文章编号：2095-2163(2022)11-0064-07 中图分类号：TP181 文献标志码：A

基于 C-K-N-Cluster 的居民出行时空特征分析

戴 兵，田 博，高心雨，严李强
（西藏大学 信息科学技术学院，拉萨 850000）

摘 要：为解决传统聚类算法在大数据轨迹信息应用中的簇类数不确定、病态初始化等问题，文章提出了一种结合 Canopy 与 K-Means++的小生境遗传智能聚类算法（C-K-N-Cluster），并应用于居民出行时空特征分析；以杭州市为例，对出租车轨迹数据进行降噪标准化等预处理，按照筛选原则提取载客点数据；提取出的数据投入智能聚类算法仿真识别城市上下载客热点地域，结合数据分析方法可视化研究城市居民出行特征。仿真结果表明：改进算法相比传统 K-Means 能够实现大数据应用场景下的簇类数与初始化自动最优化，分析了杭州市居民出行规律及出租车载客时空特征，为司乘服务和城市功能区优化提供参考。

关键词：轨迹数据；C-K-N-Cluster 算法；可视化分析；居民出行特征

Analysis of residents′ travel time-space characteristics based on C-K-N-Cluster

DAI Bing, TIAN Bo, GAO Xinyu, YAN Liqiang
(School of Information Science and Technology, Tibet University, Lhasa 850000, China)

【Abstract】This paper proposes a niche genetic intelligent clustering algorithm (C-K-N-Cluster) combining Canopy and K-Means++ and applies it to the analysis of residents′ travel space-time characteristics in order to solve the problems of uncertain cluster number and ill initialization of traditional clustering algorithms in big data trajectory information application. In the research, Hangzhou is taken as an example, where taxi track data is preprocessed by noise reduction standardization and passenger point data is collected using the screening method. The collected data is fed into an intelligent clustering algorithm, which simulates and identifies the hot regions of urban inbound and outgoing passengers, and data analysis methods are used to show the travel characteristics of urban people. The simulation results demonstrate that, when compared to the classic K-Means approach, the upgraded algorithm can achieve automated optimization of cluster number and initialization in a big data application situation. The research examines Hangzhou inhabitants′ travel habits as well as the time-space features of taxi clients, serving as a guide for optimizing driver services and urban functional regions.

【Key words】trajectory data; C-K-N-Cluster algorithm; visual analysis; travel characteristics of residents

0 引 言

地理数据以及高精度定位技术的发展，使得大量移动对象的移动、位置信息能够以轨迹数据的形式被收集下来，通过分析大量居民轨迹数据集为确定城市热点区域及提取出行时空特征信息提供了新的重要研究思路。区别于公交车、地铁等轨道交通出行载体，出租车是城市中提供给居民便捷和个性化的出行服务，对出租车行驶过程中产生的轨迹数据进行合理挖掘则可以揭示居民出行特征，让城市规划更加合理。

随着多样的应用场景和数据规模不断提高，很多经典聚类分析方法的不足难以适应大数据背景下的智能应用分析。伴随着计算机发展，基于轨迹数据的信息挖掘研究不断丰富更新。程静等人[1]利用北京市出租车 GPS 数据，结合时间序列距离度量和 K-means 聚类，研究了乘客出行的时空分布特征。刘旭等人[2]将 Canopy 结合到聚类算法中用于确定簇类值，并应用到武汉市公交车站的预测分析中。Rahman 等人[3]首次提出了智能学习 G-A-K-Clustering 遗传聚类算法，来解决聚类算法中突出的初始化种子点等问题。

传统 K-means 算法在处理小规模数据集上，相较其他聚类算法因其高效的模型结构和良好的聚类效果被广泛应用，而随着复杂大规模数据集的应用场景加入，其算法主要的弊端逐渐暴露：

基金项目：2021 年中央引导地方科技发展资金项目（XZ202101YD0014C）；西藏大学研究生“高水平人才培养计划”项目（2020-GSP-S169）。
作者简介：戴 兵（1998-），男，硕士研究生，主要研究方向：机器学习；严李强（1980-），男，教授，硕士生导师，主要研究方向：智能控制。
通讯作者：严李强 Email：158201730@qq.com
收稿日期：2022-08-04

