

基于居民出行起讫点分布特性的 集中交付点选址方法

边 军¹,丁 乔¹,渠志云²

(1.包钢矿业有限责任公司,内蒙古 包头 014000;
2.内蒙古第一机械集团有限公司,内蒙古 包头 014000)

[摘 要]集中交付点(Collection and Delivery Point, CDP)配送方式可以有效弥补上门配送的短板,但目前并没有针对CDP选址与居民日常出行规律结合的选址方法,据此提出一种基于居民出行起讫点(Origin-Destination, OD)分布特性的CDP选址方法。以居民日常出行OD点为中心,统计在一定范围内特定城市兴趣点(Point Of Interest POI)的数量和距离,据此为每个OD点赋予不同的权重值。将带有不同权重值的OD点数据进行Mean Shift聚类分析,根据聚类分析结果选取合适的聚类簇中心作为选址点。通过提取济南市1 005辆出租车一个星期的运营数据中的OD点并进行分析,完成了对济南市历下区的CDP选址。

[关键词]集中交付点;最后一公里;物流设施选址;起讫点;居民出行

[中图分类号]F252;F224

[文献标识码]A

[文章编号]1005-152X(2021)10-0043-07

Selection of Collection and Delivery Locations Based on Distribution Characteristics of Points of Origin and Destination for Resident Travel

BIAN Jun¹, DING Qiao¹, QU Zhiyun²

(1. Baotou Steel Mining Co., Ltd., Baotou 014000; 2. Inner Mongolia First Machinery Group Co., Ltd., Baotou 014000, China)

Abstract: In this paper, we proposed a method of collection and delivery point location selection based on the origin-destination (OD) distribution characteristics of the residents. Taking the daily travel OD of the residents as the center, we counted the number and distance of the points of interest (POI) in a specific city within a certain range, and assigned accordingly different weights to each OD pair. Then we had a mean shift cluster analysis on the OD pair of different weight values, and selected appropriate cluster centers as the candidate locations. At the end, by extracting and analyzing the OD pairs in the operating data of 1,005 taxis in Jinan for the duration of a week, we finished the CDP site selection in the Lixia District of Jinan.

Keywords: collection and delivery point; last mile; location selection of logistics facilities; point of origin and destination; resident travel

0 引言

随着电子商务的飞速发展,物流行业迎来难得机遇的同时也面临着巨大的挑战。快递业务量的急

剧增长使得最后一公里配送成为难题。由于快递员送货与客户收货时间窗经常无法匹配,导致出现配送延迟遭客户投诉以及多次配送增加配送工作量等问题。如何提高配送服务满意度和配送效率成为亟

[收稿日期]2021-05-31

[作者简介]边军(1984-),男,内蒙古通辽人,硕士,经济师,政工师;研究方向:企业管理;丁乔(1993-),通信作者,男,山西大同人,硕士研究生,助理工程师,研究方向:物流设施布置与规划;渠志云(1991-),女,山西大同人,硕士研究生,助理工程师,研究方向:模式识别和算法研究。

待解决的难题。针对这一难题,有研究学者提出 CDP 配送模式,如 Song L^[1]指出为了应对中国网络购物市场的迅猛发展,最后一公里问题备受关注,但上门配送模式失败后会导致运营成本增加。同时,也有研究学者对 CDP 配送模式的接受程度进行了调查,如 Yuen K F^[2]以创新扩散理论(IDT)为基础,分析了顾客是否愿意将 CDP 配送模式作为最后一公里交付方式,调查数据收集自 164 名新加坡消费者,采用层次回归分析方法进行分析,结果表明 CDP 模式明显优于其他最后一公里交付方式。CDP 配送方式都能有效地解决上门配送中快递员和客户时间不一致问题,有效地缓解了最后一公里配送这一难题。随着 CDP 模式不断推广,CDP 选址是否合理,将直接影响到客户使用的方便性、CDP 的使用率、客户满意度及运营商收益等,因此,CDP 选址成为亟待解决的问题,找到一种合适且高效的 CDP 选址方法成为解决这一问题的关键。

