

# 手机基站定位数据可视分析

李海生<sup>1,2</sup>, 黄媛洁<sup>1,2</sup>, 宋璇<sup>1,2</sup>, 杜军平<sup>3</sup>, 陈国润<sup>4</sup>, 丁富强<sup>4</sup>

1. 北京工商大学计算机与信息工程学院, 北京 100048; 2. 食品安全大数据技术北京市重点实验室, 北京 100048;  
3. 北京邮电大学计算机学院, 北京 100876; 4. 上海理想信息产业(集团)有限公司, 上海 201315

## 摘要

手机基站定位数据蕴含丰富的时空信息。设计了一种基于电信每次呼叫测量数据的可视分析方法, 给出了基于每次呼叫测量数据的基站定位方法, 提取手机用户的出行轨迹, 对凝聚层次聚类算法进行改进, 可以高效率地对轨迹进行聚类, 利用流向图和热力图表示轨迹聚类结果和手机用户在某时刻的整体分布。将该方法应用于上海电信手机基站定位数据, 取得了良好的效果。

## 关键词

手机基站定位数据; 轨迹聚类; 流向图; 可视分析

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2017008

## *Visual analyses of mobile phone base station location data*

LI Haisheng<sup>1,2</sup>, HUANG Yuanjie<sup>1,2</sup>, SONG Xuan<sup>1,2</sup>, DU Junping<sup>3</sup>,  
CHEN Guorun<sup>4</sup>, DING Fuqiang<sup>4</sup>

1. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China

3. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

4. Shanghai Ideal Information Industry(Group) Co., Ltd., Shanghai 201315, China

## *Abstract*

Mobile phone base station location data contains rich spatio-temporal information. A visual analytic method based on per call measurement data (PCMD) was designed. A base station positioning method based on PCMD to extract mobile phone users' trajectories was designed. An agglomerate hierarchical clustering algorithm was improved to cluster trajectories efficiently. Flow map and heat map were employed to present the clustering result and the overall distribution of mobile phone users at a specific time. This method was applied to the dataset of Shanghai mobile phone base station data, which has achieved satisfied results.

## *Key words*

mobile phone base station location data, trajectory clustering, flow map, visual analyses

## 1 引言

随着手机等移动终端的普及,在城市中2G/3G/4G网络已经基本实现全区域覆盖。根据国家工业和信息化部统计,截至2015年,移动电话用户已达到13亿户,移动电话用户普及率达95.5部/百人,人们开始更加关注如何利用从移动通信网络中获取的数据进行可视化研究。其中,手机定位数据作为移动通信网络数据中的一类,在分析人群移动模式<sup>[1,2]</sup>、城市功能区识别<sup>[3]</sup>以及交通网络规划<sup>[4]</sup>中都提供了很大的帮助。

通常,手机等移动终端收集到的定位数据可以来自移动通信网络、Wi-Fi接入点位置信息、移动终端的GPS定位信息等,记录了移动对象的位置、时间、速度和方向等行为特征。GPS定位数据最为精确,多由志愿者提供<sup>[5]</sup>,因此样本数量很少并且难以获取。Wi-Fi接入点数据也较为精确,但多用于室内定位<sup>[6]</sup>。移动通信网络能够定期或不定期地主动或被动地记录手机用户时间序列的基站编号,该种定位方式精确度低,数据粒度不均匀,往往需要配合其他类型数据来分析<sup>[7]</sup>,但在样本量、覆盖范围以及实施成本和周期上更具有优势<sup>[4]</sup>。本文使用的手机基站定位数据即每次呼叫测量数据(per call measurement data, PCMD)是上海电信系统用来记录每个呼叫的相关信息的数据,主要包括主叫通话、基站扇区和信号质量等信息数据。

在对轨迹进行可视化时,传统的可视化方法直接将轨迹数据一一绘制在地图上,由于相互遮挡等原因,不适用于大量数据的可视化<sup>[8]</sup>。采用聚集可视化的方法,将对象个体数据转换为聚集值,能够观察移

动对象的群体特征,同时也能减少刻画群体特征的数据量<sup>[9]</sup>。

本文设计了一种基于电信PCMD的人群流动可视分析方法。首先,对PCMD进行处理,提取用户的出行数据以得到用户的轨迹。然后根据用户选择的时间段和区域,使用轨迹层次聚类算法对用户出行轨迹进行聚类。最后,将聚类结果映射到地图中,使用基于流向图的多地图缩放级别的层次可视化方法进行可视分析。

## 2 相关工作

### 2.1 基于基站的手机定位数据可视分析

手机定位数据被广泛用于发现人群的移动模式,Zhang Y<sup>[1]</sup>使用用户上网时产生的蜂窝数据信息进行人群移动模式建模,并且能够预测出某个特定用户在给定位置可能用到的应用软件。Xiong H等人<sup>[2]</sup>发现特定的某一类人的位置信息有很强的关联性和相关性,并提出基于集体行为模式(collective behavioral patterns, CPB)的方法来预测人的轨迹,这种方法能够很好地预测某个人接下来6 h之内的位置。

除了移动模式,手机定位数据还可以用于发现人类生活中重要的位置信息,比如居住地点和工作地点等。Isaacman S等人<sup>[3]</sup>提出一种基于聚集和回归的方法,分析蜂窝网络数据,发现有意义的位置信息,计算出通勤距离,并且通过几十万匿名用户的碳排放量分析,证明了该算法可以作为有效的政策和基础设施研究的支撑。

