基于轨迹划分与密度聚类的移动用户重要地点识别方法

杨 震 王红军

(国防科技大学电子对抗学院 合肥 230037)

摘 要 移动用户轨迹数据作为新兴的空间轨迹数据,可用于分析个体或群体的行为特征、兴趣爱好,在智慧城市、交通规划和反恐维稳等领域应用广泛。为了从庞大的数据集中识别出移动用户的重要地点,提出了一种基于转角偏移度与距离偏移量的轨迹划分算法。该算法首先通过轨迹划分提取出用户的重要地点候选集,然后采用一种改进的密度聚类算法进一步对用户的候选重要地点实现聚类,从而识别出用户的最终重要地点。在Geolife轨迹数据集与Foursquare用户签到数据集上的实验表明,采用轨迹划分与密度聚类相结合的重要地点识别方法具有比现有的重要地点识别方法更高的准确率,证明了所提方法的可行性与优越性。

关键词 移动用户,轨迹划分,重要地点,密度聚类

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.004

Important Location Identification of Mobile Users Based on Trajectory Division and Density Clustering Method

YANG Zhen WANG Hong-jun

(Electronic Countermeasures College, National University of Defense Technology, Hefei 230037, China)

Abstract As emerging spatial trajectory data, mobile user trajectory data can be used to analyze individual or group behavioral characteristics, hobbies and interests, and are widely used in smart cities, transportation planning, and anti-terrorism maintenance. In order to identify the important locations of mobile user from a huge data set, this paper proposed a trajectory division method based on the angle and distance offset. The method firstly extracts the important locations candidate set by trajectory division, and then further clusters the important locations through an improved density clustering algorithm, extracting the final important location of user. The experiment on Geolife trajectory data set and Foursquare data set shows that the important location identification method combining trajectory division and density clustering has higher accuracy than other existing important location identification method, which proves the feasibility and superiority of the proposed method.

Keywords Mobile user, Trajectory division, Important locations, Density clustering

1 引言

近年来,随着移动通信技术的飞速发展和智能移动终端网络功能的日趋强大,手机、平板电脑等智能终端设备已经逐渐超越了个人计算机,成为人们使用得最多的信息设备。同时,全球导航定位系统的蓬勃发展为智能移动终端提供了精确、实时的位置信息[1]。基于位置数据的服务(Location Based Service, LBS)已成为最流行的终端信息服务之一,大量的相关研究也证明了海量位置数据中所隐含的挖掘价值[2-5]。

研究表明,人类的活动具有很强的规律性,用户流动性的潜在可预测率为93%[6]。为了从用户的历史位置数据中推断出其活动规律及偏好特征,需要对用户的重要地点进行挖掘。然而,这是一项极具挑战性的工作,主要面临以下困难:1)移动用户历史轨迹数据具有非一致性和稀疏性,因为城市建筑的遮挡、用户随意开关GPS常常导致轨迹记录的中断;2)海量轨迹数据中仅有小部分具有特殊意义,例如移动用户住址附近的轨迹。研究人员已经在该领域展开了广泛的研究,并提出了基于轨迹点密度的方法[7-9]、基于停留时间的方法[10-14]以及基于轨迹结构的方法[15-16]。现有研究能够较为有效地从海量轨迹数据中提取出具有意义的信息,但是依然存在以下问题。

Ashbrook等[7]直接采用k-means算法对轨迹点进行聚类,该方法主要存在3个缺点:1)需要用户事先指定聚类簇的个数;2)没有对原始位置点进行预处理,聚类结果中将包含噪声点;3)算法的初始聚类中心是随机选择的,不同的初始聚类中心会导致不同的聚类结果。Yang等[8]首先采用梯度阈值对噪声进行处理,然后采用DBSCAN(Density-Based Spatial Clustering of Applications with Noise)算法对位置点聚类,将各类簇输出标注为重要地点。但是,梯度阈值及DBSCAN算法输入参数的设定具有很强的主观性与敏感性,且该方法难

到稿日期:2018-07-27 返修日期:2018-12-06 本文受国家自然科学基金(61273302)资助。

杨 震(1994-),男,硕士生,主要研究方向为聚类分析、轨迹预测,E-mail:eei_yz@163.com;王红军(1968-),男,博士,教授,主要研究方向为移动通信网、认知电子战,E-mail:hongjun-wang@163.com(通信作者)。

