

基于位置签到数据的城市分层地标提取

王 明¹⁾ 胡庆武¹⁾ 李清泉^{2),3)} 秦龙焜¹⁾

¹⁾(武汉大学遥感信息工程学院 武汉 430079)

²⁾(深圳大学空间信息智能感知与服务重点实验室 深圳 518060)

³⁾(武汉大学测绘遥感信息工程国家重点实验室 武汉 430079)

摘 要 为了及时准确地获取用于智慧位置服务的城市层次性空间知识, 该文提出了一种依据签到属性显著度的差异从位置签到数据中提取城市分层地标的方法. 首先, 针对传统基于调查问卷的公众认知度计算方法存在的不足, 该文从位置签到数据中的签到次数、签到用户数和用户影响因子 3 个方面入手讨论 POI 显著度的影响因素, 进而提出了一种基于签到数据的 POI 显著度计算模型; 然后, 利用 Voronoi 图对 POI 空间作用范围进行分割并结合邻域分析与合并算法逐层迭代, 形成了基于签到 POI 显著度与 Voronoi 图的分层地标提取算法; 最后, 以北京市的位置签到数据为例, 进行了 POI 显著度计算与分时分层地标提取, 通过与现有实体、网络地图资料的比对, 验证了算法的有效性, 并进一步分析了分层、分时地标反映的空间现象和规律, 为个性化智慧位置服务提供城市层次性空间知识.

关键词 位置服务; 位置签到(Check-In); 地标; POI; 显著度; 分层; 云计算

中图法分类号 P208 **DOI号** 10.11897/SP.J.1016.2016.00405

Extracting Hierarchical Landmark from Check-In Data

WANG Ming¹⁾ HU Qing-Wu¹⁾ LI Qing-Quan^{2),3)} QIN Long-Kun¹⁾

¹⁾(School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079)

²⁾(Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060)

³⁾(State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079)

Abstract For acquiring the city hierarchical spatial knowledge to be applied in smart location service on time, a method of extracting city hierarchical landmarks from check-in data according to their check-in attribute significance is proposed. To overcome the disadvantage of traditional hierarchical landmarks extracting method of public cognition based on questionnaires, a POI significance measure model is constructed after analyzing the factors influencing the significances of POI objects from three vectors which are check-in numbers, check-in users, user impact factor in check-in data. Then, combined with neighborhood analysis and merge algorithms, the POI space scope is segmented into Thiessen polygons with Voronoi diagrams iteratively, that is the hierarchical landmark extracting algorithm based on POI significance and Voronoi diagrams. An experiment is carried out to compute the significances of the POIs selected from the area of Beijing City, and the POI in the Thiessen polygon with higher significance than its surroundings are treated as landmarks in each level at last. The paper confirms the efficiency of the algorithm by comparing the hierarchical landmarks with real objects and electronic map data, analyzes the spatial phenomenon and regularity reflected by the hierarchical time-sharing landmarks got from check-in data. This

收稿日期:2014-04-16;在线出版日期:2014-12-19. 本课题得到国家自然科学基金(61172175)资助. 王 明,男,1988年生,博士研究生,主要研究方向为众源签到数据分析与挖掘. E-mail: wangming19880910@126.com. 胡庆武,男,1975年生,博士,教授,主要研究领域为3S集成与应用. 李清泉,男,1965年生,博士,教授,主要研究领域为智能交通、3S集成与应用. 秦龙焜,男,1989年生,硕士研究生,主要研究方向为众源签到数据分析与挖掘.

can be used to provide hierarchical spatial knowledge of city for personalized intelligent location service.

Keywords location service; Check-In; landmark; Point Of Interest; significance; hierarchical; cloud computing

1 引 言

人们在认知空间位置、传递空间信息的过程中,相对于定量、精确的空间坐标定位方式,往往更倾向于定性、模糊的空间关系描述方式,即选择一个大众所熟知且具有鲜明代表性的空间实体作为参照物,通过描述目标与参照物之间的空间关系来反映目标的空间位置、传递相关空间信息.而在这一过程中,地标就是最好的参照物.

地标(Landmark)是指在一定的空间范围内,能被一定的认知人群所熟悉,并且在其心理和地理意义上都能成为标志的空间形态^[1].从其定义来看,作为一种标志性的空间形态,一方面,地标应当在大众的心理层面上为人们所熟知,只有其具有了一定的公众认知度,才能够作为一种参照物来辅助描述待定目标的空间位置、传递空间信息;另一方面,地标还应当具有一定的范围属性与地理意义,例如在路径导航案例中,系统在与用户的交互过程中发现陌生环境下用户可认知的高层级地标,并在此基础上进行的空间推理,发现导航目的地与实时位置之间的其他相关地标,从而辅助用户选择导航决策点,最终实现路径导航寻址分析,而这一过程往往依赖于建立在导航数据库中具有地理环境显著性与层次结构性的地标数据^[2].因此,如何衡量地标的公众认知度、划分地标的层级结构特征并形成一种准确有效的分层地标提取方法一直是相关领域专家学者研究的热点.