(1)初始化中的簇类数仅凭主观因素确定。直接决定最后的输出极有可能达不到理想的结果。

(2)不良初始种子点的选择会随着算法迭代对结果产生严重的影响。如何消除 K-means 算法在处理大规模数据集上的弊端并有效保留其算法优势是目前研究的重点。

遗传智能算法具有良好的寻找最优解能力，为解决传统 K-means 算法弊端提供了重要的研究思路。基于此，本文提出一种新的改进型智能遗传 C-K-N-Cluster 聚类算法，有效解决种子点数选择弊端与病态初始化等问题，并以杭州市大量出租车 GPS 轨迹点为实验数据，实现最佳簇类中心的输出和城市热点区域的挖掘研究，最后通过数据分析和可视化方法，研究杭州市城市出租车运行特征，对不同区域的居民出行规律展开分析，为城市交通管理和居民出行提供决策服务。

1　基础理论

1.1　遗传算法简介

遗传算法(Genetic Algorithm，GA)是当前应用领域最广泛、影响最深远的优化算法之一。通过特定遗传算子指标从父代染色体(解)中筛选出部分优秀的子代染色体(更优秀的解)，重复迭代，直至达到最大进化代数或结果满足收敛条件，收敛到的最终个体可能代表着问题的最优解或次要解。基本的遗传算法包含 5 个基本步骤：解的编码；种群初始化；适应度函数选择；遗传操作算子；设定遗传参数[4]。

1.2　K-Means 算法描述

聚类(Cluster)是一种常用的无监督学习，其在机器学习、数据挖掘等统计应用中备受关注与重视。经典 K-means 算法由于高效快速的显著特点已经成为聚类算法中使用广泛的算法之一[5]。常用欧氏距离作为相似性的评价指标，计算数据集中每一个数据点 x_i 与每个质心 c_j 在 m 维空间中的欧式距离 d 表示为[6]：

$$d(x_i,c_j)=\sqrt{\sum_{i=1}^{m}\| x_{ik}-c_{jk}\|_2^2} \tag{1}$$

当有一组 n 个样本的数据集 X，将其分成 k 个独立的簇 $C=(C_1,C_2,\cdots,C_k)$ 时，聚类过程可描述为每个簇内的平方误差 E 不断降低并趋向最小的过程，平方误差定义为：

$$E=\sum_{i=1}^{k}\| x-\mu_i\|_2^2 \tag{2}$$

其中，$x\in C_i$ 为样本均值，u_i 是簇类 C_i 的质心。

2　结合 Canopy-K-Means++的小生境遗传 C-K-N-Cluster算法

本文以杭州市大量出租车 GPS 轨迹数据为基础，用密度代替传统分类阈值的思想改进 Canopy，结合 K-means++算法实现初始化种子点，达到消除种子点数选择与病态初始化限制的目的，为避免智能学习过程易陷入局部最优的弊端，通过共享小生境提高遗传算法操作中的优化能力，与 K-Means 结合实现最优染色体(聚类中心)的输出和城市热点区域的挖掘研究。

2.1　算法结构设计

本文通过改进 Canopy-K-Means++的初始化种群生成方法、以准则函数的倒数作为适应度函数、遗传算子的自适应设计、小生境划分种群等改进方法完成聚类算法的改进优化，以消除局部最优、病态初始化、难确定 k 值等传统算法弊端。整体的算法设计步骤如图 1 所示。

