基于信令数据的轨迹驻留点识别算法研究

李万林[①] 王 超[①②] 许国良*[②] 雒江涛[②] 张 轩[①②]

[①](重庆邮电大学通信与信息工程学院 重庆 400065)

[②](重庆邮电大学电子信息与网络工程研究院 重庆 400065)

摘 要：针对密度聚类算法只能识别密度相近的簇类且计算复杂度高等问题，该文提出一种基于信令数据中时空轨迹信息的密度峰值快速聚类(ST-CFSFDP)算法。首先对低采样密度的信令数据进行预处理，消除轨迹震荡现象；然后基于密度峰值快速聚类(CFSFDP)算法显式地增加时间维度限制，将局部密度由2维扩展到3维，并提出高密度时间间隔以表征簇中心在时间维度上的数据特征；接着设计筛选策略以选取聚类中心；最后识别用户出行轨迹中的驻留点，完成出行链的划分。实验结果表明，所提算法适用于采样密度低且定位精度差的信令数据，相比CFSFDP算法更适用于时空数据，相比基于密度的时空聚类算法(ST-DBSCAN)召回率提升14%，准确率提升8%，同时降低计算复杂度。

关键词：信令数据；时空聚类；密度峰值快速聚类算法；驻留点识别；出行链

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2020)12-3013-08

DOI: 10.11999/JEIT190914

Research of Track Resident Point Identification Algorithm Based on Signaling Data

LI Wanlin[①] WANG Chao[①②] XU Guoliang[②] LUO Jiangtao[②] ZHANG Xuan[①②]

[①](*Institute of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

[②](*Electronic Information and Networking Research Institute, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: For the problem that the density-based clustering algorithm can only identify clusters with similar density and high computational complexity, a Clustering by Fast Search and Find of Density Peaks based on Spatio-Temporal trajectory information in mobile phone signaling data, namely ST-CFSFDP, is proposed. Firstly, the low sampling density signaling data are pre-processed to eliminate the trajectory oscillation phenomenon in the data. Then, based on the Clustering by Fast Search and Find of Density Peaks(CFSFDP) algorithm, the time dimension limitation is explicitly increased, and the local density is extended from two-dimension to three-dimension. Moreover, in order to characterize the cluster center point in the time dimension, the concept of high-density time interval is defined. Secondly, the suitable cluster center screening strategy is developed to select automatically the appropriate cluster center. Finally, the resident points are identified in the travel trajectory of individual users over a period of time and the division of the travel chains is completed. The experimental results show that the algorithm is suitable for signaling data with low sampling density and poor positioning accuracy. It is more suitable for spatio-temporal data than CFSFDP algorithm. Compared with Density-Based Spatial Clustering of Applications with Noise based on Spatio-Temporal data (ST-DBSCAN) algorithm, the recall rate is improved by 14%, the accuracy rate is increased by 8%, and the computational complexity is also reduced.

Key words: Signaling data; Spatio-temporal clustering; Clustering by Fast Search and Find of Density Peaks (CFSFDP); Residual point recognition; Travel chain

收稿日期：2019-11-14；改回日期：2020-06-09；网络出版：2020-07-16

*通信作者： 许国良 xugl@cqupt.edu.cn

基金项目：重庆市自然科学基金 (cstc2018jcyjAX0587)，新型感知技术、信息融合处理及其应用(A2017-10)

Foundation Items: The Natural Science Foundation of Chongqing (cstc2018jcyjAX0587), The New Sensing Technology, Information Fusion Processing and its Application (A2017-10)

1 引言

随着智能手机、智能手表、智能行车记录仪等移动通信设备的普及，通信运营商积累了海量的信令数据资源。近年来，“智慧城市”的提出也使得各界学者基于信令数据在城市计算领域展开了广泛且深入的研究。在驻留点识别的研究方向上，相比传统的问卷调查、电话咨询、家庭上访等形式，信令数据具有用户被动上传、可信度高、获取简便、覆盖范围广等优势。