1 研究现状

针对物流设施选址问题,国内外学者进行了大量研究,黄凤春^[3]在传统的最大覆盖模型的基础上,以快递公司投放快递柜后收益最大化为目标函数,建立数学模型,使用设计的 BPSO 算法对模型求解,完成对智能快递柜的选址。G R Sutanto, 等^[4]通过 k-means 聚类算法确定设施位置并且完成对客户的分配,然后使用三种聚类评估方法分别从选址结果的连通性、紧凑性、分离因子三个方面进行评价。黄涛^[5]提出了一种结合地理空间信息的快递自提点选址方法,通过对地理空间信息分析确定快递自提点备选地址,利用距离惩罚成本函数对 CFLP 模型进行改进建立快递自提网点选址模型,实现自提网点选址。Irawan, 等^[6]针对两阶段供应链中的多产品设施选址问题,以配送点的固定成本和运输成本最小为目标函数,建立数学模型,并设计了一种结合聚集法和精确算法的求解方法对模型进行求解,并验证了其建立模型和算法的有效性和优越性。Kedia, 等^[7]在消费者对 CDP 可接受性研究中指出:CDP 与住宅

和工作区的距离是影响 CDP 接受度的一个重要因素。Esser 和 Kurte^[8]在对 CDP 使用过程的调查中发现有 80% 的人更愿意把到 CDP 行程与其他出行行程相结合。

综上所述,居民更倾向于把到 CDP 取寄快递和其他的行程相结合,比如,在上下班的路上顺便到 CDP 取寄快递。所以,CDP 的选址应该结合居民的出行规律,尽可能使到 CDP 取寄快递的行程与居民的日常出行行程结合更为便利,但目前还没有针对 CDP 这一特性的选址方法。

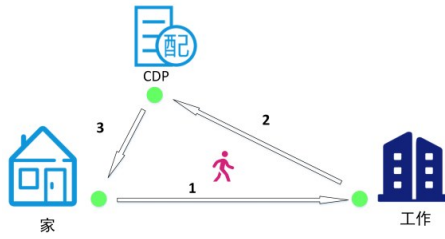
2 方法概述

在 CDP 选址研究中充分考虑居民出行 OD 点分布特性,便于居民将到 CDP 的行程与日常行程进行关联。如图 1(a)所示,将到 CDP 行程作为日常行程一部分,避免了把 CDP 单独作为出行行程造成时间和距离的浪费,如图 1(b)所示。在对 OD 样本点分析时,还应该考虑 OD 样本点的相对位置关系,如:靠近住宅区要比靠近购物中心的 OD 点对选址结果的贡献更大,因此,需要根据 OD 点对选址结果的贡献度为其设置不同的权重值。计算权重值应考虑 OD 点在一定范围内包含 POI 数量和各 POI 与其的距离。然后,使用 Mean Shift 聚类算法对不同权重值的 OD 点进行聚类分析,得到 OD 点聚类簇用于反映居民出行起讫点规律,聚类簇成员越密集,表示该区域居民日常访问越频繁,聚类簇中心表示 CDP 设施地址的备选点。

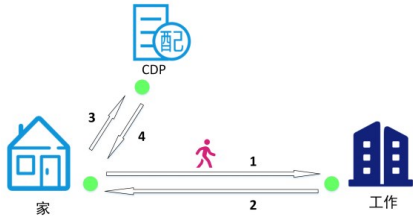
2.1 均值漂移聚类分析

Mean Shift 是基于密度的非参数聚类算法,根据当前点与给定空间内样本点的偏移向量计算出当前点的偏移均值,然后,按照该偏移均值向量更新当前点位置,并以此位置作为新的起点,重新计算偏移均值,以此方法迭代计算,直到满足算法设定的终止条件。由于其计算速度快且具有较强的鲁棒性,所以,本文选用此聚类算法完成对 OD 点的聚类分析。

假定 d 维空间 R^d 中的 n 个样本点 $x_i (i=1,2,\dots,n)$,对于任意一点 x ,其 Mean Shift 偏移向量的形式



(a) CDP行程结合日常行程



(b) CDP行程不结合日常行程

图1 居民行程对比示意图

为:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (1)$$

其中 S_h 表示在半径为 h 的区域内样本点的集合:

合:

$$S_h(x) = \{y | (y-x)(y-x)^T \leq h^2\} \quad (2)$$

k 表示落入 S_h 区域样本点的数量。

将球心 x 沿着偏移向量更新:

$$x_s = x + M_h(x) \quad (3)$$

其中 x_s 为偏移后的球心。将终止条件设为 $M_h(x) \leq \varepsilon$, ε 为容许误差。

考虑到样本点个体的差异性,不同的样本点对 $M_h(x)$ 会产生不同的贡献度,因此,在基本的 Mean Shift 算法中引入核函数和样本权重,偏移向量扩展形式为:

$$M_h(x) = \frac{\sum_{i=1}^n G_H(x_i - x)w(x_i)(x_i - x)}{\sum_{i=1}^n G_H(x_i - x)w(x_i)} \quad (4)$$

