引文格式：刘乐, 张东映. 基于连锁 POI 名称的城市商圈提取方法[J]. 地理空间信息, 2023, 21(8): 49-52. doi:10. 3969/j.issn. 1672-4623.2023.08.012

基于连锁POI名称的城市商圈提取方法

刘　乐[1]，张东映[2*]
（1. 北京朗歌科技有限公司，北京 100020；2. 华中科技大学土木与水利工程学院，湖北 武汉 430074）

摘　要：城市商圈是城市商业活动的区域中心，以兴趣点（POI）为研究对象，提出一种识别商圈及辐射范围的方法。首先以相同分店名称的POI构造多边形，利用拓扑运算实现对相交和包含多边形的处理，将其作为商圈核心范围，然后利用空间聚类思想计算商圈外围POI的隶属度，进而利用空间插值方法对隶属度进行插值拟合，实现商圈辐射区域的生成，最后以北京市为例进行了商圈的提取和分析，取得了较好的效果。主要结论为：该方法可以有效提取城市商圈的核心范围及辐射范围，对研究城市功能区具有重要价值。

关键词：商圈；兴趣点；聚类
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）08-0049-04

Urban Business District Extraction Method Based on Chain POI Names

LIU Le[1], ZHANG Dongying[2]
(1. Beijing Langge Technology Co., Ltd., Beijing 100020, China;
2. School of Civil and Hydraulic Engineering, HUGT, Wuhan 430074, China)

Abstract: Urban business district is the regional center of urban commercial activities. Taking POI as the research object,we proposed a method to identify the business district and the radiation range. Firstly, we constructed the polygon with the POI with the same branch name, and realized the processing of the intersection and inclusion polygon by topological operation, which wastaken as the core scope of the business district.Then, we used the spatial clustering idea to calculate the membership degree of peripheral POI of business district, and further used the spatial interpolation method to interpolate the membership degree togeneratethe radiation area of business district. Finally, taking Beijing for example, weextracted and analyzed the business district, with good results achieved. The method can effectively extract the core scope and radiation range of urban business districts, which is of great value for the researchof urban functional areas.

Key words: urban business district, POI, clustering

城市商圈提取本质上是对空间点的聚类。空间点聚类的方法研究一直是空间数据挖掘的研究热点。学者从不同维度出发，提出了基于划分的方法、基于分层的方法、基于密度的方法和基于格网的方法等。其中，基于划分的方法最为常用[1-2]，广泛应用于图像分割[3]、土地利用[4-5]等领域。在商圈的聚类研究中，主要是从城市规划角度研究商圈聚类分析方法，进而分析城市的商业空间结构[6-8]。近些年来，随着移动互联网的兴起，基于移动互联网积累的POI大数据在城市研究中，也逐步成为热点[9-10]，现有的研究更多的是利用空间分析的方法呈现商圈区域的空间聚集特征，如李湾湾[11]等以武汉市主城区兴趣点和道路网为基准，用核密度估计等方式对武汉市主城区商圈网点空间分布情况进行分析；王世雄[12]等使用多阶段聚类方法分析了杭州市的兴趣点聚集效应，也有提出聚类方法的探索城市商圈的工作；如胡庆武[13]等基于用户签到数据用聚类的方法进行了商圈的探索挖掘。现有的相关研究中，主要考虑利用单一空间分析方法和整体POI数据对象呈现商圈区域的空间聚集特征，而针对城市内部单一领域POI数据、利用复合空间分析方法进行城市商圈辐射区域划定的实证研究则鲜少涉及。本文利用知名的连锁品牌的分店名称作为商圈提取的基本语料，耦合聚类方法中基于划分的思想和模糊数学中的隶属度策略进行商圈提取，得到包含名称、多边形以及辐射范围的完备商圈，并以北京市为例进行了商圈的提取和分析。

收稿日期：2022-03-30；修回日期：2023-06-05。
项目来源：黄河勘测设计规划研究院项目（0231242312）。
第一作者简介：刘乐（1986—），硕士，研究方向为POI挖掘、地理大数据挖掘，E-mail：liulez@163.com 。
通信作者：张东映（1985—），博士，副研究员，研究方向为时空数据智能，E-mail：zhangdongying@hust.edu.cn。

1 研究区域及数据来源

1.1 研究区域概况

北京作为我国首都，是我国政治、经济、文化、教育中心，具有巨大的城市发展潜力，其主城区商圈的功能结构更加复杂。鉴于此，本文选取餐饮兴趣点分布相对密集的北京市五环内区域作为研究区域，进行商圈分析。