目前，关于驻留点识别的算法研究，可根据有无借助时空轨迹数据之外的信息分成两类。文献[1]利用用户在社交平台发布内容时附加的位置标签，考虑访问位置的顺序特征，得到用户轨迹在训练样本上的向量化表示，进而得到包含用户位置顺序信息的轨迹向量；文献[2]借助来自OpenStreetMap中的兴趣点(Point Of Interests, POI)信息，通过密度聚类算法识别活跃区域；文献[3]融合GPS数据和POI信息对城市范围按照功能进行划分，结合行为目的提取轨迹驻留点。当POI信息来源可靠且丰富时，上述研究是有意义的，但目前完善且能持续更新的POI信息较难获取，从而限制了此类算法的使用范围。更多的研究仅依赖轨迹数据的时空特征对轨迹点进行划分。一方面，可以基于时空特性设置阈值，来识别停留点。文献[4]使用手机数据结合土地属性，分析各基站间用户轨迹特征，设定时间、空间和频次阈值，将轨迹点划分为停留点和移动点；文献[5]基于出租车GPS数据，将城市区域栅格化，对各区域的发生事件个数分级量化提取热点区域。这一类算法基于特定场景建立规则模型，方法的普适性较差。另一方面，可以利用机器学习中的聚类算法识别驻留点[6]。文献[7]基于K-Means聚类算法对城市热点区域进行划分，但算法对K值依赖性强，不适用于驻留点个数未知的场景；文献[8,9]改进了DBSCAN算法并通过设定时间阈值来减少聚类次数提高计算效率，但算法只能识别密度相近的轨迹点，不适用于多密度环境下驻留点的识别；Birant等人[10]提出了适用于时空聚类的ST-DBSCAN算法，引入空间邻域和时间邻域，使用阈值区分距离较近的簇类，但该算法需要设置邻域半径、阈值等多个独立的参数，先验知识未知时，很难确定合适的参数；基于密度峰值快速聚类 (Clustering by Fast Search and Find of Density Peaks, CFSFDP)算法[11]解决了DBSCAN中参数的选择问题，虽然不需要迭代过程，但是缺少对时间维度的限制，且无法自动获取簇的中心点，以人工观察的方式获取聚类中心，无法直接应用于时空聚类的场景。除此之外，当前的研究大多是基于GPS数据，其采样密度高，定位较准确，但依赖用户主动上传，不适用于大规模出行轨迹数据分析的场景。

本文首先对原始数据预处理消除“轨迹震荡”干扰，然后基于CFSFDP算法提出一种适用于时空数据(CFSFDP based on Spatio-Temporal data, ST-CFSFDP)的轨迹驻留点识别算法，得到用户出行起讫信息。

2 问题描述

用户在进行数据传输时，智能终端(如手机)会借助信令与当前提供接入服务的基站进行信息注册和数据交互。根据产生方式的不同，可以将信令数据分成3类，即话单信令数据、CS(Circuit Switch)域信令数据、PS(Packet Switch)域信令数据。各类信令数据均包含用户ID、基站编号、时间戳等信息。具体如表1所示。

本文所用到的数据是原始信令数据经预处理后得到的用户出行轨迹数据$\langle x\ y\ t\ d\rangle$，其中$x,y$分别表示空间位置中的经度和纬度信息，$t$代表时间戳，$d$表示当前基站覆盖场景。为了实现对个人出行轨迹按照驻留点切分，提取模式单一的轨迹信息，本文针对现有问题给出解决方案：

(1) 针对基站定位的数据特点，提出适用于低采样密度的稀疏时空数据的ST-CFSFDP聚类算法；

(2) 增加时间约束，使改进后的ST-CFSFDP算法适用于时空聚类，解决路径重复带来的误聚类问题；

(3) 引入聚类中心权重值，解决CFSFDP算法无法自动选取聚类中心的问题；

表 1 信令数据主要字段

字段名称	字段解释	字段内容(示例)
user ID	用户身份	0001A
LAC_CID	基站位置区域码与小区识别码	13119-2056
TimeStamp	用户接入时间戳	2019-10-23 17:42:09
CoverScene	当前基站的覆盖场景	道路/学校/火车站等
Lon_Lat	当前基站经度、纬度	(106.59767, 29.40709)

(4) 针对时间、空间均有一定邻近性的不同簇类，仍能做出正确的划分。

3　基于序列模式的轨迹震荡消除方法

用户进行短信、通话、上网、位置移动等活动时都会产生信令数据，但由于基站覆盖范围重叠、位置切换、基站负载等，一次活动可能会产生多条不同位置数据[12]。设备没有移动，而信号会在不同的基站之间切换，从而产生震荡现象，如轨迹序列：$L_0-L_1-L_0$ 或$L_0-L_1-L_2-L_0-L_3-L_0$ 等，其中，L_i表示基站i的位置。L_0在轨迹序列中多次出现，形成震荡轨迹。信令数据中的位置是用户所接入基站的位置，而非用户实际活动位置[13]，在城市范围内，两者误差通常为250～900 m[14]。因此，需要对原始数据做预处理。

本文提出基于序列模式的轨迹震荡消除方法，主要分为两部分：(1)引入时间窗，结合震荡轨迹的切换特征识别震荡序列；(2)根据震荡序列中各点的访问频次和各位置的停留时长提取合适位置，修正震荡数据。

