

DOI: 10.13955/j.yzyj.2022.02.15.04

# 基于热点资源的快递揽投网点画像方法研究

刘芳, 张琳超\*, 张述平, 郑梦飞

( 中国邮政集团有限公司邮政研究中心, 北京 100096 )

**摘 要:** 快递揽投网点是直接与客户交流的窗口, 是快递企业充分参与市场竞争的“前沿阵地”。文章基于揽投网点位置、网点日均揽投业务量、热点资源分布数据进行分析挖掘, 借助 ArcGIS 工具, 采用泰森多边形、平均值分析等方法, 对北京市朝阳区 S 快递公司 73 个揽投网点进行画像。该方法有助于快递企业充分了解各揽投网点服务区域内的资源分布情况, 有针对性地进行市场开发和竞争。

**关键词:** 揽投网点; 热点资源; 网点画像; 泰森多边形

**中图分类号:** F61      **文献标识码:** A

依托于电商行业的蓬勃发展, 中国快递业保持高速增长。行业业务量已经从 2011 年的 36.7 亿件增至 2020 年的 833.6 亿件, 年复合增长率约 41%, 快递揽投网点数量近 10 年扩大了 3.5 倍, 截至 2020 年 12 月, 全国快递网点数量已达 22.4 万个。但是自 2017 年开始, 全年快递业务量增速跌破 30%, 快递行业整体增速放缓, 叠加快递公司入局者增多, 行业竞争日趋激烈, 获客成本不断增加。严峻的行业竞争形势要求快递企业的经营方更加了解市场及客户情况, 从而有针对性地进行揽投网点的布局及市场资源的开发。

揽投网点画像是根据快递网点服务范围内的热点资源分布情况对揽投网点贴标签的一种方式, 一个揽投网点可以有多重标签, 这些标签可以有包含关系、并列关系等。本文主要基于揽投网点覆盖

的热点资源数据对网点进行画像, 从而帮助快递企业更好地了解市场情况及潜在客户分布。

## 1 相关理论及研究基础

当今时代徜徉在互联网中的数据越来越多, 随着数据分析技术的成熟, 基于数据挖掘的用户画像方面的研究逐渐增多。Michael Beasley (2013) 从度量和维度两个角度先对数据进行过滤, 再专注于用户群表现出的特征行为, 或比较不同用户群间不同的设置, 从而细分客户。Deepa, 等 (2016) 利用人口统计学相关理论, 将客户细分为不同类别, 并描述出不同分类客户的典型模式, 在客户消费的过程中, 客户在消费旅程地图每一步的决策行为都会影响最后的决策点。赵琨、杨潇潇 (2018) 以某快递公司一年的消费者行为数据, 运

**基金项目:** 中国邮政集团有限公司邮政研究中心项目 (项目名称: 省会一类城市揽投网点布局及其影响因素探究; 项目编号: ZB2021-062-JD)。

**作者简介:** 刘芳 (1973 ~), 女, 黑龙江哈尔滨人, 硕士, 高级工程师, 主要从事自然语言处理、运筹学、机器翻译、信息检索、名址匹配、系统架构等研究; 张琳超 (1994 ~), 女, 河南新乡人, 通讯作者, 硕士, 主要从事快递物流研究; 张述平 (1989 ~), 女, 山东济宁人, 硕士, 工程师, 主要从事数据分析研究; 郑梦飞 (1993 ~), 女, 山西汾阳人, 硕士, 主要从事邮政大数据应用研究。

**收稿日期:** 2021-10-21

**本刊网址:** yzyj.sjzpc.edu.cn

用 kohonen 网络进行客户画像，通过聚类算法提取客户特征，将客户标签化，为商家针对营销与挖掘潜在相关客户提供参考。施文幸，曹诗韵（2021）提出了一种基于改进的萤火虫优化加权 K-means 算法的分层聚类的画像推荐模型，该模型在用户画像构建时，为提高计算速度和精度，仅就单项业务设计标签模型，通过分层聚类着重构建特征群体画像，主要应用于电力用户画像。袁绮蕊（2021）使用 RFM 模型筛选典型用户，并从事实维度、模型维度和预测维度构建在线健康社区用户画像标签，以问卷调查数据为依据，通过 K-means 聚类分析实现部分用户画像实证研究，主要应用于在线健康社区用户画像。