对手机定位数据的挖掘和分析可以帮助调整交通政策以及基础设施的建设,使得城市的居民能获得更好的出行体验。冉斌<sup>[4]</sup>提出了手机数据在交通调查以及交通

规划中的应用,通过手机话单定位数据和手机信令定位数据进行去噪、扩样等预处理,最终能够获得居民出行特征数据。根据这些特征数据,可以分析人口就业分布、通勤出行特征,还可以进一步分析城市人口的时空动态分布等。

## 2.2 基于流向图的时空轨迹数据可视分析

当轨迹数据量非常大时,在地图上显示轨迹会出现严重的视觉混乱和不清晰的问题。一种解决方法是使用边捆绑技术<sup>[10]</sup>,通过弯曲边使相似的边相互靠近形成一束,以减少相互遮挡。

Guo D等人<sup>[11]</sup>提出了一种可以从大量流数据中提取主要流模式的方法,通过一个基于向量密度的模型为每一对位置估计流密度,然后选择光滑路径的子集在流向图中表示主要的流,但是这种方法的计算复杂度非常高。

Andrienko N等人<sup>[12]</sup>提出了一种对移动数据进行空间泛化和聚集的方法,将数据覆盖的版图划分成适当的小区域。Von L T等人<sup>[13]</sup>使用了上述划分版图的方法,先对区域进行了划分,然后对移动数据线进行空间上的聚类,再进行时间上的聚类,用于展示长时间段的移动数据的时空变化情况。

## 3 可视分析算法描述

本文设计了一种基于PCMD的基站定位方法得到用户的出行轨迹,然后计算轨迹间的相似度,接下来采用改进的层次聚类算法对所有轨迹进行聚类,最后对聚类结果进行可视分析,算法技术框架如图1所示。

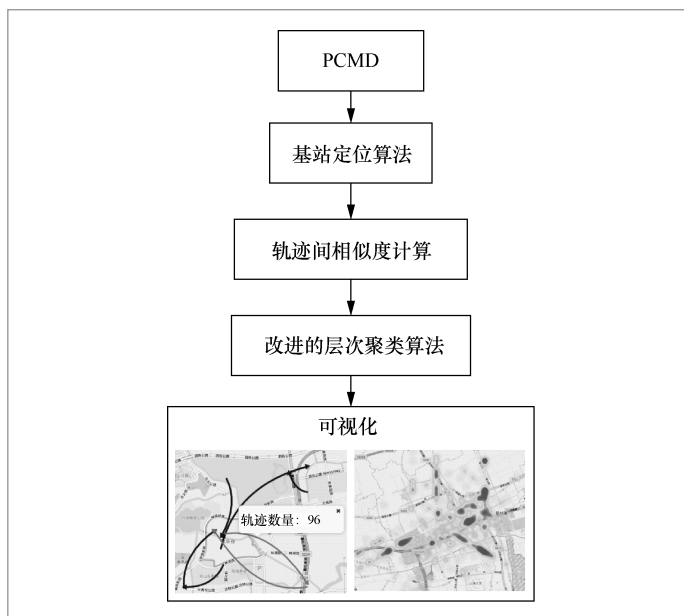


图1 算法技术框架

### 3.1 基于PCMD的基站定位方法

一条PCMD中包含两个关键时间信息,分别为初始时刻和终止时刻的时间戳,这反映了手机接入和断开网络的时间。PCMD每次获取一组信息,其中与定位相关的信息有基站号、扇区号、时延、电磁辐射场强等。定位的关键信息是场强和时延。但是场强更容易受到环境、建筑、天气、电网、屏蔽体、设备等的影响,在城市内尤甚,定位的准确度难以保证。与场强相比,时延所受的干扰更少,所以这里使用时延信息进行定位。每组信息可以由一个或多个基站产生,这些基站分为参考基站和非参考基站。本文设计了以下3种方法进行定位。

#### (1) 单基站定位

如果一条PCMD中仅包含1个基站的数据,则只能使用单个基站进行定位。由于1个基站有3个扇区,有时电波到达这3个扇区的时间是不相同的,这种情况是由于多径效应造成的。当发生这种情况时,取时延最短的扇区对应的弧,由于没有其他补充

信息,无法将用户定位到弧的具体点上,因此取弧的中点作为用户的期望位置,如图2(a)所示;当两个扇区的时延相同时,不能判定用户在哪段弧上,这时以两弧的临界点作为期望位置,如图2(b)所示;当3个扇区的时延相同时,用户可能位于一个