其中:

$$G_H(x_i - x) = |H|^{\frac{1}{2}} G\left(H^{\frac{1}{2}}(x_i - x)\right) \quad (5)$$

$G(x)$ 为单位核函数; H 是一个正定的对称 $d \times d$ 矩阵,一般称之为带宽矩阵; $w(x_i) \geq 0$ 为样本点 x_i 的

权重值。

Mean Shift 中,常用的两类核函数为:

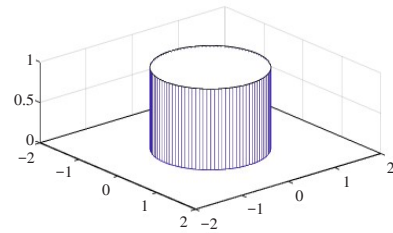
(1)单位均匀核函数,见式(6),如图2(a)所示。

$$F(x) = \begin{cases} 1, & \text{if } \|x\| < 1 \\ 0, & \text{if } \|x\| \geq 1 \end{cases} \quad (6)$$

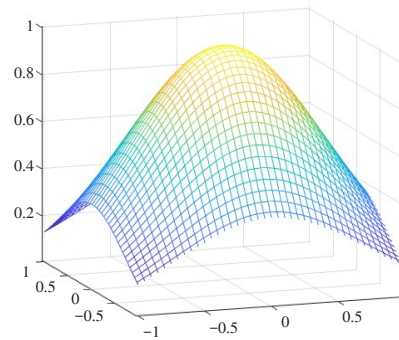
(2)单位高斯核函数,见式(7),如图2(b)所示。

$$K(x) = e^{-\|x\|^2} \quad (7)$$

在本文聚类方法研究中,距离中心点越近则对偏移向量的贡献度越大,表明不同的样本点对偏移向量的贡献度是不同的,而单位均匀核函数并不能反映这个特性,所以选择单位高斯核函数,如图2所示。



(a) 单位均匀核函数



(b) 单位高斯核函数

图2 单位核函数

Mean shift 聚类基本流程如图3所示:

(1)在样本空间随机挑选一个未分类的样本点 x 作为初始中心点。

(2)获取满足条件的样本点集合,记为 S_h 。

(3)以初始点 x 为中心,以 x 为起点和 S_h 集合的点为终点分别作向量,根据式(4)计算 $M_h(x)$ 。

(4)判断 $M_h(x)$ 是否满足终止条件,若满足,则结束循环,重新执行步骤(1)-(4),否则重复步骤(2)-(4)。

(5)直至所有样本数据都完成分类,结束算法,输出聚类分析结果。

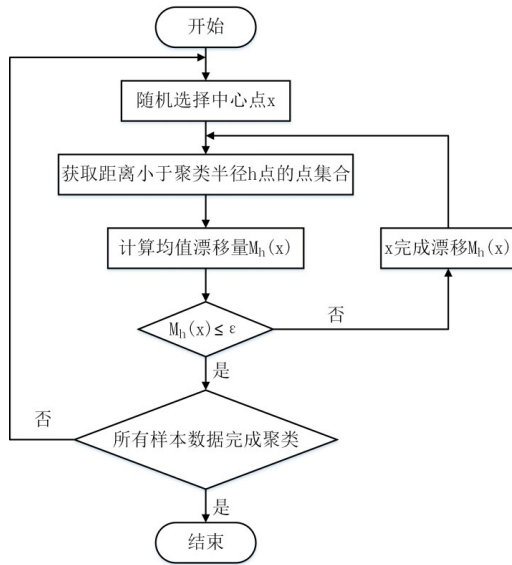


图3 Mean Shift算法流程图

2.2 居民日常出行样本数据权重值设置

住宅区作为一种参与居民日常出行最频繁的POI类型,距离住宅区的远近将直接影响居民是否方便将CDP行程与日常出行相结合,从而影响CDP的使用率,所以在为每个OD样本点赋予权重值时,主要考虑样本点在规定区域内所包含的住宅区POI数量和距离,即样本点在一定范围内所包含的住宅区POI点越多且与这些POI点的距离越近,则该样本点的权重值越大。具体计算公式见式(8)。

$$w(i) = \sum_{j \in C(i)} \left(1 - \frac{D_{ij}}{R} \right) \quad (8)$$

$w(i)$ 表示OD样本点 i 的权重值; $C(i)$ 表示OD样本点 i 在半径 R 内所覆盖住宅区POI集合; D_{ij} 表示OD样本点 i 与 $C(i)$ 中住宅区 j 的距离。

