



杭州 O2O 生活服务业的聚集性及空间特征

王世雄, 李天晨, 李泽彪

(浙江理工大学经济管理学院, 杭州 310018)

摘要: 生活服务行业是城市空间的重要组成部分, 分析生活服务行业的空间分布特征是了解一座城市的窗口, 是感知人们生产生活幸福水平的重要视角。分析生活服务行业的空间分布特征, 有助于揭示城市的空间结构并对诸如物流等影响城市运作的基础行业提供理论支撑。利用网络爬虫获取了 O2O 生活服务网站大众点评中杭州市内共 339884 条实体门店数据, 通过多阶段聚类算法识别出 431 个集群。分析发现: 杭州市整体 O2O 生活服务点的聚集效应显著, 空间实体点在西湖东侧形成一个“超级中心”, 沿着该中心向外扩散出多个小型集聚点; 门店聚集性和人口密度呈现出较高相关性, 常住人口数量对 O2O 生活服务类实体点的影响最为显著; 在人口趋于饱和的老城区和高速发展的新城, O2O 生活服务点的种类有较大差异, 区域发展的不同阶段对 O2O 生活服务种类的要求不同。

关键词: 空间格局; 聚类; 位置点; 杭州; O2O 生活服务

中图分类号: K901.2

文献标志码: A

文章编号: 1673-3851 (2019) 12-0603-08

Clustering and spatial characteristics of Hangzhou O2O life service industry

WANG Shixiong, LI Tianchen, LI Zebiao

(School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: The life service industry is an important part of urban space. Analysis of the spatial distribution characteristics of life service industry is an important perspective to perceive the happiness level of people's production and life. Analyzing the spatial distribution characteristics of the life service industry helps to reveal the spatial structure of the city and provide theoretical support for basic industries affecting the urban operation like logistics. Using web crawler technology, we obtained 339884 pieces of physical store data from Dianping in O2O life service website in Hangzhou, and identified 431 clusters through multi-stage clustering algorithm. The analysis found that the aggregation effect of Hangzhou O2O life service is significant. At first, there is a "super center" on the east side of West Lake, which spreads out a number of small gathering points. Then the store aggregation and population density show a high correlation, and permanent resident population has the most significant impact on the O2O life service stores. And at last, the types of O2O life service stores in the old town with asaturated population and the new area with high-speed development have great differences, indicating that the requirements for the types of O2O life service are different at different stages of regional development.

Key words: spatial pattern; cluster; location point; Hangzhou; O2O life service

收稿日期: 2019-07-22 网络出版日期: 2019-09-16

基金项目: 浙江省自然科学基金项目 (LY15G010010)

作者简介: 王世雄 (1976—), 男, 湖北浠水人, 副教授, 博士, 主要从事信息管理与数据挖掘、网络舆情传播方面的研究。

社会活动与城市的空间结构之间是相互联系、相互影响的,生活服务行业作为城市经济的重要行业,其空间分布特征和人类的众多社会活动密切相关。物流作为主要的社会活动之一也必然受到城市空间结构的影响。对生活服务行业的空间特征进行分析,有助于揭示城市的空间结构,从而对诸如物流等影响城市运作的基础行业提供理论支撑。此外,对当前城市商业空间格局的分析,可以预测未来的商业中心的位置^[1],从而在城市规划前期对物流配送中心的选址提供科学的建议,实现降本增效。杭州作为新一线城市,近年来不仅城市空间发展迅速,生活服务行业的线上化比例也位居全国前列。诸如美团、点评之类的 O2O 生活服务平台,利用互联网的天然优势整合了大量的生活服务类实体相关的数据。由于杭州互联网渗透率高,生活服务行业线上化程度高,故利用 O2O 生活服务业数据可以为分析杭州生活服务行业的空间分布特征以及聚集性提供一个新途径。

一、国内外研究现状

关于城市空间结构和社会活动之间的相互影响,目前已经有了了一定的研究。伦敦大学的 Hillier^[2]提出了空间句法理论,该理论提出人类对于城市空间环境的认知在很大程度上是由空间结构形态决定的,而空间认知又在一定程度上决定了人类的空间行为。在空间句法理论基础之上创建的空间句法工具,提供了从图论的角度定量研究城市形态和城市生活之间关联性的新思路。例如,刘承良等^[3]以武汉为例,通过对空间距离进行计算,以拓扑算法和句法模型为基础构建了通达性数理模型,分析了路网通达性与城市空间形态之间的量化关系。陶伟等^[4]结合句法工具中轴线图和意向图,探索了广州小洲村的空间形态与不同空间认知的使用者之间的关系,发现本地常住居民对村落深层次空间格局的认识相较于其他使用者更加容易。黄晓冰等^[5]引入信息熵模型,分析了地铁商圈零售业的熵值和均衡性特征,探讨了轨道交通对零售业种类的影响。Cró 等^[6]从影响酒店位置的因素出发,结合商圈信息发现,高档酒店一般分布在商业发达的 CBD,而中低档的酒店一般会更加寻求集聚效应。