3.1　基于时间窗的震荡轨迹检测方法

震荡轨迹数据如表2所示，特征主要有两点：(1)相邻的两段或多段位移出现“循环”现象；(2)震荡数据发生时间较短，切换距离远，切换速度远超正常移动速度。

针对震荡数据发生时间短的特征，引入时间窗T_w，考虑到用户接入某个基站后的停留时间是随机的，因此时间窗的大小由序列片段中基站位置个数N_w决定，如式(1)

$$T_w = \sum_{p=i}^{i+N_w} t_p \tag{1}$$

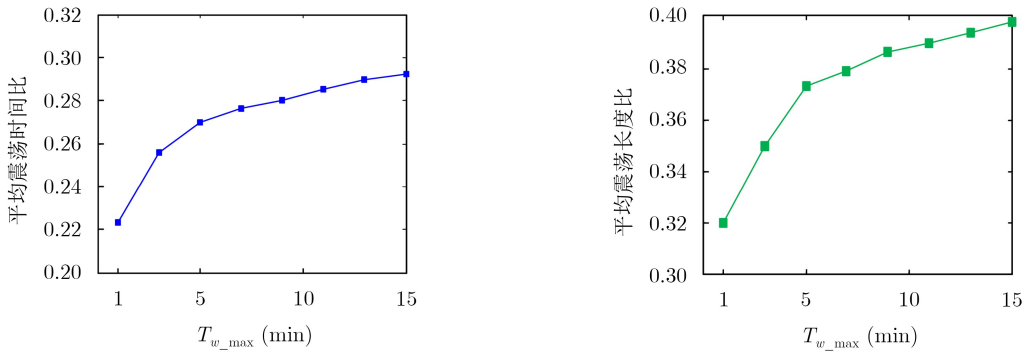
其中，t_p表示序列$L_i, L_{i+1}, \cdots, L_{i+N_w}$中轨迹点$L_p$的停留时间。基于时间窗的震荡轨迹检测方法是在$N_w$个基站的序列内判断是否产生了重复切换现象。同时为了避免误检，在检测到有重复切换现象后，仍需判断轨迹切片中震荡数据的时间间隔是否小于$T_{w_\max}$。为确定N_w和$T_{w_\max}$的值，本文引入平均震荡长度比和平均震荡时间比。其中，震荡长度比为震荡序列长度与N_w的比值，震荡时间比为震荡序列时间与T_w的比值。震荡序列中包含$L_0-L_1-L_0$，$L_0-L_1-L_2-L_0$和$L_0-L_1-L_0-L_2-L_3-L_2$等常见模式[15]。由于震荡序列中可能包含多种模式，选取所有震荡序列长度之和表示时间窗内的基站数，即N_w=15。由图1可得，当$T_{w_\max}$<5时，随着$T_{w_\max}$的增加，平均震荡比都有较大增加；当$T_{w_\max}$>5时，曲线较为平缓，此时增加$T_{w_\max}$并不能更好地检测震荡轨迹，过大的$T_{w_\max}$会对真实轨迹发生误检。当$T_{w_\max}$=5时，由图1可得平均震荡时间比约27.0%，平均震荡长度比约37.3%，该结果也符合Wang等人[12]对于震荡数据研究的结论。方法具体步骤如表3所示。

3.2　修正震荡轨迹数据

在轨迹序列中，用户真实的轨迹不会频繁切换，所以当检测出震荡序列后需要对震荡序列进行修正，剔除由信号漂移产生的错误轨迹点。由于实际位置点在震荡序列中出现的频次较多或停留时间较长，本文选取震荡序列中被访问次数最多或在震荡序列中总停留时间最长的点作为真实位置。

表 2　震荡轨迹数据示例

轨迹	位置	时间	距离 (km)	切换速度 (km/h)
D_0	L_0(106.607617, 29.530807)	08:19:35	/	/
D_1	L_1(106.602659, 29.545336)	08:20:14	1.6	147.6923
D_2	L_2(106.607617, 29.530807)	08:20:39	1.6	230.4000



(a) 平均震荡时间比　　(b) 平均震荡长度比

图 1 震荡时间最大间隔与平均震荡比的关系

表 3 基于时间窗的震荡轨迹检测方法

输入：原始轨迹数据$\boldsymbol{L}=\{L_1\cdots L_iL_{i+1}\cdots L_{i+N_w}\cdots\}$，轨迹序列切片中基站位置个数$N_w$，震荡数据最大时间阈值$T_{w_\max}$；

输出：检测到的震荡轨迹数据L_{osc}；

(1) 按顺序截取原始数据$\boldsymbol{L}$中的前N_w个位置组成序列L_w；

(2) 检测L_w中是否出现循环模式，如果出现则执行步骤(3)，否则序列点向前移1位，重新执行步骤(1)，截取后续N_w个位置的序列片段；

(3) 对检测到的震荡部分记为$(L_{\text{beg}}\cdots L_{\text{end}})$，判断该部分序列的总时间是否小于$T_{w_\max}$，如果满足，那么将该震荡序列记为$L_{\text{osc}}$，同时序列点向前移1位，返回步骤(1)；如果不满足，直接返回步骤(1)，直至遍历完$\boldsymbol{L}$内所有轨迹点。