式(8)表示样本数据权重值由OD样本点覆盖的住宅区数量和与其距离决定。通过对OD样本数据进行聚类分析,聚类簇中心可以反映出居民出行OD点分布规律,同时也能保证对住宅区的覆盖率,提高CDP使用的便利性和使用率。

2.3 最优聚类半径选取

本文采用统计分析方法选取最优聚类半径。首

先,结合实际情况和参考文献选取最大聚类半径 h_{max} 和最小聚类半径 h_{min} , 然后以 Δh 为间隔取多组聚类半径 h_i , 取每组聚类半径 h_i 将样本点进行聚类分析。对每次聚类分析结果进行统计分析,将每次聚类后所得的聚类簇按照成员数量 C_i 进行降序排列,计算每个聚类簇成员数量在总样本数据中的占比 P_i , 见式(9)。

$$P_i = \frac{C_i}{\sum_{i=1}^n C_i} \times 100\% \quad (9)$$

其中 P_i 为第 i 个聚类簇所含成员数量在总样本数据中的占比, C_i 表示第 i 个聚类簇成员点个数, n 表示形成聚类簇的数量。将 P_i 进行依次累加,见式(10),其中 P_a 为累加比例, k 为累加的聚类簇数量。

$$P_a = \sum_{i=1}^k P_i \quad (10)$$

若聚类簇成员较少,可判定该区域居民日常出行访问的频率较低,为了保证CDP使用的便捷性和使用率,则此类聚类簇中心不适合作为CDP选址点。故本文设置样本数据占比 P_T 对此类无效聚类簇进行清洗。 P_a 与 C_k 的关系示意如图4所示,即 $P_a = P_T$ 时,将对应第 k 个聚类簇成员数量 C_k 作为聚类簇的清洗条件,即当 $C_i \leq C_k$, 则该聚类簇被清洗,反之保留。

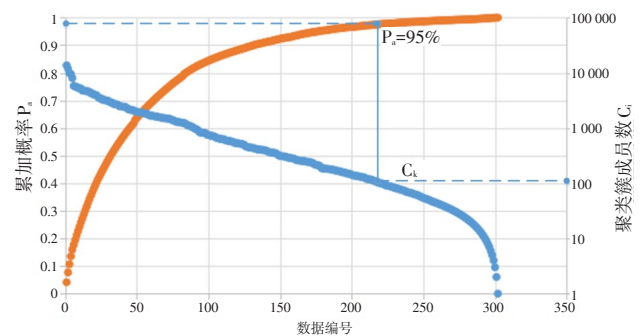


图4 P_a-C_k 关系示意图

将每组聚类半径依次进行聚类分析,统计每次聚类分析中的聚类半径 h 和对应的 C_k 值,最后将 h 和 C_k 数值进行函数拟合,得到拟合函数曲线,通过数据特性分析确定 Mean Shift 算法的最优聚类半径。

3 实例分析

3.1 数据准备

出租车作为一种重要的公共交通工具,其出行的起讫点完全由乘客决定,能够非常贴切地反映居民出行的规律。出租车与其他出行方式相比,其轨迹更易收集且数据量大,因此,本文将出租车运行GPS轨迹数据作为实例分析的样本数据来源之一。选取山东省济南市历下区1 005辆出租车一个星期的运营数据。数据的主要字段信息见表1,其中tflag表示是否有载客,即0表示空载,非0表示载客,将其数值由零变为非零或由非零变为零作为判定乘客上下车点的依据。将tflag数值突变时对应的经纬度进行提取,作为居民出行的OD数据样本点,用于聚类分析。