Sorrow 和 Hirtle^[3]最早将地标分为了视觉地标、认知地标和结构地标三类,他们通过分析三类地标的内在联系提出地标应当具有视觉、语义与结构这 3 个方面的显著特征.在此基础上,一些学者提出了用地标显著度来衡量其公众认知性并进行地标筛选的思路. Raubal 和 Winter^[4]从视觉吸引力、语义吸引力、结构吸引力 3 个方面构造了地标显著度计算模型.赵卫锋等人^[5]则从公众认知、空间分布和个体特征 3 个方面讨论影响兴趣点(Point of Interest, POI)显著性的要素,并构建了基于公众认知度、城市中心度和特征属性值的 POI 显著度计算模型.然

而这些地标显著度计算方法一般需要人为拟定参数、确定权重或者通过专家打分的方式进行估值,这就使得最终结果往往存在一定的主观性与不确定性.

此外,也有一些学者借助数据挖掘的算法思想发现数据内在趋势与联系,提取显著性的地标. Elias 和 Brenner^[6-7]采用 ID3 分类和 Cobweb 聚类算法对城市规划 GIS 空间数据进行了空间分析与地标提取. Tomko 和 Tezuka 等人^[8-9]则主要侧重于对来自互联网的海量网络数据资源进行挖掘,从而获取不同场景下的城市地标.但是这些基于数据挖掘技术的地标提取方法一方面受限于数据的来源、属性、类型等条件,很难具有通用性;另一方面往往忽视了地物要素的地理属性,造成提取出来的地标存在一定程度上的空间特征缺失.

对于地标层级结构的划分,传统基于地标显著度阈值的分层方式往往存在无法客观、合理选择各层级地标显著度值域范围与临界阈值的缺陷^[5]. Winter 等人^[10]提出了一种基于 Voronoi 图的地标分层方法,在以地标的显著度为指标的基础上采用 Voronoi 图的区域性位置描述特征划分地标层级.龚咏喜等人^[11]则以北京大学为例,探讨了基于带权 Voronoi 图与地标的空间位置描述的相关方法.相较于传统阈值分层的思想,这种通过研究地标空间作用范围变化来反映地标层级结构的思路不失为一种更加客观、合理的解决方案.

鉴于前人在分层地标提取相关领域研究的经验与不足,本文提出了一种基于位置签到数据和 Voronoi 图的城市分层地标提取算法.本文分析了位置签到数据中影响 POI 显著度的因素,讨论了以位置签到数据中的签到次数、签到用户数和用户影响因子来计算 POI 显著度的方法,并结合 Voronoi 图对地标空间作用范围的分割特征,以北京市为例展示了逐层迭代提取城市分层地标的过程.

2 基于位置签到数据的 POI 显著度计算

2.1 位置签到数据

众源地理数据(Crowd Sourcing Geographic

Data)是指由普通用户自发地借助非正式社会网络和 Web 2.0 技术采集得到并向公众开放的大众地理信息数据^[12]。由于众源地理数据是公众志愿采集与传播的地理信息,也有人称之为志愿者地理信息(Volunteered Geographic Information, VGI)。与传统的地理数据相比,众源地理数据具有数据量大、覆盖面广、内容丰富、采集效率高、成本低等特点。

近年来,伴随着无线通信技术的逐步成熟以及智能移动终端设备的广泛应用,一种基于位置服务(Location-Based Services, LBS)的互联网应用-签到(Check-In)得到了飞速的发展与普及。用户通过能够定位的移动终端设备(如手机)与无线通信技术连接互联网并分享自己所在的地理位置与信息,这一过程就积累了能够反映用户的日常生活轨迹与行为的海量位置签到数据,而这种位置签到数据正是一种典型的众源地理数据。

由于众源地理数据来源于大众的日常生活,其内在包含了这类人群的社会行为特征。通过对这种众源地理数据进行深入挖掘,能够得到反映社会事件、行为主体的空间趋势规律与特征知识。具体结合分层地标提取来看,针对位置签到数据的挖掘研究不仅能够借助相关签到属性反映其地标显著性特征,而且数据本身的地理位置属性也避免了传统数据挖掘过程中空间特征的缺失现象。

2.2 签到属性显著性表征

位置签到数据记录了一段时间内大众用户在每个 POI 上总的签到次数、签到人数信息,这些信息作为一种签到属性在一定程度上客观地反映了各个 POI 的热度特征,即签到次数多、签到人数多的 POI 相对于签到次数少、签到人数少的 POI 而言热度更高。同时,签到 POI 记录了人们日常的生活行为轨迹,其签到属性反映的是各个点的人流量与受关注程度,以其作为 POI 显著性的表征,不仅数据本身客观、量化,而且由数据反映出的信息也具有实际的社会意义,符合社会规律与普遍性。表 1 记录了来自街旁网与新浪网的北京市几处高值签到 POI 数据以及其对应区域低值签到 POI 数据。

表 1 高、低值签到 POI 对比

POI 名 (高值)	新浪 签到	街旁 签到	POI 名 (低值)	新浪 签到	街旁 签到
北京南站	11 570	13 924	北京南站地铁站	195	1513
国家体育场/ 鸟巢	11 207	13 089	凯迪克兰云天 大酒店	149	246
天安门广场	10 862	7832	天安门东地铁站	129	564
颐和园	5524	8953	国际关系学院	314	376

