

doi:10.3969/j.issn.1672-0504.2020.02.001

基于聚类的AOI出入口位置提取方法研究

钱振,周侗*,陶菲,范紫灵,刘浩,陆杰

(南通大学地理科学学院,江苏南通226007)

摘要:AOI出入口是行人寻址与车辆导航的重要地标,城市化导致出入口位置和通行状态频繁变动,通告不及时则会影响正常出行,出入口的更新问题亟须一种自动化的探测方案。由于出入口与客流的聚集点存在一定的正相关,而出租车的下客点在一定程度上可以代表周边客流量,该文据此提出了一种基于出租车下客点的最高热度出入口位置提取方法,以城市的AOI和POI作为地理背景数据,以出租车轨迹和路网作为出行行为数据,通过改进的地理位置文本聚类获取AOI位置,通过层次聚类并结合二分K均值聚类提取下客点簇,将两者进行关联计算得到出入口的具体位置。该方法既可应用于出入口位置的快速提取,也可作为出入口变动的动态监测方案。

关键词:聚类;AOI;时空数据挖掘;车辆轨迹

中图分类号:P208 文献标识码:A 文章编号:1672-0504(2020)02-0001-07

0 引言

感兴趣区域(Area of Interest, AOI)是包含多类感兴趣点(Point of Interest, POI)的面状实体, AOI及其出入口均为网络地图可视化的重要组成部分^[1,2]。近年来,城市化进程不断加快,导致AOI出入口位置变动频繁,使得网络地图需要确保该类数据能够及时更新,以保证寻址导航的准确性。现阶段网络地图对于AOI出入口的更新多通过实地勘测或GIS数据数字化等传统方法实现^[3],人力、物力消耗严重,因此,快速提取并动态更新出入口位置成为网络地图服务行业的迫切需求。

时空数据的开发利用为更新方法提供了新的可能,对轨迹数据进行挖掘分析,可感知居民出行的时空规律,进而为提取AOI出入口位置提供参考。聚类算法在轨迹数据挖掘方面被广泛使用^[4,5],其不仅可以探索地理对象的分布格局与特征,也可作为空间数据分析的预处理步骤,是当前地理信息数据挖掘的重要手段^[6,7];但常规聚类算法并不适用于具有多维属性的地理信息数据,需充分考虑数据的时间、空间和语义等信息^[8-11],故部分学者尝试改进聚类算法以满足时空数据特性^[12,13],进而挖掘事件的时空规律^[14],如探测道路状况^[15]、分析车辆行程特征^[16]、预测地区犯罪趋势^[17,18]、识别交通事故高发区域^[19]等。尽管近年来基于地理数据的聚类算法应

用广泛,但对不同几何类型的地理数据,如何选择合适的数据相似性衡量指标仍存在障碍^[20]。

本文旨在对AOI的最高热度出入口进行动态提取,将AOI、POI与车辆轨迹等数据的空间和语义信息同层次聚类、二分K均值聚类和改进的地理位置文本聚类等算法相结合,提出一种基于聚类的AOI出入口位置提取方案,以期优化网络地图服务,提高居民出行的便捷性。

1 方法设计

出租车轨迹数据具有一定行为特征,表现为上客点多位于城市主干道,下客点与目的地位置相近,隐含着居民出行的目的^[21]。本文针对AOI出入口的寻址问题,结合车辆下客点、POI和AOI的数据特征及其时空规律,利用不同的聚类算法进行位置提取。

1.1 二分K均值聚类算法

因传统K均值算法受初始聚类中心选择的影响较大,二分K均值算法对此问题进行了局部最优化^[22],具体实现过程为:开始聚类前,指定聚类个数K,将所有点作为一个簇C一分为二,放入簇集S;从簇集S中选出一个簇C_i并将其一分为二,选择总误差平方和SSE(式(1))最小的两个簇并将其放回簇集S中^[23,24]。不断重复上述基于SSE的划分过程,直至得到用户指定的簇数目K为止。

收稿日期:2019-06-10; 修回日期:2019-11-09

基金项目:国家自然科学基金项目(41301514);南通市空间信息技术研发与应用重点实验室项目(CP12016005);江苏省大学生创新创业训练计划项目(201910304113Y);国家级大学生创新创业训练项目(201910304036Z)