从研究现状看，运用数据挖掘对快递揽投网点进行画像的研究较少，还未发现以客户价值为导向的揽投网点画像研究方法。本文以北京市朝阳区 S 快递公司的 73 个揽投网点为研究对象，在获取网点位置、日均揽投量、热点资源等数据的基础上，借助 ArcGIS 工具，采用泰森多边形的方法模拟揽投网点的服务范围，统计商务写字楼及产业园区、政府机构及社会团体、购物场所、商务住宅、高校五种热点资源落入揽投网点服务范围的数目，采用平均值比较法对揽投网点进行资源类型画像。

## 2 揽投网点画像方法

### 2.1 主要方法及数学模型

#### 2.1.1 泰森多边形

泰森多边形又叫冯洛诺伊图（Voronoi diagram），是一组由连接两邻点线段的垂直平分线组成的连续多边形。一个泰森多边形内的任一点到构成该多边形的控制点的距离小于到其他多边形控制点的距离。数学描述如下：

假设平面区域 B 上有一组离散点  $(X_i, Y_j)$ ， $(i=1, 2, 3, \dots, k; j=1, 2, 3, \dots, k, k$  为离散点点数)，若将区域 B 用一组直线段分成  $k$  个互相邻接的多边形，使得：

每个多边形内含有且仅含有一个离散点；

若区域 B 上任意一点位于含离散点的多边形内，不等式

$$\sqrt{(x^1 - x_i)^2 + (y^1 - y_i)^2} < \sqrt{(x^1 - x_j)^2 + (y^1 - y_j)^2}$$

在  $i \neq j$  时恒成立；

若点  $(X^1, Y^1)$  位于含离散点  $(X_i, Y_j)$  的两个多边形的公共边上，则等式

$$\sqrt{(x^1 - x_i)^2 + (y^1 - y_i)^2} = \sqrt{(x^1 - x_j)^2 + (y^1 - y_j)^2}$$

成立。

由此得到的多边形叫泰森多边形。

#### 2.1.2 揽投网点画像模型

假设用  $M$  代表热点资源，a、b、c、d、e 分别代表商务写字楼及产业园区、政府机构及社会团体、购物场所、商务住宅、高校五种不同的资源类型。 $i$  代表快递公司在朝阳区 73 个不同的揽投网点，取值范围为  $1 \leq i \leq 73, i \in N^*$ 。 $N$  代表不同资源类型的数量。 $M_i$  则代表  $M$  型揽投网点及揽投网点的画像种类。

其中，不同揽投网点的资源占比可以表示为： $N_{M_i}/N_M$ ， $i$  揽投网点服务范围内的 5 种资源类型占比分别为： $\frac{N_{a_i}}{N_a}, \frac{N_{b_i}}{N_b}, \frac{N_{c_i}}{N_c}, \frac{N_{d_i}}{N_d}, \frac{N_{e_i}}{N_e}$ ，揽投网点的类型主要取决于  $N_{M_i}/N_M$  与  $1/i$  的关系及  $\max(\frac{N_{a_i}}{N_a}, \frac{N_{b_i}}{N_b}, \frac{N_{c_i}}{N_c}, \frac{N_{d_i}}{N_d}, \frac{N_{e_i}}{N_e})$ 。

若  $\frac{N_{M_i}}{N_M} \geq \frac{1}{i}$ ， $M \in (a, b, c, d, e)$ ，则  $M_i$  为综合型揽投网点。

否则取  $\max(\frac{N_{a_i}}{N_a}, \frac{N_{b_i}}{N_b}, \frac{N_{c_i}}{N_c}, \frac{N_{d_i}}{N_d}, \frac{N_{e_i}}{N_e})$  所对应的  $M$  作为揽投网点的画像种类，即商务写字楼及产业园区型揽投点、政府机构及社会团体型揽投点、购物场所型揽投点、商务住宅型揽投点、高校型揽投点。