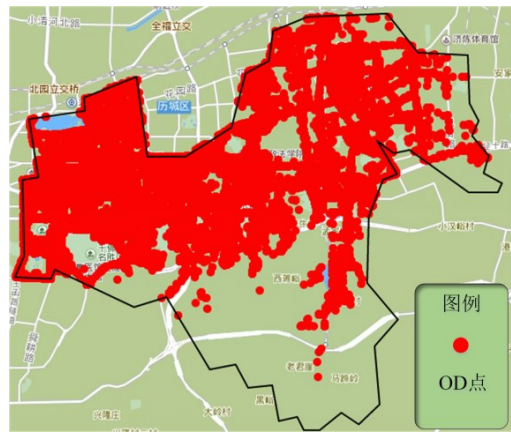
表1 字段信息表

commaddr	utc	lat	lon	tflag
车辆编号	时间戳	纬度	经度	是否载客

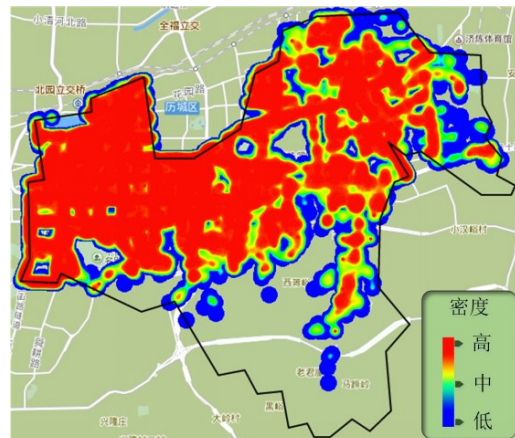
公交车作为上班族通勤最常用的交通方式,所以其运营数据能够充分体现上班族的出行规律,故选取了济南市历下区包含的共28个公交车站点的运营数据,根据乘客上下车刷卡信息,提取出乘客的上车点经纬度,与上面的出租车数据共同作为居民出行OD数据样本点。

将OD样本数据和历下区638个住宅区POI数据分别通过分布图和热力图直观地展示在百度地图上,如图5和图6所示,通过对比图5(a)和6(a)可知:OD样本点和住宅区POI点在千佛山等风景区区域中都没有分布;从大明湖公园到千佛山景区之间主要分布有商业区以及娱乐场所,所以,该区域表现为OD样本点热区,但住宅区分布却稀少。

由图5(b)和图6(b)可知,除了以上特殊区域,其余区域OD点热力分布与住宅区POI热力分布基本吻合。由此反映出OD点与住宅区分布具有相关性,以上特殊区域的分布情况也体现了二者也具有差异性。不同OD点对选址结果的影响不同,也证明了为OD点赋予不同权重值的必要性。



(a)OD点分布图



(b)OD点热力图

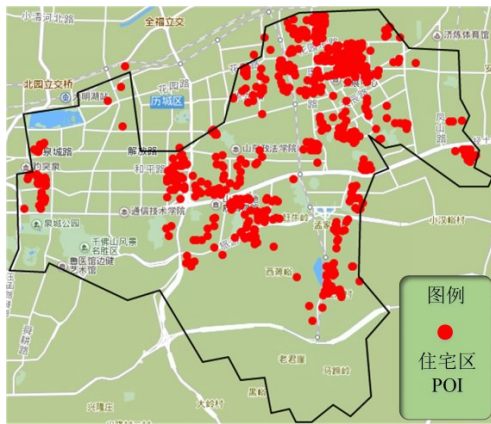
图5 历下区OD点分布情况

3.2 聚类分析

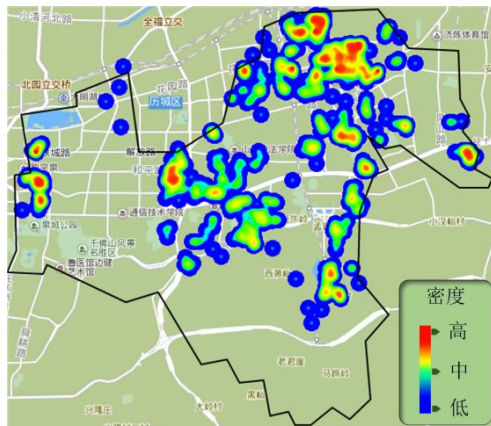
根据肖卡飞^[9]关于服务性公共设施服务半径与客户满意度关系的研究,500m为客户普遍认为的最大可接受步行距离,超出这个距离客户会产生抵触心理,据此选取范围半径 $R=500m$,代入式(8)计算每个OD样本点的权重值。设置聚类半径系列 h_i , $h_{min}=50m$ 、 $h_{max}=800m$ 和 $\Delta h=50m$ 。考虑现有普通民用GPS精度约为10m,设置容许误差 $\varepsilon=10m$ 作为算法终止条件。分别完成对聚类半径系列 h_i 的聚类分析,获取每次聚类分析后得到的 C_k 值,统计值见表2。

