

引用格式: 徐晓宇, 李梅. 基于开源大数据的北京地区餐饮业空间分布格局[J]. 地球信息科学学报, 2019, 21(2): 215-225. [Xu X Y, Li M. Analysis on spatial distribution pattern of Beijing restaurants based on open source big data[J]. Journal of Geo-information Science, 2019, 21(2): 215-225.] DOI: 10.12082/dqxxkx.2019.180437

基于开源大数据的北京地区餐饮业空间分布格局

徐晓宇, 李梅*

北京大学地球与空间科学学院 遥感与地理信息系统研究所, 北京 100871

Analysis on Spatial Distribution Pattern of Beijing Restaurants based on Open Source Big Data

XU Xiaoyu, LI Mei*

Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China

Abstract: Using big data to analyze the spatial pattern of urban service facilities has become a new research hotspot, and catering industry is a typical representative of urban service industry. Therefore, it is of great significance to study the spatial layout of urban catering industry through open source big data. The restaurants in a city can be abstracted as point objects in the geographical study, and clustering analysis is a classical data mining method that quantitatively identifies geographical clustering among objects. In this paper, Beijing is selected as the research area, and the data of 153 895 restaurants in Dianping.com are obtained by using web crawler technology. The density-based CFSFDP clustering algorithm (clustering by fast search and find of density peaks) is adopted here to analyze the geographical clustering characteristics of catering industry in terms of spatial distribution density and per capita consumption level. This approach, which is based on the idea that cluster centers are characterized by a higher density than their neighbors and by a relatively larger distance from points with higher densities, can recognize clusters regardless of their shape and the dimensionality of the space in which they are embedded, so more accurate spatial analysis results can be obtained. The results show that: (1) the spatial pattern of Beijing restaurants is imbalanced, which generally presents the characteristics of multi-center spatial distribution, and the agglomeration degree of restaurants decreases with the distance increase from the main urban area which are regarded as the core. Besides, the restaurants hot spots mainly circles around important business centers, tourist attractions as well as residential areas, and extends along the traffic line evidently. (2) The catering stores with different per capita consumption levels have the characteristics of hierarchical system. That is to say, there are the number of high-grade restaurants is few, and mainly concentrated in the commercial centers, financial centers and famous tourist attractions in Dongcheng district, Xicheng district, Chaoyang district and Haidian district, while the number of middle and low-grade restaurants is much more and their spatial distribution are more scattered. (3) The density and price of restaurants accords with consumption level of consumers. At the same time, this paper also analyses the factors influencing the

收稿日期: 2018-09-04; 修回日期: 2018-11-29.

基金项目: 国家重点研发计划项目(2016YFC0803108)。[**Foundation item:** National Key Research and Development Program of China, No.2016YFC0803108.]

作者简介: 徐晓宇(1995-), 女, 辽宁营口人, 硕士生, 主要从事地学信息可视化与数据挖掘研究。E-mail: xuxiaoyu@pku.edu.cn

*通讯作者: 李梅(1978-), 女, 陕西西安人, 博士, 副教授, 主要从事地学信息可视化, 实时GIS与应急研究。

E-mail: mli@pku.edu.cn

spatial distribution pattern of catering clusters by combining the two indicators of spatial agglomeration characteristics and consumption level, in order to provide useful reference of urban commercial spatial layout for the government planning departments.

Key words: geographical big data; CFSFDP; restaurant; spatial pattern; Beijing area

***Corresponding author:** LI Mei, E-mail: mli@pku.edu.cn

摘要:基于大数据进行城市服务设施空间格局分析已成为一种新的研究热点,而餐饮业是城市服务业的典型代表,因此通过开源大数据对城市餐饮业的空间布局进行研究具有重要的意义。本文以北京地区作为研究区,采用网络爬虫技术获取大众点评上153 895家餐饮店数据,引入基于密度的CFSFDP聚类算法从空间分布密集度和人均消费等级方面对餐饮业背后蕴含的地理聚集特征进行分析。研究发现:①北京地区餐饮店总体呈现多中心的空间分布特征,其集聚程度以主城区为核心向外逐级递减,并明显表现出围绕重要商圈、旅游景点和住宅区进行布局以及沿交通轴线扩展的趋势;②不同人均消费水平的餐饮店呈现等级体系特征,即高档餐馆少而集聚,中低档餐馆多而散的分布特点;③餐馆分布密集程度和定价表现出接近市场和消费者的特征。同时,本文综合空间集聚特征和消费水平2项指标对影响餐饮店集群空间分布格局的因素进行了分析,以期政府规划部门进行城市商业空间布局研究提供借鉴。

关键词:地理大数据;CFSFDP算法;餐饮;空间格局;北京地区

1 引言

城市空间结构一直是城市地理学及城市规划的重点研究内容,而餐饮业作为城市系统的重要组成部分,其区位特征和空间分布格局一直是学者们关注的焦点。国外学者很早就从旅游的角度来分析饭店的分布特征,主要集中在空间布局 and 影响因素方面,如Roehl等^[1]和Urtasun等^[2]发现城市旅游热门景点对于饭店空间布局具有重大影响;Wall等^[3]和Ritter^[4]研究表明饭店空间分布与交通便捷性有关,在交通便利的机场附近,饭店出现了明显的集聚性。随着研究内容的不断深化,一些学者开始研究饭店空间集聚性与城市发展的关系。如Cró等^[5]研究了葡萄牙里斯本市酒店位置的影响因素,发现高档宾馆更接近经济繁荣的商业区,而低档宾馆则倾向于积极寻求集聚效应的收益。而Lee等^[6]的研究表明,随着人类生活水平的逐步提升,饭店在城市艺术和娱乐场所周围也逐渐表现出了集聚特征。国内有关餐饮业空间布局的研究起步相对较晚,最早学者们主要研究饮食的地域差异^[7]和饮食口味受地理因素和文化传统的影响^[8],多为定性分析。随着计算机技术和数据分析软件的普及,研究逐渐转向定量分析,如梁璐^[9]、张旭等^[10]分别基于西安和南京的统计数据来分析餐饮店的空间分布,结果表明由于受到经济发展程度、人口及交通等因素的影响,城市餐饮业的发展存在着在档次、密度、种类等空间格局上的不平衡性。邬伦等^[11]采用网络K函数法分析香港岛餐饮店的空间分布模式,揭示出