作者简介:钱振(1997—),男,研究方向为时空数据挖掘。*通讯作者 E-mail:tonggis@163.com

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist(c_i, x)^2 \quad (1)$$

式中: c_i 为簇 C_i 的聚类中心; x 为该簇的样本。

1.2 层次聚类算法

数据经清洗后,会造成部分数据集地理位置关联性缺失,直接使用二分 K 均值算法进行聚类无法确定最优的 K 值。本文使用凝聚型层次聚类算法对数据进行分层处理,具体过程为:将每个元素看作一类放入簇集 S ,计算元素与元素之间的欧氏距离,并将距离最小的两个簇 C_i, C_j 合并成一个新簇 C_k ;重新计算新簇与所有簇之间的距离,不断重复,直到所有簇合并成一个簇为止^[25]。

1.3 文本聚类算法

基于同类型文本相似度较大的原理,对不同的文本数据进行聚类,具体实现过程为:对文本进行切词处理,构建词袋空间(Vector Space Model, VSM);利用 TF-IDF(TF 为特征项频率, IDF 为逆文档频率)算法将单词的出现次数转换为权值(式(2));使用 K 均值算法进行聚类,并以余弦相似度 $Similarity_{i,j}$ 作为聚类算法的距离公式(式(3)),衡量两个文本间的相似度。

$$TF-IDF(wi) = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \lg \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (2)$$

式中: wi 为 POI 文本属性中的词组单元; $n_{i,j}$ 为当前词 i 在词项 d_j 中出现的次数; $\sum_k n_{k,j}$ 为所有字词出现的总次数; $|D|$ 为语料库中的词项总数; $|\{j : t_i \in d_j\}|$ 为包含词语 t_i 的词项数目(即 $n_{i,j} \neq 0$ 的文件数目),一般情况下使用 $1 + |\{j : t_i \in d_j\}|$ 进行计算。

$$Similarity_{i,j} = 1 - \cos\theta \quad (3)$$

$$\cos\theta = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\| \times \|\mathbf{v}_j\|} \quad (4)$$

式中: $\mathbf{v}_i, \mathbf{v}_j$ 分别为两个文本特征向量; $\cos\theta$ 表示两文本间距离,相似度为 1 时距离为 0,相似度为 0 时距离为 1。

1.4 改进的地理位置文本聚类算法

POI 数据包含多维度的信息(如名称、经纬度、类型等),文本聚类算法仅能对其名称或类型等文本信息进行聚类,会造成实际地理意义缺失^[26,27]。故本文结合二分 K 均值算法和文本聚类算法对 POI 进行聚类,每一类 POI 均位于相应的 AOI 中,算法具体实现过程为:1)以 POI 经度(lon_i)、纬度(lat_i)数据集 $LL = \{(lon_1, lat_1), (lon_2, lat_2), \dots, (lon_i, lat_i), \dots, (lon_n, lat_n)\}$ 为样本,使用二分 K 均值算法进行聚类,聚类结果为序列 $Cluster_A = \{ca_1, ca_2, \dots, ca_i, \dots, ca_n\}$;2)以 POI 文本名称数据集 $N = \{name_1,$

$name_2, \dots, name_i, \dots, name_n\}$ 为样本进行文本聚类,可得序列 $Cluster_B = \{cb_1, cb_2, \dots, cb_i, \dots, cb_n\}$;3)对序列 $Cluster_A$ 或 $Cluster_B$ 进行排序,并以其中一个序列为基础,更换另一个序列中的对应元素;4)将排序后的序列 $Cluster_A$ 与 $Cluster_B$ 对应元素相乘,可得序列 $M = \{new_ca_1 * new_cb_1, new_ca_2 * new_cb_2, \dots, new_ca_i * new_cb_i, \dots, new_ca_n * new_cb_n\}$,创建序列 $Cluster_C = \{1, 0, 0, \dots, 0\}$ (容量为 n)和类别变量 $label = 1$,遍历 M ,若 M 中的元素发生改变,则 $label = label + 1$,对应 $Cluster_C$ 的值为 $label$,得到最终序列 $Cluster_C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$ (c_i 为第 i 个兴趣点对应的聚类区域类别, n 为兴趣点数据中包含的兴趣点数量)(表 1)。

表 1 步骤 4 算法样例
Table 1 Algorithm examples of step 4

ID	Cluster _A	Cluster _B	M	Cluster _C
0	1	1	1	1
1	1	2	2	2
2	1	2	2	2
3	2	2	4	3
4	3	3	9	4
5	4	3	12	5
6	4	3	12	5

1.5 AOI 出入口提取方案

定义 1 车辆轨迹:设 p_i 为一辆车的第 i 个 GPS 信号点,则描述该车辆从 p_i 点移动到 p_n 点的 GPS 点集轨迹 $VT = \{p_1, p_2, \dots, p_i, \dots, p_n\}$ ^[28]。