算法结束

4 模型建立

4.1 CFSFDP聚类算法

CFSFDP算法[11]是一种基于密度峰值的聚类算法，相比DBSCAN算法避免了多次迭代，具有聚类速度快、实现简单、参数维度单一等特点。算法假设聚类中心是簇中密度最大的点，并且不同簇之间距离较大。

定义1：局部密度ρ_i，表示相距目标点的距离d_{ij}小于截断距离(cutoff distance)d_{c}的点的个数，如式(2)

$$\rho_i=\sum_j\chi(d_{\text{c}}-d_{ij}) \tag{2}$$

其中，$\chi(\Delta d)$为指示函数。当$\Delta d>0$时，$\chi(\Delta d)=1$，其它情况$\chi(\Delta d)=0$。

定义2：高密度间距δ_i，表示目标点到比该点局部密度ρ_i更大的点的距离最小值，如式(3)

$$\delta_i=\min_{j:\rho_j>\rho_i}(d_{ij}) \tag{3}$$

特殊地，对于密度最高的点，高密度间距δ_i定义为目标点到其它点的距离最大值，即$\delta_i=\max_j(d_{ij})$。

CFSFDP算法根据ρ_i和δ_i绘制决策分析图，选取图中ρ_i和δ_i均较大的点作为聚类中心。随机生成的40个2维空间内的原始数据点，数据分布如图2(a)所示。这些数据点可以组成3个簇类，并使用不同颜色表示。在决策分析图2(b)中，选取ρ_i和δ_i均较大的点作为簇的中心点，即数据点15, 25和26。

4.2 改进后的ST-CFSFDP算法

经典的CFSFDP算法缺少对时间维度的显式限制。如果直接利用该算法对时空数据进行聚类，在计算局部密度ρ_i时，d_{c}由2维变为3维，时间约束被隐式地增加到d_{c}中，使得d_{c}更难确定。除此之外，CFSFDP算法聚类中心需人为对决策分析图的观察得到，难以扩展到海量用户。本文提出ST-CFSFDP算法。具体地，将局部密度由2维扩展到3维；为了体现聚类中心在时间维度上的特点提出高密度时间间隔；引入聚类中心权值，并结合基站覆盖场景信息制定聚类中心筛选策略，自动选取聚类中心。

定义3：局部时空密度ρ_i，表示在空间维度相距目标点小于截断距离d_{c}，同时在时间维度与目标点间隔时间小于截断时间t_{c}的数据点的个数，如式(4)

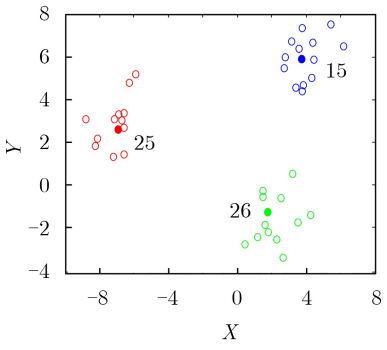
$$\rho_i=\sum_j\chi[\text{sgn}(d_{\text{c}}-d_{ij})+\text{sgn}(t_{\text{c}}-t_{ij})] \tag{4}$$

其中，$\text{sgn}(x)$为符号函数，当$x>0$时，$\text{sgn}(x)=1$；当$x=0$时，$\text{sgn}(x)=0$；当$x<0$时，$\text{sgn}(x)=-1$。$\chi(\Delta d)$为指示函数，当$\Delta d>0$时，$\chi(\Delta d)=1$，其它情况$\chi(\Delta d)=0$。

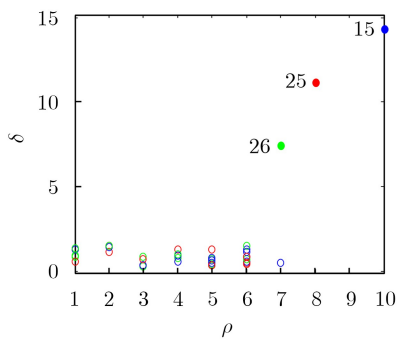
定义4：高密度空间距离δ_i，指目标点距离局部时空密度ρ_i更大的点的空间距离最小值，如式(5)

$$\delta_i=\min_{j:\rho_j>\rho_i}(d_{ij}) \tag{5}$$

特殊地，对于局部时空密度最高的点，δ_i定义为目标点到其它点的空间距离最大值，即$\delta_i=\max_j(d_{ij})$。



(a) 原始数据分布



(b) 决策分析

图 2 原始数据分布与CFSFDP算法决策分析

定义5：高密度时间间隔τ_i，指目标点距离局部时空密度ρ_i更大的点的时间间隔最小值，如式(6)

$$\tau_i = \min_{j:\rho_j>\rho_i}(t_{ij}) \tag{6}$$

特殊地，对于局部时空密度最高的点，τ_i定义为目标点到其它点的时间间隔最大值，即$\tau_i = \max_j(t_{ij})$。

随机生成了50个3维数据点，如图3(a)所示，原始数据被分成3簇。X轴和Y轴用以表示平面位置，Z轴用以表示时间维度。通过设置截断距离d_c、截断时间t_c，依次遍历各个数据点，计算各数据点的ρ_i，δ_i和τ_i。根据决策分析图3(b)，选取ρ_i，δ_i和τ_i均较大的点作为簇的中心点，即数据点3，23和32。