餐饮业选址实际更加倾向于靠近交通网络中的交通站点。总之,传统研究多基于官方统计数据或实地调研数据对餐饮业空间布局进行探讨,对城市规划与发展具有一定的指导意义。但受限于海量数据采集困难且时效性差等问题,往往只能局限在宏观尺度上对城市空间进行研究,缺乏精细化研究^[12]。因此,有必要基于大量样本数据点来识别和提取城市餐饮业空间集群,从而更准确地描述餐饮业在城市不同区域的空间分异特点。

近年来,随着互联网技术和移动终端技术的飞速发展,蕴含地理位置信息的网络数据呈现大规模的爆炸式增长,这为城市空间结构这一基础性研究提供了新的研究机会和挑战^[13]。大数据时代的到来,在一定程度上解决了传统城市空间研究中耗时费力的数据采集问题。在这一背景下,融合开源大数据进行城市空间结构的分析已经成为当下城市空间研究中的热点^[14]。而餐饮业作为城市服务业的一大重要分支,是城市空间组织的重要环节。因此,用餐饮大数据研究其空间集聚特征及影响其布局的社会、经济、文化等因素,可以辅助政府部门和规划部门审视和剖析处于快速发展中的城市空间结构,指导城市规划建设和产业选择,具有十分重要的研究意义。

Prayag等^[15]运用ArcGIS技术对新西兰Hamilton城市1996–2008年的餐饮业进行了动态研究,发现CBD附近餐饮店存在明显集群,且由于酒吧、购物中心和电影院等其他互补性商业的存在,使该区域对顾客更具吸引力。Yang等^[16]基于Logit模型研究

了北京市酒店的分布格局,发现其明显受到公共服务设施、地铁可达性和旅游景点可达性的影响,且高档酒店对可达性更加敏感。秦萧等^[17]基于大众点评数据得到南京市餐饮业集聚于主城区典型商圈的结论。谭欣等^[18]基于互联网餐饮数据,采用核密度分析法从空间分布、人均消费和网络口碑 3 个视角来分析北京主城区餐馆的空间分布,结果表明北京主城区内餐馆呈现从城市中心向外围递减的空间格局。杨帆等^[19]基于 DBSCAN 空间聚类来提取广州市的餐饮集群,得到高等级餐饮集群集中于中心城区且不同等级的集群在数量上符合中心地理论模型的结论。曾璇等^[20]利用网络核密度估计来分析广州市海珠区餐饮店 POI 的空间分布特性,结果表明餐饮业总体呈现多中心的分布特征,且在较小尺度下餐饮店的分布与公交站具有显著的集聚关系。

总的来看,在大数据逐渐引领城市空间研究的背景下,学者们开始利用数据挖掘技术从 POI、大众点评和美团等网站上获取大量数据,数据来源趋于多样化。研究方法也从基于 ArcGIS、GeoDa 等软件的热点分析^[15]、核密度分析^[17-18,20]等逐步扩展到用空间聚类如基于密度的 DBSCAN 算法^[19,21]进行城市餐饮业空间布局的系统分析。但是还存在数据量相对不完整,如文献[18]中仅包括北京主城区 17 512 家餐饮店,文献[19]中包含广州市 27 037 家餐饮兴趣点,以及当数据集密度变化较大时,传统聚类方法失效等问题。

针对这些问题,本文以北京地区为研究对象,从互联网上实时获取大量餐饮店位置、价格信息,引入基于密度的 CFSFDP(Clustering by Fast Search and Find of Density Peaks)聚类算法^[22]来探究北京地区餐饮业空间分布的规律特征。与传统聚类算法不同,CFSFDP 算法的优势在于聚类中心为附近区域内餐饮店分布最密集的点,并利用距离属性避免在某一热门区域内存在多个密度相近的聚类中心,错把原本属于一类的数据变成了多类的情况,从而确保聚类结果的合理性,且该算法可以处理任意形状的簇。此外,本文还根据人均消费对北京地区餐饮店的空间分布做了进一步的细化分级研究,结果发现基于密度计算得到的聚类中心和基于人均消费分级得到的中心在某些区域存在非一致性,本文进一步的探究了高密度聚类中心和高消费区域之间的关系,对影响餐饮店分布的因素进行了合

理推断和解释,以期政府城市规划和商业地理的研究提供有益借鉴。

2 数据源

本文以北京地区为研究对象,利用网络爬虫技术获取大众点评上的餐饮店数据,包括名称、经纬度、详细地址、人均消费等属性信息,数据获取时间为 2018 年 7 月 1-30 日,原始数据共计 178 585 家餐饮店铺。为了确保结果的有效性和真实性,首先进行了数据清洗和预处理,去掉了重复和位置有误的数据,然后采用空间坐标转换技术将其经纬度转换为百度坐标,最终筛选出 153 895 家餐饮店用于本文研究,在百度地图上可视化后如图 1 所示。在数量分布特征上,朝阳区餐馆数量最多,共 35 612 家;海淀区餐馆数量位居第二,共 20 719 家;丰台区、昌平区、大兴区、东城区、西城区次之;餐馆数量最少的城区是门头沟区,仅有 1539 家。进一步计算各城区餐馆分布密度,得到东城区餐馆密度最高,为 253 家/km²;其次为西城区,密度为 194 家/km²;接下来依次为朝阳区 76 家/km²,丰台区 52 家/km²,海淀区 48 家/km²;餐馆密度最低的是延庆区,约为 1 家/km²。研究发现,北京中心城区餐馆数量多且分布密集,而四周城区餐馆数量明显少于中心城区,地理位置也相对分散,呈现出以中心城区为核心向外聚集程度逐级递减的特征。

