



## (12) 发明专利申请

(10) 申请公布号 CN 104657424 A

(43) 申请公布日 2015. 05. 27

(21) 申请号 201510028409. 9

(22) 申请日 2015. 01. 21

(71) 申请人 段炼

地址 530001 广西壮族自治区南宁市西乡塘  
区明秀东路 175 号

申请人 胡宝清

(72) 发明人 段炼 胡宝清 李峙 闫妍

(51) Int. Cl.

G06F 17/30(2006. 01)

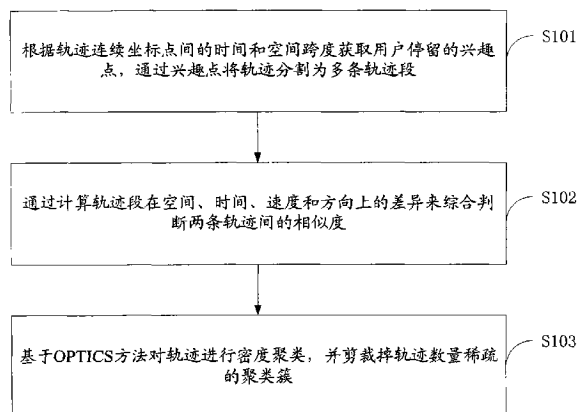
权利要求书2页 说明书7页 附图2页

### (54) 发明名称

一种多时空特征融合下的兴趣点轨迹聚类方法

### (57) 摘要

本发明公开了一种多时空特征融合下的兴趣点轨迹聚类方法,首先根据轨迹连续坐标点间的时间和空间跨度获取用户停留的兴趣点,通过兴趣点将轨迹分割为多条轨迹段;然后通过计算轨迹段在空间、时间、速度和方向上的差异来综合判断两条轨迹间的相似度;最后基于 OPTICS 方法对轨迹进行密度聚类,并剪裁掉轨迹数量稀疏的聚类簇。本发明将轨迹转换为兴趣点序列,综合考虑连续兴趣点间形成的轨迹段的速度、方向和时空特性,进行轨迹聚类,挑选出显著簇以获取能反映全局重要性的轨迹聚类形态。实验结果表明,该方法在聚类中保留了轨迹原有的时空和移动属性特性,能更全面地反映移动对象的运动和行为模式,聚类准确性高。



1. 一种多时空特征融合下的兴趣点轨迹聚类方法,其特征在于,该多时空特征融合下的兴趣点轨迹聚类方法利用轨迹的运动信息和兴趣点停留信息,更加准确表达具有社会行为相似性的轨迹集合;

具体包括以下步骤:

步骤一、根据轨迹连续坐标点间的时间和空间跨度获取用户停留的兴趣点,通过兴趣点将轨迹分割为多条轨迹段;

步骤二、通过计算轨迹段在空间、时间、速度和方向上的差异来综合判断两条轨迹间的相似度;

步骤三、基于 OPTICS 方法对轨迹进行密度聚类,并剪裁掉轨迹数量稀疏的聚类簇。

2. 如权利要求 1 所述的多时空特征融合下的兴趣点轨迹聚类方法,其特征在于,充分利用轨迹的运动信息和兴趣点停留信息,表达具有社会行为相似性的轨迹集合,计算轨迹段时空相似度的具体方法为:

步骤一、基于空间四叉树索引,以轨迹 S 中各轨迹段为查询对象,获取轨迹 R 中与 S 各轨迹段空间距离最接近的相应的轨迹段,两轨迹段间的空间距离小于阈值,则构成轨迹段对 SP;

步骤二、在同一 SP 内的轨迹段进行后续相似性计算,轨迹段之间的相似性计算包括 4 方面:空间相似性 spatialSIM、时间相似性 tempoSIM、方向相似性 OrientSIM 和速度相似性 velocitySIM。

3. 如权利要求 2 所述的多时空特征融合下的兴趣点轨迹聚类方法,其特征在于,充分利用轨迹的运动信息和兴趣点停留信息,表达具有社会行为相似性的轨迹集合,计算轨迹段时空相似度的中,spatialSIM 为两轨迹段 U、V 始末兴趣点中心位置的差异值为:

$$\Delta d_{u,v} = |p_{u,s} - p_{v,s}| + |p_{u,e} - p_{v,e}|$$

时间相似性 tempoSIM 包括两轨迹段始末坐标点的时间差异及轨迹段间隔时间差异为:

$$\Delta t_{u,v} = |t_{u,s} - t_{v,s}| + |t_{u,e} - t_{v,e}| + |\Delta t_u - \Delta t_v|$$