定义6：聚类中心权值γ_i表示数据点作为聚类中心的可能性，为ρ_i, δ_i和τ_i归一化后乘积，如式(7)

$$\gamma_i = \tilde{\rho}_i \cdot \tilde{\delta}_i \cdot \tilde{\tau}_i \tag{7}$$

其中，$\tilde{\rho}_i$, $\tilde{\delta}_i$和$\tilde{\tau}_i$分别为ρ_i, δ_i和τ_i按照max-min标准归一化后的值。数据点的ρ_i, δ_i和τ_i3个特征的乘积越大，该数据点就越可能是原始时空数据中的聚类中心点。

图4是表示聚类中心权重分布，图中红线表示数据点权重的平均值。由于大部分数据点为非中心点，其权重接近0，尤其在数据量较大时，平均值和非聚类中心的权重差距更大。因此，本文将聚类中心权重大于平均值的数据点作为初始的聚类中心候选点。结合图3(a)原数据分布可以发现，数据点6，9和32实际上属于同一类，聚类中心候选点中存在冗余点。为解决该问题，利用信令数据中基站覆盖场景信息，对分类结果进行“剪枝”。具体算法流程如表4。

5　模型验证与结果分析

5.1　实验准备

为验证模型结果的有效性和准确性，本文利用运营商提供的有出行标注的轨迹数据(共3020条)进行验证。在进行数据预处理后，使用CFSFDP算法、ST-DBSCAN算法分别与本文算法进行实验对比。首先将CFSFDP算法与本文算法在时空聚类的场景下进行对比，然后从准确度和复杂度方面与ST-DBSCAN算法进行对比。

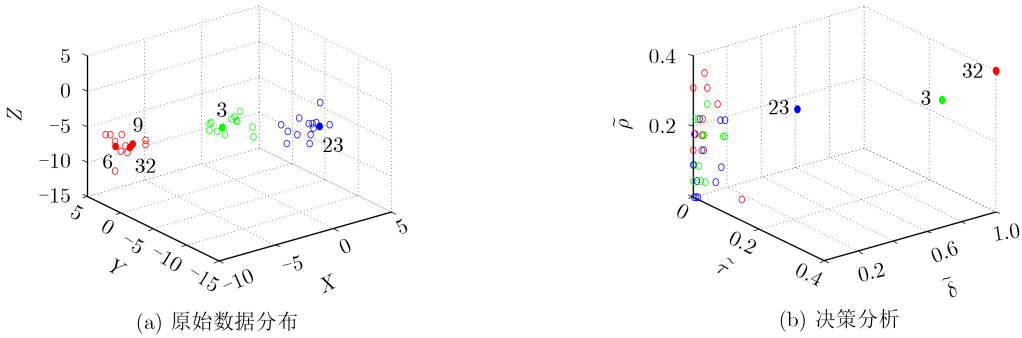
5.2　实验结果与分析

实验结果对比如表5所示。从识别个数的角度对比，实际全部驻留点有5个，CFSFDP与ST-DBSCAN算法分别识别出3个驻留点，而本文ST-CFSFDP算法识别出5个用户驻留点。从识别驻留点与真实驻留点的距离误差角度对比，ST-DBSCAN算法和本文算法两者性能较为接近，均优于CFSFDP算法。

从计算复杂度的角度分析，本文ST-CFSFDP算法复杂度在于计算ρ_i, δ_i和τ_i，需要全局遍历1次计算任意两点之间的距离d_{ij}和时间间隔t_{ij}，其计算复杂度为$O(N^2)$。通过建立索引对数据存储方式进行优化，可将复杂度降低为$O(N\lg N)$。而ST-DBSCAN算法在计算d_{ij}和t_{ij}的基础上，需要迭代地聚集直接密度可达的核心对象，并涉及大量的密度可达簇的合并。因此，本算法计算复杂度优于ST-DBSCAN。另外，相比CFSFDP算法，本算法增加了τ_i的计算，但未增加遍历次数，因此计算复杂度并未显著增加。