利用MATLAB分析软件将得到的 $h_i - C_k$ 数据进行曲线拟合,拟合函数使用三阶多项式函数。所得拟合函数见式(11),拟合曲线如图7所示。

$$y=0.000\ 002x^3-0.002\ 1x^2+0.674\ 5x-42.063\ (11)$$



(a)住宅区分布图



(b)住宅区热力图

图6 历下区住宅区分布情况

表2 $h_i - C_k$ 值实验数据结果

h_i	C_k	h_i	C_k
50	3	450	50
100	6	500	91
150	10	550	93
200	15	600	98
250	21	650	133
300	27	700	145
350	39	750	192
400	53	800	371

由图7可知, C_k 随着 h_i 整体呈递增趋势,但在不同 h_i 范围,函数的斜率不同,反映了 C_k 随 h_i 变化的快慢,当拟合函数的一阶导数为零时,表示函数变化趋于稳定,即此时 h_i 小幅度变化, C_k 变化量很小,故选择拟合函数一阶导数为零时的 h_i 作为最优聚类半径。通过计算求得当拟合函数一阶导数为零时对应

的 h_i 分别为 $h_1 = 249.5m$ 和 $h_2 = 450.4m$ 。以这两个聚类半径分别进行聚类分析,得到的聚类结果见表3。通过将聚类结果进行对比分析,可知这两个聚类半径下聚类簇数量相差多达153个,但对500m范围内住宅区覆盖率相差很小,仅2.50%。考虑对物流体系建设成本的控制,故选择450.4m为最优聚类半径,可以较低的成本,实现对住宅区较高的覆盖率。

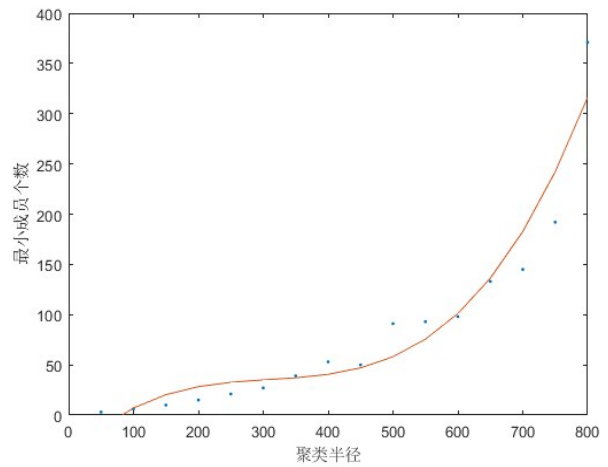


图7 $h_i - C_k$ 拟合曲线

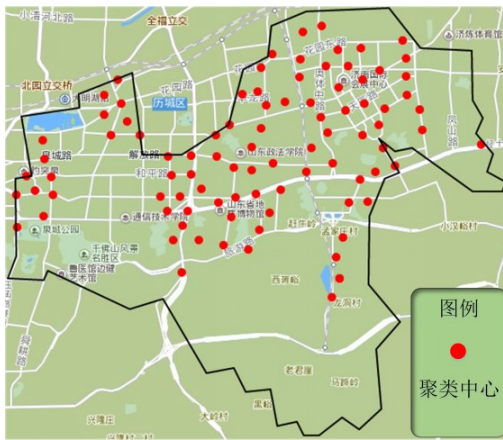
表3 聚类结果数据对比

聚类半径(h_i)	有效聚类中心数目(C_k)	500m住宅区覆盖率
$h_1 = 249.5m$	224	96.71%
$h_2 = 450.4m$	91	94.21%