定义 2 路网:在计算机中以图结构描述,由交叉路口 $V = (v_1, v_2, \dots, v_i, \dots, v_n)$ 、路段 $E = \{(v_i, v_j) | 0 \leq i, j \leq n \text{ 且 } i \neq j\}$ 构成,包含了道路的等级、宽度、长度等信息。其中,以交叉路口为结点,可将路网分割为若干基本单元。

定义 3 AOI 周边道路:首尾节点连接后,面域能够拓扑包含 AOI 的路段集合,表示为 $RG = <V, E>$ 。为提高方案提取的精确度,各周边道路闭合面域所包含的 AOI 应仅有一类(图 1)。

定义 4 车辆下客点:客运车辆运载时,会在某个位置 dp_i 放下乘客,车辆下客点集合 $DP = \{dp_1, dp_2, \dots, dp_i, \dots, dp_n\}$ ($0 \leq i \leq n$, i 为车辆订单的序号),通过车辆轨迹可提取出车辆下客点。

定义 5 反距离数量权重 (Inverse Distance Quantity Weight, IDQW):为本文针对出入口方向问题提出的指标,计算公式为:

$$IDQW_i = Q_j / D_{i,j}^p \quad (5)$$

$$D_{i,j} = \sqrt{(dp_{c_{jx}} - poic_{ix})^2 + (dp_{c_{jy}} - poic_{iy})^2} \quad (6)$$

式中: Q_j 为下客点第 j 个簇中所有元素的数量; $D_{i,j}$ 为聚类后的车辆下客点每类簇中心^[29]与对应 AOI

面内POI每类簇中心的距离; p 为反距离幂参数(距离权重),其值越大,越能强调近邻点对出入口方向的影响,适当取值可减少周边建筑的影响; dpc_{jx} 、 dpc_{jy} 和 $poic_{ix}$ 、 $poic_{iy}$ 分别为下客点簇中心和POI簇

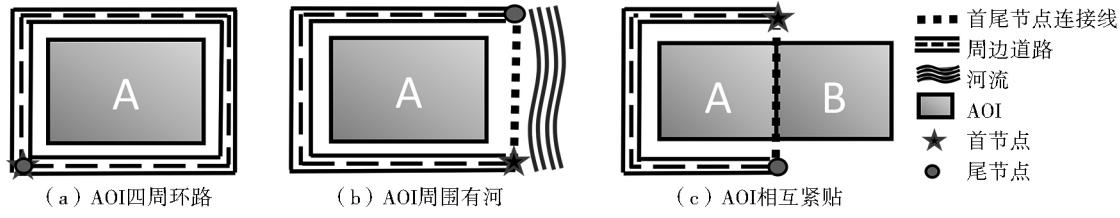


图1 AOI周边道路
Fig. 1 Surrounding roads of AOI

综上,本研究出入口提取方案具体实现过程为:
1) 提取AOI周边道路,筛选车辆下客点并进行数据清洗;2) 以清洗处理后的车辆下客点 DP 为数据集,以球面余弦定理(式(7))为距离公式进行层次聚类,并对各层次数据进行二分 K 均值聚类,计算每个最终类别的簇中心 dpc_j ;3) 以POI的名称数据集 N 、经纬度数据集 LL 为样本,进行改进的地理位置文本聚类,并计算每个类的簇中心 $poic_i$;4) 以 $poic_i$ 为起点、 dpc_j 为终点,计算向量集 $v_{i,j}$ (式(8));5) 出租车下客数据隐含一定的出行目的性,本文以AOI周边下客点簇中的元素数量以及该簇至AOI的距离为参数,构建反距离数量权重,以此衡量每簇感知AOI出入口方向的贡献度;6) 在向量集 $v_{i,j}$ 中,模为对应簇之间的反距离数量权重,模越大,代表该下客点簇对出入口的感知越强,对该向量集求和作为各类AOI的出入口方向向量 $Direction_i$ (式(9));7) 出入口往往建设在AOI边界上,因此,计算出入口方向后,延伸其方向向量与AOI边界的交点即可认为是该AOI的出入口位置。

$$S = R \cdot \arccos[\cos\beta_1 \cos\beta_2 \cos(\alpha_1 - \alpha_2) + \sin\beta_1 \sin\beta_2] \quad (7)$$