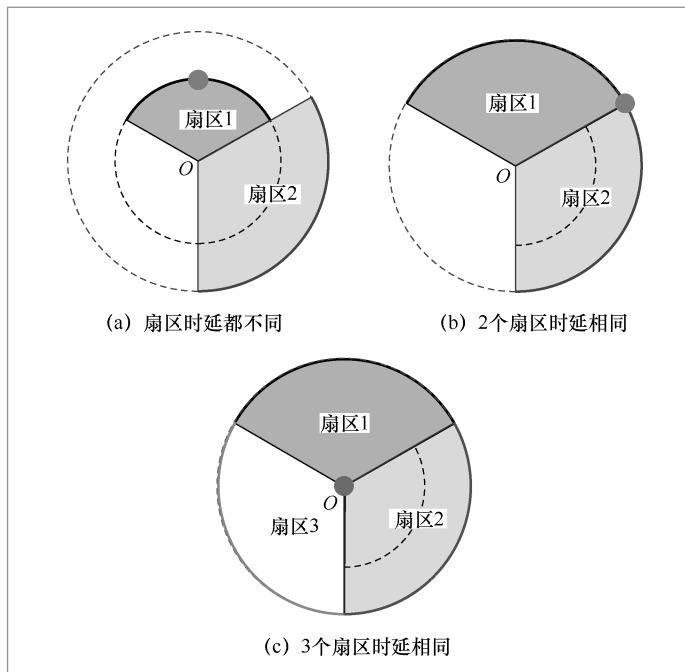


图2 单基站定位

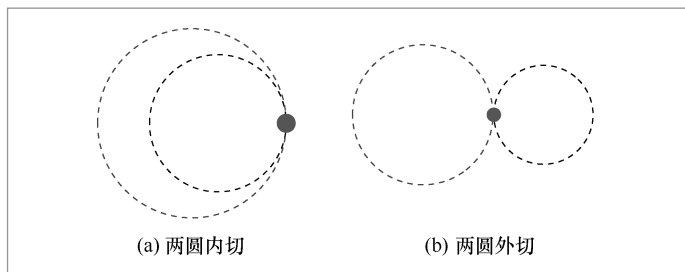


图3 找到符合时延条件的一个点

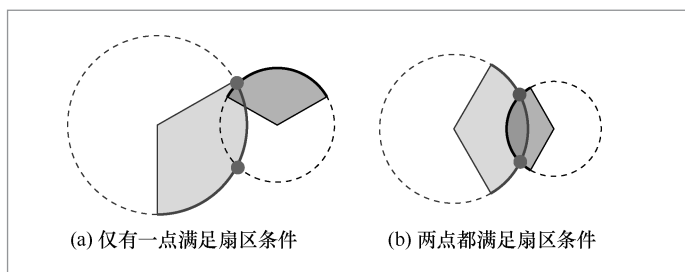


图4 找到符合时延条件的两个点

圆周的任意位置,这时以基站的位置作为期望位置,如图2(c)所示。

### (2) 两点定位

当一条PCMD中包含2个基站的数据时,使用两点定位方法。与单基站定位的情形相似,用户到达某个基站不同扇区的时延可能不完全相同,为减少多径效应的影响,仍使用到达各个基站的最小时延作为计算依据。当找到符合时延条件的一个点时,该点作为用户的期望位置,如图3所示。图4表示找到符合时延条件的两个点的情况,如果两点中只有一点满足扇区条件,如图4(a)所示,则取该点为用户的期望位置;如果两点都满足或不满足扇区条件,如图4(b)所示,则取与两个基站有效扇区正方向的总误差更小的点的位置。如果找不到符合时延条件的点,如图5所示,则取时延总误差最小的点作为用户的期望位置。

### (3) 3点及多点定位方法

当一条PCMD中包含3个或更多基站的数据,则可以进行较为准确的定位。基站越多,定位精度越高。本文使用到达时间差(time difference of arrival, TDOA)/到达角度测距(angle of arrival, AOA)混合定位算法<sup>[14]</sup>。

单基站定位方法不可能定位到准确位置。一条PCMD包括两个时刻的信息,因此对一条PCMD中两个时刻的信息交叉使用,某些情况下可以提高定位的准确度。当两个时刻的信息来自于同一基站时,定位的两点位于以基站为圆心的两个同心圆弧上,如图6所示。将这两个圆弧的中心连线的中点作为在这个时段内用户位置的估算;当两个时刻的信息来自于不同基站时,使用前面叙述的两点定位方法对用户的位置进行估算,并选择其中一点作为这个时段内位置的估算。

通过上述基站定位方法,可以得到每

条PCMD对应的用户的位置和时间信息。然后将一天的时间划分为长度相等的时间片段,得到每个用户在每个时间段对应的起始位置和结束位置。时间段的长度基于PCMD的获取频率和用户的需求来选择。由于空间数据具有空间位置、非结构化、空间关系、分类编码、海量数据等特征<sup>[15]</sup>,为了有效地进行空间查询,使用PostgreSQL数据库中的PostGIS<sup>①</sup>来存储数据。将用户的出行数据按照每天进行分区,保证数据的访问效率。

### 3.2 轨迹间相似性度量方法

本文使用Lee J G等人<sup>[16]</sup>提出的轨迹间的相似性度量方法,该距离是3种距离的加权和表示,分别是其垂直距离 $d_{\perp}$ 、平行距离 $d_{\parallel}$ 和角度距离 $d_{\theta}$ 。给3种距离赋予相同的权重,即轨迹间的距离 $d=d_{\perp}+d_{\parallel}+d_{\theta}$ 。轨迹 $L_i$ 和 $L_j$ 间的3种距离如图7所示,其中, $s_i$ 、 $s_j$ 、 $e_i$ 、 $e_j$ 分别表示轨迹 $L_i$ 和 $L_j$ 的起点和终点; $p_s$ 和 $p_e$ 分别表示 $s_j$ 和 $e_j$ 在轨迹 $L_i$ 上的投影; $l_{\perp 1}$ 、 $l_{\perp 2}$ 、 $l_{\parallel 1}$ 、 $l_{\parallel 2}$ 则分别表示图7中对应端点间的欧氏距离,  $\|L_j\|$ 表示轨迹 $L_j$ 的长度; $\theta$ 表示两条子轨迹的夹角( $0^{\circ} \leq \theta \leq 180^{\circ}$ )。