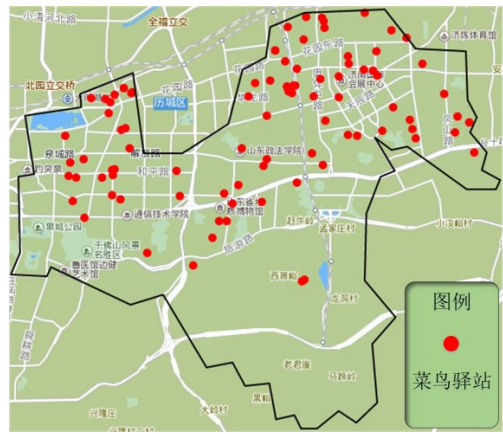
3.3 选址结果对比分析

将最优聚类半径下的聚类簇中心通过百度地图进行可视化展示,如图8(a)所示,将该结果与不为OD样本点赋予权重值的聚类簇分布结果进行比较,如图8(b)所示,可知加权值的聚类簇中心更加趋向于住宅区密集区域,极大地提高了对住宅区的覆盖率,满足本文CDP选址的需求。不加权重值的聚类簇分布趋于平均分布,即使在住宅区分布的商业娱乐区依然有聚类簇中心点分布,违背了本文选址需求的初衷,也从侧面反映出赋予权重值的必要性。

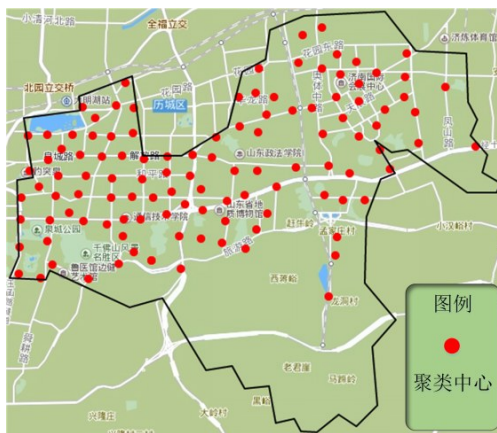
根据 Liu C^[10]提出的趋近于便利店和交通便利区域的调整原则,将不合理的CDP设施地址备选点进行局部调整后,得到CDP选址点91个。菜鸟驿站作



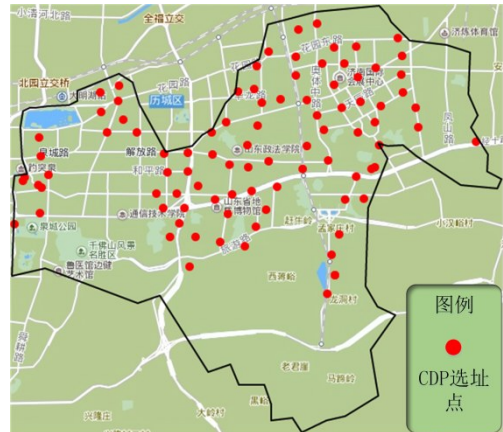
(a)加权重值分布图



(a)菜鸟驿站站点分布



(b)无权重值分布图



(b)CDP选址点分布

图8 聚类结果分布图

图9 选址点分布对比图

为目前我国最后一公里服务的典型代表,主要为消费者提供多元化的最后一公里配送业务,故选取了本文选址区域内所有菜鸟驿站站点,共93个,与本文选址结果进行数据对比分析,二者站点分布图如图9所示,对住宅区覆盖率对比数据见表4。经对比分析可知,本文选址结果中选址点的数量与菜鸟驿站站点数量相差不大,但相较菜鸟驿站本文选址结果对住宅区具有更高的覆盖率。

据作为分析居民日常出行OD分布规律的试验数据,结合POI分布进行了Mean Shift聚类分析,最终实现济南市历下区的CDP地址的选取。本文将居民日常出行OD分布规律与CDP选址相结合,提高CDP的使用率,节约居民日常出行时间和行程。通过将本文的选址结果与菜鸟驿站网点分布情况进行对比,在相同的服务半径条件下,得到本文选址结果对POI的覆盖率更高,选址点数量也更少。最后,通过调整不合理的聚类簇中心点,在不改变聚类中心数量的情况下,提高了POI覆盖率,确定CDP选址点。

表4 住宅区覆盖率对比表

网点类型	网点数量	500m住宅区覆盖率	300m住宅区覆盖率
CDP备选点	91	94.21%	62.07%
菜鸟驿站	93	65.20%	44.20%

本文为最后一公里集中交付点选址提供了新的思路,也存在不足之处,同时也是以后致力要实现的,由于缺少相关数据的支撑,本文没有对CDP选址点容量进行研究,因为CDP的容量和它所覆盖区域POI使用人数和频率的数据有关,本(下转第102页)

4 结语

本文通过使用济南市出租车和公交车的运营数