式中: α 、 β 分别为经、纬度角; R 为地球半径。

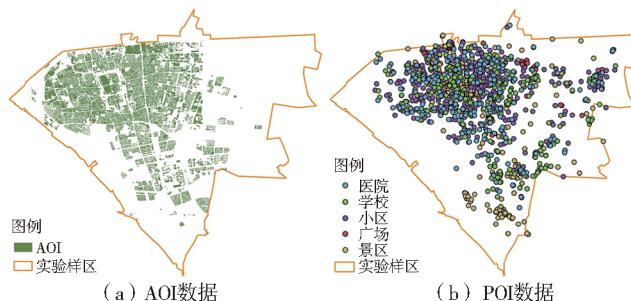
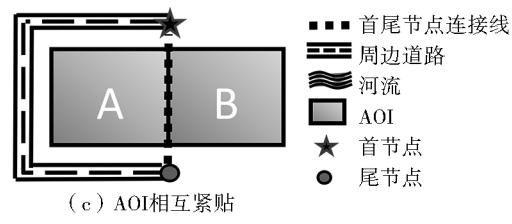


图2 原始数据
Fig. 2 Raw data

3 实验流程

本文实验环境为ArcGIS10.6+Python2.7.14+PyCharm,实验流程包括数据预处理、聚类分析、

中心的坐标。当 $D_{i,j}$ 越小、 Q_j 越大时,则认为 dpc_j 对感知当前POI簇所在的AOI出入口方向贡献度越大。



$$v_{i,j} = \left\{ \left(\frac{dpc_{jx} - poic_{ix}}{D_{i,j}} \times IDQW_i, \frac{dpc_{jy} - poic_{iy}}{D_{i,j}} \times IDQW_i \right), (0 \leq i \leq m, 0 \leq j \leq n) \right\} \quad (8)$$

式中: m 、 n 分别为POI和下客点聚类中心数量。

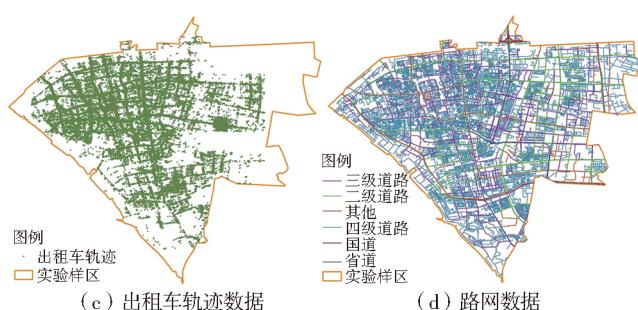
$$Direction_i = \sum_{j=0}^n v_{i,j} \quad (9)$$

2 实验数据

本文数据(图2,彩图见封3)的地理范围为江苏省南通市崇川区,其中,POI数据、AOI数据和路网数据分别利用网络爬虫通过高德地图API获取,以医院、学校、景区、居民小区和购物广场的AOI为例(表2),在ArcGIS中利用Python提取并验证;车辆数据为2018年11月1455辆出租车运营轨迹数据,来自南通市出租车管理系统。

表2 实验对象
Table 2 Experimental objects

类别	实验对象
医院	南通大学附属医院等4所三甲医院
学校	南通大学等10所大中小学
景区	海底世界等9处自然/人文景点
购物广场	文峰城市广场等6家购物广场
居民小区	中南世纪城等12个居民小区



出入口提取和方法验证(图3)。

3.1 数据预处理

根据车辆、POI、AOI以及路网数据的特征和地理属性进行数据清洗,具体包括:1) 车辆数据清洗,

剔除车辆记录为空且地理位置有误的轨迹数据;2)利用原始轨迹数据的状态变化提取下客记录;3)文件格式转换,将下客记录数据与POI数据等文本类型文件转换为图形文件(ShapeFile),便于空间分析和显示;4)地图配准, GIS 数据具有空间坐标系信息,将所有数据转换为相同地理坐标系,并进行几何

校正;5)以POI聚类结果为参考,确定AOI实验样例,并以遥感影像为参考,进行AOI边界的编辑和修正;6)筛选出道路等级为Ⅱ、Ⅲ、Ⅳ级的道路,并编辑成为AOI周边道路;7)以AOI周边道路为基准,对其做半径为50 m的缓冲区面域,保留被拓扑包含于该面域内的下客点。