3 研究方法

地理集聚是识别餐饮业空间分布集群的主要依据,即在地理空间上显著临近的餐饮店可归属为一个集群,而空间聚类是定量识别对象间地理集聚性的经典数据挖掘方法。聚类算法大致可分为 5 类:① 基于划分的方法,如经典聚类算法 K-means;② 基于层次的方法,如 CURE、Chameleon 算法;③ 基于密度的方法,如 DBSCAN 算法;④ 基于网格的方法,如 CLIQUE、STING、Wave-Cluster 算法;⑤ 基于模型的算法,如 Gaussian Mixed Model 算法。现有针对城市服务业的聚类分析,多采用经典 K-means 聚类算法^[23]或基于密度的 DBSCAN 算法^[19,21]。然而,北京地区餐饮店数量多、空间分布复杂、类簇形状多样且餐饮店密集程度变化较大,而 K-means 算法处理大量数据点时间开销过大,DBSCAN 算法对密度变化较大的数

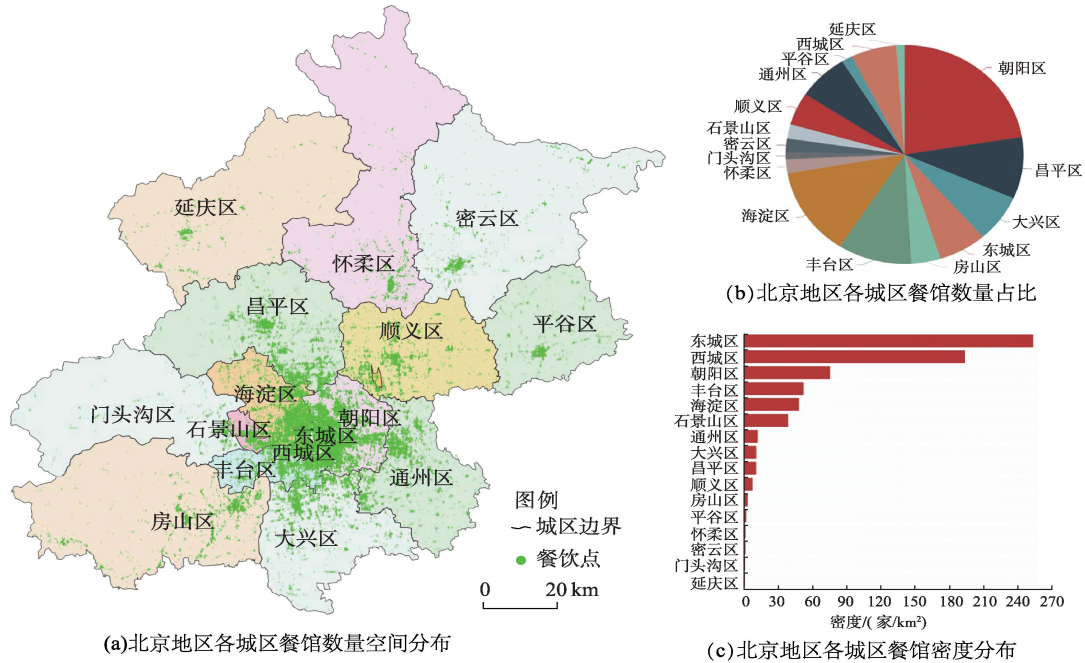


图1 北京地区餐饮店分布
Fig. 1 The distribution of restaurants in Beijing

据集聚类效果不理想。可见,传统聚类方法在处理密度分布不均的大量点时都存在不足。

针对上述问题,本文采用Alex Rodriguez等^[22]提出的一种快速搜索寻找高密度区的CFSFDP算法来提取餐饮店的聚类中心。该算法原理简单,运行效率高,可以处理任意形状的簇。由于此算法只考虑点与点之间的距离,因此不需要将点映射到一个向量空间中,大大提升了计算效率,适用于大量数据点的处理。此外,该算法在选取聚类中心时结合了密度和距离的双重信息,从而克服了传统DBSCAN算法由于高密度区域间有粘连而错把原本属于一类的数据变成了多类的缺点,确保了聚类结果的合理性^[24]。同时,针对原始CFSFDP算法需要人工辅助确定聚类中心的不足,本文借助归一化后的密度与距离的乘积排序图斜率变化趋势实现聚类中心的自动选择,避免了主观臆断对聚类结果的影响。

3.1 CFSFDP算法原理

CFSFDP算法^[22]核心是对聚类中心的刻画。聚类中心具有2个特点:① 聚类中心周围都是密度比其低的点;② 这些点距离该聚类中心的距离相比于其他聚类中心来说是最远的。基于聚类中心的上述特征,每一个数据点都要计算2个属性值:点的局部密度 ρ_i 和该点到具有更高局部密度的最近点的

距离 δ_i 。点的局部密度计算公式为:

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \quad (1)$$

式中: ρ_i 是以数据点 i 为圆心; d_c 为半径范围内数据点的个数; d_c 为预先给定的截断距离; $d_{ij} = \text{dist}(i, j)$ 是数据点 i 和 j 间的距离。若 $d_{ij} < d_c$,则 $\chi(d_{ij} - d_c) = 1$,否则为0。CFSFDP算法对于 d_c 参数的选择具有较强的鲁棒性。

将所有数据点按局部密度值降序排列,对于局部密度值 ρ_i 最大的点, δ_i 为点集中距点 i 最远的数据点与点 i 之间的距离,计算公式为:

$$\delta_i = \max_j(d_{ij}) \quad (2)$$

而对于其他非局部密度最大点, δ_i 为所有局部密度大于点 i 的数据点中,与点 i 距离最小的数据点与点 i 之间的距离,计算公式为:

$$\delta_i = \min_{j: \rho_j > \rho_i}(d_{ij}) \quad (3)$$

3.2 自动确定聚类中心

文献[22]给出可根据 $\rho-\sigma$ 决策图选出合适的聚类中心点。但通过决策图来判断聚类中心采用的是定性分析,易受主观因素影响,当数据量较大时,人为确定聚类中心更为困难。为了综合考虑局部密度 ρ_i 和距离 δ_i 的影响,本文定义决策变量