随着互联网的发展,海量空间位置数据的累积对经济地理学产生了一定的影响^[7],使得关于城市空间特征的研究正在逐渐从以生活圈^[8]、商圈^[9]以及轨道交通^[5]等作为支撑,通过问卷调查^[10]、实地

走访^[11-12]、模拟仿真^[13]以及统计年鉴^[11,14-15]获取数据的方式,转变为以基于位置服务为依托,通过互联网获取海量 O2O 实体点数据的方式进行分析和研究。杨帆等^[16]选取广州作为研究案例地,通过百度地图 API 获取广州市区餐饮类 POI 点的空间数据,通过引入空间点密度聚类算法识别了广州市区餐饮集群,并以此为依据分析了广州的空间特征。沈体雁等^[17]利用通过网络爬虫获取百度地图的 POI 数据集,结合空间句法和核密度估计法,对北京市的服务行业区位选择和交通指向之间的关系进行了探索。曾璇等^[18]基于从大众点评平台获取的门店点数据,一方面通过核密度估计法探究了广州珠海地区的生活服务门店的空间分布特征以及城市内热点路段的分布情况,另一方面通过网络双变量 K 函数法对生活服务点的分布与公交站和小区之间的相关性进行了分析。张晓瑞等^[19]利用空间句法和基于位置服务的大数据,研究了合肥的人口分布空间格局并提出了合肥市的人口发展建议。徐晓宇等^[20]以北京地区作为研究对象,采用网络爬虫技术获取了餐饮门店数据,引入基于密度的 CFSFDP 聚类算法,从北京市餐饮店空间分布密集度和人均消费等级等方面对餐饮业背后蕴含的地理聚集特征进行了分析。随着研究的深入,针对当前数据量大、使用传统算法处理速度慢且效率低等问题,有学者基于空间点的位置大数据提出了更加符合处理场景的算法。Rodriguez 等^[21]提出了一种基于点密度快速搜索聚类中心的算法 CFSFDF,它通过只考虑点与点之间的空间位置来对距离进行计算,不需要将点映射到一个向量空间中从而大大提高了计算效率,适用于处理数据量较大的空间点数据。于彦伟等^[22]提出了一种面向位置大数据的快速密度聚类算法 CBSCAN,该算法首先基于 cell 网格将空间点按照位置进行分区,在分区的基础之上改进 DBSCAN 算法,形成基于网格的密度簇的相邻和包含关系,可以极大地减少高密度区域的距离计算,利用位置数据的内在特性提高了聚类效率。

上述各类研究为城市基础设施及生活服务实体的研究累计了一定的成果,但是利用大数据开展生活服务行业的空间特征研究尚处于探索阶段。近几年,杭州数字经济、信息经济发展迅猛,城市范围进一步拓展,空间结构不断调整,与之相应的生活服务业也处于快速重构的过程中。在此期间,在杭州范围内以互联网为依托的 O2O 生活服务业衍生和积累了大量可用于解析杭州生活服务业的数据。本文

利用从网络获取的大量空间位置数据,对杭州市的 O2O 生活服务行业空间分布特征进行分析,通过各个聚类中心所在的具体位置进行横向对比以确定影响聚集性的主要因素,进而对不同发展阶段的区域之间的服务种类空间分布差异性进行分析。研究结果可以看作杭州生活服务行业空间结构整体的一个映射,将有助于政府有关部门更高质量地做好城市规划,有助于相关企业更好地做出商业规划和决策,有助于居民更好地获得生活服务的资讯和信息,进而提高整座城市的幸福感。

二、数据获取及其处理

(一) 数据获取

目前对于杭州空间特征的研究中,研究数据一般是通过实地走访^[23]、问卷调查^[24]等传统的方式获得。但由于城市生活服务行业空间点具有分布广泛、数据量大等特点,用传统方式获取的数据不能全面、有效地量化城市空间格局的实际情况。爬虫技

术是一种为了从网络中获取数据的技术,它可以通过某些特定的规则以滚雪球的方式对互联网上的数据进行抓取,XPath 路径类似于文件系统中的文件路径,通过对 html 页面中的元素进行 XPath 解析可定位特定的元素以获取相应的数据。本文利用 python 的 request 模块获取页面,之后使用 XPath 路径对页面中的元素进行解析,将解析之后的数据存储到 HDFS 中以便后续使用。开发过程中使用语言为 python3.6.5,环境为 hadoop2.6.4。

本文利用网络爬虫技术从大众点评网获取了杭州 10 个区共 339884 条 O2O 生活服务类实体门店的数据,爬取时间为 2018 年 11 月 17 日到 2019 年 2 月 1 日,其空间密度分布情况如图 1 所示。数据属性包括店铺名称、一级分类、二级分类、店铺地址以及营业时间等,由于计算距离时需要用到各个实体点的经纬度信息,所以在获取原始数据之后,根据其地址详细字段信息,通过高德地图开放 API 获取各个位置点的经纬度信息。