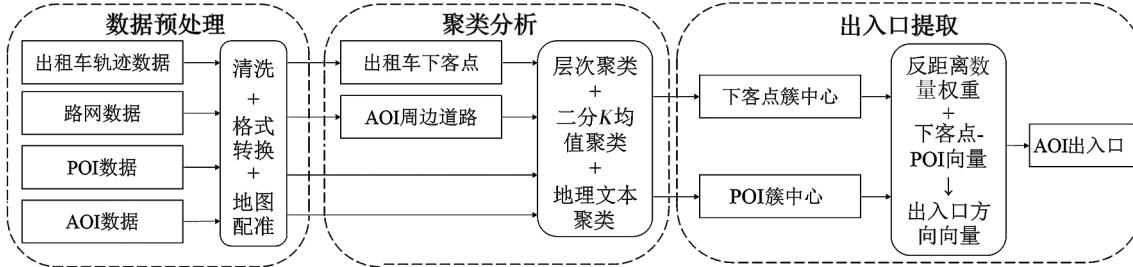
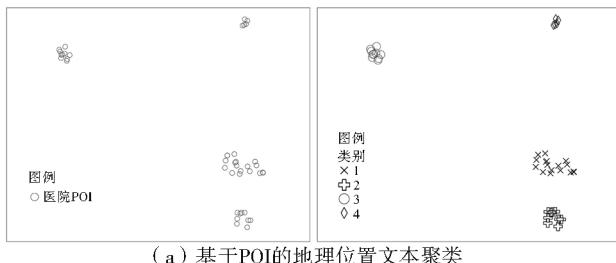


图3 实验流程
Fig. 3 Flow chart of the experiment

3.2 聚类分析

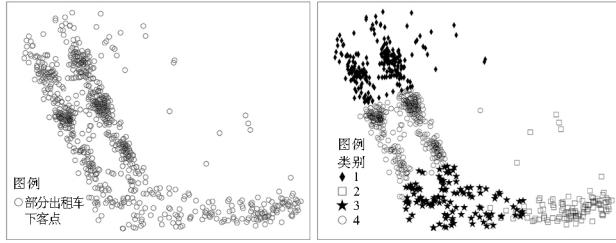
以医院为例对预处理后的下客点以及POI数据进行聚类(图4)。在使用二分K均值算法对层次聚类后的下客点数据SSE与K的折线图(图5)可确定肘部对应的K值为4,故对每一层数据而言,最佳聚类数为4。以该方法确定其他类别POI的较优K值,具体为: $K_{\text{学校}}=5, K_{\text{景区}}=4, K_{\text{购物广场}}=5, K_{\text{居民小区}}=6$ 。



(a) 基于POI的地理位置文本聚类



(b) 基于出租车下客点的层次聚类



(c) 基于出租车下客点的二分K均值聚类

图4 聚类结果对比
Fig. 4 Comparison of clustering results

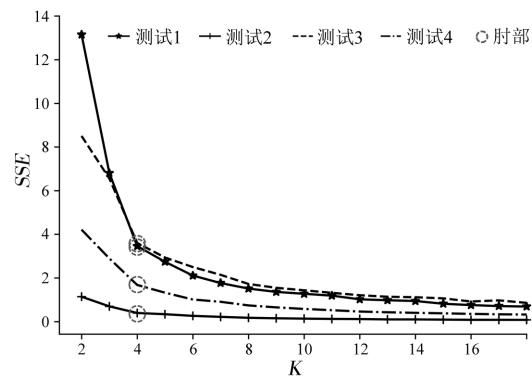


图5 “手肘法”K值选取曲线
Fig. 5 Curve of selecting K value by "elbow method"

3.3 出入口提取

在提取每类AOI的出入口位置时, IDQW中的距离权重 p 需根据数据特性,经对比实验选择合理数值。以医院为例,分析不同距离权重 p 提取的出入口位置与实际位置之间的偏差(表3)可知,当 p 取3或4时,位置整体偏差较小。图6、图7分别为 $p=3$ 时的提取过程及提取结果,提取结果的地图可视化效果较好,且偏差在地图用户的可接受范围内。因此,本文以距离权重 p 为3,以较优 K 值为基础,提取全部类别AOI的出入口,其位置偏差如表4所示。由表4可知,本文提取的出入口平均偏差为27 m,个别AOI出入口的偏差较高,尤其是学校的偏差最高(35 m)。主要原因为:1)因使用不同来源的地

表3 不同距离权重提取的出入口位置与实际位置之间的距离偏差

Table 3 Distance deviation between the entrance and exit positions extracted by different distance weights and the actual positions m

p	第一人民医院	中医院	南通大学附属医院	第三人民医院
1	43	41	18	56
2	7	24	14	46
3	3	14	9	12
4	7	21	8	14
5	11	23	13	23