以识别多密度区域。Montoliu 等[9]先采用基于时间的聚类算法提取停留点，然后采用基于网格的聚类算法提取停留区域，并将其输出标注为重要地点。该方法的效率较高，但是网格聚类在一定程度上牺牲了识别的精度。基于停留时间的方法[10-14]通过设置阈值来进行停留点的判定，该类方法的时间消耗较少，但是阈值设置是否合理会极大地影响最终的识别结果。Kreveld 等[15]首次将时间维度引入轨迹分析中，通过轨迹采样点之间的时延计算轨迹相似度。但是，该方法将轨迹看作一个整体，忽略了局部特征的提取。袁冠等[16]提出了基于结构相似度的轨迹分析方法，其通过计算轨迹的时间、位置、方向、转角、速度等特征来提取高相似度的频繁轨迹。但是，该方法的复杂性较高，且难以解决轨迹数据集的稀疏性问题。Khetarpaul 等[17]提出的算法是近几年提出的基于停留时间的改进算法，其在 SMOT（Spatio-Temporal Clustering Method）[10]算法的基础上加入了新的判定条件，一定程度上解决了 GPS 信号丢失时导致的误判问题，因此本文将该算法作为对比算法。

针对上述研究存在的问题，本文提出了一种基于轨迹转角偏移度与距离偏移量的轨迹划分算法（Trajectory Division Algorithm Based on Angle and Distance Offset，TD-ADO）。首先设定偏移阈值，将超过阈值的轨迹采样点作为候选重要地点（Candidate Important Locations，CIL），然后对其进行聚类，以识别移动用户最终的重要地点。

2 轨迹划分

2.1 轨迹转角偏移度与距离偏移量

轨迹划分的目的在于提取出轨迹中用户行为变化较大的采样点，并将提取出的轨迹采样点划入候选重要地点集中，每个候选重要地点作为前一段子轨迹的终点与下一段子轨迹的起点，将一条完整轨迹划分为若干条连续子轨迹。

为了更好地描述后续算法，采用文献[3]中对轨迹的定义。设 TD（Trajectory Dataset）为一个轨迹数据集，且该数据集由 n 条轨迹组成，即 $TD=\{TJ_1,TJ_2,\cdots,TJ_n\}$。轨迹（$TJ$）是由若干个轨迹采样点按时间先后顺序组成的一个序列，即 $TJ_i=\{P_1,P_2,\cdots,P_m\}(1\leqslant i\leqslant n)$。其中，$P_j(1\leqslant j\leqslant m)$是由经纬度及采样时间组成的轨迹采样点，即$\langle Lat_j,Lng_j,T_j\rangle$，其中 Lat_j,Lng_j,T_j 分别表示采样点 P_j 的纬度、经度以及采样时间。

定义 1（子轨迹，sub-trajectory） 轨迹 TJ_i 的子轨迹表示为 $STJ_i=\{P_{s1},P_{s2},\cdots,P_{sk}\}(1\leqslant s1<s2<\cdots<sk\leqslant m)$。本文中提到的子轨迹均为连续子轨迹，即从 P_{s1} 到 P_{sk} 为连续轨迹采样点。

定义 2（转角偏移度，angle offset） 如图 1 所示，以轨迹 TJ_i 为例，P_1 为初始轨迹点，$\boldsymbol{P_1P_2}$ 为初始移动行为。转角偏移度是指移动行为 $\boldsymbol{P_iP_{i+1}}$ 与 $\boldsymbol{P_{i+1}P_{i+2}}(1\leqslant i\leqslant m-2)$尾端相连时的夹角 θ_i，如图 1 中 $\boldsymbol{P_1P_2}$ 与 $\boldsymbol{P_2P_3}$ 尾端相连时的夹角 θ_1 以及 $\boldsymbol{P_2P_3}$ 与 $\boldsymbol{P_3P_4}$ 尾端相连时的夹角 θ_2。本文设定顺时针转角为正值，逆时针转角为负值，即 θ_1 为正值，θ_2 为负值。为了计算转角偏移度的大小，首先通过轨迹采样点的经纬度值计算出两点之间的距离大小：

$$\begin{cases}A=\sin^2\left(\dfrac{Lat_j-Lat_i}{2}\right)\\ B=\cos(Lng_i)\cos(Lng_j)\sin^2\left(\dfrac{Lng_j-Lng_i}{2}\right)\\ d(P_i,P_j)=2R\arcsin\sqrt{A+B}\end{cases}\tag{1}$$