γ_i , 由于 ρ_i 和 δ_i 两类值可能处于不同的数量级, 为了消除数量级的影响, 分别将其进行归一化得到归一化后的局部密度 ρ_i^* 和距离 δ_i^* , γ_i 的计算公式如式(4)所示。

$$\gamma_i = \rho_i^* \times \delta_i^* \quad i \in I_s \quad (4)$$

显然, γ_i 值越大, 越有可能是聚类中心。将所有数据点的 γ_i 值降序排列, 对排序后的点重新进行标号, 用 Index 表示。以 Index 为横轴, γ 值为纵轴绘制 γ -Index 图(图2), 称之为“ γ 降序排列图”。非聚类中心的 γ 值较为平滑, 从非聚类中心过渡到聚类中心时, γ 值应有一个明显的跳跃, 可以帮助确定聚类中心的个数。图2(a)是包括局部密度最大点的 γ -Index 图, 为了突出非局部密度最大点的 γ 值变化幅度, 图2(b)只给出了点 1-1000 的 γ -Index 图。

然而, 当数据量较大时, γ -Index 图上存在多个从平滑到跳跃的点, 衡量 γ_i 究竟大于何值才能作为聚类中心还需进一步判定。由于 γ_i 值越大, 越有可能成为聚类中心, 王洋等^[25]提出了一种根据斜率变化程度来自动确定聚类中心的方法, 即根据 γ 降序排列图, 按式(5)依次计算点 i 与点 $i+1$ 间线段的斜率 k_i , 按式(6)计算所有数据点相邻两点间斜率差之和的平均值 A_k 。

$$k_i = \frac{\gamma_i - \gamma_{i+1}}{i - (i+1)} \quad (5)$$

$$A_k = \frac{\sum_{i=1}^{n-2} ||k_i| - |k_{i+1}||}{n-2} \quad (6)$$

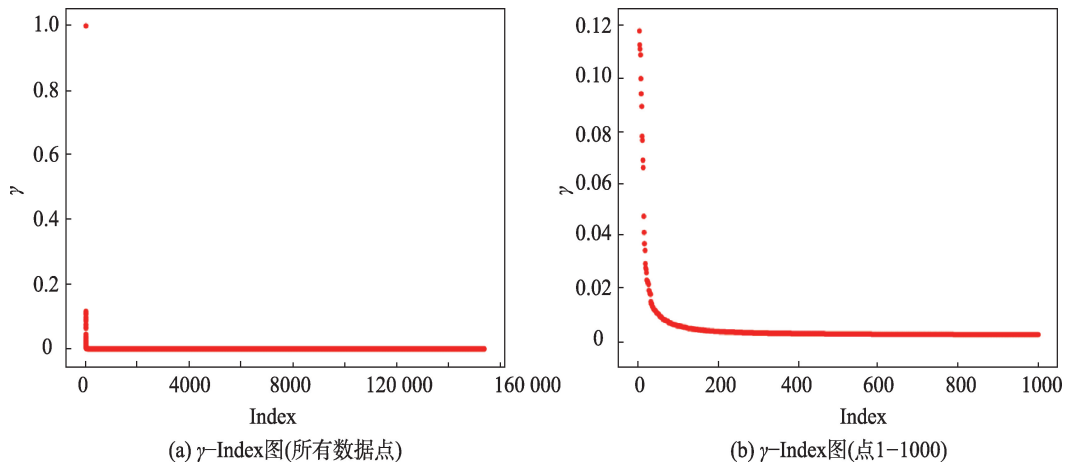


图2 γ 降序排列

Fig. 2 The value of γ sorted in decreasing order

寻找斜率差值不低于均值 A_k 中序号最大的点, 视为临界点 I 。由于临界点 I 后的点 γ_i 值相差较小, 因此不可能成为聚类中心。将序号不高于临界点 I 的点作为备选聚类中心, 但备选聚类中心中可能存在比较接近的点, 即原本应属于同一簇的点均划分到了备选聚类中心中去, 基于此, 依次计算备选中心点彼此之间的距离, 若存在两点距离小于截断距离 d_c , 则剔除 γ 值小的点, 最终得到合适的聚类中心。

找到聚类中心后, 剩余点依次被归属到它的具有更高密度的最近邻聚类中心。该算法剩余点归类简单, 无需对目标函数进行迭代优化, 大大提高了计算效率。

4 结果与分析

4.1 基于密集程度的空间分布特征

用 CFSFDP 算法对北京地区餐饮店进行聚类分析, 结果如图3所示。可以发现, 聚类中心点多集中于东城区、西城区、朝阳区及海淀区, 呈圈层式分布, 具有从中心城区向外密集程度逐级递减的趋势。本文按聚类中心的点密度将聚集区分为4类: ① 高密度区域主要集中于三里屯、工人体育场、朝外商圈、王府井商圈、西单大悦城、通州北苑等地; ② 中高密度聚类区域逐步向外扩展到学院路、新街口、亮马桥、燕莎、昌平区政府、沙河地铁站、回龙观、顺义区政府、大兴区政府等地; ③ 中密度聚类区域则进一步延伸到怀柔区政府、平谷区政府、石景



图3 北京地区餐饮服务业聚类结果

Fig. 3 The cluster results of restaurants in Beijing area

山区政府、八通线商业带(果园-梨园)等地;④中低及低密度聚集区更多更分散,除部分位于中心城区外,大多分散在各周边郊县城区。可以发现,餐饮店密度分布受到其周围经济繁华程度、是否靠近热门商圈、是否靠近住宅区以及交通便捷性的影响。繁华的商业区附近餐馆数量最多;交通便利的位置人流量较大,所以餐馆的分布较密集。朝阳区因地域面积优势和繁华程度高,包含更多的聚类中心,其他非中心城区餐饮店多集中在其区中心附近。

点密度排名前20的聚类中心如图4所示。餐饮店分布密集的区域可分为以下4类:

(1)热门商圈附近,如三里屯&工体、王府井、朝外街道、光华路、西单大悦城、国贸、望京、中关村、五道口等。北京地区餐饮店高密度区域大多位于热门商圈附近,且围绕商圈形成了不同等级的集聚