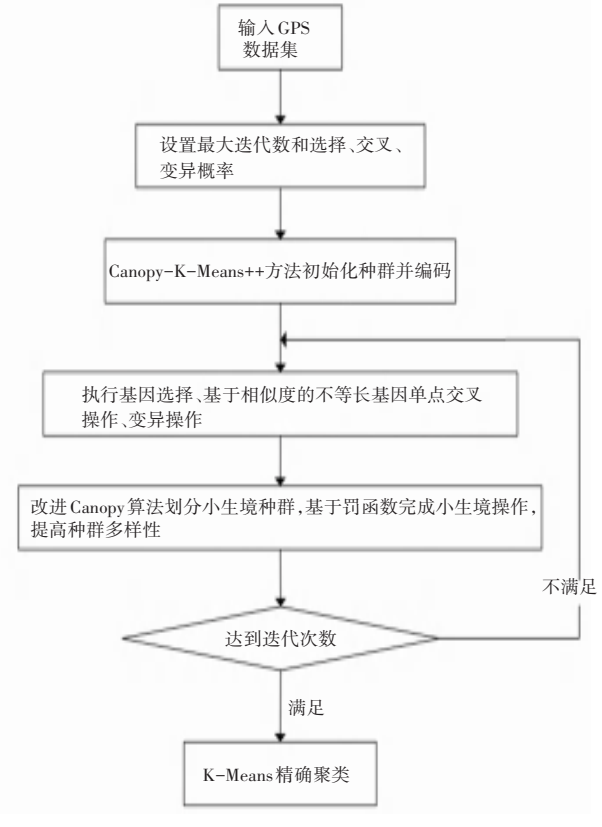


图 1　改进算法流程图

Fig. 1　Flow chart of the improved algorithm

2.2　染色体编码及种群初始化

若 GPS 数据集有 n 个数据点，每个数据点都有 m 个字段特征，假设种群中有 NI 条染色体(种群规模为 NI)，每条染色体上的基因数(聚类中心数)为

$K_1=\{k_1,k_2,\cdots,k_{K1}\}$。将 m 个字段属性综合起来进行编码,则该种群可以描述为:

$$pop:\ NI\leftarrow\begin{bmatrix}CHR_1\leftarrow K_1*m\\CHR_2\leftarrow K_2*m\\\cdots\\CHR_{NI}\leftarrow K_{NI}*m\end{bmatrix}\tag{3}$$

其中, NI 为种群规模;pop 表示种群;CHR_1、CHR_2、…、CHR_{NI} 是种群中的各条染色体;K_1 表示 CHR_1 染色体含有的初始化聚类中心数量。

同时,在实际操作过程中,需要对该数据集通过式(4)进行归一化处理,使数据的范围值都转换控制在[0,1]范围内:

$$P_{nor}=\frac{p-p_{\min}}{p_{\max}-p_{\min}}\tag{4}$$

种群初始化过程中,初始聚类中心点的选择对最后的聚类效果影响颇大,传统 K-Means 算法的随机生成与改进后的 K-Means++距离分散初始点生成方法较为常见,但前者结果的最优值有着不稳定性,后者并不适用于数量多而富含信息量少的冗余数据集[7]。Canopy 算法是一种结构简单、使用距离测度的聚类数生成方法,已广泛出现在各类聚类应用中[8]。本文利用密度概念对 Canopy 算法进行改进并与 K-Means++算法结合完成种群初始化[9]:

(1)设置一组密度,用于生成不同基因数目(种子点数量)的种群:

$Den=[d_1,d_2,\cdots,d_{NI}]=[0.000\ 1,0.000\ 5,0.001,0.002,0.003,0.004,0.005,0.006,0.007,0.008,0.009,0.1,\cdots]$。

(2)基于以上 NI 个不同半径生成 NI 条不等长的染色体:简化 Canopy,将 2 个阈值 $T1$ 与 $T2$ 简化为一个阈值集合 Den。

(3)向列表 $List$ 中放入向量化的数据集,从 $List$ 中随机选取一点 P_1 作为第一个 Canopy。

(4)从 $List$ 中随机选取的点 P_i 开始快速计算点 P_i 与所有已有 Canopy 之间的距离半径,如果点 P_i 与某个 Canopy 距离在 T 以内,则将点 P_i 加入到这个 Canopy,同时点 P_i 从 $List$ 中删除。

(5)重复步骤(3),当 $List$ 中每个点都删除完毕,输出该 Den 值下的一条染色体长度(Canopy 数量)。

(6)重复执行步骤(3)~(5)各 NI 次,使用 K-Means++方法初始化种群作为每条染色体上的基因(聚类中心),直到成功生成 NI 条长度不同的染色体为止。

2.3 适应度函数设计

考虑到聚类过程即是每个簇内的平方误差不断降低并趋向最小的过程,将平均标准函数作为该过程的准则函数:

$$E=\sum_{i=1}^{k}\sum_{p\in c_i}\|p-\mu_i\|^2\tag{5}$$

其中, E 表示数据集在不同中心点下的准则函数,也称聚类代价, E 值越小,代表聚类效果越好,各簇内相似度越高。为与遗传算法结合,本文利用准则函数的倒数作为适应度函数:

$$fitness(X_i)=\frac{1}{E_{xi}}\tag{6}$$

2.4 遗传操作

(1)基于相似度的基因交叉。由于种群中染色体长度不一致,无法与其他相关等长染色体一样简单地实现局部交叉操作[10],本文利用基因重排技术[3],根据余弦定理计算不同染色体中的相似性,将相似性高的作为交叉操作中的参考体[11],具体操作步骤如下:

① 迭代中及时存储适应度最好的染色体 CHR_{best}。

② 通过赌轮盘方法[12-13]从种群随机挑选 2 条染色体 CHR_i、CHR_j, 轮盘赌方法实现概率问题、即适应值越大,选出的概率就越大,计算方法见式(7):

$$P(CHR_i,CHR_j)=\frac{fitness(CHR_i,CHR_j)}{\sum_{i=1,j=i+1}^{NI}fitness(CHR_i,CHR_j)}\tag{7}$$

③ 用余弦相似度计算 CHR_i、CHR_j、CHR_{best} 之间的相似度,对于长度不一致的染色体,通过将短的染色体后端自动补零来实现等长计算相似度操作。将最好和最差的分别作为基因交叉中的参考和目标染色体,实现适应值大的染色体基因对适应度差的染色体基因的替代。

④ 对选出的参考染色体和目标染色体按照式(8)进行基因单点交叉操作[14],即 2 个父代染色体通过预设的交叉概率生成新的 2 个子代染色体,保障种群的多样性。需要注意的是,完成一次交叉操作后需要将父代染色体移除,以保证交叉后得到的新染色体对父代染色体的迭代更新。这里的式(8)可表示为:

$$Spc(CHR_i,CHR_j)\begin{Bmatrix}CHR_i^1=(1-rdm)*CHR_j+rdm*CHR_i\\CHR_j^1=(1-rdm)*CHR_i+rdm*CHR_j\end{Bmatrix}\tag{8}$$

其中, CHR_i^1、CHR_j^1 表示新产生的染色体,rdm 是一个范围为(0,1) 的随机数。

(2)基于密度划分小生境操作。本文引入密度参数划分小生境,并通过预选择机制完成后续操作,达到提高种群的多样性和全局优化水平的目的。具体步骤如下:

① 预设一个合适的密度参数 R, 通过式(4)对预处理数据进行归一化。

② 处理后的数据按照 R 进行可达性计算(从一个顶点到另一个顶点的容易程度),记录各分类数量。

③ 统计各分类内部点的数量,把数量最多的类当成一个小生境。

④ 重复步骤③,直到所有点都被选择过,进行小生境划分。

⑤ 对划分的小生境执行预选择操作。

3　实验数据与预处理方法

杭州市是中国沿海地区较为重要的交通枢纽和长江三角洲中心城市之一,人口规模约一千二百余万。对其城市交通的热点区域和城市出行特征展开研究分析具有必要性。本文采用杭州市 2019 年 9 月 24 日全市约一万辆正常运营出租车一天的 GPS 定位数据为实验数据集,基本覆盖全天段杭州市出租车整体运营状况,数据样本集能够很好地反映该时间段杭州市居民出行的实际情况。分布情况显示,大部分市内中心路网已被覆盖。

本文选用数据集包含车辆编号、定位时间戳、经度、纬度、瞬时速度、行驶方向、车辆状态的属性,数据描述见表 1。

表 1　轨迹数据集描述

Tab. 1　Description of track data set

属性	数据	描述
车辆编号	222977	临时编号
定位时间	15:01:00	—
经度	120.205 467	120°20′54″67 E
纬度	30.349 653	30°34′96″53 N
瞬时速度	31.4	31.4 km/h
方向	148	顺时针正北方向夹角 148°
车辆状态	1	空车 0/重车 1

3.1　预处理操作

本文对研究采用的原始出租车轨迹数据需要处理的异常数据,拟做阐释分述如下:

(1)超出研究区范围:本文以杭州市及其周边为研究对象,经纬度范围为东经 120°00′至 120°43′和北纬 30°16′至 30°50′。

(2)异常数据:主要为该时间段内速度始终为零的数据、行驶过程中速度超过市区最大限速 1.5 倍的不合格数据。

(3)冗余数据的化简:使用模型简单、运算速度快的 Douglas-Peucker 算法,以采样轨迹点与前后相邻采样轨迹点间的平均距离和速度来识别筛选。

3.2　上下客源轨迹点筛选

根据出租车轨迹数据的车辆状态字段,其值为“0”和“1”时,分别代表不同的车载状态“空车”和“重车”,当该字段的数值发生变化时,表示车辆载客状态的改变,即该地点为一个上下客轨迹点。但实际数据中会存在上下两个客源点位置与实际上下客点位置距离过大的异常数据,本文将 7.50 km/h 的时速设置为速度阈值:

(1)当 2 个相邻轨迹点的行驶速度均小于该速度阈值时,则表示该 2 个数据点位置与真正的上下客点距离误差较小,可以视为有效的上下客轨迹点。

(2)对于车载状态发生改变、但并不满足该速度阈值条件的 2 个轨迹点,则将轨迹点中速度较小的作为有效上下客点进行保留。

提取上下载客点坐标数据步骤如图 2 所示。

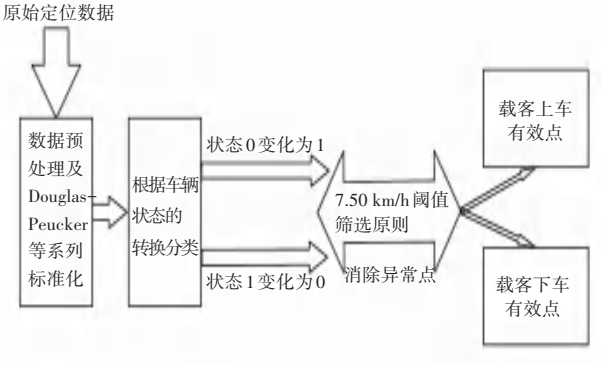


图 2　上下客源有效点提取

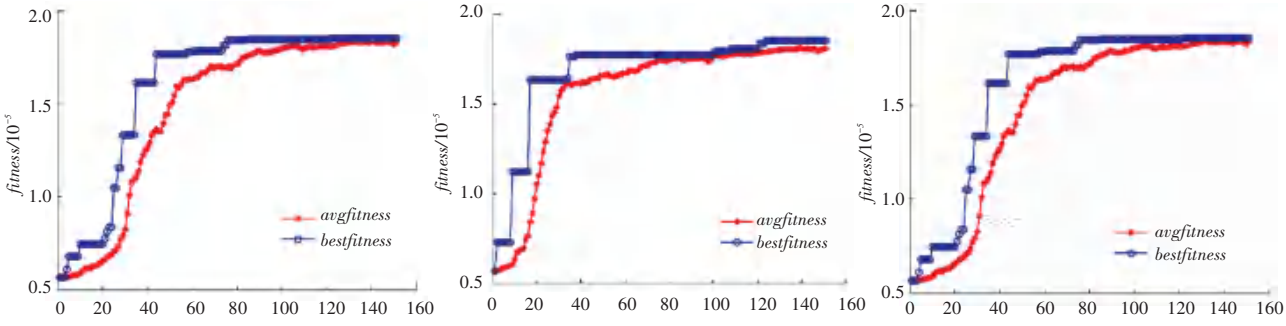
Fig. 2　Effective point extraction of up and down passenger sources

4　实验与数据可视化分析

4.1　城市时段热点区域分析

通过对杭州市一天出租车轨迹数据研究城市出行热点区域的划分识别。文中利用软件 Matlab2020a 实现上述小生境 C-K-N-Cluster 算法,设置种群规模 $sizepop$ = 100,初始种群密度半径划分 $Den = [d_1, d_2, \cdots, d_{NI}]$ = [0.000 1,0.000 2,0.000 3,0.000 4,0.000 5,0.001,0.002,0.003,0.004,0.005,0.006,0.007,…]共 100 个,代表初始化 100 个不同长度大小的个体。设置交叉概率 $P(cross)$ = 0.6,变异概率 $P(mutation)$ = 0.01,小生境划分密度 R = 0.4,对预处理上下载客点数据进行导入,对具有代表性的 3 个不同时间段数据分别绘制迭代适应度变化曲线,其中 2 条曲线分别反映了平均适应度 $avgfitness$ 和最佳适应度 $bestfitnessd$ 的值随迭代次数

的变化情况,为避免因次数过少导致的优化不足以及次数过多所带来的浪费计算时间,本文从曲线变化规律中设置迭代次数 *Margen* = 150 次。3 个时间段载客迭代适应度变化曲线如图 3 所示。