1.2 数据来源

实验数据来源于高德地图北京市兴趣点数据集POI数据集，抽取五环内的餐饮类的连锁POI信息作为实验数据，共计54 303个[14]。

验证数据为局部验证使用北京市政府网站关于王府井商圈的官方说明[15]；整体验证使用北京市政府公布的2022年各区的GDP数据[16]。

2 商圈提取方法

连锁POI的分店名称一般为所在区域的商圈名称或地标名称。本文利用知名的连锁品牌的分店名称作为商圈提取的原始数据，首先对POI名称进行分店名规范化和噪声过滤的预处理；然后进行商圈生成，利用相同分店名的多个品牌的POI计算外接多边形，作为一个商圈的初始化范围，然后通过空间相交运算，获得初始范围内其他分店名称集合；最后进行辐射区生成，首先计算初始范围内外POI的隶属度，并利用隶属度值进行空间插值，拟合为辐射区域的等值线图。

2.1 连锁POI名称预处理

连锁POI名称一般包含品牌名称和分店名称，如“稻香村（西单店）”，其中“稻香村”为品牌名称，“西单店”为分店名称，其中“西单”即为目标商圈名称。实际中，分店名称内涵较为丰富，且不同品牌在同一地区的分店名称可能会有不同的命名标准，因此需要对分店名称建立规则进行预处理。

分店名称规范化处理。由于商圈一般具有自己的正式名称，部分商圈有俗称和别名，因此，采用商圈作为品牌名称时，不同的品牌具有不同的选择标准，对于此类情况，需要建立名称映射关系，首先将分店名称规范化，然后进行后续处理，如“宏状元粥店（东四店）”，“麦当劳（东四十条店）”规范化为“东四十条”。另外，部分商圈面积较大，一些品牌在一个商圈可能有多个分店，此类也需要建立规范化规则处理，如“汉拿山（望京店）”“汉拿山（望京二店）”规范化为“望京”。

分店名称噪声过滤。分店名称一般为各个品牌自发命名，命名的对象多种多样，如商场（如“稻香村（天虹商场店）”）、大厦（如“麦多馅饼（枫蓝国际店）”）、小区（如“稻香村（吉祥里小区店）”）、道路（如“稻香村（志新西路店）”）、车站（如“味多美（北京站店）”）等。部分连锁店的分店名以连锁商场名称命名，如“百盛”、“华威”等，此类分店名称会在多个商圈存在，需要过滤；部分分店名称不可以作为大众意义上的商圈，需要根据实际情况分析建立规则库，将明确不可作为商圈的分店名称排除，作为验证商圈的基础。

按照规则将分店名处理完成后，根据分店名称构成的点集，利用凸包算法实现多边形构造。

2.2 商圈生成

由于分店名称中大多包含商场、小区、大厦等地标，在对同一分店名称下的所有品牌的POI进行外接多边形计算后，会出现多边形区域叠加或相交的现象。此时，通过多边形的空间拓扑关系来处理。

如果初始多边形之间为包含关系，则取较大的多边形。如果相交，则根据多边形的质心C和最大内切圆半径D来决定取舍关系，如图1所示。

对于多边形a与b，b的质心包含于a，则保留a，隶属于b的POI融合到a；对于多边形a与c，c的质心不属于a，此时需要计算a和c的内径，D（a）>nD（c），则保留a，隶属于c的POI融合到a；对于多边形a与d，d的质心不属于a，且D（a）<nD（d），则a和d都保留。

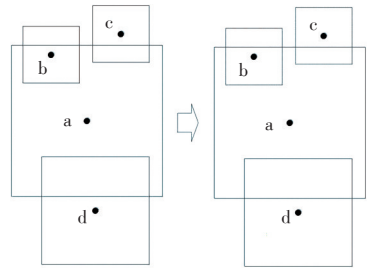


图1 商圈相交关系

该种处理方式能够有效地将商圈中包含的地标（商场，大厦等）有效地去除，并且能够保留完整的商圈区域，仅有选择的剔除重叠的区域。其中的n是合并商圈的因子，取值越小，对周边多边形的合并能力越强，反之，越弱。

算法流程如下：

初始化商圈多边形集合S，计算多边形质心坐标和多边形内径；

遍历S中的多边形，对于Si，计算其质心与Sj

（$j>i$）的拓扑关系：若Sj包含Si，则Si标识剔除，标识已处理。若Si包含Sj，则Sj标识剔除，标识已处理。若相交，则比较两多边形内径大小：若$Dj>nDi$，则Si标识删除，标识已处理；若$Dj<nDi$，则Si标识保留，标识已处理；若$Di>nDj$，则Sj标识删除，Si标识保留，标识已处理；若$Di<nDj$，则Si标识保留，标识已处理，Sj标识保留；$i=i+1$，直到处理完成，输出标识保留的多边形。

2.3　辐射区域的生成

商圈的生成完成了商圈核心部分的提取，然而，未参与商圈多边形计算的点的商圈隶属仍然未定。这部分点一部分已包含在多边形中，一部分在多边形外部。本文借助基于划分的聚类的思想，将已经完成的商圈多边形质心作为初始化的聚类中心，计算商圈多边形外部POI点对各个商圈多边形的隶属度[17]，取隶属度较大的商圈多边形作为归属商圈。