区,一般情况下,商圈等级越高,餐饮店分布越密集。

(2)热门旅游景点附近,如天安门、崇文门、前门大街、北新桥簋街、什刹海公园等。旅游热门景点客流量大,密集的流动人口带动了餐饮业的繁荣。

(3)大型住宅区附近,如通州北苑、天通苑、双井青年公寓等。通州北苑附近社区众多,且东侧万达项目的落成带来了大量的居住人口,因此,通州北苑成为仅次于三里屯和朝外街道的餐饮店高密度区域。天通苑是北京地区大型经济适用房居住区,作为典型的近郊城乡结合部地区,吸引了大量中心城区的务工人员居住于此,因此餐饮店分布十分密集。

(4)交通线附近,如六里桥客运站东等,六里桥客运主枢纽是全国一级客运枢纽,出京去往北京房山区和石家庄沿途的车辆都经由此地。交通的便利性为其带来了大量的流动人口,一般越靠近交通

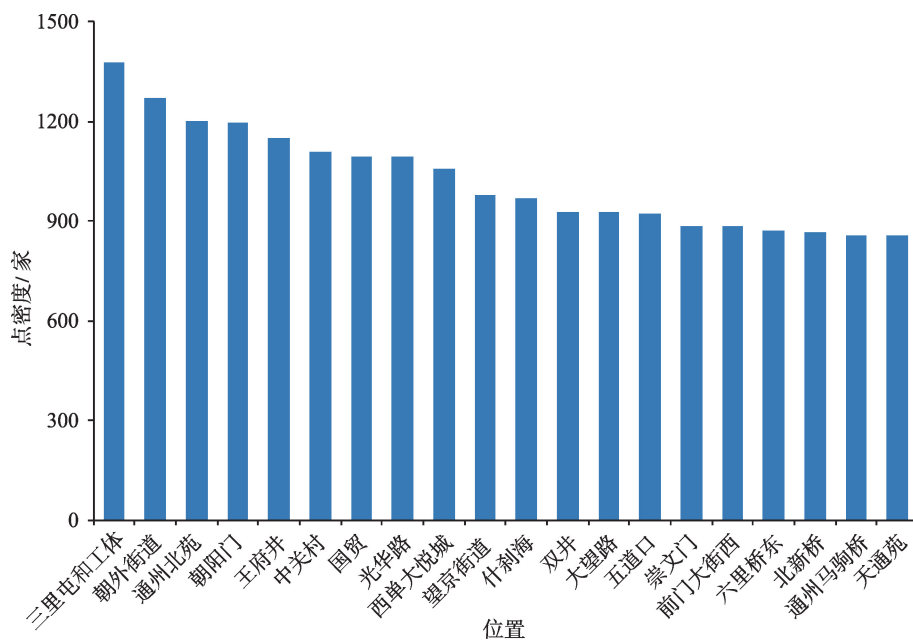


图4 点密度 TOP20 的餐饮服务业聚类中心

Fig. 4 The cluster centers whose local density is on TOP20

枢纽的地方,餐馆分布越密集。

总体来看,北京地区餐饮服务业空间格局整体分布较为均衡,大多数商铺都处于北京的主城区,呈现出从主城区中心向外聚集度圈层递减特征和沿交通线扩散的布局特征。据官方统计资料,北京地区2017年朝阳区常住人口密度为8415人/km²;海淀区常住人口密度为8575人/km²;西城区常住人口密度为25 602人/km²;东城区常住人口密度为22 073人/km²。将餐饮服务业空间分布与各城区常住人口的分布进行对比,发现二者有较高的一致性,这表明餐馆在空间分布上具有明显的接近市场和消费者的特征。餐饮服务业明显表现出围绕主城区重要的商圈进行布局的趋势,并受商圈规模的影响;此外,交通条件也是影响餐馆空间分布的重要因素。

4.2 基于人均消费等级的空间分布特征

餐饮店人均消费一定程度上能够代表该地区的消费水平。经统计发现,大众点评上餐饮店的人均消费多集中在20~500元区间内。为了更好地区分餐饮店的空间分布特征,本文将人均消费额划分为4个等级:人均消费200元以上的为一类餐厅;100~200元的为二类餐厅;30~100元的为三类餐厅;人均消费30元以下为四类餐厅。对原始数据进行进一步的清洗,去掉了价格为缺省值的商铺,共余46 659家商铺数据,餐馆价格热力图如图5所示。

人均消费不低于200元的一类餐厅共计1391家,约占总数的2.98%,如图5(a)所示。从图上可知,一类餐厅仅集聚在朝阳区、东城区、西城区以及海淀区内,呈现东多西少的空间特征,且明显表现出围绕城市中重要商圈进行布局的趋势。根据一类餐厅的集聚地可得到北京地区消费水平较高的商圈主要有三里屯、工人体育场、朝外街道、国贸、大望路、亮马桥;王府井世纪大厦和新天地、东单附近;西单大悦城和金融街等。

人均消费位于100~200元的二类餐厅共计5289家,约占总数的11.34%。由图5(b)可看到,二类餐厅数量明显增多,其分布聚集地由一类餐厅聚集地逐步扩展到了朝阳区的望京、四元桥、机场高速、朝阳大悦城;东城区的前门、崇文门、东直门东四十条及北新桥附近;西城区的什刹海公园、宣武门、地铁三号线附近;海淀区的新中关、五道口、五棵松以及学校等地,但主要还是集聚在中心城区中,呈现东多西少的分布特征,而在其它城区数量则较少。

人均消费位于30~100元的三类餐厅共计25 029家,占比53.64%。三类餐馆小尺度集聚区更多更分散,逐渐到了四环以外,除了在东西城区及朝阳区出现了更多的聚集地,范围也逐步扩展到了天通苑、管庄常营、通州北苑、昌平区政府、顺义区政府附近,且表现出了沿交通线扩展的集聚特征。