CFSFDP算法结果如图5所示。由于CFSFDP算法缺少对时间条件的显式约束，结果缺少驻留点2和4，无法识别轨迹中出现“A-B-A”重复路径的情况；ST-DBSCAN算法结果如图6所示，缺少驻留点4和5。驻留点4未识别是因为ST-DBSCAN算法将驻留点3和4判定为密度可达关系，因此划分为同一簇类；驻留点5未识别是因为该位置无法形成核心点。如果减小MinPts的值，则会将同密度的噪声点误判为核心点。这也说明了ST-DBSCAN无法在密度相差较大的数据中识别驻留点。

ST-CFSFDP算法结果如图7所示。计算各个轨迹点的ρ_i, δ_i和τ_i，绘制决策分析图7(a)；根据聚类中心权值得到未剪枝情况中心点分布图7(b)；结合基站覆盖场景等信息可得剪枝后的聚类中心分布图7(c)。



(a) 原始数据分布　　(b) 决策分析

图 3 原始数据分布与改进的CFSFDP算法决策分析

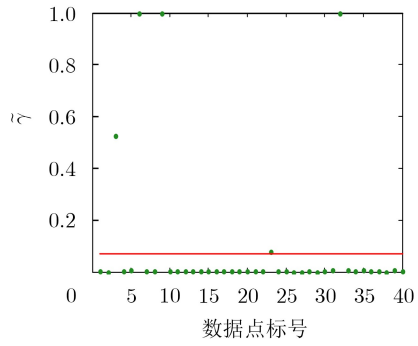


图 4 聚类中心权重分布

为了对算法准确性进行分析，本文将距离误差在50 m范围内的聚类中心记为正确标记，未找到或者距离误差大于50 m的中心点记为错误标记，并选取多组数据对两类算法分别进行了50次重复试验。从算法结果的召回率、准确率与F1值对比了ST-DBSCAN和本文ST-CFSFDP两算法，如图8所示。算法对比结果显示，在本数据集中针对驻留点识别方面，本文ST-CFSFDP算法召回率与准确率均有明显提升。具体地，ST-DBSCAN算法的召回率为72%，准确率为81%；本文算法的召回率为86%，准确率为89%。此外，本文算法中的参数只包含空间和时间约束，参数设定比ST-DBSCAN算法更为简单。

6 结束语

本文利用以移动信令为代表的低精度、低采样密度的时空数据，提出一种基于密度峰值的ST-CFSFDP算法，以识别时空轨迹中的驻留点。相比电话采访、问卷调查或GPS定位等数据，信令数据具有覆盖范围广、被动上传、可信度高、获取成本相对较低等优势。相比CFSFDP算法，本文所提ST-CFSFDP算法更加适用于时空数据且能自动筛选聚类中心，相比ST-DBSCAN算法能够识别多种密度簇类，简化参数的设置，避免多次迭代过程，降低了计算复杂度。此外，驻留点的距离误差仍有待改善，将是下一步的研究重点。

表 4 ST-CFSFDP聚类算法

输入：原始空间数据$P\langle x\ y\ t\ d\rangle$；截断距离$d_c$；截断时间$t_c$；数据点覆盖场景的描述$d$

输出：该原始数据的聚类集合C_k, $k=1,2,\cdots,n$；

(1) 计算每一个数据点的局部时空密度ρ_i；

(2) 依照定义4与定义5计算每个数据点的高密度空间距离δ_i、高密度时间间隔τ_i；

(3) 计算各个数据点的聚类中心权值，将聚类中心权值的平均值作为阈值，将大于该阈值的数据点放入聚类中心候选点集合C_c中；

(4) 合并候选点中覆盖场景相同且空间距离小于d_c或时间间隔小于t_c的“近邻数据点”，保留聚类中心权重较高的点；

(5) 将剩余的数据点，按照最近邻思想分配到各个聚类中心所代表的簇中。

算法结束

表 5 算法距离误差对比(m)

编号	驻留点坐标(Lon, Lat)	CFSFDP算法的距离误差	ST-DBSCAN算法的距离误差	ST-CFSFDP算法的距离误差
1	106.601230, 29.5339600	44.8	42.2	34.6
2	106.602061, 29.5343564	\	43.5	48.3
3	106.496737, 29.6166844	48.8	35.3	37.4
4	106.496729, 29.6166840	\	\	50.6
5	106.546322, 29.6203120	52.6	\	46.4

