:

$$ga = \{(x, y, z), \text{term}\} \quad (1)$$

其中, (x, y, z) 表示地理注记的空间位置; term 为结构化词汇或数值,代表地理注记的内容。地图服务环境,结构化的 term 难以适应其描述需要,为此,用非结构化的文本替代 term,并将这种地理注记形式称为宽泛地理注记(broad geographic annotation, BGA),描述为:

$$bga = \{(x, y, z), \text{text}\} \quad (2)$$

式中, (x, y, z) 表示宽泛地理注记的空间位置;而 text 为文本内容,通常是以自然语言描述的非结构化文字段落。在英文行文中,单词之间是以空格作为自然的分界符,而中文只是在字、句、段层次上能够通过明显的分界符来划界,在词汇层次上没有一个形式上的分界符。因此,本文通过中文分词过程,将宽泛地理注记文本内容 text 描述为结构化的单词序列形式:

$$\text{bga} = \{(x, y, z), (\text{term}_1, \text{term}_2, \dots, \text{term}_n)\}, n \in \mathbf{N} \quad (3)$$

式中,由 term_i 统计词汇出现频率,被认为是 text 的维度特征,能够有效支持 text 分类^[8-9],即 bga 属于什么类型。事实上,构成文本的词汇数量是相当庞大的,因此,表示文本的向量空间也会很大。一些通用并普遍存在的词汇对分类的贡献较小,而某些词汇出现频率则相对不稳定,在不同文本类型中的特征差异明显,对分类的贡献较大。空间上, bga 的作用范围是局部的,其空间位置 (x, y, z) 是构建局部范围的基准。

2 投放模型

在陆地表面绘制等高线是合适的,同样的绘制动作如果置于水体表面就不恰当或是不正确;类似地,在水体表面出现岛状地理实体是合适的,而如果出现代表建筑物实体的规则形状则是不恰当的。以上现象解释了投放过程的两个关键要素:投放对象与投放区域。评价投放过程的适宜性,必需将两个关键要素相结合,而不能割裂开来。将地理笔记操作视为投放过程,投放模型如下所述。

2.1 宽泛地理笔记词频分类

依据式(3),可以根据词汇出现的频率建立笔记文本分类。即根据已有宽泛地理笔记集 BGA ,得到基础词汇表(basic term table, BTT),形如 $\text{BTT} \langle \text{term} \rangle$ 。对于任意 $\text{bga} (\text{bga} \notin \text{BGA})$,将 bga. text 中出现的所有词汇 $(\text{term}_1, \text{term}_2, \dots, \text{term}_n)$ 顺序地组成一个向量 $e[\text{ext}_1 \text{ ext}_2 \dots \text{ext}_n]$,其中, ext_i 表示词汇 term_i 是否存在于 BTT 中,如果存在,那么 $\text{ext}_i = 1$,否则 $\text{ext}_i = 0$ 。至此,宽泛地理笔记的文本词汇存在分类可以被定义为:

$$\text{ec}(\text{bga}, \text{BGA}) = \sum_{i=1}^n \begin{cases} 1, \text{ext}_i = 1 \\ 0, \text{ext}_i = 0 \end{cases} \quad (4)$$

函数 $\text{ec}()$ 将 ext_i 存在值为 1 的次数累加即可以确定所属分类,对于不同类型的 BGA_i 与 BGA_j ,则宽泛地理笔记类型是函数值大的 BGA 类型。值得注意的是,存在分类只是考虑词汇存在与否的简单情况,将出现一次与出现 n 次的词汇对分类的影响视为相同作用,显然,出现 n 次的词汇在分类上的作用要大于出现一次的词汇。扩展基础词汇表为由已知类型宽泛地理笔记集 BGA_i 所得到的 BTT_i ,形如 $\text{BTT}_i \langle \text{term}, \text{freq} \rangle$ 。对于任意 $\text{bga} (\text{bga} \notin \text{BGA})$,将 bga. text 中出现的所有词汇 $(\text{term}_1, \text{term}_2, \dots, \text{term}_n)$ 顺序地组成一个向量 \mathbf{v}