其中，$d(P_i,P_j)$为轨迹采样点 P_i 与 P_j 之间的距离，R 为地球半径。

轨迹转角偏移度的计算公式为：

$$\alpha_i=\arccos\frac{(|\overrightarrow{\boldsymbol{P_iP_{i+1}}}|^2+|\overrightarrow{\boldsymbol{P_{i+1}P_{i+2}}}|^2-|\overrightarrow{\boldsymbol{P_iP_{i+2}}}|^2)}{2|\overrightarrow{\boldsymbol{P_iP_{i+1}}}|^2|\overrightarrow{\boldsymbol{P_{i+1}P_{i+2}}}|^2}\tag{2}$$

$$\theta_i=\begin{cases}\alpha-\pi, & \overrightarrow{\boldsymbol{P_iP_{i+1}}}\times\overrightarrow{\boldsymbol{P_{i+1}P_{i+2}}}<0\\ \pi-\alpha, & \overrightarrow{\boldsymbol{P_iP_{i+1}}}\times\overrightarrow{\boldsymbol{P_{i+1}P_{i+2}}}\geqslant 0\end{cases}\tag{3}$$

定义 3（距离偏移量，distance offset） 距离偏移量是指轨迹采样点到初始移动行为所在直线的垂直距离。如图 1 所示，采样点 P_3 与 P_4 到初始移动行为 $\boldsymbol{P_1P_2}$ 延长线的垂直距离 d_1 与 d_2 即为距离偏移量，其计算公式为：

$$d_i=|\overrightarrow{\boldsymbol{P_2P_{i+2}}}|\sin\angle P_1P_2P_{i+2}\tag{4}$$

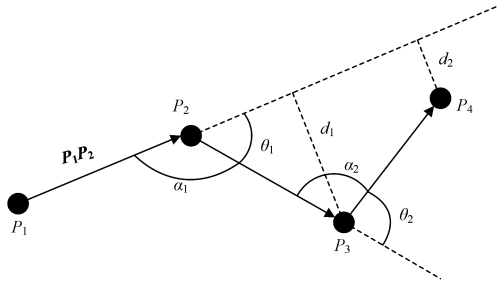

图 1 轨迹转角偏移度与距离偏移量示意图

Fig. 1 Diagram of angle and distance offset

2.2 轨迹划分算法

对于轨迹 TJ_i，首先从采样点 P_3 开始，计算其转角偏移度 θ_1 与距离偏移度 d_1，若 θ_1 的绝对值不超过转角偏移度阈值 θ_{th}，且 d_1 也不超过距离偏移度阈值 d_{th}，则继续计算采样点 P_4 的偏移度，依此类推。若采样点 P_i 的转角偏移度或距离偏移度超过了相应的阈值，则将 P_i 划入候选重要地点集中，将 $\boldsymbol{P_iP_{i+1}}$ 定义为新的初始移动行为，并从采样点 P_{i+2} 开始继续执行上述步骤。最后，轨迹 TJ_i 将被划分为若干个连续子轨迹。本文的轨迹划分算法的时间复杂度为 O(n)，其中 n 为轨迹采样点的个数。

采用转角偏移度与距离偏移量相结合的轨迹划分算法能够有效避免以下两种情况。1）若只通过设置转角偏移度阈值进行轨迹划分，则无法提取出图 2 中的候选重要地点。图 2 中的每个转角偏移度均较小，但是由于多次顺时针旋转的叠加效应，P_1-P_6 轨迹段的形状近似为一个半圆，显然将 P_1-P_6 这 6 个采样点归为非候选重要地点是不合理的，这样会导致重要地点的遗漏。2）若只通过设置距离偏移量阈值进行轨迹划分，同样可能导致重要地点的漏选。图 3 中，由于轨迹有规律地顺-逆时针转向，导致采样点 P_3-P_5 均在初始移动行为所在直线的附近，但是相连移动行为之间的转角已经接近 90°，此时将这一轨迹段的所有采样点归为非候选重要地点显然也是不妥的。