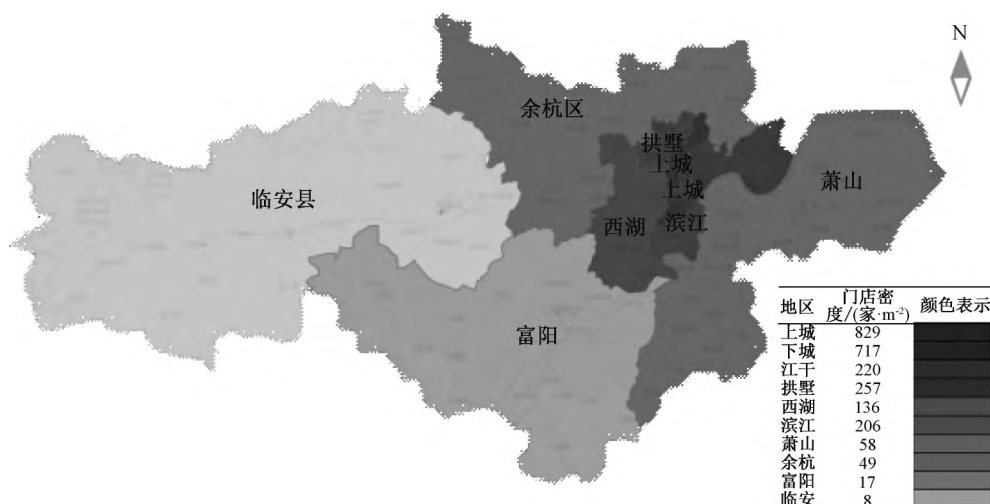


图 1 杭州市 O2O 生活服务类实体空间密度分布

(二) 算法描述

1. 算法选取

分析城市生活服务业聚集性是典型的聚类问题,可以利用聚类算法进行研究。通过对 O2O 生活服务实体位置数据进行聚类分析并将其与城市规划进行对比,可以反映出杭州生活服务行业的空间特征和生活服务业整体布局。生活服务业的海量数据是多维的、非凸集的,考察其聚集性和空间特征需要准确地获取各个聚类簇的中心位置。

对于形状任意的空间点数据,运用 DBSCAN 密度聚类算法^[25]从样本密度的角度出发来考察样本之间的可连接性,通过样本之间的可连接性不断聚类,可以在聚类过程中屏蔽其形状的任意性从而获

得较优的结果。DBSCAN 算法将数据集划分为不同的簇,研究聚类簇和城市区域间的关联性需要对比簇中的点和城市相应的商圈、生活圈之间的关系。一个聚类簇中的点往往随机的分散在簇区域范围内的各个位置,聚类的中心点可以代表整个聚类簇所处的位置。在通过 DBSCAN 求出了各个簇之后,运用将参数 k 值设置为 1 的 k -means 算法可以求出聚类中心点,因为聚类簇数 k 是 k -means 算法的唯一需要确定的一个参数,所以在 k 为确定值的情况下 k -means 算法的效率较高且模型可解释性较强。

DBSCAN 和 k -means 两个算法均需要计算任意两个节点的距离,在海量数据的情况下,传统的欧拉距离公式等计算方法时间复杂度高,运行效率偏

低。因此论文引入 Geohash 作为计算地理位置距离的算法,Geohash^[26]是一种基于空间 Z 阶曲线的数据结构,它利用空间填充曲线中的 Z 阶曲线把空间划分为网格,在网格精度确定了的情况下,Geohash 字符串的长度也就确定了,在距离越近的地方的 hash 字符串公共前缀的长度总是越长,所以 Geohash 可以用来表示一定精度范围内的一个点,并且该点可用于基于空间距离的近邻点搜索。其具体的编码方式为:a)分别根据坐标点的经纬度计算二进制编码,可以得出一个经度编码序列和一个纬度编码序列;b)将经纬度二进制编码序列进行组合:经度放在偶数位,维度放在奇数位,生成一个新的编码串;c)将产生的二进制序列串转化成十进制之后使用 base32、base64 等编码方式进行编码,完成经纬度串的 Geohash 编码。Geohash 算法将经纬度转码成为了一个可以用来表示空间范围的较短的序列串,通过判断编码串公共前缀的长度即可判断点之间的距离,从而降低计算的复杂度。Geohash 算法将空间划分为 Z 阶曲线的编码过程示意图如图 2 所示。