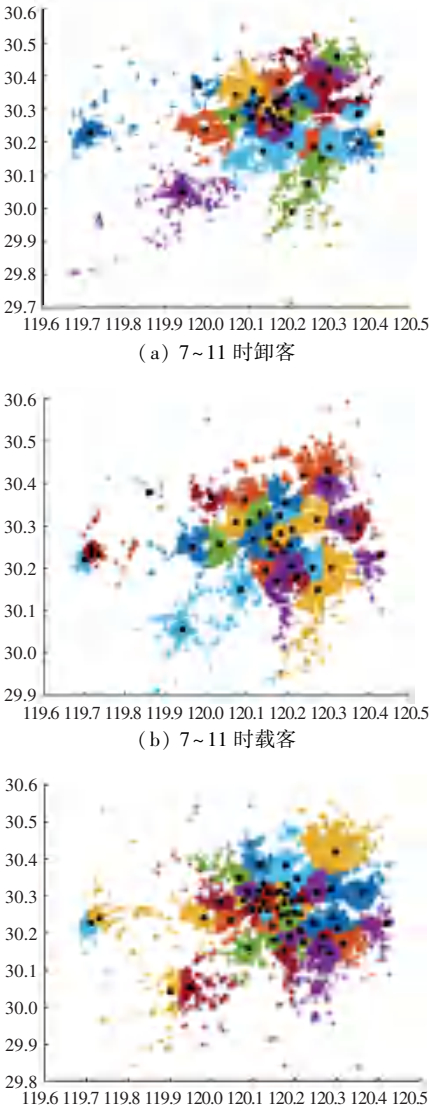


(a) 0~6 点 (b) 7~11 点 (c) 11~15 点

图 3 3 个时间段载客迭代适应度变化曲线

Fig. 3 Change curve of passenger carrying iteration fitness during three time periods

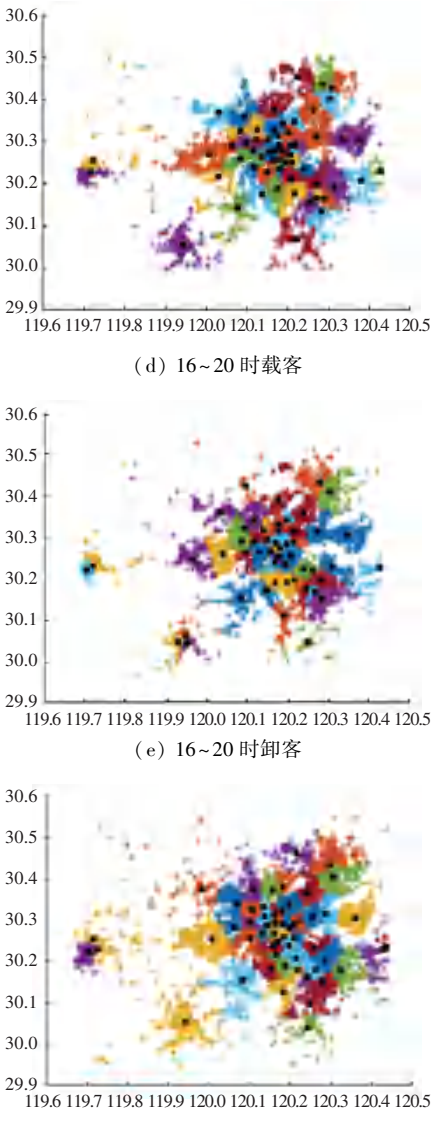
基于改进 Canopy 算法种群初始方法完成种群初始化,迭代结束自动计算出 5 个时间段上下车簇类数,结果进行可视化显示,如图 4 所示。



(a) 7~11 时卸客

(b) 7~11 时载客

(c) 11~16 时卸客



(d) 16~20 时载客

(e) 16~20 时卸客

(f) 20~24 时卸客

图 4 各时间段载/卸客簇类结果可视化

Fig. 4 Visualization of passenger load/unload cluster results in each time period

由结果可知，居民出行较为集中，而中午出行较分散，这也与居民正常出行特征相吻合。表 2 统计了研究范围内几个典型时段内上下客热点区域大致分布，这些聚类中心区域周边大量分布着商业、交通、城市服务等居民需求地点。围绕这些城市热点展开的城市功能区规划和出租车分布管理将极大地方便城市管理与居民出行需求。