$[\text{freq}_1 \text{ freq}_2 \dots \text{freq}_n]$,其中, freq_i 表示词汇 term_i 在 BTT_i 中的频率。如果存在多种类已知分类宽泛地理笔记集 $\{\text{BGA}_1, \text{BGA}_2, \dots, \text{BGA}_n\}$,那么,宽泛地理笔记的词频分类可以被定义为:

$$\text{fc}(\text{bga}, \text{BGA}_k) = \text{true}, k = \underset{i}{\text{argmax}} | \mathbf{v}_i (\text{freq}_{i1}, \text{freq}_{i2}, \dots, \text{freq}_{im}) |, \text{freq}_{i*} \in \text{BTT}_i \quad (5)$$

词频分类函数 $\text{fc}()$ 的物理意义是,宽泛地理笔记类型由距离其最近的已知笔记集合类型所决定,距离则由向量 \mathbf{v} 的模长度量。需要说明的是,如果 bga. text 的长度过长,会导致维数 n 过大,多个 freq 值较小的词汇会使模长值显著变小,导致不同类型之间的差异变小,影响分类精度,因此需要对 bga. text 进行降维处理,即选择 freq 值在不同 BTT 中差异较大的词汇构成向量。

关于已知宽泛地理笔记集 BGA 类型,可参考《国家语委语料库》中的分类体系构建。语料库由 3 个一级分类(人文与社会、自然、综合)以及约 40 个二级分类构成,如果二级分类不够细致,可参考其他分类资料对其进行细分类。

2.2 Voronoi k 阶邻近投放邻域

地理学第一定律指出“任何事物都是相关的,但相近的事物关联更紧密”^[10]。邻域范围内的地理实体分布具有相似性,称其为空间自相关性。更加具体地,空间自相关性是指空间目标的属性受它的空间位置影响,是空间目标属性值聚集程度的度量。与此同时,邻域之间会出现行为不确定现象,如居民购买力与居民收入的关系,不同地区(经济发达、经济落后)之间差异较大,称其为空间异质。宽泛地理笔记的空间分布亦是服从这样的规律,在邻域内部由相同或相近的标注类型组成,表现出空间自相关性;在邻域之间,则存在标注类型的差异,表现出空间异质性。邻域内部的笔记类型相同或相近,表现出空间自相关性;邻域之间的笔记类型有所差异,表现出空间异质性。将宽泛地理笔记的存在环境视为局部邻域,宽泛地理笔记相互之间的位置构成邻域的自相关性,如图 1 所示。

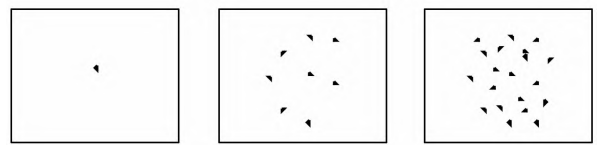


图 1 空间自相关邻域形成过程

Fig. 1 Sequential Process of Spatial Auto Correlation Neighborhood

空间目标之间的位置邻近描述主要存在欧氏

距离邻近与 Voronoi 邻近两种方法,相对于欧氏距离邻近, Voronoi 邻近在邻近描述与邻近度量方面更具优势^[11-12]。

定义1 Voronoi 邻近:存在空间目标集合 $O = \{o_1, \dots, o_n\} \subset \mathbf{R}^2, 1 < n < \infty, \forall o_i, o_j (i \neq j), \text{vor}(o_i), \text{vor}(o_j)$ 分别为目标 o_i 与 o_j 的 Voronoi 区域,如果下式成立:

$$\text{vor}(o_i) \cap \text{vor}(o_j) = \emptyset \quad (6)$$

则称 o_i 与 o_j 为 Voronoi 邻近。

Voronoi 邻近关系具有传递特性,可由多个步骤的 Voronoi 邻近-Voronoi 距离定义,进而归纳出普遍意义下的 Voronoi k 阶邻近。

定义2 Voronoi k 阶邻近:存在空间目标集合 $O = \{o_1, \dots, o_n\} \subset \mathbf{R}^2, 1 < n < \infty, \forall o_i, o_j (i \neq j)$,如果目标 o_i 经过最小 Voronoi 邻近步数 $k (k \in \mathbf{N})$ 到达目标 o_j ,则称 o_i 与 o_j 为 Voronoi k 阶邻近。