隶属度μ的取值依赖于商圈多边形的最大内切圆半径D以及当前POI到商圈多边形质心的距离d。因多边形内径越大，表明商圈覆盖范围越广，对于到达2个商圈边界的距离相等的POI，商圈半径越大则隶属度越高。计算公式如下：

$$\mu(\mathrm{p})=\max\frac{Di}{d(\mathrm{p},i)} \tag{1}$$

式中，$\mu(\mathrm{p})$为p点的商圈隶属度；Di为第i个商圈的内径；$d(\mathrm{p},i)$为当前POI点到第i个商圈的距离。

算法流程如下：

初始化商圈多边形集合S，计算多边形质心坐标和多边形内径；

遍历样本所有POI点，对于Pi，依次判断与商圈多边形Sj的拓扑关系：若包含，则结束循环，隶属度$\mu(i)=1$；若不包含，计算当前点与多边形质心之间的距离，计算隶属度$\mu(i)$；选取隶属度最大的，记录$\mu(i)$和所属商圈；$i=i+1$，直到完全处理完成。

利用计算获得的POI的商圈隶属度，进行反距离权重插值，得到商圈整体分布及辐射情况。

3　实验与分析

本文以北京市五环内餐饮类连锁分店POI为实验数据，利用本文方法对数据进行了处理。在处理时，设定品牌个数大于等于10的分店名称为初始化商圈。下面首先聚焦局部，针对王府井商圈进行具体的效果对比分析；然后针对北京市商圈整体分布进行宏观分析。

表1为数据处理过程中的一些数据的统计，包括处理前的POI个数，利用POI第一次构造商圈多边形完成后多边形的个数，以及完成多边形合并后商圈多边形的个数。

表1　处理过程数据统计

区域	POI个数	多边形构造后	多边形合并后
王府井商圈	120	2	1
北京市整体	11 722	255	70

3.1　商圈生成效果分析

本文选择王府井商圈来验证商圈的生成效果，王府井商圈是南起东长安街，北至灯市口大街，包含金鱼胡同和东安门大街的多边形区域[15]，从商圈的生成过程和与官方的商圈范围对比两个角度。

首先验证商圈生成的过程，选取官方公开的王府井商圈作为边界，分析商圈的生成过程。图2是该区域POI点的分布情况，红色边线是王府井商圈的官方范围，蓝色为构造多边形后的情况，可以看到包含2个高度重叠的多边形；图3为合并结果，由2个初始商圈合并为1个，可以看到生成的商圈轮廓与官方范围高度重叠。

其次，将生成结果与公开商圈范围对比，验证商圈的辐射区域。图4为王府井商圈合并的结果及插值后的辐射区域空间叠加效果，对比红色边线的实际区域，挖掘多边形覆盖了商圈的核心区域；等值线展示了商圈的辐射能力，可以看到，它以挖掘多边形为核心，向外围延伸，辐射能力超过官方商圈本身，南向的辐射能力比较明显，经验证，南向区域为政府机构和公园区域，商业设施稀少，王府井商圈辐射到这片区域，符合预期。

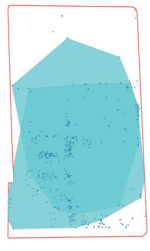


图2　王府井官方范围及POI分布

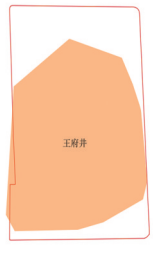


图3　合并结果（王府井商圈）

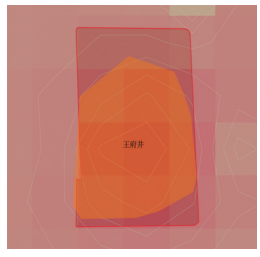


图4　插值结果（王府井商圈）

3.2　整体分布分析

图5是北京市整体处理完成后，商圈多边形的整体分布和利用隶属度经过插值生成的POI辐射区域。由图5可见：从整体看，北京市商圈整体成“回”状分布，以故宫为中心，故宫处无商圈，五环内边界除望京外几乎无商圈，这2个环形区域之间商圈密集。

从各方位看，各方位均有代表性的簇状聚集商圈，但分布并不均衡，东部国贸、双井区域，东北部望京、酒仙桥区域，北部亚运村、五道口、中关村区域，南部方庄、刘家窑区域，西部五棵松、大成路区域均成聚集簇状；但东部商圈密度最大，北部次之，南部和西部相对较少。