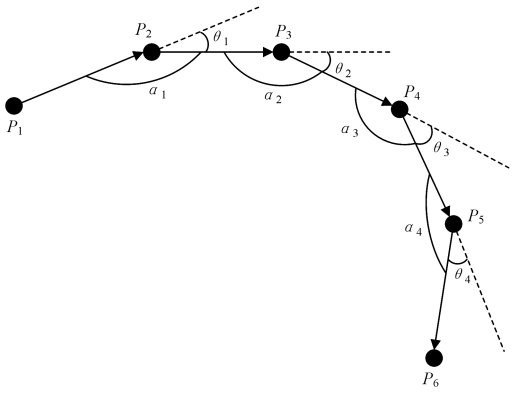

图 2　轨迹连续顺时针转向示意图

Fig. 2　Diagram of continuous clockwise steering trajectory

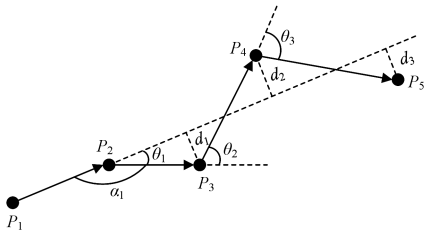

图 3　轨迹规律顺-逆时针转向示意图

Fig. 3　Diagram of regular clockwise-anticlockwise steering trajectory

轨迹划分算法的完整伪代码如算法 1 所示。

算法 1　TD-ADO

Input:轨迹 $TJ_i=\{P_1,P_2,\cdots,P_m\}$,转角偏移度阈值 θ_{th},距离偏移量阈值 d_{th}

Output:轨迹候选重要地点集 CILD

```
1. P1→CLID;
2. i=1;
3. repeat
4.    initial movement behavior=P_iP_{i+1};
5.    for j=1 to m−2 do
6.       if θ_j>θ_th ‖ d_j>d_th then
7.          P_{j+2}→CLID;
8.          i=j+2;
9.          break;
10.      else
11.         if j= m−2 then
12.            P_{j+2}→CLID;
13.         end if
14.      end if
15.   end for
16. until Pm→CLID.
17. return CLID.
```

算法 1 将每条轨迹的初始采样点和最终采样点都加入到候选重要地点集中,这是因为轨迹的起点和终点通常具有特殊意义。例如,用户每天出门时开启手机或关闭飞行模式,这时轨迹的起点往往是用户的住所。同理,用户晚上睡觉前关闭手机或打开飞行模式,便终止了一天的轨迹记录。

这样,将轨迹集中的各条轨迹按照上述算法进行划分,每条轨迹均被划分为若干条子轨迹,从中提取出的候选重要地点组成了候选重要地点集。

3　候选重要地点聚类

根据轨迹采样点的密度进行聚类,从而识别出重要地点,是研究者们常用的方法。但是,常用的密度聚类算法通常存在两个问题。1)算法复杂性高,如经典的密度聚类算法 DBSCAN,其算法复杂度为 $O(n^2)$,n 为轨迹采样点的个数,且该算法需要输入多个参数。高复杂度的聚类算法难以完成海量轨迹点的重要地点识别任务。2)全局唯一阈值。密度聚类算法通常需要预先设置好输入参数,从而根据采样点的密度状况来将其判定为类簇或噪声点。但是,移动用户的轨迹采样点存在着密度差异大的状况,例如,用户在住所附近时,轨迹采样点可能集中在很小的区域内,密度极高;而用户在逛商场时,其密度就小得多。此时,如何设置合理的密度阈值就成了最大的问题。