Voronoi 邻近是 Voronoi k 阶邻近的特例,即当 $k=1$ 时, Voronoi 一阶邻近等价于 Voronoi 邻近。本文中,将宽泛地理注记的存在环境视为由 Voronoi k 阶邻近构成的局部自相关邻域,而自相关性的度量依据局部 Moran 指数统计量得到^[13]。Moran 指数是用于识别空间邻域自相关的有效方法,结合宽范地理注记类型向量,局部 Moran 指数统计量计算为:

$$I_i = \frac{(|v_i| - \bar{v})}{S^2} \sum_j w_{ij} (|v_j| - \bar{v}) = \frac{n(|v_i| - \bar{v}) \sum_j w_{ij} (|v_j| - \bar{v})}{\sum_i (|v_i| - \bar{v})^2} \quad (7)$$

其中, n 为 Voronoi k 阶邻域中的空间目标数; $|v_i|$ 是类型向量的模, \bar{v} 是类型向量模的平均值; w_{ij} 是权重矩阵 \mathbf{W} 在 i 行 j 列的取值。观察可知, Moran 指数的分子为邻域中所有注记目标类型向量的协方差之和,而其分母为所有注记目标的标准差之和,比值位于 $(-1, 1)$ 范围。当 $I_i > 0$ 时,表示邻域为正的空间自相关,且 I_i 值越大相关性越大;当 $I_i < 0$ 时,表示邻域为负的空间自相关;当 $I_i = 0$ 时,则表示邻域为随机独立分布,化简式(7),得:

$$I_i = \frac{n z_i \sum_j w_{ij} z_j}{z_i^T z} = z_i' \sum_j w_{ij} z_j' \quad (8)$$

式中, z_i' 与 z_j' 是经过标准差标准化的注记向量; w_{ij} 是权重矩阵 \mathbf{W} 在 i 行 j 列的值,对于 w_{ij} 取值,为了保证空间权重矩阵 \mathbf{W} 的外延性,根据注记目标之间的 Voronoi 邻近距离来度量,依据定义2,将 w_{ij} 赋值为注记目标 bga_i 与 bga_j 的 Voronoi k 阶

邻近距离,即

$$w_{ij} = vk(i, j) \quad (9)$$

对邻域局部 Moran 统计值的计算结果采用随机分布或者近似正态分布进行验证,用标准化统计量 Z 来检验 n 个邻域地理注记是否存在空间自相关,统计量 Z ^[14] 的计算公式为:

$$Z(I_i) = \frac{I_i - E(I_i)}{\sqrt{\text{VAR}(I_i)}} \quad (10)$$

其中, $E(I_i)$ 为均值; $\text{VAR}(I_i)$ 为方差。当 $Z > 0$ 且 $I_i > 0$ 时,表明存在正空间自相关,即是说相同类型地理注记在邻域中趋于聚集态势;当 $Z < 0$ 且 $I_i > 0$ 时,表明存在负的空间自相关,说明相同类型的地理注记在邻域中趋于分散态势;当 $Z = 0$ 时,则表明呈独立随机分布。

假设地理注记存在于具有正空间自相关性质的局部邻域中,由上述计算过程,可根据 Voronoi k 阶邻近依据阶数 k 值由小到大调整来动态构建邻域,收敛条件为 $I_i > 0$ 且 $Z(I_i) > 0$ 成立的最小 k 值,建立投放邻域(board geographic annotation region, BGAG)描述为:

$$\text{BGAG} = \{ \text{bga} \mid vk(\text{bga}_i, \text{bga}_j) = k, k \leq \text{argmin}(I_i(\text{BGAG}) > 0 \wedge Z(I_i) > 0) \} \quad (11)$$