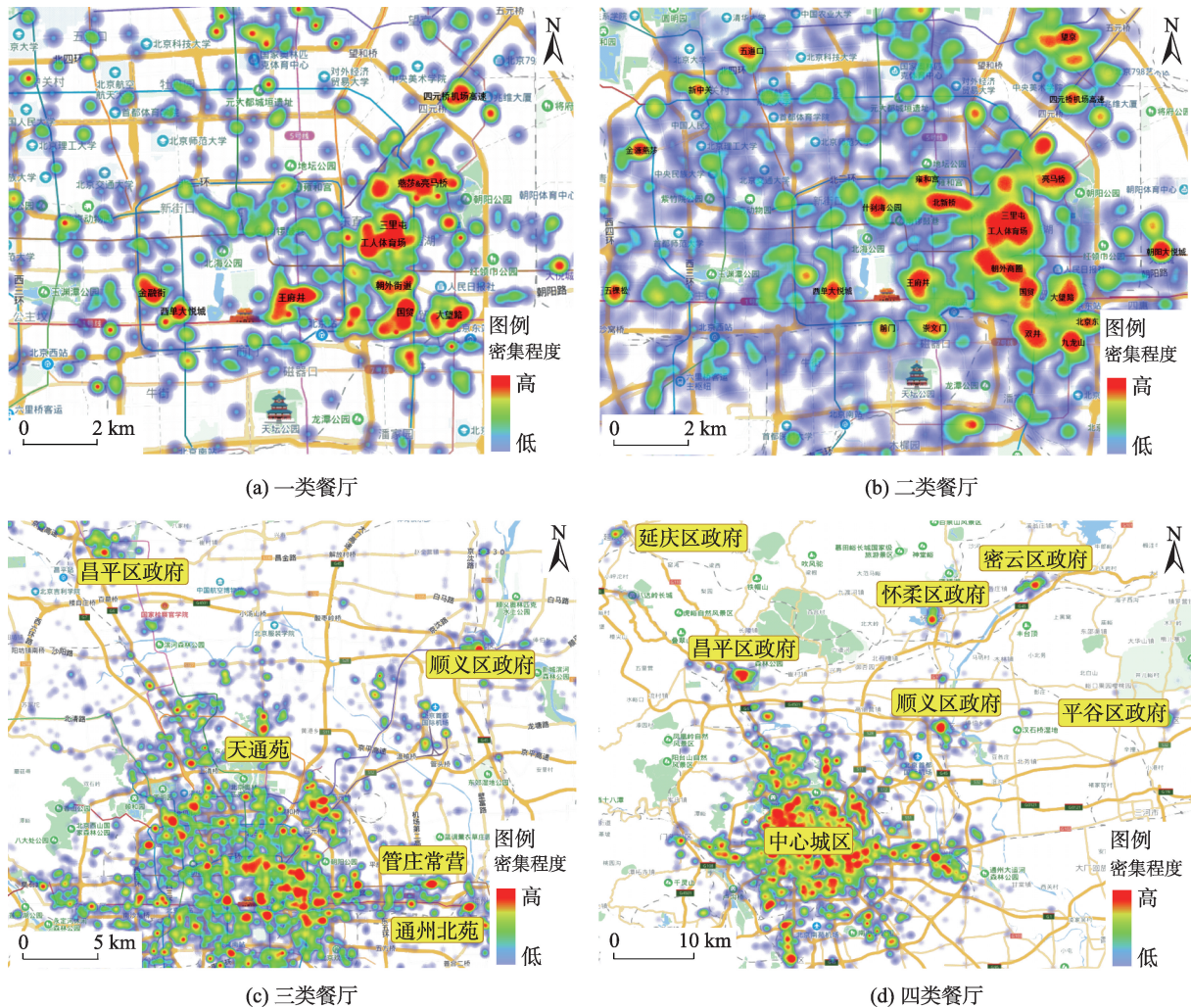


图5 北京地区餐厅价格热力图

Fig. 5 Thermodynamic chart of restaurants' prices in Beijing area

人均消费不高于30元的四类餐厅则有14 950家,约占总数的32.04%。四类餐厅分布基本覆盖北京地区所有中心城区,在怀柔区、密云区、延庆区等周边城区的区中心也出现了集聚中心。

总的来说,不同消费档次餐馆的空间分布特征明显存在差异,且表现出一定的中心地等级分布特征。北京地区中低档餐馆数量占总数的绝大部分,而高档餐馆数量则较少,且多集中在东西城区CBD周边和朝阳区热门商圈和旅游景点附近。随着档次降低,餐馆的集聚特性呈现出均匀、分散的特点,且更易受交通便捷性的影响。

4.3 影响因素分析

从上述分析可知,某地区餐厅密集程度和人均消费等级都可以作为衡量该地区繁华程度的指标。一般来说,餐厅数量越多,空间分布越密集,人

均消费越高,该地区就越繁华,越有可能位于繁华商圈或热门旅游景点附近。但第4.1、4.2节的分析结果却出现了非一致性,如通州北苑及中关村等地的餐厅分布虽十分密集,但其人均消费却并不高,而金融街等虽然餐厅数量相对不多,但是其人均消费等级却很高。为了探究这种现象出现的原因,本文以所有聚类中心点的点密度为横坐标,归属于该聚类中心类别的餐饮店均价为纵坐标,绘制散点图如图6所示。同时,为了更好地解释基于餐饮店分布密集程度和人均消费得到的结果的差异性,本文选取了散点图中具有代表性的区域将其划分为4类:高密度、高消费区域;高密度、低消费区域;低密度、高消费区域;低密度、低消费区域。在地图上找到该集群所处的位置和空间区位特征,进而得到产生这种差异性的原因,详细如表1所示。

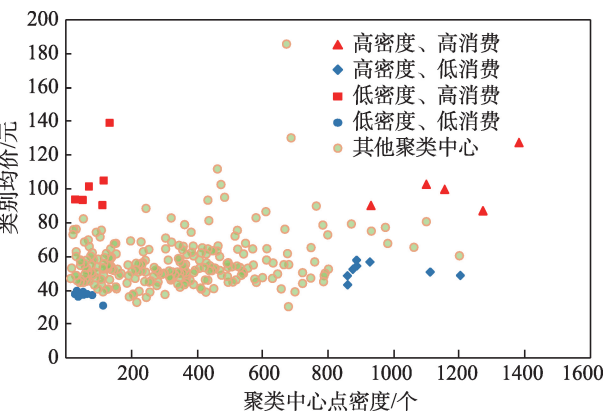


图6 聚类中心散点图

Fig. 6 Scatter diagram of all clustering centers

第一类中的三里屯是北京地区最繁华的商圈之一,汇集了大批高收入、高消费的白领群体,同时也是热门的旅游休闲场地。王府井大街作为北京最著名的商业街,俨然已成为国内外游客和中高收入白领阶层的休闲娱乐购物场所。朝外街道作为北京CBD的“西大门”,包括朝外SOHO、SOHO尚都、光华路SOHO等办公场所,同时,世贸天阶、悠唐生活广场等购物中心也吸引了更多人在朝外休闲消费。高密度的人流量导致该类地区餐饮服务业集聚,且目标客群的高消费力导致这类区域餐饮店的定价也较高。