$$d_{\perp} = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\parallel 1} + l_{\parallel 2}} \quad (1)$$

$$d_{\parallel} = \min(l_{\parallel 1}, l_{\parallel 2}) \quad (2)$$

$$d_{\theta} = \|L_j\| \times \sin \theta \quad (3)$$

### 3.3 改进的层次聚类算法

给定时间段 $[t_o, t_d]$ ,定义手机用户 $i$ 在该时间段的轨迹为 $T_i = \{O_i, D_i\}$ ,其中 $t_o$ 为起始时刻, $t_d$ 为结束时刻, $O_i$ 为用户 $i$ 在该时间段有最早记录的位置, $D_i$ 为用户 $i$ 在该时间段有最晚记录的位置。定义 $T = \{T_i\}$ 为在给定时间段下,所有捕获到的手机用户轨迹的集合。定义 $O = \{O_i\}$ 为所有在 $T$ 中用户轨迹的

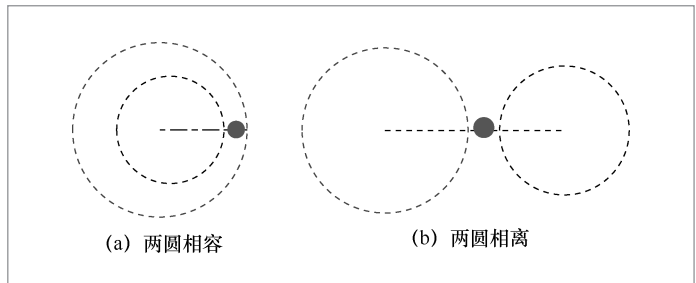


图5 找不到符合延时条件的点

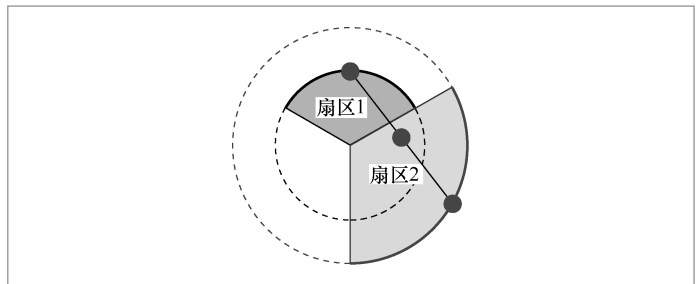


图6 同基站整合

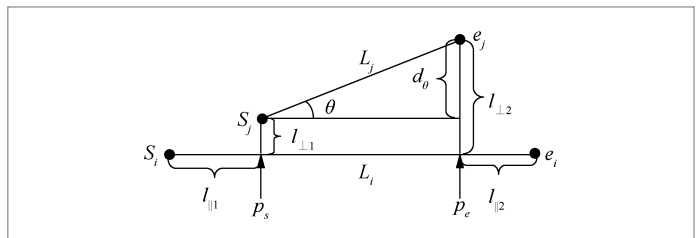


图7 轨迹间的3种距离

起始位置的集合,  $D = \{D_i\}$ 为所有在 $T$ 中用户轨迹的结束位置的集合。

定义 $kNN(O_i, k)$ 为属于集合 $O$ 并且距离点 $O_i$ 最近的 $k$ 个点。同理, $kNN(D_i, k)$ 为属于集合 $D$ 并且距离点 $D_i$ 最近的 $k$ 个点。

**定义1** 轨迹的 $kNN$ 邻近轨迹。一条轨迹 $T_p$ 的 $kNN$ 邻近轨迹 $FN(T_p, k) = \{T_q \in T \mid O_q \in kNN(O_p, k) \wedge D_q \in kNN(D_p, k)\}$ ,其中 $O_p$ 、 $D_p$ 分别是轨迹 $T_p$ 的起始位置和结束位置, $O_q$ 、 $D_q$ 分别是轨迹 $T_q$ 的起始位置和结束位置。

计算所有轨迹间的距离会十分耗时并且效率低,因此,只计算给定时间段下的每条轨迹和它的 $kNN$ 邻近轨迹的距离。为了能够快速找到每条轨迹的起始位置的

① <http://www.postgis.org/>

$k$ NN邻近点和结束位置的 $k$ NN邻近点,对所有起始位置 $O$ 和所有结束位置 $D$ 分别建立 $k-d$ 树。 $k-d$ 树是一种分割 $k$ 维数据空间的数据结构,主要应用于多维空间关键数据的搜索,如范围搜索和最近邻搜索。在本文中,位置信息为经纬度坐标,因此为二维空间, $k$ 为2。

层次聚类算法需要一个类间最大距离阈值来判断两个聚类是否合并。在判断聚类 $C_x$ 和 $C_y$ 是否合并时,使用基于共享近邻(shared nearest neighbor, SNN)的个数的方法计算 $SNN(C_x, C_y)$ <sup>[17]</sup>。与第3.2节提出的轨迹间距离计算方法不同, $SNN(C_x, C_y)$ 只用于判断两个聚类是否合并。改进的凝聚层次聚类算法步骤如下。