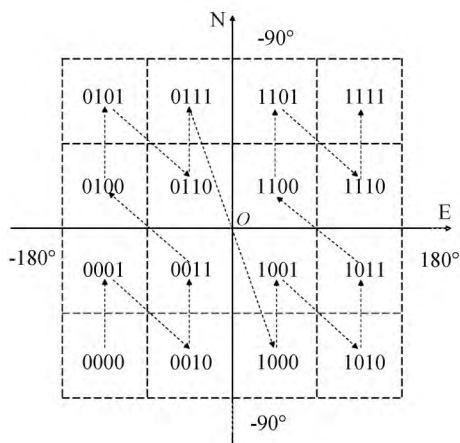


图2 Geohash 算法编码过程示意图

2. 算法改进

对于一个给定的数据集 $D = \{X_1, X_2, \dots, X_m\}$, DBSCAN 算法存在如下几个基本定义:a)对于数据集内任意的点 $x_j \in D$, x_j 的 ϵ 半径范围称为点 x_j 的 ϵ 邻域,集合内任意点的 ϵ -邻域范围内的点与该点的距离小于 ϵ ,即 $N(x_j) = \{x_j \in D \mid \text{dist}(x_i, x_j) \leq \epsilon\}$; b)若 x_j 的 ϵ -邻域至少包含 $MinPts$ 个样本,即 $|N(x_j)| \geq MinPts$, 则称 x_j 是一个核心对象;c)若 x_j 位于 x_i 的 ϵ -邻域中,且 x_i 是核心对象,则称 x_j 由 x_i 关于 $(\epsilon, MinPts)$ 可密度直达;d)对于 x_i 和 x_j , 若存在样本序列 p_1, p_2, \dots, p_n , 其中 $p_1 = x_i$,

$p_n = x_j$ 且 p_{i+1} 可以由 p_i 密度直达,则称 x_j 由 x_i 密度可达;e)对 x_i 与 x_j , 若存在 x_k 使得 x_i 与 x_j 均由 x_k 密度可达,则称 x_i 与 x_j 密度相连。基于以上概念,给出 DBSCAN 算法中聚类簇的定义为由密度可达关系导出的最大密度相连的样本集合。算法步骤描述如下:a)任意选择一个没有类别的对象,若该对象的 ϵ 范围内的点小于 $MinPts$ 个,则标记该点为噪声点,否则将该点标记为核心点,找到所有这个核心点能够密度可达的样本集合,即为一个聚类簇;b)重复过程 a)直到所有对象都被标记为止,算法结束,输出聚类簇。

k -means^[27]算法是一种对给定的数据按照距离划分集群的聚类算法,其基本思想是将样本集划分为 k 个簇,让簇内点之间的距离尽量小,同时让簇间的距离尽可能地大。其算法步骤描述为:a)确定聚类中心的数量 k ,并随机选取 k 个点作为聚类中心;b)将样本中的点按照最短距离分配给最邻近的聚类;c)在每一个聚类中寻找新的聚类中心;d)重复 b)、c)直到聚类中心收敛于某一点不再发生变化,聚类完成,得出中心点的位置。

针对 DBSCAN 算法和 k -means 算法的优势和不足,本文设计了基于 Geohash 编码的多阶段聚类算法,具体算法如表 1 所示。

表1 基于 Geohash 编码的多阶段聚类算法

input: Dataset D
parameter: Parameter $(\epsilon, MinPts)$
algorithm:
① cluster C
② for each $x_j \in D$:
③ for each $x_i \in D$:
④ $N(x_j) = \{x_i \in D \mid \text{dist}(x_i, x_j) \leq \epsilon\}$
⑤ If $ N(x_j) \geq MinPts$ then $p_1 = x_i, p_n = x_j$
⑥ If $\text{dist}(p_i, p_{i+1}) \leq \epsilon$ then
⑦ $C_j.append(x_i)$
⑧ Return C_j
⑨ for each $p_i \in C_j$:
⑩ for each $p_j \in C_j$:
⑪ $d_{ij} = \text{Geohash}(p_i, p_j)$
⑫ $d_j += d_{ij}$
⑬ for each d_j :
⑭ If $d_j > d_{j+1}$ then $d_j = d_{j+1}$
else $p_j = d_j$
⑮ Return p_j
Repeat ①~⑮ return C_i
output: Dataset C_j , Center Point p_j , dataset C_i

表 1 中算法步骤为:a)利用 DBSCAN 算法将空间中所有点分为若干个聚类簇;b)对每一个簇设定其聚类簇数 k 为 1,通过 k -means 算法求出其聚类

中心点;c)以质心为输入再次利用 DBSCAN 算法将质心进行聚合,按照簇中质心数量的多少划分该簇所属的等级。其中位置点之间的距离利用 Geohash 进行计算,公共编码的前缀相似性越高,则其距离越近。相较于利用传统的欧氏距离来衡量点之间的距离,Geohash 通过在数据预处理就进行编码,在算法运行过程中只需要对长度较短的一维编码进行比较从而大大提高了算法的运行效率。