理数据,坐标配准过程中存在一定误差,如部分路网数据偏离了实际位置且存在变形;2)不同的AOI之

间存在影响,一些人流量高的地区会影响其周边AOI出入口的提取。

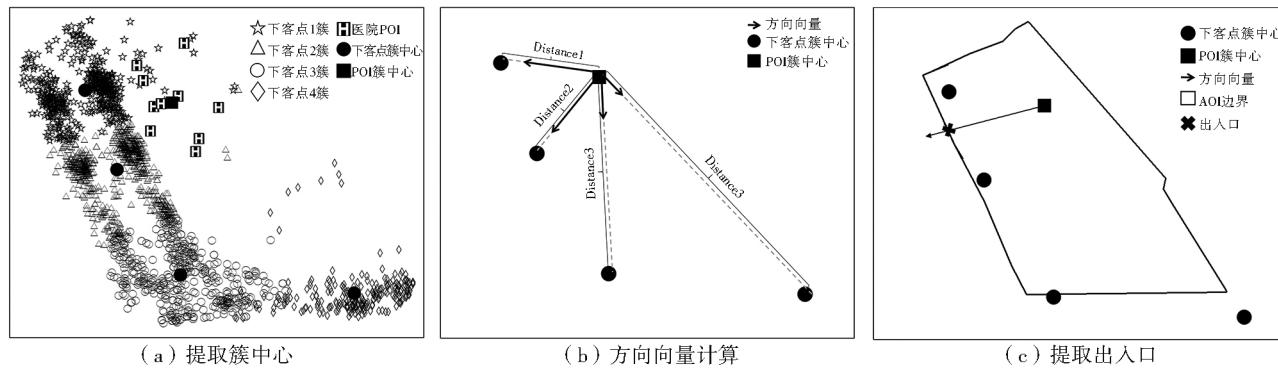


图6 加权向量提取出入口
Fig. 6 Extraction of entrances and exits by weighted vector

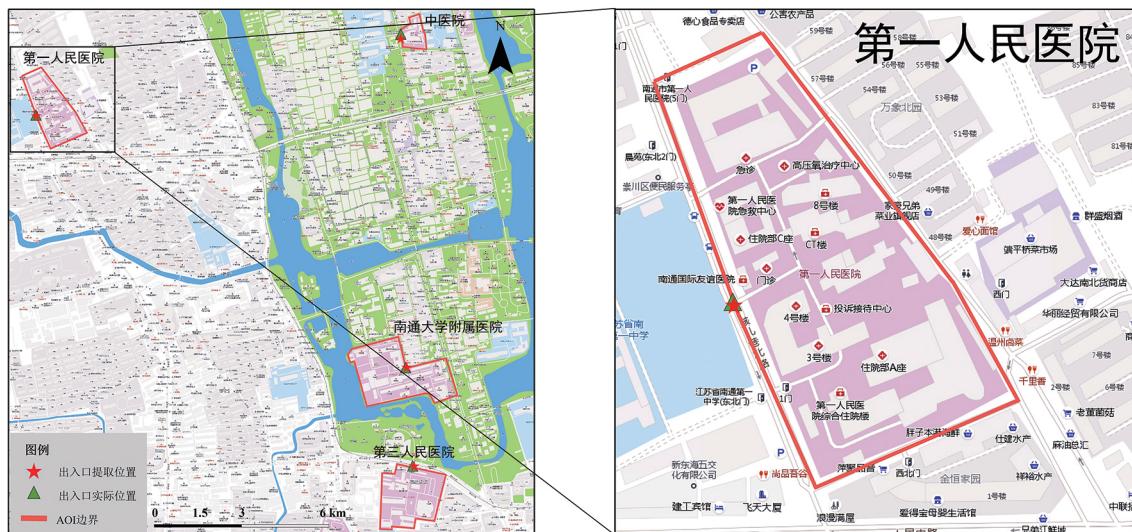


图7 医院出入口提取效果对比
Fig. 7 Comparison of the extraction effects of hospital entrances and exits

表4 各类AOI出入口位置距离偏差

Table 4 Distance deviation of entrance and exit positions for various AOI

类别	学校	景区	购物广场	居民小区	m
簇 I	5	24	3	31	
簇 II	16	12	7	48	
簇 III	9	19	24	36	
簇 IV	78	63	11	11	
簇 V	17	11	19	93	
簇 VI	16	21	47	23	
簇 VII	35	14	—	20	
簇 VIII	31	11	—	11	
簇 IX	17	83	—	19	
簇 X	127	—	—	23	
簇 XI	—	—	—	31	
簇 XII	—	—	—	11	

3.4 优化方案验证

进一步对提取方案进行优化,数字化周边道路并减小其缓冲面域的半径。以学校为例,使用“手肘法”测得最优K值为3,并将缓冲区域降为20 m,进行出入口提取,结果(表5)显示,改进后偏差有所降低,学校的平均偏差由原来的35 m降至17 m,南通