**算法1** 凝聚轨迹聚类算法。

**输入:** 指定时间段的轨迹数据集 $T=\{T_i|1\leq i\leq n\}$ ,计算距离时邻近轨迹的个数 $k$ 。

**输出:** 聚类结果 $C=\{C_m|1<m\leq n\}$ 。

**步骤1** 为 $T$ 的所有起始位置 $O$ 和所有结束位置 $D$ 分别建立 $k-d$ 树,并得到每条轨迹的 $k$ NN邻近轨迹。

**步骤2** 按照第3.2节计算距离的方法计算每条轨迹和它的 $k$ NN邻近轨迹之间的距离,并根据距离升序排列。

**步骤3** 将每一条轨迹初始化为一个聚类。

**步骤4** 对按距离排序过后的每一个轨

迹和它的邻近轨迹 $(p, q)$ 。首先找到 $p$ 和 $q$ 分别所在的聚类 $C_x, C_y$ ,然后计算 $C_x$ 和 $C_y$ 之间的距离,如果 $x\neq y$ ,并且 $SNN(C_x, C_y)<1$ ,则 $C_x=C_x\cup C_y, C=C-C_y$ 。

在计算两个聚类 $C_x$ 和 $C_y$ 之间的距离时,按照平均连接(average-linkage)算法聚类法,应该计算 $C_x$ 和 $C_y$ 的轨迹之间的平均距离,但是这样十分耗时。因此,使用近似但是效率高的方法计算聚类 $C_x$ 和 $C_y$ 之间的距离,计算过程如图8所示,计算步骤如下。

**算法2** 类间距离计算算法。

**输入:** 聚类 $C_x$ 和聚类 $C_y$ 。

**输出:** 聚类 $C_x$ 和聚类 $C_y$ 之间的距离。

**步骤1** 分别计算聚类 $C_x$ 和 $C_y$ 的起始位置的质心 $O_{cx}$ 和 $O_{cy}$ 以及结束位置的质心 $D_{cx}$ 和 $D_{cy}$ 。

**步骤2** 从起始位置集 $O$ 中找到最接近 $O_{cx}$ 和 $O_{cy}$ 的点 $O'_{cx}$ 和 $O'_{cy}$ ,从结束位置集 $D$ 中找到最接近 $D_{cx}$ 和 $D_{cy}$ 的点 $D'_{cx}$ 和 $D'_{cy}$ 。

**步骤3** 生成两个中间轨迹 $\langle O'_{cx}, D'_{cx} \rangle$ 和 $\langle O'_{cy}, D'_{cy} \rangle$ 表示聚类 $C_x$ 和 $C_y$ 。

**步骤4** 使用 $SNN(C'_{cx}, C'_{cy})$ 计算轨迹 $\langle O'_{cx}, D'_{cx} \rangle$ 和 $\langle O'_{cy}, D'_{cy} \rangle$ 之间的距离,用来近似表示聚类 $C_x$ 和 $C_y$ 之间的距离。

### 3.4 轨迹可视化

通过上述轨迹聚类算法对用户给定时间段下的手机用户轨迹进行聚类,得到了一组聚类结果。每个类用中间轨迹来代替该类,使用流向图的方法将每个类的代表轨迹画在地图中,如图9所示,显示至少包含70条轨迹以上的类。其中,原始数据为上海电信手机用户在顾村公园和欢乐谷两个区域某天全天的24 278条轨迹数据,如图9(a)所示。设置 $k=150$ ,使用聚类算法聚成了2 917个类,最大的类包含了355条轨迹。其中90%以上的轨迹可以至少找到