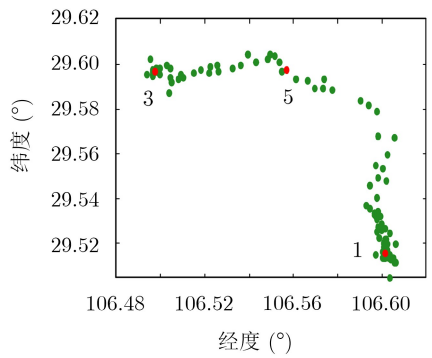


图 5 CFSFDP算法结果图

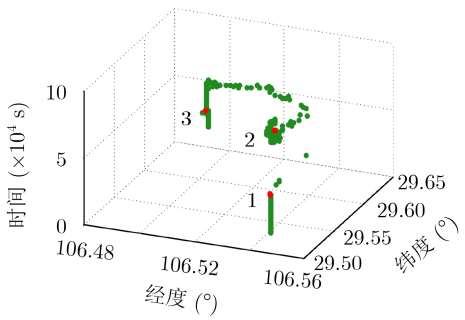
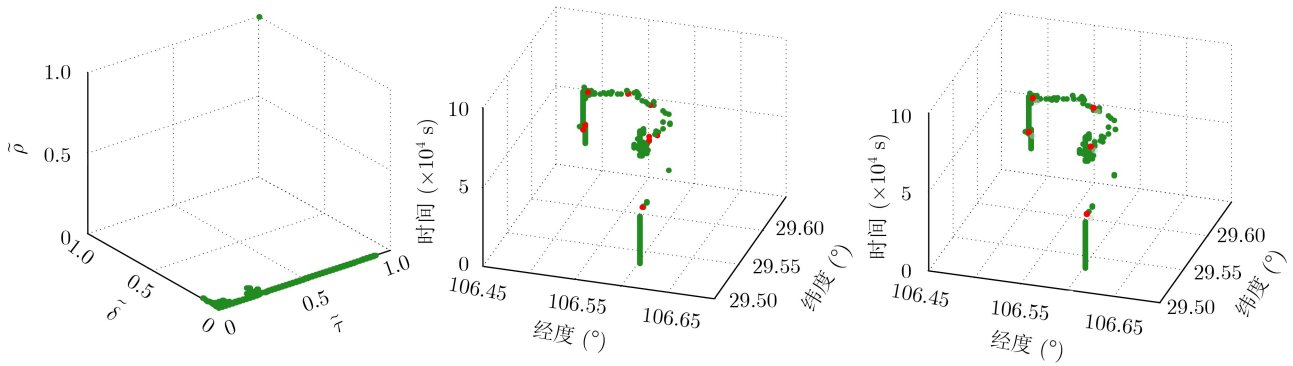


图 6 ST-DBSCAN算法结果图



(a) 决策分析　　(b) 聚类后数据分布-未剪枝　　(c) 聚类后数据分布-已剪枝

图 7 ST-CFSFDP算法结果图

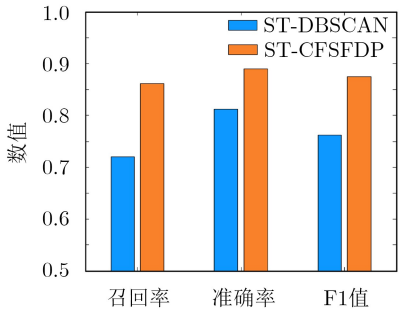


图 8 ST-DBSCAN与ST-CFSFDP算法对比

参 考 文 献

[1] 陈鸿昶, 徐乾, 黄瑞阳, 等. 一种基于用户轨迹的跨社交网络用户身份识别算法[J]. 电子与信息学报, 2018, 40(11): 2758–2764. doi: 10.11999/JEIT180130.

CHEN Hongchang, XU Qian, HUANG Ruiyang, *et al*. User identification across social networks based on user trajectory[J]. *Journal of Electronics & Information Technology*, 2018, 40(11): 2758–2764. doi: 10.11999/JEIT180130.

[2] 彭大芹, 罗裕枫, 江德潮, 等. 基于移动信令数据的城市热点识别方法[J]. 重庆邮电大学学报: 自然科学版, 2019, 31(1): 95–102. doi: 10.3979/j.issn.1673-825X.2019.01.013.

PENG Daqin, LUO Yufeng, JIANG Dechao, *et al*. Urban hotspots identification method based on mobile signaling data[J]. *Journal of Chongqing University of Posts and Telecommunications*: *Natural Science Edition*, 2019, 31(1): 95–102. doi: 10.3979/j.issn.1673-825X.2019.01.013.

[3] 罗孝羚, 蒋阳升. 基于出租车运营数据和POI数据的出行目的识别[J]. 交通运输系统工程与信息, 2018, 18(5): 60–66. doi: 10.16097/j.cnki.1009-6744.2018.05.010.

LUO Xiaoling and JIANG Yangsheng. Trip-purpose-identification based on taxi operating data and POI data[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18(5): 60–66. doi: 10.16097/j.cnki.1009-6744.2018.05.010.

[4] 鲍冠文, 刘小明, 蒋源, 等. 基于改进DBSCAN算法的出租车载客热点区域挖掘研究[J]. 交通工程, 2019, 19(4): 62–69. doi: 10.13986/j.cnki.jote.2019.04.010.