第二类主要以通州北苑,中关村,五道口为代表。通州北苑辖内社区众多,密集在社区人口推动了餐饮业行业的飞速发展。同时,地铁通州北苑站也吸引了部分朝阳、望京周边客流,便利的交通条件进一步加快了附近餐饮业的繁荣,因此该地区餐饮业分布十分密集。但通州北苑附近多为社区居民和务工人员,因此这一地区餐饮店的定价并不高,符合目标客群的实际购买力。中关村和五道口周围是鳞次栉比的IT电子卖场和海淀高教园区,因此餐饮店分布密集,同时人均消费不高,符合IT从

业者、高校等消费理念。而在北京城市发展与CBD东扩的双重压力下,越来越多的人不断的沿着朝阳北路、朝阳路以及南面的京通快速逐步向东部扩张。管庄就位于这样的一个地理位置上,因此管庄作为未来CBD的产业功能补充区,餐饮店在这一地区慢慢出现了集聚性,由于该地区正在发展中,因此均价相对也并不高,属于餐饮店密度高,但消费相对较低的地区。

第三类是以朝阳区裕京花园别墅区为代表的餐饮店分布稀疏,但人均消费较高的区域。北京地区一些繁华的别墅区汇集了大量高收入人群,其餐饮店分布虽不密集,但却是典型的高消费地区。

第四类是大兴安定镇为代表的低密度、低消费区域,其多为北京地区经济水平相对不发达的城区中的村庄等。综上可得出餐饮店空间集聚程度和定价明显表现出接近市场和消费者的特征。

5 结论

本文通过数据挖掘技术获取互联网上北京地区大量餐饮店信息,采用快速搜索发现高密度区的思想提取餐饮店聚类中心,之后对无类别剩余点进行归类从而实现聚类。与传统算法相比,本文采用的算法不仅计算效率高,适用于大量数据点的处理,且通过密度和距离的双重信息来选取聚类中心,克服传统算法高密度区域间有粘连的缺陷。最后,从其空间分布密集程度和人均消费方面对其蕴含的地理特征和规律进行了分析探讨。研究发现:

(1)北京地区餐饮店整体空间分布格局为东多西少,呈现出以主城区为核心向外聚集程度逐级递减的特征和沿交通线扩展的特点,并明显表现出围绕重要商圈、旅游热门景点和住宅区进行布局的趋势。同时,餐馆的空间分布受城市商圈规模,人口密度和交通便捷性影响较大。

表1 北京地区典型集聚区等级体系

Tab. 1 Hierarchical system of typical gathering areas in Beijing

等级	集聚区代表	空间区位特征
高密度、高消费区域	三里屯、朝外街道、王府井大街、国贸、天安门广场、大望路等	繁华商圈/热门旅游地点,人流量大;周围包括大量企业、外国商社和高端楼盘,汇集了大批高收入、高消费的白领群体
高密度、低消费区域	通州北苑、中关村、五道口、六里桥东、西直门等	辖内社区众多/邻近高等院校;具备地理优势和便利的交通条件
低密度、高消费区域	朝阳区裕京花园别墅附近、顺义区龙安别墅附近等	别墅区
低密度、低消费区域	大兴安定镇、房山琉璃河镇、顺义区河南村等	周边城区分散的居民区等

(2)北京地区呈现等级体系特征,高档餐厅少,中低档餐厅多,且高档餐厅在东西城区、朝阳区及海淀区出现小范围内大规模的聚集分布,而中低档等餐饮店集聚区更多更分散,并受交通便利性的影响。

(3)餐馆分布密集程度和定价表现出接近市场和消费者的特征。繁华商圈和旅游热门地主要面向商务活动、高收入人群和游客,餐饮业呈现出高密度高消费的特点。而居住区、交通沿线、学校附近等的餐饮店则多呈现高密度低消费的特征,以契合大众百姓的消费水平。

互联网上的餐饮业信息具备数据量大、时效性强的特点,在一定程度上能够完整准确地反映出城市餐饮店分布的空间规律。但本研究还存在一定的局限性:①原始数据集的偏差性,如大众点评上餐饮店的消费水平偏低且部分小型餐饮店并未在大众点评上注册商户等,今后的研究可以通过融合多源数据来弥补单一数据源所产生的偏差;②由于城市的餐饮店更新速度快,因此本文只能针对近一段时间内的餐馆数据进行探索和分析,今后可通过长时间追踪城市餐饮业信息,以期从不同时空尺度来探究餐饮业空间分布的变化规律,从而为餐饮经营者选址和定价提供一定的指导和借鉴。

参考文献(References):