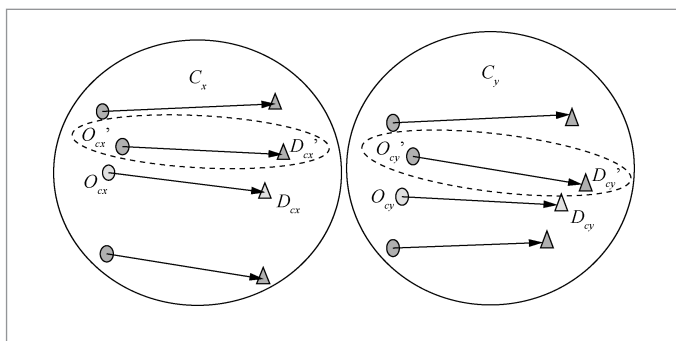


图8 计算聚类 $C_x$ 和 $C_y$ 距离示意

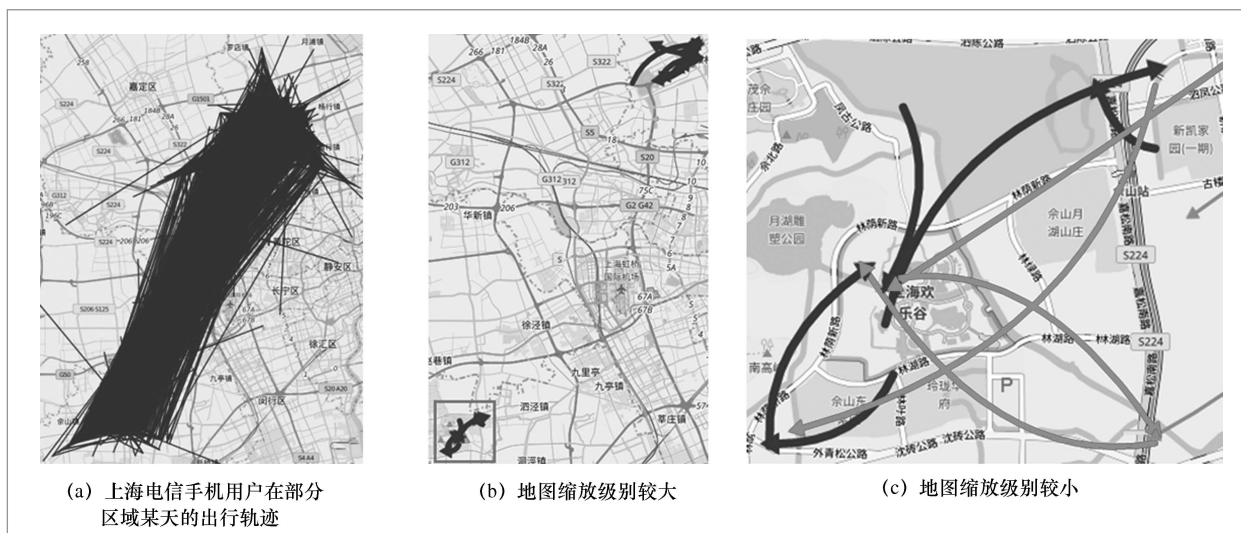


图9 轨迹聚类结果可视化

一条邻近轨迹,每个轨迹平均有7条邻近轨迹。有1 321条轨迹无法找到任何邻近轨迹,会自己形成一个类,在轨迹可视化时会去除这些单独的类。

本文设计了一种多地图缩放级别的层次可视化方法,根据地图的缩放级别,显示不同聚类大小的轨迹。当地图缩放级别较小时,只显示包含轨迹数量较大的类,如图9(b)所示。当扩大地图缩放级别时,增加显示其他包含轨迹数量较小的类,如图9(c)所示,该图表示的区域与图9(b)方框所示区域相同。其中,颜色越深的线表示包含轨迹数量越多的类;反之,颜色越浅的线表示包含轨迹数量越少的类。

使用热力图的方法表示用户选择的时间段的结束时刻的手机用户分布情况,如图10所示,图10为14:00—14:05用户的移动轨迹和用户在14:05时所在位置的热力图。热力图可以显示大规模个体的整体状况,颜色越深表示数目越大。

### 3.5 参数选择与算法对比

在轨迹聚类时,若参数 $k$ 设置过小,结果会产生许多很小的类;反之若 $k$ 设置过

大,结果则会产生较大的类,并且计算量也会非常大。给定一个合适的类簇指标,只要假设的类簇的数目等于或者高于真实的类簇的数目时,该指标上升会很缓慢,而一旦试图得到少于真实数目的类簇时,该指标会急剧上升。本文类簇指标选择类簇的轨迹数量加权平均值,图11表示选择不同 $k$ 值对应的类簇的轨迹数量加权平均值。可以看到,当 $k$ 值取150左右时,类簇指标的上升趋势开始加快,通过蚁群优化算法可以自动得到最优 $k$ 值。

图12(a)为 $k=100$ 的聚类结果,图12(b)为 $k=200$ 的聚类结果, $k=150$ 的结果在



图10 热力图

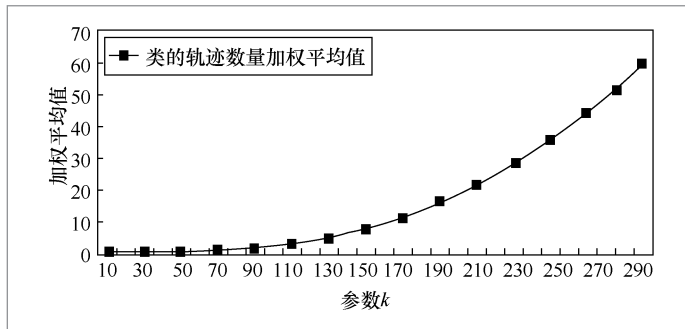
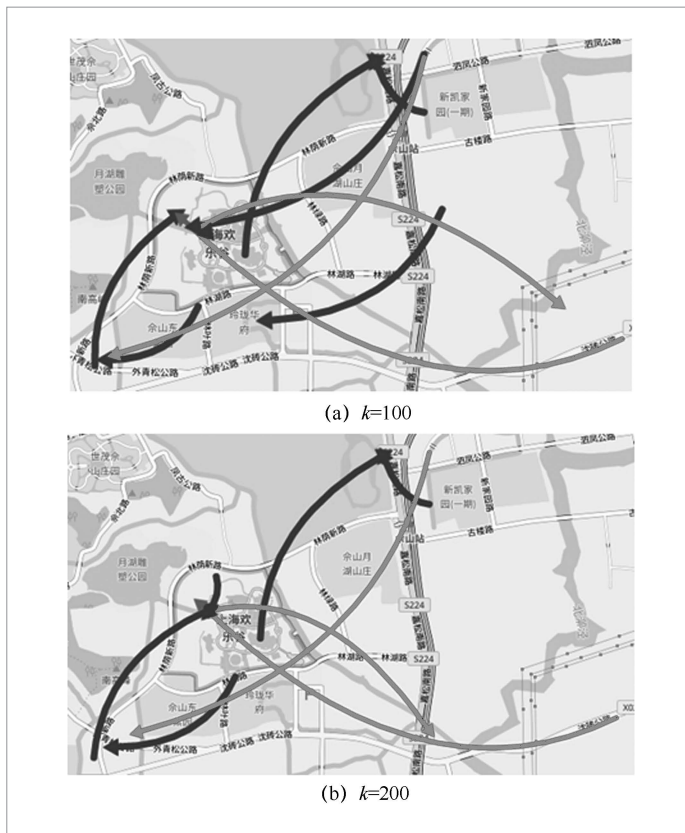
图 11 不同  $k$  值对应的类簇的轨迹数量加权平均值图 12 不同  $k$  值的聚类结果