BAO Guanwen, LIU Xiaoming, JIANG Yuan, *et al*. Research on mining taxi pick-up hotspots area[J]. *Journal of Transportation Engineering*, 2019, 19(4): 62–69. doi: 10.13986/j.cnki.jote.2019.04.010.

[5] 李岩, 陈红, 孙晓科, 等. 基于热点探测模型的城市居民出行特征分析[J]. 交通信息与安全, 2019, 37(1): 128–136. doi: 10.3963/j.issn.1674-4861.2019.01.017.

LI Yan, CHEN Hong, SUN Xiaoke, *et al*. An analysis of travel characteristics of urban residents based on hot spot detection model[J]. *Journal of Transport Information and Safety*, 2019, 37(1): 128–136. doi: 10.3963/j.issn.1674-4861.2019.01.017.

[6] 张海霞, 李腾腾, 李东阳, 等. 基于车辆行为分析的智能车联网关键技术研究[J]. 电子与信息学报, 2020, 42(1): 36–49. doi: 10.11999/JEIT190820.

ZHANG Haixia, LI Tiantian, LI Dongyang, *et al*. Research on vehicle behavior analysis based technologies for intelligent vehicular networks[J]. *Journal of Electronics & Information Technology*, 2020, 42(1): 36–49. doi: 10.11999/JEIT190820.

[7] 李浩, 王旭智, 万旺根. 基于位置数据的居民出行时空特征研究——以上海市为例[J]. 电子测量技术, 2019, 42(19): 25–30. doi: 10.19651/j.cnki.emt.1902923.

LI Hao, WANG Xuzhi, and WAN Wanggen. Research on temporal and spatial characteristics of residents' travel based on location data—A case of Shanghai[J]. *Electronic Measurement Technology*, 2019, 42(19): 25–30. doi: 10.19651/j.cnki.emt.1902923.

[8] 周洋, 杨超. 基于时空聚类算法的轨迹停驻点识别研究[J]. 交通运输系统工程与信息, 2018, 18(4): 88–95. doi: 10.16097/j.cnki.1009-6744.2018.04.014.

ZHOU Yang and YANG Chao. Anchors identification in trajectory based on temporospatial clustering algorithm[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18(4): 88–95. doi:

10.16097/j.cnki.1009-6744.2018.04.014.

[9] 方琪, 王山东, 于大超, 等. 基于出租车轨迹的居民出行特征分析[J]. 地理空间信息, 2019, 17(5): 128–130. doi: 10.3969/j.issn.1672-4623.2019.05.034.
FANG Qi, WANG Shandong, YU Dachao, *et al*. Analysis of resident trip characteristics based on taxi trajectory[J]. *Geospatial Information*, 2019, 17(5): 128–130. doi: 10.3969/j.issn.1672-4623.2019.05.034.

[10] BIRANT D and KUT A. ST-DBSCAN: An algorithm for clustering spatial–temporal data[J]. *Data & Knowledge Engineering*, 2007, 60(1): 208–221. doi: 10.1016/j.datak.2006.01.013.

[11] RODRIGUEZ A and LAIO A. Clustering by fast search and find of density peaks[J]. *Science*, 2014, 344(6191): 1492–1496. doi: 10.1126/science.1242072.

[12] WANG Feilong and CHEN C. On data processing required to derive mobility patterns from passively-generated mobile phone data[J]. *Transportation Research Part C*: *Emerging Technologies*, 2018, 87: 58–74. doi: 10.1016/j.trc.2017.12.003.

[13] CHEN C, BIAN Ling, and MA Jingtao. From traces to trajectories: How well can we guess activity locations from mobile phone traces?[J]. *Transportation Research Part C*: *Emerging Technologies*, 2014, 46: 326–337. doi: 10.1016/j.trc.2014.07.001.

[14] HARD E, CHIGOY B, SONGCHITRUKSA P, *et al*. Synopsis of new methods and technologies to collect Origin-Destination (O-D) data[R]. FHWA-HEP-16-083, 2016.

[15] LEE J K and HOU J C. Modeling steady-state and transient behaviors of user mobility: Formulation, analysis, and application[C]. The 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Florence, Italy, 2006: 85–96.

李万林：男，1963年生，教授、博士生导师，研究方向为新一代网络技术、自动驾驶，车联网及移动大数据等.

王 超：男，1994年生，硕士生，研究方向为移动大数据、机器学习.

许国良：男，1973年生，教授、硕士生导师，研究方向为光电传感与检测、通信网络设计与规划、大数据分析挖掘.

雒江涛：男，1971年生，教授、博士生导师，研究方向为移动大数据、新一代网络技术、通信网络测试与优化等.

张 轩：男，1991年生，硕士生，研究方向为移动大数据、机器学习.

责任编辑：马秀强