本文采取改进后的密度峰值聚类算法对候选重要地点进行聚类[18],从中识别出最终的重要地点。密度峰值聚类算法是 Alex 等[19]于 2014 年在 *Science* 上发表的一种新的密度聚类算法。与经典的密度聚类算法相比,该算法的原理简单,仅需要一个输入参数,且无须迭代,可对各种形状的点簇进行聚类分析,但需要经过决策图人工选取聚类中心,这不仅增加了该算法的冗余性,而且存在主观的隐患。文献[18]对密度峰值聚类算法进行了改进,使其能够自适应地确定截断距离与聚类中心。

假设某个移动用户共有 N 条历史轨迹,采用 TD-ADO (Trajectory Division Algorithm Based on Angle and Distance offset)算法提取出 $m_i(1\leqslant i\leqslant N)$ 个候选重要地点,将其按时间顺序连接,从而形成了由候选重要地点联结而成的简化轨迹。图 4 给出了用户的 5 条简化轨迹,对其中的候选重要地点进行聚类后,形成了 4 个类簇 C_1-C_4,未被划入圈中的候选重要地点则被判定为噪声点。这样,最终的重要地点就由所有圈中的采样点组成。

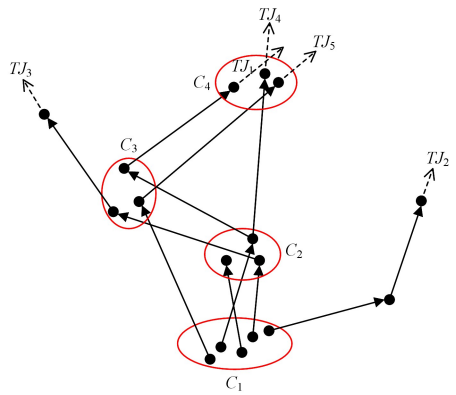

图 4　候选重要地点聚类图

Fig. 4　Diagram of candidate important location clusters

4　实验分析

4.1　实验数据集

首先用 Geolife 轨迹数据集[20-22]来验证所提方法的可行性,然后在 Foursquare 用户签到数据集[23]上分别计算了本文

方法与 Khetarpaul 提出的方法的重要地点识别准确性。数据集的基本属性如表 1 和表 2 所列，实验环境为 Windows10 64 位操作系统，Intel Core i7-6700HQ @2.60GHz CPU，8GB 内存，硬盘 1TB。

表 1 Geolife 轨迹数据集的基本属性

Table 1 Basic properties of Geolife trajectory dataset

数据集编号	采集时间	用户数量	轨迹数量
Data1	2007.04－2012.08	182	17624

表 2 Foursquare 签到数据集的基本属性

Table 2 Basic properties of Foursquare check-in dataset

数据集编号	采集时间	用户数量	签到数目
Data2	2010.04－2010.10	18107	2073740
Data3	2011.01－2011.12	11326	1385223

4.2 可行性分析

本实验选择了 Geolife 轨迹数据集目标用户 7 天内的位置数据来验证本文方法的重要地点识别性能。借助谷歌地图进行标注，初始点有 1293 个，通过轨迹划分筛选出候选重要地点后，再通过 ADPC(Adaptive Density Peaks Clustering)算法[18]聚类后的点有 854 个，聚类率为 66%，如图 5 和图 6 所示。

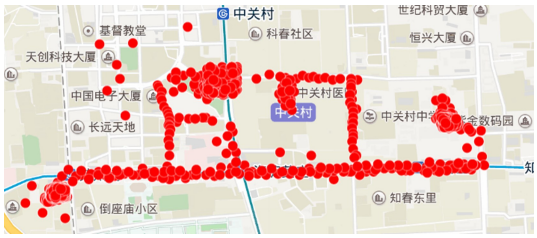

图 5 目标用户 7 天内的原始位置点

Fig. 5 Original location point of target user in seven days

图 6 经过轨迹划分及密度聚类的位置点

Fig. 6 Location points after trajectory division and density clustering

从图 6 中可以看出，本文方法将目标用户的重要地点分为 4 个类簇 R_1－R_4，各类簇的详细信息如表 3 所列，利用 GoogleMap 的 geocoder 服务进行逆向地理编码，解析得到重要地点对应的语义名称，从而完成重要地点的语义化。为了量化评估识别结果，计算各类簇质心点距该类簇对应语义地点的距离误差 。