三、结果及其分析

(一)聚类结果概述

对杭州地区的 O2O 生活服务点进行密度聚类分析,其中 DBSCAN 算法参数设置为 $\epsilon=100\text{ m}$ 、 $\text{MinPts}=25$,基于空间点的经纬度进行聚类,在研究范围内合计识别出了 431 个簇,根据聚类后的数据做出门店分布热力图,如图 3 所示。从图 3 中可以看出,服务点的分布呈现出沿着一个超级中心向外发散的特征。根据聚类中心位置再次将密度聚类的集群中心点的聚集等级分为三级,第一等级是重叠的上下城区为中心向东北方向辐射的超大规模集群,第二等级是向东延伸的下沙区域和向南延伸的滨江区域以及萧山政府所在区域等,第三等级是大杭州范围内的富阳、临安区域以及余杭政府所在区域、古镇塘栖和机场附近的瓜沥等。

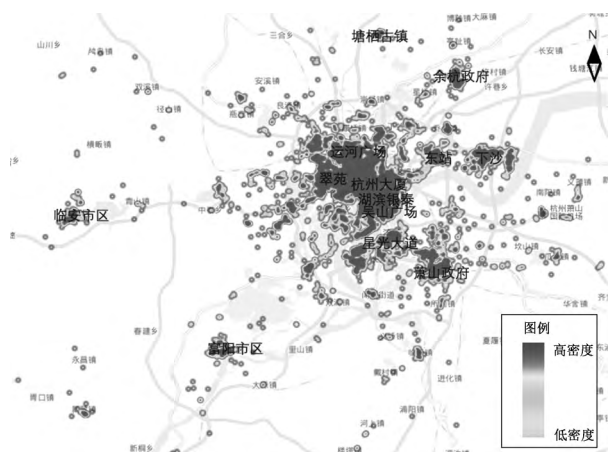


图3 O2O 服务点分布热力图

对 O2O 生活服务聚类中心的密度值分析后发现,杭州市的 O2O 生活服务点分布出现沿着一个超级中心向外延伸的结构,其中最大的中心在西湖风景区以东的方向,由南至北对应于上下城区两个行政区划中的河坊、湖滨以及武林三个主要商圈,和商业综合体的现状格局^[28]趋于一致。除了与老城区重叠的西湖风景区外,位于北边的塘栖古镇风景区也形成了集聚效应,其周围一级分类为美食的门店

占比为 44.01%,远高于全市美食平均占比 30.28%,说明在热门景区大量流动的游客带动了餐饮行业的发展。除了与老城区重叠的城站火车站以外,杭州东站和萧山国际机场也形成了小范围的聚集效应,和风景区相似,交通枢纽附近一级分类为美食的占比也相对较高。下面分别从空间点分布密度和服务种类两方面对比分析杭州 O2O 生活服务业的聚集性和空间特征。

(二)基于空间密度的聚集性分析

根据杭州市统计局公布的《2018 年杭州统计年鉴综合篇》第 2 卷人口和就业人员^①可知 2017 年末杭州市的土地面积和人口密度分布的数据,据其做出以区政府为中心点的人口分布密度热力图,如图 4 所示。对比分析服务点分布热力图和人口分布热力图可以发现,二者空间特征呈现出较高的一致性,即以上下城区为中心,向四周扩散的聚集特征。二者之间存在相似性可以说明人口的分布是影响服务点分布的因素之一。聚类分析所得 O2O 生活服务点规模排名前 15 的聚类中心如表 2 所示。聚集等级为通过密度聚类算法求出的中心点聚集效应的分级,聚集等级越高说明该中心点附近的中心点位置越多,从而说明该中心点附近的聚集效应显著。人口密度为从统计年鉴中获取的中心位置所属的行政区域中人口的密度。从表 2 可以看出,聚类中心规模和聚集等级与人口密度的数量级呈现出较高的一致性,设人口密度为自变量 $density$,聚类中心规模为因变量 $scale$,利用 SPSS22 软件对规模与人口密度做出双对数回归分析得出幂函数 $\ln(scale)=0.6941\ln(density)-0.2733(R^2=0.7263)$,可知常住人口密度对 O2O 生活服务业规模的影响显著。此外,富阳和临安区域的行政区域面积较大但是大部分区域居住人口较少,所以其行政区划整体人口密度较低,但是在人口聚集的区中心其空间服务实体聚集程度较高,这进一步说明了人口对服务业规模的影响较大。

同样处于人口密度量级最大的下城区以及轨道交通 1 号线沿线,西湖文化广场以北的德胜新村、胜利河美食街以及朝晖小区三处均存在聚类中心点的位置,其对应的聚类簇规模分别为 128、135 和 97,相较于向南区域西湖沿线聚类规模为 954 的湖滨银泰聚集点,可以看出风景名胜是区是影响 O2O 生活服务点空间分布的因素之一。

① 数据来源:2018 年杭州统计年鉴, http://tjj.hangzhou.gov.cn/art/2018/10/30/art_1653175_35022585.html。