通过对比可以明显看出:在一个邻域内高值签到 POI 相对于低值签到 POI 而言,无论在语义层面还是功能层面都具有更大的影响力,更突出的显著性,更加符合地标的空间认知要求。因此,在衡量 POI 显著程度时,位置签到数据的签到属性可以作为一个合适的定量描述指标,这与传统的地标显著性度量方法相比,避免了在打分、定权方面可能存在的主观因素影响,使得提取的结果更加客观、真实。

2.3 签到用户影响因素

由于 POI 的签到属性来源于大众用户的签到行为,不同用户的签到行为往往具有不同的现实意义,因而在以 POI 签到属性作为地标显著度指标时还应当考虑不同用户签到行为的差异。

现实生活中,长期生活在一个城市的本地居民相对于外来人员、游客而言会更加熟悉当地的实际情况,这是因为人们对特定环境的认识往往是伴随着其经验、阅历的丰富而逐渐加深的。在衡量 POI 显著度的过程中亦是如此:在特定环境中,阅历、经验更加丰富用户的签到行为相对于阅历、经验较少用户的签到行为而言应该更具有代表性与说服力,因此来自不同用户的签到行为应当根据用户本身的影响因素具有不同的显著意义。

郑宇和谢幸^[13]曾提出可以参考类似 HITS 网页排名算法的相关思想,通过统计不同用户的信息计算不同用户的经历经验值,从而使不同用户的空间行为具有不同的影响力。本文在此基础上,主要通过对详细签到记录集中的用户签到记录进行统计分析,确定每个用户的个人影响因素(*User_IF*)。表 2 记录了详细签到记录集中每次签到记录所包含的属性。

表 2 详细签到记录属性表

属性类型	属性说明	属性类型	属性说明
Body	详细签到信息	User_City	签到用户所在地
Create_On	创建签到时间	User_Name	签到用户昵称
Client_Type	签到客户端类型	User_Sex	签到用户性别
Guid	签到 POI 唯一标识	User_Birthday	签到用户生日
Type	签到类型	User_Id	签到用户唯一标识

首先,详细签到记录集中的每次签到记录凭借 Guid 与 User_Id 两个属性分别与签到 POI 以及签到用户一一对应,通过 User_Id 属性可以统计出每个用户已经签到过的 POI 个数指标(*POI_Num*),而 *POI_Num* 这一指标可以定量地反映出在特定环境中用户签到的阅历、经验情况,即签到 POI 个数越多,用户签到的阅历、经验越丰富。

然后,User_City 属性记录了签到用户设置的用

户所在地,通过这一属性与 POI 所在地进行匹配可以判断签到用户是否为 POI 所在地的当地用户,从而定性地区分用户个人影响权重($User_Weight$),若为本地用户则其权重为 1,若为非本地用户则其权重减半。

最后,通过定性、定量两方面的分析结果得到签到用户个人影响因子的计算公式:

$$User_IF = POI_Num \times User_Weight.$$

此外,还可以适当调整不同地域用户权重的权重比例,以适应不同生活环境、用户群体可能存在的实际差异。

2.4 签到 POI 显著度计算

综上,本文以位置签到数据为基础,主要从签到用户的签到属性($Check_Num$,表示用户在某 POI 签到的次数; $User_Num$,表示在某 POI 签到的用户人数)及其影响因子($User_IF$)两个方面入手:一方面以客观实际的数据定量反映了 POI 的显著性程度;另一方面也考虑到了用户在签到过程中可能存在的人为主观影响因素,最终提出了一种基于 POI 经验得分($POIExp$)的 POI 显著度的计算模型:

$$POIExp = \sum_{User_Num} User_IF \times Check_Num.$$

通过这种基于 POI 经验得分的显著性指标可以直观地反映 POI 的显著程度以及多个 POI 之间的显著关系,表 3 记录了北京市 10 个签到高值 POI 基于经验得分的显著度统计情况。

表 3 基于签到属性与经验得分的 POI 显著度

POI 名	签到人数	签到排名	POI 显著度	显著度排名
北京南站	13924	1	699910	8
国家体育馆/鸟巢	13089	2	941101	1
798 艺术区	10772	3	829957	2
宜家家居(四元桥店)	10131	4	785213	3
故宫博物院/故宫	9988	5	743298	6
颐和园	8953	6	698969	9
万事达中心	8841	7	780721	4
后海	8399	8	760842	5
北京西站	7994	9	504730	34
中国国家大剧院	7948	10	741520	7

从表 3 的统计结果来看,基于经验得分的 POI 显著度指标反映出的显著地标 POI 与签到属性趋势大体一致,差别主要存在于北京南站、北京西站两个大型铁路交通站点上面。作为北京市两个主要的铁路客运枢纽,其在现实生活中都是人流量非常大的地方,从签到人数属性反映出的人流量信息来看,基本符合这一实际情况。

然而,从 POI 本身的功能特征来看,由于火车

站运输的客流中有很大一部分是外来非北京客源,这些签到用户在北京这一特定环境下的用户影响因子较低,从而影响了 POI 总体的显著度指标。此外,作为运营京津、京沪高铁的主要站点,北京南站承载了更多的北京客源,而作为北京与全国各地客运集散的主要节点,北京西站的客流则以外来人员为主,这也就导致即使同为大型铁路客运枢纽站点,二者 POI 显著度受到的影响情况也不尽相同。