2.3 投放过程

在已知地理注记数量较多的情况下,假设在局部邻域中,地理注记的分布呈注记文本正空间自相关形态,投放邻域由某一种类型的地理注记依照式(11)构建。对于投放域空间范围内只存在一种类型的地理注记,称为单类型邻域;而如果存在多种类型地理注记,则称为复合类型邻域。对于一个新投放的地理注记,按其是否与邻域类型(或复合邻域类型之一)相同分为两种情况,而邻域为单类型或复合类型又分为两种情况,则可分为4种完备的情况。如图2所示,图2(a)中新注记投放到与其类型一致的单类型邻域;图2(b)中新注记投放到与其类型不一致的单类型邻域;图2(c)中新注记投放到复合邻域,并且与复合邻域中某类注记类型一致;图2(d)中新注记投放到复合邻域,并且与复合邻域中的任一种注记类型均不一致。

图2(a)是图2(c)的特例,新注记投放到与其类型相同的自相关邻域中,在类型数值的分布上,新注记的类型值不会使邻域类型均值、方差出现明显变化。所以,图2(a)、2(c)投放过程是正常与合理的。图2(b)是图2(d)的特例,由于新注记的类型不同于邻域类型,因此,用空间自相关性去评价是没有意义的。根据地理学第一定律,认为

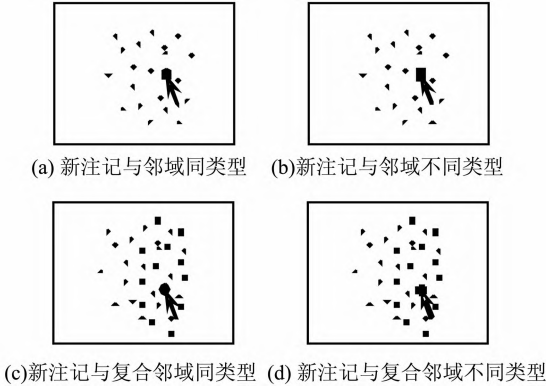


图 2 投放过程的 4 种情况

Fig. 2 Four Situations of Putting Progress

新笔记降维后类型值改变即是新笔记与邻域相关,并以此为依据建立情况图 2(b)、2(d)的评价。

根据式(5),在词频文本分类中,将 bga. text 中出现的所有词汇($term_1, term_2, \dots, term_n$)顺序地组成一个向量 $v = [freq_1 \ freq_2 \ \dots \ freq_n]$ 。其中,每个词汇在分类中的贡献是不同的,出现频率高的词汇对当前文本类型的作用大,相反,出现频率低的词汇作用小。根据 $freq_i$ 值的大小,对 $(freq_1, freq_2, \dots, freq_n)$ 作频率递减排列,得到向量 $v' = [freq'_1 \ freq'_2 \ \dots \ freq'_n]$,做降维函数 $ds()$:

$$ds_n[freq'_1 \ freq'_2 \ \dots \ freq'_n] = [freq'_2 \ \dots \ freq'_n], m = 1 \quad (12)$$

其中, m 为每次降低的维数,此处 $m = 1$ 。降维函数 $ds()$ 通过删除频率最高的项,使得 bga. text 分类最大可能地由当前类型发生改变。

设邻域的类型集合为 BGAG. CLS,对于任意新地理笔记 bga,满足下式条件的新地理笔记将不能被投放。

$$bga. cls \notin BGAG. CLS \wedge ds_n(bga) \notin BGAG. CLS \quad (13)$$

式(13)的物理意义是:如果新笔记类型不同于邻域类型,那么,对新笔记降维处理;如果降维后的类型仍不同于邻域类型(如果相同,虽然新笔记类型值不变,但降维后的词频向量类型发生改变,即由新笔记类型转变为邻域类型,表明新笔记与邻域有相关性,即其在这个邻域中的存在是合理的),那么,新笔记不能被投放到这个邻域。