### 2.2 揽投网点画像过程及结果分析

#### 2.2.1 数据来源

本文研究所需数据主要包括：北京市朝阳区 S 快递揽投网点的位置信息（经纬度）及日均揽派业务量，北京市朝阳区内商务写字楼及产业园区、政府机构及社会团体、购物场所、商务住宅、高校五种热点资源的名称、位置（经纬度）及数量。数据预处理过程如下。

S 公司揽投网点数据来源于官方网站，在北京市边界点范围内，将其分成 15 120 个  $1.5\text{km} \times 1.5\text{km}$  的网格，基于 Python 语言，以每个网格点经纬度为基础，从 S 快递公司官方网站查询该网格点周边 3km 范围内所有的 S 公司网点数据，对抓取到的数据进行筛选、去重等预处理操作，最终得到 S 快递公司在朝阳区共有 73 个揽投网点信息。

S 公司揽派业务量主要通过草根调研及相关技术手段进行获取。本次主要使用的是 S 公司 2021 年 4 月朝阳区各网点的揽投量。

热点资源类型数据来源于高德地图开放平台，具体类型包括：代码、大类名称、中类名称、小类名称等属性信息，本文使用了产业园区、楼宇、商务住宅相关、住宅区、学校等中类作为研究对象。具体信息包括：热点资源所属城市、所属区县、经纬度、地址信息等。

### 2.2.2 揽投网点画像

步骤一：在 ArcGIS 工具中导入 73 个揽投网点的经纬度信息，将地理坐标转为投影坐标，使用泰森多边形做出揽投网点的覆盖范围，并裁剪出朝阳区的揽投网点覆盖情况。

步骤二：将产业园区、楼宇、商务住宅相关、住宅区、学校五种资源的经纬度信息导入 ArcGIS，统计落入每个揽投网点服务范围的热点资源数量。

步骤三：统计落入各揽投网点的资源类型数量，朝阳区共有商务写字楼及产业园区 2 958 个、政府机构及社会团体 7 962 个、购物场所 18 405 个、商务住宅 4 741 个、高校 496 个。部分揽投网点服务范围内的热点资源分布情况如表 1 所示。

步骤四：分别计算揽投网点内热点资源数量占对应总热点资源的比例（见表 2），若存在揽投网点对应五种资源类型占比均大于资源类型比例的均值，则认为该揽投网点为综合型揽投网点，否则取五种资源类型中占比最大的资源类型作为揽投网点的资源画像。

步骤五：对 S 公司在朝阳区的 73 个揽投网点的资源类型进行统计，将揽投网点划分为写字楼/园区型网点、高校型网点、购物型网点、政府机关型网点、住宅型网点与综合网点六大类。部分结果如表 3 所示。

### 2.2.3 揽投网点画像结果分析

根据 S 公司揽投网点业务量数据和揽投业务量排名前 20 的揽投网点名称及资源画像结果数据可以看出：在揽收方面，排名前 20 名的网点主要为商务写字楼及产业园区型网点（占 35%）、政府机构及社会团体型网点（占 15%）、购物型网点（占 30%），见表 4；在投递方面，排名

表 1 揽投网点服务范围内资源类型统计（单位：个）

揽投网点代码	商务写字楼及产业园区 (a)	政府机构及社会团体 (b)	购物场所 (c)	商务住宅 (d)	高校 (e)
010CP	176	247	540	98	0
010CN	54	210	100	119	47
010NQ	56	179	539	116	5
010SQ	27	251	180	94	6
010FR	23	226	449	137	0
010NK	189	208	612	121	15

表 2 揽投网点服务范围内资源类型占比统计

揽投网点代码	商务写字楼及产业园区 (a)	政府机构及社会团体 (b)	购物场所 (c)	商务住宅 (d)	高校 (e)	网点类型
010CP	5.95%	3.10%	2.93%	2.07%	0.00%	写字楼/园区
010CN	1.83%	2.64%	0.54%	2.51%	9.48%	高校
010NQ	1.89%	2.25%	2.93%	2.45%	1.01%	购物
010SQ	0.91%	3.15%	0.98%	1.98%	1.21%	政府机关
010FR	0.78%	2.84%	2.44%	2.89%	0.00%	商务住宅
010NK	6.39%	2.61%	3.33%	2.55%	3.02%	综合网点