表 3 各类簇的基本属性

Table 3 Basic properties of various clusters

类簇编号	位置点数量	地点语义	距离误差/m
R_1	307	中关村航天社区	35.7
R_2	234	航天精密大厦	63.5
R_3	181	中关村步行街	102.7
R_4	132	中关村医院	85.2

从表 3 中可以看出，各类簇质心点距对应语义地点的误差距离较小，且该距离与类簇内的位置点数目及类簇范围有关，位置点数目越多，范围越小，得出的重要地点识别结果就越精确。从图 5、图 6 及表 3 可以看出，采用轨迹划分对原始位置点进行预处理能够有效筛选出候选重要地点，改进的密度峰值聚类算法由于加入了参数及聚类中心的自适应特性，能够选择最合适的聚类中心，并将剩余点分配至各簇；逆向地理位置编码则实现了从坐标到语义地点的转换。综上，本文方法在移动用户重要地点识别中具有可行性。

4.3 准确性评估

本实验采用 Foursquare 用户签到数据集作为实验数据，签到地点 *ID* 作为评估准确性的标签，从识别准确率及 F-measure[24]两方面来评估本文方法及 Khetarpaul 提出的方法的准确性，实验结果如图 7 所示，横轴代表性能参数，纵轴代表两种方法在不同参数上的百分比。

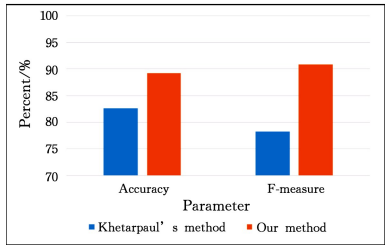

图 7 准确率与 F-measure 的对比

Fig. 7 Comparison of accuracy and F-measure

从图 7 中可以看出，无论是识别准确率还是 F-measure，本文方法均高于 Khetarpaul 提出的方法，这主要是由于 Khetarpaul 所提方法首先通过设置时间与距离阈值来滤除噪声点，然后设置密度阈值来完成重要地点的识别。这种预处理方法只能筛除掉一小部分存在明显异常的位置点，最终的识别结果还包含了许多无意义点，而且这种通过设置全局唯一密度阈值的识别方法对于那些密度分布不均匀的数据集的效果并不理想。相比之下，本文提出的 TD-ADO 算法更有针对性地筛选出了用户的特征位置点，且结合了自适应的密度聚类算法，对形状、分布不均匀的数据集具有更强的适应性，因此本文方法的准确性高于 Khetarpaul 提出的方法的准确性。

结束语 本文利用轨迹划分与聚类相结合的方法，尝试在移动用户位置数据中进行重要地点的识别，并在真实用户轨迹数据集与签到数据集上进行了验证分析。结果表明，本文所提方法能够有效识别出用户的重要地点，并结合谷歌逆向地理编码技术获取重要地点对应的语义信息，这些结果可以为用户个人意图探求、商业兴趣点推荐以及用户下一步位置追踪等领域提供信息保障。

参 考 文 献

[1] MENG X X. Research on Location based Mobile Information Service [D]. Changsha: University of Defense Technology, 2013. (in Chinese)
孟祥旭. 基于位置的移动信息服务技术与应用研究[D]. 长沙：学技术大学，2013.

[2] YUAN J, ZHENG Y, XIE X, et al. T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence[J]. IEEE Transac-

tions on Knowledge & Data Engineering,2013,25(1):220-232.

[3] ZHENG Y,XIE X,MA W Y. GeoLife:A Collaborative Social Networking Service among User,Location and Trajectory[J]. Bulletin of the Technical Committee on Data Engineering,2010,33(2):32-39.

[4] LU F,DUAN Y,ZHENG N. A Practical Route Guidance Approach Based on Historical and Real-time Traffic Effects[C]// International Conference on Geoinformatics. IEEE,2009:1-6.

[5] YUAN J,ZHENG Y,XIE X. Discovering Regions of Different Functions in A City Using Human Mobility and POIs[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2012:186-194.

[6] SONG C,BARABASI A L. Limits of Predictability in Human Mobility[J]. Science,2010,327(5968):1018-1021.