表 2　典型时段上下客轨迹点热点区域统计

Tab. 2　Statistics of hot spot areas of boarding in typical time periods

时段	上客点主要热点区域	下客点主要热点区域
7:00~11:00	星光大道、金沙印象城、浙江大学紫金巷校区、杭州市民中心、德胜中路、西湖花园、紫林公寓、复兴南苑、紫阳农贸市场、润达花园、新江花园、高桥小区等	西湖文化广场、杭州市西湖风景名胜区、杭州嘉里中心、太子湾公园、杭州佛学院景区、杭州东站、浙江省立同德医院、杭州市第一人民医院等
16:00~20:00	西湖文化广场、杭州市西湖风景名胜区、英特集团、和仁科技、光宇集团、西湖文化广场、杭州大厦购物中心、萧山人民广场、杭州百货大楼等	天竺法净禅寺、西湖灵隐景区、秦望广场、宝盛世纪中心、绿都港汇中心、杭州钱柜量贩 ktv、杭州颐高数码广场等
20:00-24:00	银乐迪涌金旗舰店、好又多科技世界、东方茂购物中心、万华广场、美影国际影城、烟浔足浴店附近、太平洋影城附近等	特色文化广场、新天地购物中心、胡桃里音乐广场、黄龙体育中心、景帝商业街、东方一品、回澜北苑、明辉花园、运河文化广场附近、蓝色嘉园等

4.2　城市客流量与运营分析

通过对所有上下载客点数据进行了时间划分，以 1 h 作为时间间隔，统计 24 个时间段内的出行量分布[12]，对一天中不同时段的客流量的变化特征进行分析，居民不同时域出行量分布如图 5 所示。

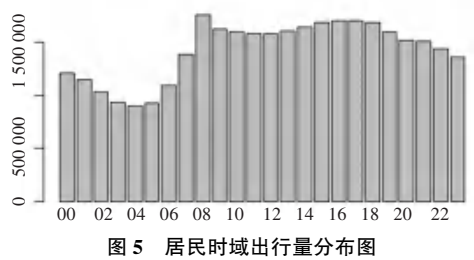


图 5　居民时域出行量分布图

Fig. 5　Distribution of residents′ travel volume in time domain

定义早 7:00 到 10:00 和晚 17:00 到 19:00 为高峰时段，经纬度数据通过逆解析完成区域划分，并分别计算各区域高峰与平时时段的出租车空载率和载客率占比，选择具有代表性的西湖区、江干区、下城区、拱墅区进行分析。分析结果如图 6 所示。由图 6 可以看到，4 个象限从绿色象限开始顺时针分别代表“高峰空载”、“平时空载”、“高峰载客”、“平时载客”，数值代表此时段出租车运营所占百分比。

该分布显示出租车空载率在高峰时间段明显小于载客率，符合实际情况且未拉开过大差距，反映出杭州市市区出租车服务数量并不短缺，能够基本满足高峰时间段居民的出行需求。但在非高峰时间段存在出租车空载率较高的问题，综合反映出杭州市出租车数量方面较为饱和，同时江干区高峰时间段出租车需求最大，考虑该时段将周边出租车调度进来服务将得到有效缓解。

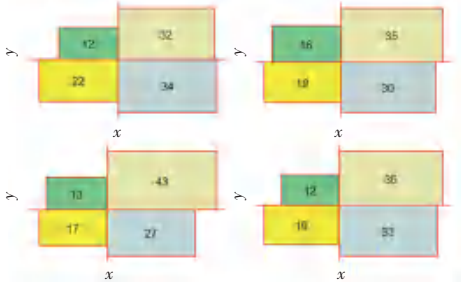


图 6　西湖、江干、下城、拱墅各区高峰与平时运营比例(左上顺时针)

Fig. 6　Proportion of peak and normal operation in West Lake, Jianggan, Xiacheng and Gongshu districts (upper left clockwise)

4.3　出租车行驶轨迹与主要跨区流动分析

忽略天气和节假日等其他因素影响，仅从单日的出租车轨迹数据分析，将行驶路程较远的出租车轨迹数据记录标记出来，具体如图 7 所示。

整理图 7 复杂的轨迹，进一步得到图 8。图 8 代表各区之间的跨区出租车流动网络，线条的粗细直观展示了区间流动量的大小。图 8 中显示西湖区、拱墅区、江干区之间的车辆流动量较大，西湖区、江干区、拱墅区和下城区是出租车最密集的区域。出租车跨区行驶的情况存在较大差异，西湖区到江干区，西湖区到下城区、拱墅区的数量较多。从数据中准确地计算发现，23%的出租车集中在西湖区，江