3 算例与分析

实验地理笔记共两类,分别为云南省昆明市一环内 574 个餐饮、372 个商务类型笔记,每种类型预留 50 个笔记作为分类验证。笔记空间位置为 WGS 84 坐标系,采集形式为百度地图与手持

GPS 设备。笔记文本来源于美食评论网与百度搜索,文本最小长度 15 个汉字、最大长度 52 个汉字,平均长度为 34 个汉字。中文基础词汇来源于搜狗输入法,最后更新日期为 2013 年 7 月,共计约 110 万个汉语词汇,包含最新出现的网络用词汇。经分词处理,餐饮类地理笔记中文基础词汇为 100 015 个,单个词汇出现次数的范围为 $[375, 1]$,即词频范围为 $[0.716, 0.001]$;商务类地理笔记中文基础词汇为 95 310 个,单个词汇出现次数的范围为 $[146, 1]$,即词频范围为 $[0.453, 0.001]$ 。根据式(5),分别用 6、10、14、18、22 个词频构成的类型向量对两种预留笔记做分类验证,分类准确率均大于 94%,其中,14 词频向量准确率为 100%,说明选用词率值较大的 14 个词汇能够对两种地理笔记类型做到最优的分类。

在图幅范围内,随机生成 50 个空间位置样点,并根据式(11)做基于 Voronoi k 阶邻域的收敛性分析,如表 1 所示。50 个样点中的 42 个在 Voronoi 3 阶邻近以内条件下达到收敛,构成的邻域具有显著正空间自相关性,也就是说,地理笔记的空间分布并非是完全随机的,而是表现出相似类型之间的空间集聚。更加具体地,空间分布的联系特征是具有较高类型值的笔记相对地趋于与较高类型值的笔记相邻,或者较低类型值的笔记相对地趋于与较低类型值的笔记相邻。需要说明的是,样点中有 8 个位于图幅相对边缘,当邻近阶数增大时,邻近目标个数增加较少,致使其满足收敛条件时的邻近阶数 k 值较大,并且存在类型值大小的差异较大,其平均 I_i 值相对较小。

表 1 样点邻域收敛性分析

Tab. 1 Convergence Analysis of Sample Point Neighborhood

收敛 k 值	样点个数	平均 Voronoi 邻近目标数	\bar{I}_i	\bar{Z}
1	9	5.6	0.53	3.41
2	17	14.2	0.54	3.43
3	16	40.7	0.54	3.46
4	8	20.3	0.50	3.21

对投放模型进行验证,收集实验区范围内实际存在的 132 个住宿类笔记,并做分词,构成 24 500 个基础词汇,选取其中 50 个笔记进行 14 词频向量投放计算。依据式(12)降维后,50 个住宿类笔记均可以在实验区域内投放,其中,单个笔记最多需要 8 次迭代降维后,其词频向量类型即可变化至邻域类型。如图 3 所示,降维处理后的类型向量值发生显著变化,向量的邻域类型值均大于住宿类型值,由住宿类型指向其投放的邻域

类型,根据式(13),可以被投放。

收集 150 个某特种工业的文本描述,做分词构成 31 000 个基础词汇,并构造 50 个特种工业类注记(在实验区域范围内实际并不存在),同样做 14 词频向量投放计算,如图 4 所示。随机投放

位置选择与降维处理,相对于其所投放的邻域类型,类型向量中的词频值非常低(接近 0),虽然经过多次迭代降维,向量的邻域类型值均小于特种工业类型值,即不能通过降维后的类型向量与其邻域产生联系,因此,不能被投放。