表 1 聚类算法运行时间对比

轨迹数量/条	AGNES/ms	CURE/ms	本文算法/ms
500	180	529	526
1 500	4 244	3 789	3 800
3 000	31 416	12 524	4 856
5 000	207 431	34 751	6 348
10 000	1 441 618	143 220	11 956

图 9 (c) 中。对比这 3 张地图可以发现, 尽管最大的类包含的轨迹数量不同、显示的聚类结果有些细微的不同, 但是总体的模式是相似的。结果表明  $k$  值的设定对聚类结果的影响和整体的分析不是十分敏感, 当需要看整体的流动情况时, 用户可以选择较大的  $k$ ; 当需要看局部区域的流动情况时, 用户可以选择较小的  $k$ 。

为了验证改进算法的效率, 本文分别使用传统的凝聚层次聚类算法 (agglomerative nesting, AGNES)、使用代表点的层次聚类算法 (clustering using representatives, CURE)<sup>[9]</sup> 以及本文改进的凝聚层次聚类算法对不同条数的轨迹进行聚类, 结果见表 1 和图 13。实验结果表明, 当轨迹数量较少时, AGNES 聚类算法效率比较高, CURE 和本文改进的聚类算法效率相对较低; 当轨迹数量较多时, CURE 聚类算法的效率略好于 AGNES 聚类算法, 但相比之下本文改进的聚类算法效率最高, 并且运行时间呈线性增长。

## 4 结束语

本文设计了一个基于大规模 PCMD 的可视分析方法, 使用基于 PCMD 的基站定位方法得到手机用户的出行数据, 对用户的出行轨迹进行聚类, 将结果呈现在可视分析系统中。用户可以从时间和空间上对手机用户进行分析, 发现其中隐含的规律。流向图因箭头本身的指向性让分析人员可以容易地判断出手机用户整体的移动方向, 线条颜色的深浅可以清楚地表达流量的大小。热力图可以清晰地表示手机用户在某时刻整体的分布情况。本文提出的轨迹聚类算法适用于大规模数据, 效率高, 可以将本文算法应用到实时在线数据分析中, 下一步将围绕轨迹聚类算法结果优劣

的评价方面展开进一步的工作。

## 参考文献:

- [1] ZHANG Y. User mobility from the view of cellular data networks [C]//IEEE INFOCOM 2014-IEEE Conference on Computer Communications, April 27-May 2, 2014, Toronto, Canada. Piscataway: IEEE Computer Society Press, 2014: 1348-1356.
- [2] XIONG H, ZHANG D, ZHANG D, et al. Predicting mobile phone user locations by exploiting collective behavioral patterns[C]//The 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, September 4-7, 2012, Fukuoka, Japan. [S.l.:s.n.], 2012: 164-171.
- [3] ISAACMAN S, BECKER R, CÁCERES R, et al. Identifying important places in people's lives from cellular network data[C]//The 9th International Conference on Pervasive Computing, June 12-15, 2011, San Francisco, USA. Berlin: Springer Press, 2011: 133-151.
- [4] 冉斌. 手机数据在交通调查和交通规划中的应用[J]. 城市交通, 2013, 11(1): 72-81.  
RAN B. Use of cellphone data in travel survey and transportation planning[J]. Urban Transport of China, 2013, 11(1): 72-81.
- [5] ZHENG Y, CHEN Y, XIE X, et al. GeoLife2.0: a location-based social networking service[C]//The 2009 10th International Conference on Mobile Data Management: Systems, Services and Middleware, May 18-20, 2009, Taipei, China. New Jersey: IEEE Press, 2009: 357-358.
- [6] 王睿, 赵方, 彭金华, 等. 基于Wi-Fi和蓝牙融合的室内定位算法[J]. 计算机研究与发展, 2011, 48(S2): 28-33.