为进一步验证上述结论，使用研究区域内各区的国民生产总值与各区的商圈数量进行相关分析。各区的国民生产总值使用北京市公开的2022年数据[16]，首先按照五环内的面积占比进行折算结果表明，各区的年度国民生产总值也存在与商圈数量相关的不均衡，即从东部、北部的朝阳和海淀较高，西部南部的丰台、大兴较低。

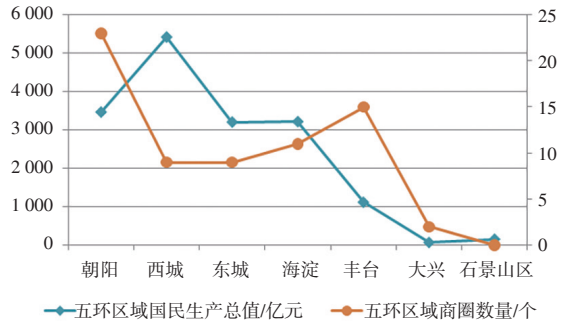


图5 五环区域各区国民生产总值与商圈数量关系图

4 总结与展望

本文提出了一种基于连锁POI名称提取城市商圈和生成辐射区的方法，并运用该方法对北京市的POI数据进行了处理、分析及验证。研究结果表明：①本文基于知名连锁品牌的分店名称单一领域POI数据、利用复合空间分析方法进行城市商圈提取的王府井商圈与官方商圈一致性高；②耦合聚类方法中基于划分的思想和模糊数学中的隶属度策略，利用相同分店名称POI的外接多边形质心作为划分方法的初始化分类中心，根据多边形的大小来计算当前点对多边形的隶属度，有效地提取了POI对商圈的隶属关系，并生成商圈的辐射区。

然而，商圈是一个受多种因素综合作用形成的聚集区域，本文所提方法虽一定程度上简化了问题求解过程，但仍有改进空间：①本文仅使用了POI数据中较有代表性的餐饮数据作为POI提取的材料，现实世界中，真正的商圈根据定位的不同，可能所包含的POI有所不同。因此商圈的提取也应综合考虑多种因素，利用多因素的综合指数进行商圈拟合；②分店名称的精确处理是商圈提取的基础，名称处理规则和语料库需要不断完善才能进行更精确的处理；③空间范围上，城市商圈一般小于区县级的行政区划，但具体的POI的位置在范围更小的楼宇、小区上，在本文基础上，可以探索建立商圈与楼宇、小区等地标的隶属关系，实现POI更加精确的定位。

参考文献

[1] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(1):48-61

[2] 张敏,于剑. 基于划分的模糊聚类算法[J]. 软件学报,2004,15(6):858-868

[3] 左浩,李雯. 改进的PCM聚类算法在图像分割中的应用[J]. 计算机与数字工程,2010,38(11):148-151

[4] 檀满枝,陈杰,郑海龙,等. 模糊C均值聚类法在土壤重金属污染空间预测中的应用[J]. 环境科学学报,2006,26(12):2 086-2 092

[5] 毛丽丽,于静洁,张一驰. 模糊c均值聚类方法在黑河下游土壤属性制图中的初步应用研究[J]. 干旱区资源与环境,2013,27(1):197-203

[6] 李新延,李德仁. DBSCAN空间聚类算法及其在城市规划中的应用[J]. 测绘科学,2005,30(3):51-53

[7] 仵宗卿,戴学珍. 北京市商业中心的空间结构研究[J].城市规划,2001,25(10):15-19

[8] 马晓龙. 西安市大型零售商业空间结构与市场格局研究[J]. 城市规划,2007,31(2):55-61

[9] 杨振山,苏锦华,杨航,等.基于多源数据的城市功能区精细化研究:以北京为例[J].地理研究,2021,40(2):77-94

[10] 陈蔚珊,柳林,梁育填.基于POI数据的广州零售商业中心热点识别与业态集聚特征分析[J].地理研究,2016,35(4):703-716

[11] 李湾湾,董燕,王迪.基于POI数据商业空间分布与集聚特征分析:以武汉市为例[J].中国水运,2020,20(12):141-143

[12] 王世雄,李天晨,李泽彪.杭州O2O生活服务业的聚集性及空间特征[J].浙江理工大学学报,2019,42(6):603-610

[13] 胡庆武,王明,李清泉.利用位置签到数据探索城市热点与商圈[J].测绘学报,2014,43(3):314-321

[14] 国家信息中心,高德地图兴趣点POI(Point of Interest)数据[EB/OL]. (2021-09-19)[2022-04-20]. https://doi.org/10.18170/DVN/WSXCNM,北京大学开放研究数据平台,V2,2017

[15] 北京市人民政府.北京市王府井步行街地区管理规定[S].

[16] 北京市人民政府.北京区域统计年鉴[S].

[17] Krishnapuram R, Keller J M. A Possibilistic Approach To clustering[J]. IEEE Trans. on Fuzzy Systems, 1993, 1(2):98-110