由此可见,基于 POI 经验得分的 POI 显著度指标能够更加充分地反映出 POI 本身的实际情况与功能特性,更加客观合理地体现出签到用户真实的日常生活特征与社会行为模式,更加全面、综合地表征地标的显著程度。

3 基于 Voronoi 图的空间分割与地标分层

3.1 Voronoi 图

Voronoi 图是计算几何中一种几何结构,也是一种空间分割的方法。二维平面上的 Voronoi 图中的多边形通常被称为泰森多边形。在几何空间中, Voronoi 图中的泰森多边形代表了各种子点相应的空间描述或作用范围。Voronoi 图的定义如下:假设 $V = \{V_1, V_2, \dots, V_n\}$, $N \geq 3$ 是欧几里得平面 R 上的一个点集,并且这些点不共线,任意 4 点不共圆。用 $d(v_i, v_j)$ 表示 v_i 和 v_j 之间的欧几里得距离,设 P 为平面上的点,则区域 $V_{(i)} = \{P \in R^2 \mid d(P, V_i) \leq d(P, V_j), i, j = 1, 2, \dots, n, i \neq j\}$ 称为泰森多边形,各点的泰森多边形共同组成 Voronoi 图。

平面上的 Voronoi 图可以看做是点集 V 中的每个点作为生长核,以相同的速率向外扩张,直到彼此相遇为止而在平面上形成的图形。除了最外层的点形成的开放的区域外,其余每个点都形成一个凸多边形(泰森多边形)。在 Voronoi 图中,任意一个泰森多边形中的任意一个内点到该泰森多边形控制点 V_i 的距离都小于该点到其他任何控制点 V_j 的距离。

3.2 Voronoi 图地标空间范围分割

在 Voronoi 图中,泰森多边形通常用于生成“领地”或控制区域。而且由于 Voronoi 图中的每个控制点唯一的被一个泰森多边形包含,因此可以用 Voronoi 图明晰地表达控制点之间的空间相邻关系。

结合人类日常行为习惯与社会活动准则来看,人们通常会选择最短时间、最近距离、最低成本以及最优路线来解决生活中遇到的问题,这就反映出人

类所遵循一种的空间行为准则. 从这一层面来说, 如果将 Voronoi 图中的种子点理解为人类日常行为或活动的出发点或目的地, 那么 Voronoi 图内的各个泰森多边形就反映了其对应行为活动的空间参考或影响的范围. 王新生等从这个角度出发, 结合河南省城市中心强度参数, 探讨了将 Voronoi 图应用于确定城市空间影响范围的效果与意义^[14].

以此类推, 若将 Voronoi 图中的种子点理解为在一定环境条件与层次条件下的地标点, 那么 Voronoi 图内的各个泰森多边形就反映了作为种子点的地标在空间上所能够描述或影响的范围. 龚咏喜等人^[11]以北京大学为例, 探讨了基于带权 Voronoi 图的地标空间位置描述方法, 通过空间认知实验验证了 Voronoi 图对地标空间范围的描述.

基于 Voronoi 图的地标空间作用范围分割一方面实现了 POI 在最短距离原则下所覆盖的空间范围的划分, 另一方面也很好处理了 POI 在局部高值聚类情况下地标空间作用范围分割与均衡化的问题, 为后续地标分层提取提供了空间连续的作用范围邻域要素.

3.3 地标邻域分析与分层提取

分层地标是指一个地标可以同时隶属于不同的地标层级, 其在较高级别下的空间作用范围要远大于在较低层级下的空间作用范围. 这种层级性结构不是一个孤立分割的结构, 而是较高级别地标作为低层级地标的子集存在, 而在不同层级表现出不同的空间影响力. 这种层级性结构在空间搜索、认知、导航的过程中发挥着重要的作用: 在一个陌生的环境里, 显著性较大、层级较高的地标为用户提供一个粗略的全局认知, 可以指定大体的导航方向与区域范围; 显著性较小、层级较低的地标则为用户提供了一个精确的局部认知, 用于确定具体的导航目的地与路径.

由分层地标的这一特点出发, 对于一个显著度较高的地标而言, 地标本身层级自下而上的增加势必会导致其空间作用范围的逐步扩大, 而这种空间作用范围的扩大其本质就是对邻域内低显著度地标空间作用范围的空间合并与邻域增长. 因此, 便可以从地标作用范围空间邻域的角度入手研究地标的层级性划分.