注：加框数据为五种资源类型比例中的最大值

前 20 名的网点主要为住宅型网点（占 30%）、购物型网点（占 30%），见表 5。从中可以看出，S 公司的揽收业务主要来自于商务文件、行政类文件及部分商品，投递主要服务的客户为住宅区居民。

## 3 结论及应用价值

本文首次利用五类热点数据，对快递末端揽投网点进行资源类型画像，借助 ArcGIS 的泰森多边形模拟快递公司的服务范围，同时结合揽投网点的揽派业务量进一步分析了快递公司揽收和投递的主要客户类型，有利于快递公司充分了解各揽投网

表3 揽投网点画像结果

揽投网点代码	网点类型
010CP	写字楼/园区型网点
010CN	高校型网点
010NQ	购物型网点
010SQ	政府机关型网点
010FR	住宅型网点
010NK	综合型网点
.....	.....

表4 S 快递公司收件量排名前20的揽投网点情况统计

收件前 20名的 网点类型	商务写字 楼及产业 园区	政府机构 及社会团 体	购物 场所	商务 住宅	高校	综合
网点	7	3	6	2	2	0
网点类 型占比	35%	15%	30%	10%	10%	0

表5 S 快递公司投递量排名前20的揽投网点情况统计

投递前 20名网 点类型	商务写字 楼及产业 园区	政府机构 及社会团 体	购物 场所	商务 住宅	高校	综合 网点
网点	4	1	6	6	3	0
网点类 型占比	20%	5%	30%	30%	15%	0

点服务区域内的资源分布情况，有针对性地进行市场开发。

在应用价值方面，一是本研究所使用的泰森多边形可用于模拟揽投网点的服务范围，特别是在无法获取竞争对手或者邮政自身网点电子围栏的情况下，该方法具有较大的应用价值；二是揽投网点画像模型采用定量对比分析的方法，可以避免靠直觉、靠经验进行网点资源类型判断所带来的不准确问题；三是研究中使用的热点资源数据来自于开放平台，可通过合法途径持续获取，随着快递揽投网点的变化进行持续性的画像研究；四是对揽投网点进行画像，除了有利于邮政进行市场开发外，对客户资源为导向的揽投网点布局也有一定的价值，邮政可以根据竞争对手及自身网点的资源分布情况合理调整揽投网点的位置。因此，该研究对邮政末

## 英国政府发布《净零战略》 致力改变“最后一英里”投递

英国政府发布《净零战略》，阐明了英国如何兑现2050年实现净零排放的承诺。

该文件指出，英国将支持和鼓励货运模式从公路运输转向更可持续的替代方式，如铁路、货运自行车和内陆水路。

为实现净零目标，英国政府将推出一揽子支持政策，具体包括：投资铁路货运网络能力建设，推进电气化改造；制定转变模式税收支持计划和水路货运补助计划；制定铁路货运增长目标；探索支持城市地区更可持续的“最后一英里”货运措施。

《净零战略》承诺：通过采取新的投递模式，实现重型运输车辆的零排放和无碳投递，彻底改变“最后一英里”投递模式，并通过精准的数据分析和数字化创新，不断提高投递效率。

(王旭译)

端揽投网点的市场开发和揽投网点位置的优化均具有一定的实践价值。

### 参 考 文 献

- [1] Michael Beasley. Segmentation [M]. San Francisco: Morgan Kaufmann, 2013
- [2] Deepa, Itimani, Tholath, et al. Customer Journey Maps for Demographic On line Customer Profiles [J]. Internal Journal of Virtual Communities and Social Networking (IJVC-SN), 2016
- [3] 赵琨, 杨潇潇. 基于 Kohonen 网络的客户行为研究 [J]. 现代商贸工业, 2018 (24)
- [4] 施文幸, 曹诗韵. 基于萤火虫 K-means 聚类的电力用户画像构建和应用 [J]. 计算机系统应用, 2021 (8)
- [5] 袁绮蕊. 基于 K-MEANS 的在线健康社区用户画像模型构建 [J]. 科技情报研究, 2021 (4)