大学的3个校区及南通职业大学偏差较低,而中小学偏差普遍较高,其中,南通第三中学偏差最高。分析其原因:1)大学校内人数多,出入自由,从出入口车辆数量能反映出人流量;2)南通第一中学紧邻南通第一人民医院,提取过程中受医院周围的下客点影响较大,从而偏差较大;3)第三中学在居民小区内,因该类AOI自身特性,周边道路车辆轨迹数据较少。

表5 方法改进前后的效果对比
Table 5 Comparison of the effects before and after the method improvement

名称	原始偏差	改进后偏差
南通职业大学	5	1
南通大学主校区	16	3
南通大学钟秀校区	9	4
南通第一中学	78	41
南通师范第一附属小学	17	14
南通大学启秀校区	16	6
南通中学	35	13
南通师范第三附属小学	31	19
南通实验中学	17	21
南通第三中学	127	52

本文将优化方法拓展至所有类别AOI的出入口提取,显示平均偏差降至15 m。将提取的出入口位置与地籍相应原始出入口位置进行关联分析,动态更新位置,最终地图表现效果有所提升。

本文提取方法不仅可以提取出热度最高的出

入口,还可拓展到多出入口的提取,即每次提取出一个出入口后,清除该出入口对应AOI边界的下客点中心,利用其余下客点中心重复执行算法,直到下客点中心被清除完为止,多出入口提取结果如图8所示。

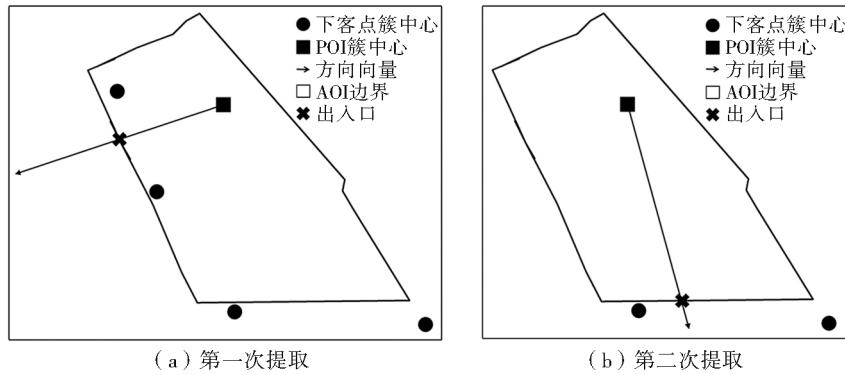


图8 多出入口提取
Fig. 8 Extraction of multiple entrances and exits in AOI

4 结语

本文提出了基于聚类的AOI出入口提取方案,针对不同类型的AOI开展一系列实验,对某些提取效果不佳的AOI类别进行分析,进一步改进了提取方案,平均偏差减小12 m,提取效果较好,与地籍相应原始出入口位置关联分析,实现动态更新,最终可达到网络地图可视化的要求。因某些类别的AOI(如小学、中学)无法通过下客点精确描述其周边人流量,从而无法有效提取该类出入口。同等数据精度下,将来可以对提取偏差较大的样例进行深入研究,寻找满足不同类别AOI出入口提取的高精度个性化方案。此外,将结合大数据分析以及其他机器学习算法,探讨AOI出入口规律性问题。

参考文献:

- [1] LI A, AXHAUSEN K W. Trip purpose imputation for taxi data [A]. 18th Swiss Transport Research Conference (STRC 2018) [C]. Ascona, Switzerland, 2018.
- [2] ZHANG Y P, LIU L. Understanding temporal pattern of human activities using Temporal Areas of Interest[J]. Applied Geography, 2018, 94: 95–106.
- [3] 徐立,肖强,陈换新,等.利用GIS数据进行地图制图表达与更新[J].测绘科学技术学报,2012,29(5):386–390.
- [4] 米春蕾,彭玲,姚晓婧,等.基于浮动车轨迹数据的路网快速提取[J].地理与地理信息科学,2019,35(1):12–19.
- [5] 刘文凯,唐建波,蔡建南,等.面向城市交通应用的移动对象聚类算法比较研究[J].地理与地理信息科学,2016,32(6):69–74.
- [6] 李志林,刘启亮,唐建波.尺度驱动的空间聚类理论[J].测绘学报,2017,46(10):1534–1548.
- [7] 李德仁,王树良,李德毅,等.论空间数据挖掘和知识发现的理论与方法[J].武汉大学学报(信息科学版),2002,27(3):221–233.
- [8] 赵竹珺,吉根林.时空轨迹分类研究进展[J].地球信息科学学报,2017,19(3):289–297.
- [9] PEI T, GONG X, SHAW S L, et al. Clustering of temporal event processes[J]. International Journal of Geographical Information Science, 2013, 27(3): 484–510.
- [10] ZHANG D Z, LEE K, LEE I. Mining hierarchical semantic periodic patterns from GPS-collected spatio-temporal trajectories [J]. Expert Systems with Applications, 2019, 122: 85–101.
- [11] CATLETT C, CESARIO E, TALIA D, et al. Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments[J]. Pervasive and Mobile Computing, 2019, 53: 62–74.
- [12] BIRANT D, KUT A. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. Data & Knowledge Engineering, 2007, 60(1): 208–221.
- [13] ZHANG D Z, LEE K, LEE I. Hierarchical trajectory clustering for spatio-temporal periodic pattern mining [J]. Expert Systems with Applications, 2018, 92: 1–11.
- [14] 张俊涛,武芳,张浩.利用出租车轨迹数据挖掘城市居民出行特征[J].地理与地理信息科学,2015,31(6):104–108.
- [15] 谭祥爽,王静,宋现锋,等.基于浮动车数据的路口探测方法[J].地理与地理信息科学,2015,31(5):34–38.
- [16] VAN LEEUWEN D, BOSMAN J, DUGUNDJI E. Spatio-temporal clustering of time-dependent origin-destination electronic trace data[J]. Procedia Computer Science, 2018, 130: 359–367.
- [17] 孙晨星,周侗,盖宸德,等.面向社区矫正的敏感区域行为分析方法[J].地理与地理信息科学,2018,34(3):114–118.
- [18] GRUBESIC T H, MACK E A. Spatio-temporal interaction of urban crime[J]. Journal of Quantitative Criminology, 2008, 24(3): 285–306.

- [19] SHEN L, LU J, LONG M, et al. Identification of accident black-spots on rural roads using grid clustering and principal component clustering [J]. Mathematical Problems in Engineering, 2019(4):1—12.
- [20] 犬乃夏,徐玉静,张恒才,等.移动轨迹聚类方法研究综述[J].测绘通报,2018(1):1—7.
- [21] 吴健生,李博,黄秀兰.小城市居民出行行为时空动态及驱动机制研究[J].地球信息科学学报,2017,19(2):176—184.
- [22] 刘广聪,黄婷婷,陈海南.改进的二分K均值聚类算法[J].计算机应用与软件,2015,32(2):261—263.
- [23] MAHMUD M R, MAMUN M A, HOSSAIN M A, et al. Comparative analysis of K-means and bisecting K-means algorithms for brain tumor detection [A]. 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) [C]. IEEE, Bangladesh, 2018. 1—4.
- [24] ZHAO W L, DENG C H, NGO C W. K-means: A revisit [J]. Neurocomputing, 2018, 291:195—206.
- [25] YUE H Q, GUAN Q F, PAN Y T, et al. Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: A case study of mainland China [J]. International Journal of Geographical Information Science, 2019, 33(5):1082—1105.
- [26] HU Y J, MAO H N, MCKENZIE G. A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements [J]. International Journal of Geographical Information Science, 2019, 33(4):714—738.
- [27] CARLOTTA M J. Text attributes and processing techniques in geographical information systems [J]. International Journal of Geographical Information Systems, 1995, 9(6):621—635.
- [28] 王晟郭,红钰.社区矫正人员位置信息挖掘[J].计算机应用与软件,2019,36(3):38—41.
- [29] 朱振国,冯应柱.基于数据场的类簇中心选取及其聚类[J].计算机工程与应用,2018,54(8):131—136.

Cluster-Based Extraction Method of Entrances and Exits in AOI

QIAN Zhen, ZHOU Tong, TAO Fei, FAN Zi-ling, LIU Hao, LU Jie

(School of Geography, Nantong University, Nantong 226007, China)

Abstract: As the important landmarks for pedestrian addressing and vehicle navigation, entrances and exits in AOI are changed frequently in their location and traffic status due to urbanization. Daily travelling will be affected without the timely updating of notification. So the update of the entrances and exits need to be automatically detected. There is a positive correlation between the entrances and exits and the passenger flow which can be represented by the taxi drop-off points. Based on this, this paper proposes a method to extract the location of entrance and exit by taxi drop-off point data. This method takes the AOI and POI of the city as the geographic background data, the taxi trajectory and road network as the travel behavior data. The improved geographic location text clustering is used to get the AOI location, and hierarchical clustering combined with bisecting K-means clustering are used to extract drop-off points cluster, then the locations of the entrances and exits can be calculated by correlation analysis of the AOI location and drop-off points cluster. This method can be used to quickly extract the entrance and exit positions and monitor the changes of entrances and exits.

Key words: clustering; AOI; spatio-temporal data mining; vehicle trajectory