本文借助 Voronoi 图对低层级 POI 进行空间作用范围分割, 然后遍历每个 POI 所在泰森多边形的最小邻域(直接邻接), 通过比较邻域内 POI 的显著度指标的大小, 选择一个相对于当前邻域而言显

著度最高的 POI 作为当前邻域的下一层级地标, 继而逐步迭代这一过程, 最终实现自下而上的分层地标提取. 地标提取的计算公式如下式所示

$$\begin{cases} POIEx p_{l_{k,\max}} = \max\{POIEx p_{l_j} \mid dist(l_j - l_k) \leq 1, \\ l_j \in L^i\} \\ L^{i+1} = \{l_r^{i+1}\} = \{l_{k,\max} \mid l_k \in L^i\} \end{cases}$$

式中: $POIEx p_{l_j}$ 表示地标 l_j 的显著度指标值; $dist(l_j - l_k)$ 表示地标 l_j 与 l_k 之间基于狄洛尼三角形的拓扑距离, 其距离不大于 1 说明地标 l_j 是地标 l_k 的邻域地标; L^i 表示第 i 层所有地标的集合; $POIEx p_{l_{k,\max}}$ 表示地标 l_k 的邻域地标的显著度的最大值, $l_{k,\max}$ 表示地标 l_k 邻域中显著度最突出的地标; $\{l_{k,\max} \mid l_k \in L^i\}$ 表示第 i 层中每个地标的邻域中显著度最突出的地标的集合; L^{i+1} 表示第 $i+1$ 层所有地标的集合. 分层地标的提取流程如图 1 所示.

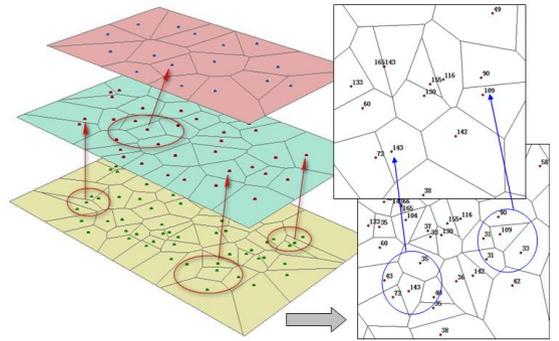


图 1 基于 Voronoi 图的地标分层模式

4 实验与分析

4.1 分层地标提取实验

综合前文关于 POI 显著度与地标空间作用范围分割方面的分析研究, 本文提出了一种基于 POI 显著度和 Voronoi 图的分层地标提取思路, 具体算法流程如图 2 所示.

其实现步骤如下:

1. 数据预处理. 首先, 结合详细签到记录集中的签到记录对原始 POI 数据进行更新, 确保 POI 数据的准确性与现势性; 然后, 对 POI 数据的完整性进行验证, 补充 POI 类别信息; 对于用户在签到过程中产生的同义 POI 数据, 可以通过 POI 归并的方式消除数据冗余; 最后, 筛选符合实验目的与要求的 POI 数据完成数据准备工作.

2. 计算用户影响因子. 针对详细签到记录集中的用户信息进行分类统计, 得到用户的签到 POI 数与个人影响权重, 进而计算出每位用户的影响因子指标.

3. 计算 POI 显著度. 对于经过预处理得到的 POI 数据, 通过 GUID 关联其签到用户信息, 并依照 POI 经验得分

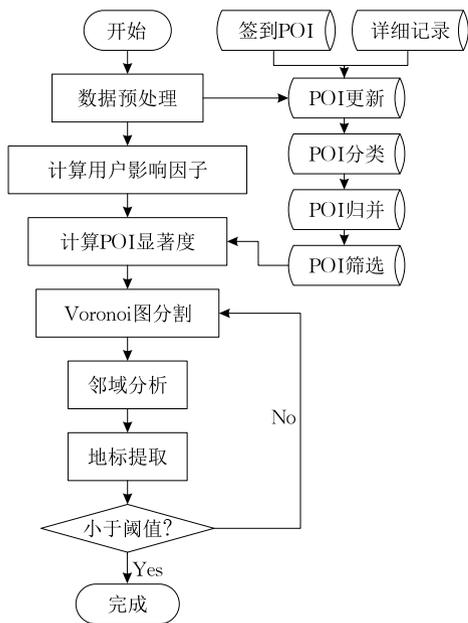


图2 分层地标提取算法流程

计算模型计算各 POI 显著度指标。

4. Voronoi 图分割。以当前层级 POI 作为种子点生成 Voronoi 图,实现对地标空间作用范围的分割。

5. 邻域分析与合并。遍历每个 POI 的空间邻域,通过比较邻域中 POI 显著度指标,选择显著度最高的 POI 作为当前邻域的地标 POI。

6. 提取下一层级地标。提取各邻域中显著度最高的 POI 作为下一层级地标 POI。

7. 阈值判断。如果提取得到的下一层 POI 个数小于预先设定的阈值,则地标提取工作结束;否则回到步 4 重新对下一层 POI 进行 Voronoi 图分割、邻域分析与迭代提取。

本文选择北京市来自街旁网与新浪微博的位置签到数据作为实验数据,按照上述算法流程进行分层地标提取实验,得到分层地标结果如表 4 所示,其中较高层级地标如图 3 所示。

表 4 分层地标提取结果

地标层级	地标个数
1	904
2	268
3	106
4	46
5	24
6	15
7	7

首先,从总体实验提取结果来看,原始 904 个地标 POI 被抽取为 7 个层级,各层级数量分别为 904 个、268 个、106 个、46 个、24 个、15 个和 7 个,基本达到了地标分层的目的与效果。北京南站、国家体育馆、798 艺术区、故宫、颐和园、万事达中心、天安门广场、工人体育场、北京大学等 15 个较高层级地标