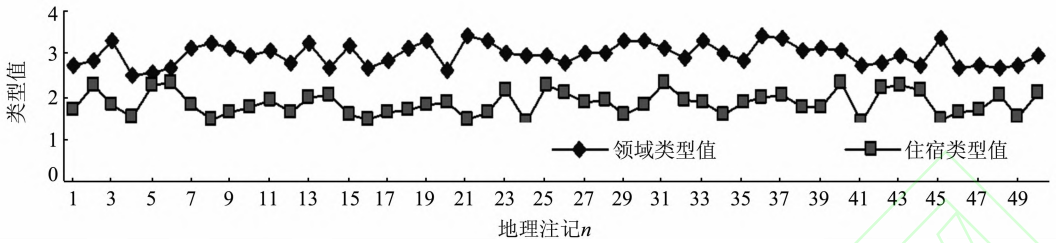


图 3 住宿类地理注记类型向量值

Fig. 3 Vector Value of Accommodation Data Set

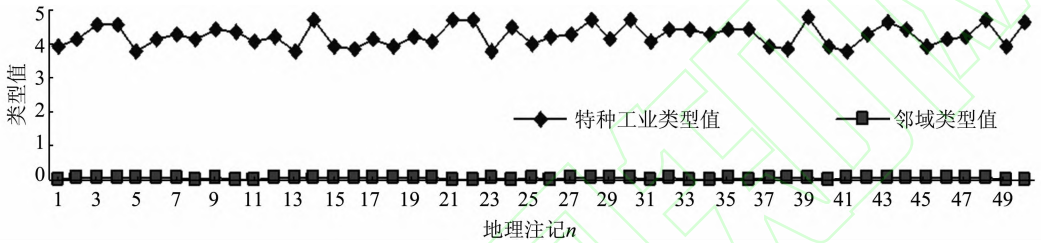


图 4 特种工业类地理注记类型向量值

Fig. 4 Vector Value of Special Industry Data Set

4 结 语

针对网络地图服务环境中的地理注记适宜性评价问题,本文首先将传统地理注记结构化内容扩展为非结构化的文本,形式化宽泛地理注记,并用投放过程来模型化网络地图环境中的标注行为,依据词频分类方法确定地理注记的类型,根据 Voronoi k 阶邻域建立投放域定量空间相关性描述,进而结合文本分类与空间自相关构建基于地理注记类型的投放模型。实验结果表明,注记存在的自相关邻域是成立的,并且,在两种已知地理注记情况下投放模型能够有效地对新增地理注记进行适宜性评价。

在实际应用环境中,地理注记的类型是多样的,注记文本的收集会导致词汇的不均衡,使得注记分类变得不准确,需要进一步研究。

参 考 文 献

[1] Ma Yaofeng, Hu Wenliang, Zhang Anding, et al. Cartography Theory [M]. Beijing: Science Press, 2004 (马耀峰,胡文亮,张安定,等. 地图学原理 [M]. 北京:科学出版社,2004)

[2] Mirko H, Wilko H, Andrea M, et al. Automatic

Annotation of Geographic Maps [C]. The 10th International Conference on Computers Helping People with Special Needs, Linz, Austria, 2006

[3] Zhang Xueying, Zhu Shaonan, Zhang Chunju. Annotation of Geographical Named Entities in Chinese Text [J]. *Acta Geodaetica et Cartographica Sinica*, 2012, 41(1): 115-120 (张雪英,朱少楠,张春菊. 中文文本的地理命名实体标注[J]. 测绘学报, 2012,41(1):115-120)

[4] Zhang Xueying, Zhang Chunju, Zhu Shaonan. Annotation of Geographical Spatial Relations in Chinese Text [J]. *Acta Geodaetica et Cartographica Sinica*, 2012, 41(3): 468-474 (张雪英,张春菊,朱少楠. 中文文本的地理空间关系标注[J]. 测绘学报, 2012,41(3):468-474)

[5] Wang Zhao, Wu Zhongheng, Fei Lifan, et al. Automatic Name Placement of Area Feature: A Metric Information Approach [J]. *Acta Geodaetica et Cartographica Sinica*, 2009, 38(2): 183-188 (王昭,吴中恒,费立凡,等. 基于几何信息熵的面状要素注记配置[J]. 测绘学报, 2009,38(2):183-188)