- [1] Roehl W S, Van Doren C S. Locational characteristics of American resort hotels[J]. *Journal of Cultural Geography*, 1990,11(1):71-83.
- [2] Urtasun A, Gutierrez I. Hotel location in tourism cities-Madrid 1936-1998[J]. *Annals of Tourism Research*, 2006, 33(2):382-402.
- [3] Wall G, Dudycha D, Hutchinson J. Point pattern analyses of accommodation in Toronto[J]. *Annals of Tourism Research*, 1985,12(4):603-618.
- [4] Ritter W. Hotel location in big cities. In *Big City Tourism* [M]. Berlin: Reimer, 1986:355-364.
- [5] Cró S, Martins A M. Hotel and hostel location in Lisbon: looking for their determinants[J]. *Tourism Geographies*, 2018,20(3):504-523.
- [6] Lee K H, Kang S, Terry W C, et al. A spatial relationship between the distribution patterns of hotels and amenities in the United States[J]. *Cogent Social Sciences*, 2018.1444918. DOI:https://doi.org/10.1080/23311886.
- [7] 陈传康.中国饮食文化的区域分化和发展趋势[J].*地理学报*,1994,49(3):226-235. [Cheng C K. The culture of Chinese diet: Regional differentiation and developing trends [J]. *Acta Geographica Sinica*, 1994,49(3):226-235.]
- [8] 蓝勇.中国饮食辛辣口味的地理分布及其成因研究[J].*地理研究*,2001,20(2):229-237. [Lan Y. The reasons and distribution of pungent flavour districts in China's dietetics[J]. *Geographical Research*, 2001,20(2):229-237.]
- [9] 梁璐.城市餐饮业的空间格局及其影响因素分析——以西安市为例[J].*西北大学学报:自然科学版*,2007,37(6):925-930. [Liang L. The distribution in space of urban catering and its factors: Xi'an as an example[J]. *Journal of Northwest University (Natural Science Edition)*, 2007,37(6):925-930.]
- [10] 张旭,徐逸伦.南京市餐饮设施空间分布及其影响因素研究[J].*热带地理*,2009,29(4):362-367. [Zhang X, Xu Y L. Study on the distribution in space of urban caterings and its influencing factors: A case study of Nanjing[J]. *Tropical Geography*, 2009,29(4):362-367.]
- [11] 郭伦,刘亮,田原,等.基于网络K函数法的地理对象分布模式分析——以香港岛餐饮业空间格局为例[J].*地理与地理信息科学*,2013,29(5):7-11. [Wu L, Liu L, Tian Y, et al. Spatial pattern analysis of geographic features using network K-Function methods with a case study of restaurant distribution in Hong Kong Island[J]. *Geography and Geo-Information Science*, 2013,29(5):7-11.]
- [12] 赵智勇.融合多源地理大数据的杭州市功能区识别和空间优化研究[D].杭州:浙江大学,2018. [Zhao Z Y. Discovering zones of different functions and research on spatial optimization based on multisource big data fusion in Hangzhou City[D]. Hangzhou: Zhejiang University, 2018.]
- [13] 牛方曲,刘卫东.基于互联网大数据的区域多层次空间结构分析研究[J].*地球信息科学学报*,2016,18(6):719-726. [Niu F Q, Liu W D. Identifying the hierarchical regional spatial structure using Internet big data[J]. *Journal of Geo-information Science*, 2016,18(6):719-726.]
- [14] 吴志强,叶锺楠.基于百度地图热力图的城市空间结构研究——以上海中心城区为例[J].*城市规划*,2016,40(4):33-40. [Wu Z Q, Ye Z N. Research on urban spatial structure based on Baidu heat map: a case study on the central city of Shanghai[J]. *City Planning Review*, 2016,40(4):33-40.]
- [15] Prayag G, Landre M, Ryan C. Restaurant location in Hamilton, New Zealand: clustering patterns from 1996 to 2008 [J]. *International Journal of Contemporary Hospitality Management*, 2012,24(3):430-450.
- [16] Yang Y, Wong K K F, Wang T. How do hotels choose their location? Evidence from hotels in Beijing[J]. *International Journal of Hospitality Management*, 2012,31(3):675-685.
- [17] 秦萧,甄峰,朱寿佳,等.基于网络口碑度的南京城区餐饮业空间分布格局研究——以大众点评网为例[J].*地理科学*,2014,34(7):810-817. [Qin X, Zhen F, Zhu S J, et al.

- Spatial pattern of catering industry in Nanjing urban area based on the degree of public praise from Internet: a case study of Dianping.com[J]. *Scientia Geographica Sinica*, 2014,34(7):810-817.]
- [18] 谭欣,黄大全,赵星烁.北京市主城区餐馆空间分布格局研究[J].*旅游学刊*,2016,31(2):75-85. [Tan X, Huang D Q, Zhao X S. A study on the spatial distribution pattern of restaurants in Beijing's main urban area[J]. *Tourism Tribune*, 2016,31(2):75-85.]
- [19] 杨帆,徐建刚,周亮.基于DBSCAN空间聚类的广州市区餐饮集群识别及空间特征分析[J].*经济地理*,2016,36(10):110-116. [Yang F, Xu J G, Zhou L. Cluster identification and spatial characteristics of catering in Guangzhou based on DBSCAN spatial clustering[J]. *Economic Geography*, 2016,36(10):110-116.]
- [20] 曾璇,崔海山,刘毅华.基于网络空间点模式的餐饮店空间格局分析[J].*地球信息科学学报*,2018,20(6):837-843. [Zeng X, Cui H S, Liu Y H. Analysis on spatial distribution characteristics of restaurant based on network spatial point model[J]. *Journal of Geo- information Science*, 2018,20(6):837-843.]
- [21] 李江苏,梁燕,王晓蕊.基于POI数据的郑东新区服务业空间聚类研究[J].*地理研究*,2018,37(1):145-157. [Li J S, Liang Y, Wang X R. Spatial clustering analysis of service industries in Zhengdong New District based on POI data [J]. *Geographical Research*, 2018,37(1):145-157.]
- [22] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. *Science*, 2014,344(6191):1492-1496.
- [23] 张庆,彭震伟.基于空间聚类分析的杭州市生产性服务业集聚区分布特征研究[J].*城市规划学刊*,2016(4):46-53. [Zhang Q, Peng Z W. Spatial distribution of producer services in Hangzhou central city based on spatial clustering analysis[J]. *Urban Planning Forum*, 2016(4):46-53.]
- [24] 陈宁,彭霞,黄舟.社交媒体地理大数据的旅游景点热度分析[J].*测绘科学*,2016,41(12):167-171,216. [Chen N, Peng X, Huang Z. Popularity analysis of tourist attraction based on geotagged social media big data[J]. *Science of Surveying and Mapping*, 2016,41(12):167-171,216.]
- [25] 王洋,张桂珠.自动确定聚类中心的密度峰值算法[J].*计算机工程与应用*,2018,54(8):137-142. [Wang Y, Zhang G Z. Automatically determine density of cluster center of peak algorithm[J]. *Computer Engineering and Applications*, 2018,54(8):137-142.]