图3 北京高层级地标分布

(最高层、次高层)均匀的分布于北京市五环线以内东城、西城、海淀、朝阳、丰台、石景山六大主要中心城区,反映了高层级地标的全局空间认知、描述。

其次,从高层地标的类别组成来看,主要包括了交通枢纽类地标(如北京南站)、商业购物类地标(如宜家家居)、教育院校类地标(如北京大学)以及旅游景点类地标(如故宫、颐和园),种类较为丰富。但是,也明显存在着一些种类地标的缺失,如机关单位、医疗机构类地标等。虽然这些类型的 POI 也存在于签到 POI 的数据库中,但是往往其签到的相关属性较低,造成这些类型的地标显著性评价受到影响。

此外,为了验证分层地标提取实验结果的正确性,本文选择从正式出版的实体地图和互联网电子地图上提取地标以用于对比验证^[15]。实体地图方面,本文选择的是由北京市测绘设计研究院编辑、地质出版社出版的《2011 北京交通旅游地图》,通过对地图上地标的字体大小、颜色等标注风格来筛选不同层级的地标要素;而互联网电子地图方面,本文选择了受商业推广影响较小的谷歌在线地图服务,通过不同比例尺的地图缩放来提取不同层级下的地标要素。在与两种地图载体上高层级地标要素的比对之后发现:实验中提取的交通枢纽、教育院校、旅游景点三类地标较为丰富且匹配程度较高,效果较好;商业购物、生活服务类地标结果基本能够匹配,但数量明显不足,占总体比例不高;而机关单位、医疗机构、住宅小区类地标则缺失严重,而且整体基于签到属性的 POI 显著度指标较低。产生这种现象的原因是由于位置签到数据产生的主体是学生和年轻的白领阶层:(1)他们比较有活力,喜欢外出玩乐,热衷于通过签到分享自己的生活,因而签到数据中交通枢纽、教育院校和旅游景点的地标会比较丰富;(2)他们由于经济条件的制约,在消费时都会有所选择和侧重,比如较少去高档场所消费、比较倾向于去连锁超市购物等,从而导致商业购物、生活服务类地标的数量会有所不足;(3)他们中住住宅小区和在机关单

位或医疗机构工作的很少,而且由于身体健康状况良好,去医疗机构看病的比例也很低,所以这一类地标在签到数据中缺失比较严重。

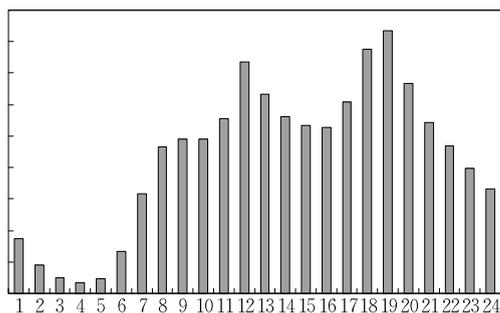
由此验证了本文分层地标提取思路的可行性,同时也清楚地认识到基于 POI 经验得分的地标提取方法由于签到数据在数据完整性、准确性方面存在的缺陷的影响有其局限性,因此比较适合作为现有地标构建方法的验证与补充。

4.2 分时签到趋势与地标规律发现实验

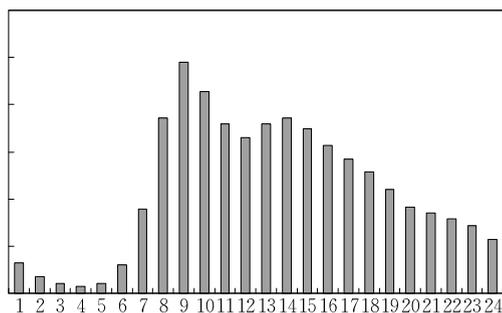
一方面,由于用户在签到过程中其签到行为本身就具有一个时间戳,而作为用户签到行为的产物,签到数据也必然包含着一定的时间特征,即用户签到行为、属性应当具有一定的时间趋势。图 4(a)为针对北京市 24 小时的签到时间趋势统计图,从图上反映的情况来看:夜晚时段总体签到量较少,而上午、下午、晚上 3 个时段签到量依次呈上升、稳定和下降的趋势,其中中午 12 点、下午 19 点为变化的两

个峰值,这一总体的签到量变化趋势完全符合日常生活中人们普遍的作息规律^[16]。

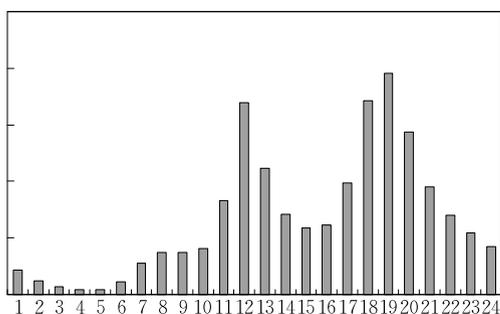
另一方面,POI 根据其本身语义与功能的定义一般可以被归并为不同的种类,比如餐饮酒店类、旅游景点类、商业购物类、交通枢纽类等。这种类别特性使得不同类别的 POI 往往具有不同的时间特征,即同一个 POI 在不同的时间段内可能会存在不同的显著度。图 4(b)、图 4(c)、图 4(d)分别针对办公单位、餐饮酒店、酒吧夜店 3 类有明显时间特征的代表性 POI 进行了签到量的统计。从统计结果来看,办公单位类 POI 峰值出现在上午 9 点,在工作时段签到量较高且稳定,而非工作时段签到量较低;餐饮酒店类 POI 全天签到量变化幅度较大,其中峰值分别出现在中午 12 点与晚上 19 点,与调查所得北京人的餐饮时间一致^[17];酒吧夜店类 POI 在夜晚时段(21 点~3 点)签到量较高,而日间时段签到量则很少。