干区、下城区和拱墅区的出租车数量分别 16%、13%、12%。

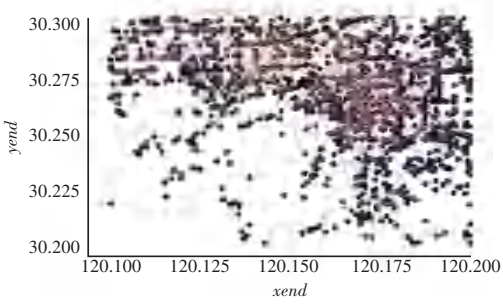


图 7 行驶轨迹

Fig. 7 Driving track of the Taxi

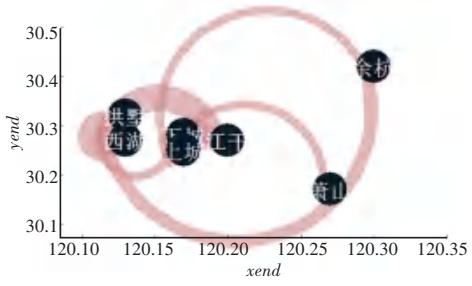


图 8 跨区流动网络

Fig. 8 Cross-region network

5 结束语

本文利用杭州市一天内的出租车载客轨迹数据挖掘城市的活动信息,用密度概念代替传统分类阈值的思想改进 Canopy,提出基于结合 Canopy-K-Means++的小生境遗传策略聚类算法,达到消除种子点数选择、病态初始化和局部最优等弊端的目的,最后成功结合数据可视化完成挖掘载客热点和城市出行时空特征信息。但仍需指出的是,由于实验数据数量具有局限性,所分析的数据时间跨度仅为一天,难以保证实验结果的普适性。此外,出租车数据是随着时间动态变化的[15],若引入在交通运输系统中适用的动态聚类算法框架可以对其进行改进。故分析长期的载客热点区域与居民通勤模型是下一步需解决的研究问题。

参考文献

[1] 程静,刘家骏,高勇. 基于时间序列聚类方法分析北京出租车出行量的时空特征[J]. 地球信息科学学报,2016,18(09):1227-1239.

[2] 刘旭,陈云波,施昆,等. 结合 Canopy-K-means 算法和出租车轨迹数据的公交车站预测方法[J]. 测绘通报,2018,4(11):63-68.

[3] RAHMAN M A, ISLAMM Z. A hybrid clustering technique combining a novel genetic algorithm with K - Means [J]. Knowledge-Based Systems, 2014, 71: 345-365.

[4] 马通. 基于遗传算法的并行化 K-means 聚类算法研究[D]. 杭州:浙江理工大学,2018.

[5] 刘鑫. 融入改进的 K-means 聚类的协同过滤算法的研究与应用[J]. 软件,2021,42(03):97-99.

[6] 杜淑颖. 基于大型数据集的聚类算法研究[J]. 软件,2016,37(01):132-135,138.

[7] 曾怡苗. 基于环形数据集的改进 K-means 聚类算法[J]. 软件,2021,42(11):74-76.

[8] SHANG TAO,ZHAO Zheng, GUAN Zhenyu, et al. A DP canopy K-means algorithm for privacy preservation of Hadoop platform [C]// Proceedings of the International Symposium on Cyberspace Safety and Security. Xi'an,China:dblp, 2017:189-198.

[9] PATWARY M M A, PALSETIA D, AGRAWAL A, et al. Scalable parallel OPTICS data clustering using graph algorithmic techniques[C]//Proceedings of the 2013 International Conference for High Performance Computing, Networking, Storage and Analysis (SC). Denver,CO., USA: IEEE, 2013:1-12.

[10] 周相兵. 位置数据智能聚类算法研究[M]. 北京:科学出版社,2021.

[11] 杨超,石连栓,施承尧,等. 求解高维函数优化的反馈多智能体遗传算法[J]. 软件,2020,41(07):81-90.

[12] LIU Yongguo, WU Xindong, SHEN Yidong. Automatic clustering using genetic algorithms[J]. Applied Mathematics and Computation,2011, 218 (4): 1267-1279.

[13] MUKHOPADHYAY A, MAULIK U. Towards improving fuzzy elustering using support vector machine: Application logene expression data[J]. Pattern Recognition, 2009, 42 (11): 2744-2763.

[14] 陈凡. 基于遗传算法的故障诊断方法研究[J]. 软件,2021,42(07):118-122.

[15] 王松,黄柯棣,杨妹. 基于动态数据驱动的交通在线决策[J]. 计算机仿真,2019,36(01):167-170.