[7] ASHBROOK D,STARNER T. Using GPS to Learn Significant Locations and Predict Movement across Multiple Users[J]. Personal & Ubiquitous Computing,2003,7(5):275-286.

[8] YANG P,ZHU T,WAN X,et al. Identifying Significant Places Using Multi-day Call Detail Records[C]// 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE,2014:360-366.

[9] MONTOLIU R,BLOM J,GATICA-PEREZ D. Discovering Places of Interest in Everyday Life from Smartphone Data[J]. Multimedia Tools and Applications,2013,62(1):179-207.

[10] ALVARES L O,BOGORNY V,KUIJPERS B,et al. A Model for Enriching Trajectories with Semantic Geographical Information[C]// ACM International Symposium on Advances in Geographic Information Systems. ACM,2007:1-8.

[11] LIN K W,HSIEH M H,TSENG V S. A Novel Prediction-Based Strategy for Object Tracking in Sensor Networks by Mining Seamless Temporal Movement Patterns[J]. Expert Systems with Applications,2010,37(4):2799-2807.

[12] WANG Z,BULUT E,SZYMANSKI B K. Distributed Energy-Efficient Target Tracking with Binary Sensor Networks[J]. Acm Transactions on Sensor Networks,2010,6(4):1-32.

[13] KHALIL E A,ATTEA B A. Energy-Aware Evolutionary Routing Protocol for Dynamic Clustering of Wireless Sensor Networks[J]. Swarm & Evolutionary Computation,2011,1(4):195-203.

[14] HE W,LI D Y,AN L F,et al. Regular Route Mining Algorithm Based on GPS Trajectories [J]. Journal of Jilin University (Engineering and Technology Edition),2014,44(6):1764-1770. (in Chinese)
何雯,李德毅,安利峰,等. 基于 GPS 轨迹的规律路径挖掘算法[J]. 吉林大学学报(工学版),2014,44(6):1764-1770.

[15] KREVELD M V,LUO J. The Definition and Computation of Trajectory and Sub-trajectory Similarity[C]// ACM International Symposium on Geographic Information Systems,Acm-Gis 2007. Seattle,Washington,Usa,DBLP,2007:1-4.

[16] YUAN G,XIA S X,ZHANG L,et al. Trajectory Clustering Algorithm Based on Structural Similarity [J]. Journal on Communications,2011,32(9):103-110. (in Chinese)
袁冠,夏士雄,张磊,等. 基于结构相似度的轨迹聚类算法[J]. 通信学报,2011,32(9):103-110.

[17] KHETARPAUL S,CHAUHAN R,GUPTA S K,et al. Mining GPS Data to Determine Interesting Locations[C]// International Workshop on Information Integration on the Web:in Conjunction with WWW. ACM,2011:1-6.

[18] YANG Z,WANG H J,ZHOU Y. A Clustering Algorithm by Adaptive Cut-off Distance and Cluster Centers [J]. Data Analysis and Knowledge Discovery,2018,2(3):39-48. (in Chinese)
杨震,王红军,周宇. 一种截断距离和聚类中心自适应的聚类算法[J]. 数据分析与知识发现,2018,2(3):39-48.

[19] ALEX R,ALESSANDRO L. Clustering by fast search and find of density peaks[J]. Science,2014,344(6191):1492-1496.

[20] ZHENG Y,ZHANG L,XIE X,et al. Mining Interesting Locations and Travel Sequences from GPS Trajectories[C]// International Conference on World Wide Web. ACM,2009:791-800.

[21] ZHENG Y,LI Q,CHEN Y,et al. Understanding mobility based on GPS data[C]// International Conference on Ubiquitous Computing. ACM,2008:312-321.

[22] ZHENG Y,XIE X,MA W Y. GeoLife:A Collaborative Social Networking Service among User,Location and Trajectory[J]. Bulletin of the Technical Committee on Data Engineering,2011,33(2):32-39.

[23] YUAN Q,CONG G,MA Z,et al. Time-aware Point-of-Interest Recommendation[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM,2013:363-372.

[24] POWERS D M. Evaluation:From Precision,Recall and F-measure to ROC,Informedness,Markedness and Correlation[J]. Journal of Machine Learning,2011,1(2):37-63.