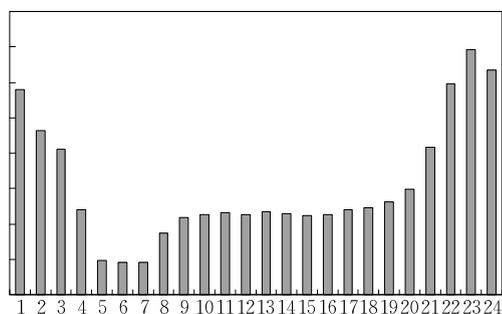
(a) 总体分时签到趋势分布图



(b) 办公单位类POI分时趋势图



(c) 餐饮酒店类POI分时趋势图



(d) 酒吧夜店类POI分时趋势图

图 4 基于分时特征的签到趋势分布图

由此可见,POI 的签到属性具有明显的时间趋势,通过对这种时间特征的挖掘即可获得具有分时特征的 POI 显著度,进而可以实现基于分时 POI 显著度的地标提取。

本文根据签到时间戳将签到数据分为 3 个时段(1 点~7 点、8 点~18 点、19 点~24 点),分别计算各时段基于签到属性的 POI 显著度指标,进而实现具有时间特征的分时分层地标提取。对于提取得到

的各时段高层地标结果,剔除所有在 3 个时段都会出现的常见综合性高层级地标(如北京南站、北京西站、国家体育馆、颐和园、圆明园等),最终得到 3 个时段所特有的高层地标如表 5 所示。

从表 5 的统计结果来看,3 个时段特有高层级地标的类别、功能特性与各时段人们日常生活的行为习惯基本一致。1 点~7 点时段的特有高层地标主要为酒店、迪厅、酒吧,因为这里是人们夜生活的主

要场所;8点~18点时段的特有高层地标基本都是重要的旅游景点、公园,而这些旅游景点、公园大多只会在白天开放;19点~23点时段的特有高层地标类别相对较多,包括影城剧院、运动休闲、商业购物等,不过这些地标也都与人们夜晚休闲娱乐的生活节奏相适应。

表 5 各时段特有高层级地标

时间段	高层级地标
1点~7点	Begins、School 酒吧、雅乐轩酒店、时态空间酒吧、奇遇花园咖啡馆、Club Bing、All-Star Sport Bar、唐会酒吧俱乐部、MIX Club、工人体育馆
8点~18点	宜家家居、国家博物馆、故宫、石景山游乐园、天坛公园、香山公园、北京植物园、烟袋斜街、北京动物园、北京大学
19点~24点	宜家家居、水立方、国家大剧院、工人体育馆、西单大悦城、耀莱成龙国际影城、万达国际影城、School 酒吧、Begins、蜂巢剧场、奇遇花园咖啡馆

5 结 论

综上所述,本文通过对位置签到数据的空间数据挖掘与利用,构造了基于签到属性的 POI 显著度计算模型,并在此基础上结合 Voronoi 图的空间分割功能,实现了基于 Voronoi 图与 POI 显著度的分层地标提取算法。实验部分对北京市签到 POI 进行分层地标提取,并与现有相关研究成果以及常用地图资料进行了对比验证。实验结果表明,本文提出的分层地标提取方法在交通枢纽、教育院校、旅游景点、商业购物和生活服务类等与大众生活密切相关的地标上提取效果很好,但是在机关单位、医疗机构、住宅小区类的地标提取上存在缺失,这又是由签到数据的局限性决定的。同时,在此基础上本文分析了签到数据所具有的时态趋势特征,通过提取不同时段的特有地标要素,分析了 POI 的语义、功能、类别属性与时间特征的关系,进而通过分时特征地标反映了人类日常生活行为习惯特征。