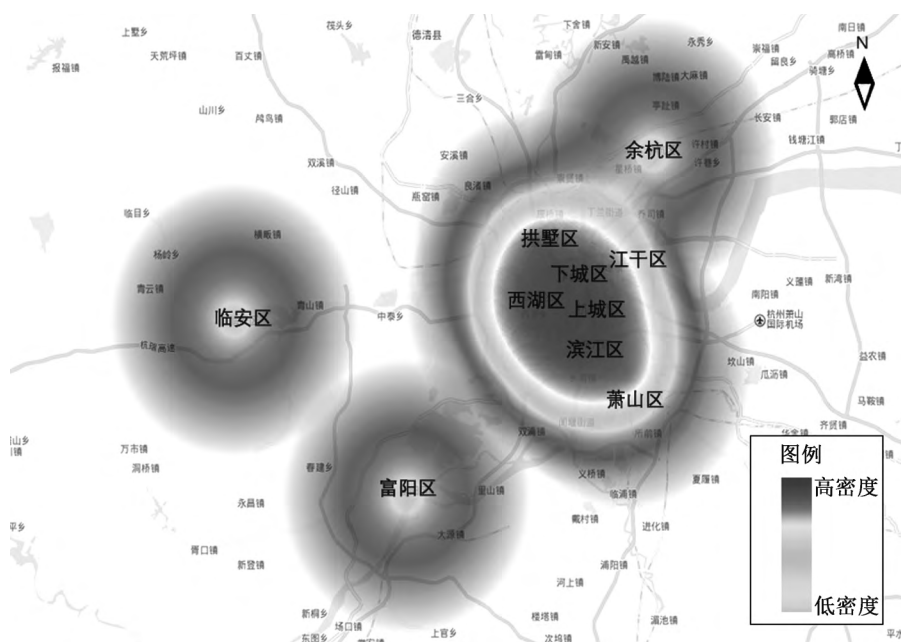


图4 人口分布热力图

表2 聚类中心规模及其所属区域人口密度

规模/家	中心位置	聚集等级	人口密度/(人·km ⁻²)	规模/家	中心位置	聚集等级	人口密度/(人·km ⁻²)
1342	武林广场	1	17129	211	临安市区	3	189
1021	嘉里中心	1	19333	209	骆家庄	2	3270
954	湖滨银泰	1	19333	183	客运中心	2	5371
671	河坊街	1	19333	176	高沙商业区	2	5371
438	城站火车站	1	19333	155	临平	2	1208
318	星光大道	2	4808	151	萧山政府	3	1423
235	翠苑	1	3270	137	滨江高教园	2	4808
226	富阳市区	3	409				

规模前15的位置中除富阳、临安区域外的13个中心位置均有地铁站点,根据集群的中心地点所在位置可以看出杭州O2O服务行业空间点分布的“超级中心”处于地铁1号线沿线的武林广场、嘉里中心、湖滨银泰和河坊街一带。地铁2号线的翠苑、骆家庄以及4号线的萧山区政府也存在聚集规模较大的聚集中心,但其规模远不如1号线沿线中心点的规模。由于地铁1号线的通车时间为2012年,早于2014年开通的2号线和2015年开通的4号线,说明城市内的轨道交通影响O2O生活服务点分布,并且其发展也会带动周边O2O生活服务点的发展。

(三)基于服务种类的形态特征分布

O2O生活服务点的聚集程度在一定程度上可以代表地区的便利程度,从而间接代表了其发达程度。不同种类的O2O生活服务点也代表了该地区不同种类的形态特征。选取湖滨银泰、城站、翠苑以及高沙几个样例聚类中心分别代表景区、交通枢纽、小区以及大学城进行对比分析发现,所有样本簇中

一级分类为美食的餐饮空间点占比均高于其他类别的空间点,其中湖滨银泰以40.89%位列第一,反映了流动人口多的景区O2O服务行业的结构会更倾向于餐饮;翠苑和城站分别为32.91%和35.01%更加接近杭州整体水平,说明服务点的种类分布和空间成熟度密切相关。位于大学城的高沙家装占比为3.62%,远低于平均水平8.00%,说明人口流动性对服务也存在影响,人口流动性越大家装所占比例就越小。与杭州市O2O服务行业的整体分布情况更加相似的区域为翠苑和城站火车站,这两个聚类中心人口密度较大,进一步说明了人口是影响服务点空间分布的首要因素。

分别对一级分类为美食、购物、家政和家装的位置点进行密度分析并画出海量点分布图,发现各类O2O生活服务点的整体走势类似,但其局部热度存在一定差异。图5(a)—(d)分别代表美食、购物、家政以及家装四类O2O生活服务实体的空间海量点分布图,可以看出四类O2O生活服务中美食服务点的聚集程度最高,家装服务点的聚集程度最低,购物

和家政服务分布的聚集性处于二者之间。聚集程度最高的美食类实体数量最多合计 103069 个, 最少的家装类实体共 28658 个服务点。对该格局形成的因素可以从两个方面进行分析。一方面, 人口趋于饱和的中心城区聚集点如大学路新村、朝晖等形成较早, 生活空间形态和住所趋于稳定所以其对家装的需求也较少, 美食、购物和家政等在稳定的生活中需求量更大的 O2O 生活服务类别便相应上升。余杭往临安方向的留下、五常, 离萧山政府较远的机场往东南方向的瓜沥等正处于高速发展的新兴区域, 家装类型的 O2O 服务点占比较高说明城市空间发展的初期必定伴随着住所的发展, 而美食、购物以及家