[6] Wu Changbin, Lv Guonian, Liu Yujun. Automated Numeric Placement for Land Utilization Map Based on Rule Database and Grid Algorithm [J]. *Acta Geodaetica et Cartographica Sinica*, 2008, 37(2): 250-255(吴长彬,闫国年,刘昱君. 基于规则库和网

- 格算法的土地利用现状图自动数字注记[J]. 测绘学报, 2008, 37(2): 250-255
- [7] Chen Rui, Zhang Zuxun, Zhang Jianqing. Identification of Digital Elevation Annotation in Scanned Map [J]. *Geomatics and Information Science of Wuhan University*, 2002, 27(2): 194-198 (陈睿, 张祖勋, 张剑清. 扫描地形图中数字高程注记的提取与识别[J]. 武汉大学学报·信息科学版, 2002, 27(2): 194-198)
- [8] Xie Chongfeng, Li Xing. A Sequence-Based Automatic Text Classification Algorithm [J]. *Journal of Software*, 2002, 13(4): 783-789 (谢冲锋, 李星. 基于序列的文本自动分类算法[J]. 软件学报, 2002, 13(4): 783-789)
- [9] Zhou Qian, Zhao Mingsheng, Hu Ming. Study on Feature Selection in Chinese Text Categorization [J]. *Journal of Chinese Information Processing*, 2004, 18(3): 17-23 (周茜, 赵明生, 扈旻. 中文文本分类中的特征选择研究[J]. 中文信息学报, 2004, 18(3): 17-23)
- [10] Tobler W. A Computer Movie Simulating Urban Growth in the Detroit Region [J]. *Economic Geography*, 1970, 46(2): 234-240
- [11] Chen Jun. Voronoi-Based Dynamic Spatial Data Model [M]. Beijing: Publishing House of Surveying and Mapping, 2002 (陈军. Voronoi 动态空间数据模型[M]. 北京: 测绘出版社, 2002)
- [12] Zhao Renliang. Voronoi Methods for Computing Spatial Relations in GIS [M]. Beijing: Publishing House of Surveying and Mapping, 2006 (赵仁亮. 基于 Voronoi 图的 GIS 空间关系计算[M]. 北京: 测绘出版社, 2006)
- [13] Zhang Chaosheng, Luo Lin, Xu Weilin, et al. Use of local Moran's I and GIS to Identify Pollution Hot-spots of Pb in Urban Soils of Galway, Ireland [J]. *Science of the Total Environment*, 2008, 398(1-3): 212-221
- [14] Shi Yimin. Mathematical Statistics (3rd Edition) [M]. Beijing: Science Press, 2009 (师义民. 数理统计(第三版)[M]. 北京: 科学出版社, 2009)

Putting Model for Broad Geographic Annotation

LI Jiatian¹ KANG Shun¹ LI Xiaojuan¹ ZHANG Lan¹ LUO Fuli¹ BAI Jianjun²

1 Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China

2 College of Tourism and Environmental Sciences, Shanxi Normal University, Xi'an 710062, China

Abstract: In view of the geographic annotation evaluation problem under the network map service environment, this paper extended the content of traditional geography annotation with the unstructured text form, and defined a broad geographic annotation. Using the putting process to model the annotation behavior, first of all, need confirm the geographic annotation text classification according to the frequency class, and establish a quantitative convergence description about the spatial correlation in the putting neighborhood according to the Voronoi k order adjacency relationship, proceed to the next step, build a putting model based on the geographical annotation types and type transformation combined with the geographic annotation text and the existing geographical annotation neighborhood. Through experimental verification, in the actual case of already known two typological geographic annotation sets, putting model can effectively evaluate the newly increased geographic annotation in a reasonable way.

Key words: geographic annotation; unstructured text; frequency class; Voronoi k order adjacency; spatial auto correlation

First author: LI Jiatian, associate professor, PhD, specializes in calculation of dynamic spatial relationship. E-mail: ljtwcx@163.com

Foundation support: The National Natural Science Foundation of China, Nos. 41161061, 40901197.