参 考 文 献

- [1] Bu Wei, Zhang Ling-Ling. Space hierarchical structures of urban landmark system planning: Taking Wuhan city as the example. *Huazhong Architecture*, 2009, 27(1): 136-137(in Chinese)
(卜巍, 张伶俐. 城市地标系统规划的空间层次研究—以武汉市为例. *华中建筑*, 2009, 27(1): 136-137)
- [2] Hu Huo-Sheng, Gu Dong-Bing. Landmark-based navigation of industrial mobile robots. *International Journal of Industry Robot*, 2000, 27(6): 458-467
- [3] Sorrows E, Hirtle S. The nature of landmarks for real and electronic spaces//Freksa C, Mark M eds. *Spatial Information Theory*. Berlin: Springer, 1999: 37-50
- [4] Raubal M, Winter S. Enriching wayfinding instructions with local landmarks//Egenhofer J, Mark M eds. *Geographic Information Science*. Berlin: Springer, 2002: 243-259
- [5] Zhao Wei-Feng, Li Qing-Quan, Li Bi-Jun. Extracting hierarchical landmarks from urban POI data. *Journal of Remote Sensing*, 2011, 15(5): 981-988(in Chinese)
(赵卫锋, 李清泉, 李必军. 利用城市 POI 数据提取分层地标. *遥感学报*, 2011, 15(5): 981-988)
- [6] Elias B. Extracting landmarks with data mining methods//Kuhn W, Worboys F, Timpf S eds. *Spatial Information Theory*. Berlin: Springer, 2003: 375-389
- [7] Brenner C, Elias B. Extracting landmarks for car navigation systems using existing GIS databases and laser scanning//*Proceedings of the ISPRS Workshop 'Photogrammetric Image Analysis'*. Munchen, Germany, 2003: 131-136
- [8] Tomko M. Case study-assessing spatial distribution of web resources for navigation services//*Proceedings of the 4th International Workshop on Web and Wireless Geographical Information System W2GIS 2004*. Goyang, Korea, 2004: 90-104
- [9] Tezuka T, Tanaka K. Landmark extraction: A web mining approach//Cohn G, Mark M eds. *Spatial Information Theory*. Berlin: Springer, 2005: 379-396
- [10] Winter S, Tomko M, Elias B, Sester M. Landmark hierarchies in context. *Environment and Planning B: Planning and Design*, 2008, 35(3): 381-398
- [11] Gong Yong-Xi, Liu Yu, Wu Lun, et al. Locality description based on multiplicatively weighted Voronoi diagram and landmarks. *Geography and Geoinformation Science*, 2010, 26(4): 21-26(in Chinese)
(龚咏喜, 刘瑜, 邬伦等. 基于带权 Voronoi 图与地标的空间位置描述. *地理与地理信息科学*, 2010, 26(4): 21-26)
- [12] Heipke C. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010, 65(6): 550-557
- [13] Zheng Yu, Xie Xing. Enable smart location-based services by mining user trajectories. *Communications of the CCF*, 2010, 6(6): 23-30(in Chinese)
(郑宇, 谢幸. 基于用户轨迹挖掘的智能位置服务. *中国计算机学会通讯*, 2010, 6(6): 23-30)
- [14] Wang Xin-Sheng, Li Quan, Guo Qing-Sheng, et al. The generalization and construction of Voronoi diagram and its application on delimitating city's affected coverage. *Journal of Central China Normal University (Natural Sciences)*, 2002, 36(1): 107-111(in Chinese)
(王新生, 李全, 郭庆胜等. Voronoi 图的扩展、生成及其应用于界定城市空间影响范围. *华中师范大学学报(自然科学版)*, 2002, 36(1): 107-111)
- [15] Beijing Institute of Surveying and Mapping; Beijing Tourist Map 2011. Beijing: Geological Publishing House, 2011(in Chinese)

(北京市测绘设计研究院. 2011 北京交通旅游地图. 北京: 地质出版社, 2011)

- [16] Wang Qi-Yan, Lei Tao, Shi Lei. Sleeping time of residents—life time distribution survey of Beijing residents. *Beijing Statistics*, 2003, 156(2): 92-93(in Chinese)
(王琪延, 雷涛, 石磊. 居民睡眠用时多少——北京居民生活

时间分配调查系列报告. 北京统计, 2003, 156(2): 92-93)

- [17] Wang Qi-Yan, Lei Tao, Shi Lei. Dining time of residents—life time distribution survey of Beijing residents. *Beijing Statistics*, 2003, 155(1): 35-36(in Chinese)
(王琪延, 雷涛, 石磊. 一日三餐用时多少——北京居民生活
时间分配调查系列报告. 北京统计, 2003, 155(1): 35-36)



WANG Ming, born in 1988, Ph. D. candidate. His research interests include data analysis and mining of crowd sourcing check-in data.

HU Qing-Wu, born in 1975, Ph. D., professor. His research interests include integration and application of GPS, RS and GIS.

LI Qing-Quan, born in 1965, Ph. D., professor. His research interests include intelligent transportation, integration and application of GPS, RS and GIS.

QIN Long-Kun, born in 1989, M. S. candidate. His research interests include data analysis and mining of crowd sourcing check-in data.

Background

Landmark is a recognizable natural or man-made feature that stands out from its near environment and is often visible from long distance. As one essential element of city, landmark plays an important role in the domains of travel, traffic, culture, etc. Extracting landmarks from environment is a grand issue of spatial cognition and researchers have proposed lots of ways to extracting hierarchical landmark in the past decades.

LBS rose abruptly in the recent years and have provided massive check-in data produced by mobile users since the smart phone comes out. This crowd sourcing check-in data offers a new possibility for us to extract hierarchical landmark. This paper reviewed important achievements made by

landmark researchers on hierarchical landmark extracting in recent years. Then a method of extracting city hierarchical landmarks from check-in data according to their check-in attribute significances is proposed. Finally, we validate the effectiveness of this method by experiments and point out the existential problems of the method.

This work is supported by the National Natural Science Foundation of China (General Program) under Grant No. 61172175. This project aims to find out a new way to update the urban road network database with crowd sourcing GPS data automatically.