政等类别的空间实体点会随着区域的发展逐步发展。另一方面, 不同的服务领域具有不同的特性也是导致其空间分布不同的因素之一。从经济学的角度来说, 餐饮行业更类似于生活必需品, 所以餐饮行业的分布和人口的分布密切相关, 且对人口是否流动敏感度较小但对人口数量的要求较高。家装相比餐饮明显更加趋向于寻找稳定的客源, 而家装服务点的客源是新发展区域中需要装修的住宅, 显然其对人口的数量敏感度较小而对户数较为敏感。通过海量点图中的分布可以看出, 在人口相对还较少但住宅区已经发展起来的新发展区域, 扩展增量市场已经成为其增加稳定客源的重要手段。

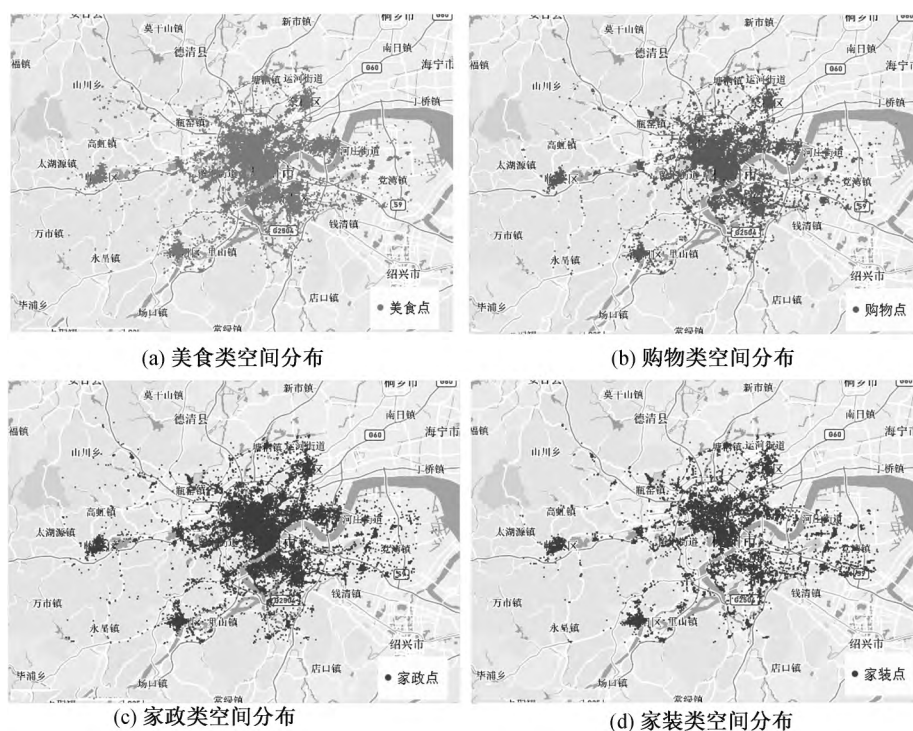


图5 按种类划分的 O2O 生活服务类实体空间分布

四、结 语

本文利用基于 Geohash 编码计算距离的多阶段聚类算法, 对通过爬虫技术获取的大众点评网站 O2O 生活服务实体点的数据进行聚类分析发现: 杭州市整体生活 O2O 服务实体点的聚集效应显著, 具体格局为 O2O 生活服务空间实体点在西湖东侧武林一湖滨一吴山商圈连线形成一个“超级中心”, 沿着中心向外扩散出多个小型集聚点, 聚集点的分布和常住人口的分布呈现出较高的一致性。通过对集群规模大小进行分析, 发现其分布明显和人口密度分布呈现出较高的一致性; 对规模与人口密度做出双对数回归分析可知, 常住人口密度对 O2O 生活服

务业规模的影响显著。除人口因素外, 风景区与轨道交通也对 O2O 生活服务实体点空间分布存在一定的影响。不同种类的 O2O 生活服务在不同的区域有所差异, 美食类的服务点在所有 O2O 生活服务实体中的聚集性最高, 且风景区、大学城附近一级分类为美食的点分布密度略高于其他聚集点; 固定常住人口密度远大于其他区域且空间发展程度较为成熟的上下城区中, 一级分类为美食和家政的分布点密度相较于其他区域较高, 说明城市空间在发展的过程中会伴随着空间服务实体类别的更迭。此外, 服务行业的空间格局和城市空间的发展有着密切的关系, 在城市空间格局相对更加成熟的地方, 餐饮和家政类别的实体点占比相对较高; 而在城市空间发