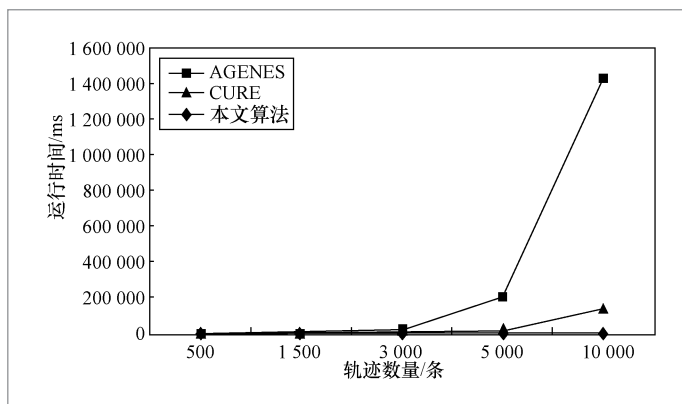


图 13 聚类算法运行时间对比

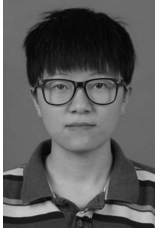
- [7] 汪飞, 张繁, 吴斐然, 等. 面向多源城市出行数据的可视化查询模型[J]. 计算机辅助设计与图形学学报, 2016, 28(1): 25-31.  
WANG F, ZHANG F, WU F R, et al. A visual query model for multi-source urban mobility data[J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(1): 25-31.
- [8] 王祖超, 袁晓如. 轨迹数据可视分析研究[J]. 计算机辅助设计与图形学学报, 2015, 27(1): 9-25.  
WANG Z C, YUAN X R. Visual analysis of trajectory data[J]. Journal of Computer-Aided Design and Computer Graphics, 2015, 27(1): 9-25.
- [9] 廖帅, 陈萃, 熊伟, 等. 面向位置服务的大规模移动对象时空聚集分析原型系统[J]. 计算机研究与发展, 2015(S1): 107-111.  
LIAO S, CHEN L, XIONG W, et al. A LBS-oriented massive moving objects spatiotemporal aggregation analysis prototype system[J]. Journal of Computer Research and Development, 2015(S1): 107-111.
- [10] ZHOU H, XU P, YUAN X, et al. Edge

- bundling in information visualization[J]. Tsinghua Science and Technology, 2013, 18(2): 145-156.
- [11] GUO D, ZHU X. Origin-Destination flow data smoothing and mapping[J]. IEEE Trans on Visualization & Computer Graphics, 2014, 20(12): 2043-2052.
- [12] ANDRIENKO N, ANDRIENKO G. Spatial generalization and aggregation of massive movement data[J]. IEEE Trans on Visualization & Computer Graphics, 2011, 17(2): 205-219.
- [13] VON L T, BRODKORB F, ROSKOSCH P, et al. MobilityGraphs: visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering[J]. IEEE Trans on Visualization & Computer Graphics, 2016, 22(1): 11-20.
- [14] 邓平, 李莉, 范平志. 一种TDOA/AOA混合定位算法及其性能分析[J]. 电波科学学报, 2002, 17(6): 633-636.  
DENG P, LI L, FAN P Z. A hybrid TDOA/AOA location algorithm and its performance analysis[J]. Chinese Journal of Radio Science, 2002, 17(6): 633-636.
- [15] 张大鹏, 张锦, 郭敏泰, 等. 开源WebGIS软件应用开发技术和方法研究[J]. 测绘科学, 2011, 36(5): 193-196.  
ZHANG D P, ZHANG J, GUO M T. Application development technology and methods of open-source WebGIS software[J]. Science of Surveying & Mapping, 2011, 36(5): 193-196.
- [16] LEE J G, HAN J, WHANG K Y. Trajectory clustering: a partition-and-group framework[C]//The 2007 ACM SIGMOD International Conference on Management of Data, June 11-14, 2007, Beijing, China. New York : ACM Press, 2007: 593-604.
- [17] ZHU X, GUO D. Mapping large spatial flow data with hierarchical clustering[J]. Transactions in Gis, 2014, 18(3): 421-435.
- [18] RAJARAMAN A, ULLMAN J D. Mining of massive datasets[M]. Cambridge: Cambridge University Press, 2012.
- [19] MIN Y, LI Y. Vehicles recognition based on the size characteristics and the CURE clustering algorithm[C]// The 2015 IEEE International Conference on Signal Processing, Communications and Computing, September 19-22, 2015, Ningbo, China. New Jersey: IEEE Press, 2015: 7-11.

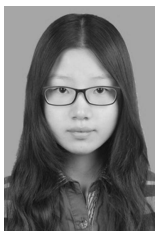
## 作者简介



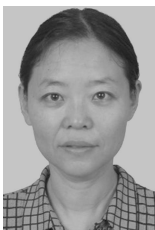
李海生(1974-),男,博士,食品安全大数据技术北京市重点实验室、北京工商大学计算机与信息工程学院教授、研究生导师,主要研究方向为数据可视化与可视分析、计算机图形学、科学可视化、三维模型检索等。



黄媛洁(1992-),女,食品安全大数据技术北京市重点实验室、北京工商大学计算机与信息工程学院硕士生,主要研究方向为数据可视化与可视分析。



宋璇(1993-),女,食品安全大数据技术北京市重点实验室、北京工商大学计算机与信息工程学院硕士生,主要研究方向为数据可视化与可视分析。



杜军平(1963-),女,博士,北京邮电大学计算机学院教授、博士生导师,主要研究方向为运动图像处理、视觉与信息获取等。



陈国润(1978-),男,上海理想信息产业(集团)有限公司电信支撑软件部研发部经理,主要研究方向为云计算管理、分布式数据处理。



丁富强(1973-),男,博士,上海理想信息产业(集团)有限公司研发中心主任,主要研究方向为电信大数据应用研究。

收稿日期: 2016-11-25

基金项目: 国家自然科学基金资助项目(No.61532006); 北京市自然科学基金资助项目(No.4162019); 北京市科技计划课题资助项目(No. Z161100001616004)

Foundation Items: The National Natural Science Foundation of China (No.61532006), The Beijing Municipal Natural Science Foundation (No.4162019), Beijing Science and Technology Project (No. Z161100001616004)