展的初期,家装占比相对较高,说明服务点的分布格局和城市的空间结构之间相互影响。

通过对基于大数据技术的 O2O 生活服务点数据的定量分析,本文在一定程度上能够反映杭州不同种类的生活服务实体点的分布情况,有助于深化人们对城市空间结构的理解,为城市物流中心点的选址规划、路线优化等提供参考。然而从互联网获取的数据在一定程度上存在偏差性,虽然在分析之前已经对数据进行了清洗和预处理,但是仍不能保证数据的完全准确,在今后的研究中可以通过数据质量模型的构建以及多源数据融合等方式来优化。此外,本文主要通过分析店铺的分类和空间密度两个相对宏观维度数据,以展现空间实体点的特征和聚集性,今后可以更加深入地从单个店铺的评分、均价以及销量等更加微观的维度来分析,以期能够得出更加深层次的结论,从而对选址、末端配送等更加具有指导意义。

参考文献:

- [1] 王德,张晋庆.上海市消费者出行特征与商业空间结构分析[J].城市规划,2001(10):6-14.
- [2] Hillier B. The common language of space: A way of looking at the social, and environmental functioning of cities on a common basis[J]. Journal of Environmental Sciences, 1999, 11(3):344-349.
- [3] 刘承良,余瑞林,熊剑平,等.武汉都市圈路网空间通达性分析[J].地理学报,2009,64(12):1488-1498.
- [4] 陶伟,陈红叶,林杰勇.句法视角下广州传统村落空间形态及认知研究[J].地理学报,2013,68(2):209-218.
- [5] 黄晓冰,陈忠暖.基于信息熵的地铁站点商圈零售业种结构的研究:以广州15个地铁站点商圈为例[J].经济地理,2014,34(3):38-44.
- [6] Cró S, Martins A. Hotel and hostel location in Lisbon: Looking for their determinants[J]. Tourism Geographies, 2017:1-20.
- [7] 杨振龙,龙瀛,Nicolas D. 大数据对人文-经济地理学研究的促进与局限[J].地理科学进展,2015,34(4):410-417.
- [8] 周素红,林耿,闫小培.广州市消费者行为与商业业态空间及居住空间分析[J].地理学报,2008(4):395-404.
- [9] 吕波.基于商圈发展动力机制的商业地产选址模式研究[D].北京:北京交通大学,2009:7.
- [10] Teller C, Reutterer T. The evolving concept of retail attractiveness: What makes retail agglomerations attractive when customers shop at them? [J]. Journal of Retailing and Consumer Services, 2008, 15(3):127-143.
- [11] Borgers A, Vosters C. Assessing preferences for mega shopping centers: A conjoint measurement approach [J]. Journal of Retailing and Consumer Services, 2011, 18(4): 322-332.
- [12] 王士君,浩飞龙,姜丽丽.长春市大型商业网点的区位特征及其影响因素[J].地理学报,2015,70(6):893-905.
- [13] 刘明广,李高扬.区域产业集群的演化模型及仿真分析[J].统计与决策,2012(24):53-56.
- [14] 于伟,郭敏,宋金平.北京市新型零售业空间特征与趋势研究[J].经济地理,2012,32(5):49-53.
- [15] 何永达,赵志田.我国零售业空间分布特征及动力机制的实证分析[J].经济地理,2012,32(10):77-82.
- [16] 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.
- [17] 沈体雁,周麟,王利伟,等.服务业区位选择的交通网络指向研究:以北京城市中心区为例[J].地理科学进展,2015,34(8):947-956.
- [18] 曾璇,崔海山,刘毅华.基于网络空间点模式的生活服务店空间格局分析[J].地球信息科学学报,2018,20(6):837-843.
- [19] 张晓瑞,华茜,程志刚.基于空间句法和LBS大数据的合肥市人口分布空间格局研究[J].地理科学,2018,38(11):1809-1816.
- [20] 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.
- [21] Rodriguez A, Laio A. Clustering by fast search and find of density peaks [J]. Science, 2014, 344(6191): 1492-1496.
- [22] 于彦伟,贾召飞,曹磊,等.面向位置大数据的快速密度聚类算法[J].软件学报,2018,29(8):2470-2484.
- [23] 金敏.基于等级加权Voronoi图模型的商圈划分研究[D].杭州:浙江大学,2016:5.
- [24] 刘阳.餐饮服务食品安全公众满意度及其影响因素研究[D].杭州:杭州师范大学,2016:4.
- [25] 周志华.机器学习及其应用[M].北京:清华大学出版社,2009:211-214.
- [26] 金安,程承旗,宋树华,等.基于Geohash的面数据区域查询[J].地理与地理信息科学,2013,29(5):31-35.
- [27] 何清,李宁,罗文娟,等.大数据下的机器学习算法综述[J].模式识别与人工智能,2014,27(4):327-336.
- [28] 崔万珍,武前波.消费社会背景下杭州商业综合体的时空格局及其趋势特征[J].热带地理,2017,37(6):844-851.

(责任编辑:陈丽琼)