;

$$+ |\Delta tp_{u,s} - \Delta tp_{v,s}| + |\Delta tp_{u,e} - \Delta tp_{v,e}|$$

方向相似性 OrientSIM 为:

$$\Delta \theta_{u,v} = |\theta_u - \theta_v|;$$

速度相似性 velocitySIM 为:

$$\Delta \bar{v}_{u,v} = |\bar{v}_u - \bar{v}_v|;$$

采用 LCSS 的倒数获得轨迹段间各采样点的相似性 ( $\Delta L_{u,v}$ );

轨迹 U 和 V 间的时空距离公式为:

$$DIS(U, V) = \frac{\exp\{\frac{I[U] + I[V]}{2}\}}{\min(|U|, |V|)} \times \sum_{\langle u, v \rangle, u \in U, v \in V} \frac{\Delta d_{u,v} \Delta L_{u,v} \Delta \theta_{u,v} \Delta t_{u,v}}{\min(|u|, |v|)}。$$

4. 如权利要求 1 所述的多时空特征融合下的兴趣点轨迹聚类方法,其特征在于,利用轨迹 U 和 V 间的时空距离公式,基于 OPTICS 算法对轨迹聚类,让最终轨迹聚类结果能反映

出具有重要全局意义的主干路径分布,具体方法为:聚类簇  $C$  中包含的轨迹数目为簇基数  $n_{cb}$ ,给定阈值  $\tau$ ,定义簇基数  $n_b$  高于  $\tau$  聚类簇称为显著簇,其余的聚类结果为非显著簇,最终被裁剪掉。

## 一种多时空特征融合下的兴趣点轨迹聚类方法

### 技术领域

[0001] 本发明属于轨迹聚类算法领域,尤其涉及一种多时空特征融合下的兴趣点轨迹聚类方法。

### 背景技术

[0002] 随着卫星、互联网和跟踪设备的发展,大量移动物体的轨迹数据被捕获,如车辆移动、动物移动、台风走向、人员移动等。这些大量积累的轨迹数据记录了移动对象的位置和时间的记录序列,蕴藏了丰富的时空知识,具有巨大的应用价值。通过分析轨迹数据,有助于对人类行为模式、交通物流、应急疏散管理、动物习性、市场营销、计算几何以及模拟仿真等各个领域进行研究。通过对各种时空轨迹数据进行聚类分析,可以提取时空轨迹数据中的相似性与异常特征,并有助于发现其中有意义的时空模式。

[0003] 近年来,基于轨迹的行为分析方法引起了国内外学者的广泛关注。但是这些方法大多没有结合速度、方向等多种移动特征信息进行聚类。此外,已有方法将轨迹拆分为轨迹段,对轨迹段进行聚类,聚类结果无法从全局角度反映轨迹的时空特征和运动趋势。

### 发明内容

[0004] 本发明实施例的目的在于提供一种多时空特征融合下的兴趣点轨迹聚类方法,旨在解决目前轨迹聚类算法没有综合考虑轨迹的全局时空相似性和局部时空相似性,导致城市活动轨迹聚类难以体现社会活动相似性的问题。

[0005] 本发明是这样实现的,一种多时空特征融合下的兴趣点轨迹聚类方法包括:

[0006] 步骤一、根据轨迹连续坐标点间的时间和空间跨度获取用户停留的兴趣点,通过兴趣点将轨迹分割为多条轨迹段;

[0007] 步骤二、通过计算轨迹段在空间、时间、速度和方向上的差异来综合判断两条轨迹间的相似度;

[0008] 步骤三、基于 OPTICS 方法对轨迹进行密度聚类,并剪裁掉轨迹数量稀疏的聚类簇。

[0009] 进一步,计算轨迹段时空相似度的具体方法为:

[0010] 步骤一、基于空间四叉树索引,以轨迹 S 中各轨迹段为查询对象,快速获取轨迹 R 中与 S 各轨迹段空间距离最接近的相应的轨迹段,如果两轨迹段间的空间距离小于阈值,则它们构成轨迹段对 SP;

[0011] 步骤二、在同一 SP 内的轨迹段进行后续相似性计算,轨迹段之间的相似性计算包括 4 方面:空间相似性 spatialSIM、时间相似性 tempoSIM、方向相似性 OrientSIM 和速度相似性 velocitySIM。

[0012] 进一步,所述的计算轨迹段时空相似度的中,spatialSIM 为两轨迹段 U、V 始末兴趣点中心位置的差异值为:

[0013] 
$$\Delta d_{u,v} = |p_{u,s} - p_{v,s}| + |p_{u,e} - p_{v,e}|$$

[0014] 时间相似性 tempoSIM 包括两轨迹段始末坐标点的时间差异及轨迹段间隔时间差异为：

$$[0015] \quad \Delta t_{u,v} = |t_{u,s} - t_{v,s}| + |t_{u,e} - t_{v,e}| + |\Delta t_u - \Delta t_v|$$

[0016] ；

$$[0017] \quad + |\Delta tp_{u,s} - \Delta tp_{v,s}| + |\Delta tp_{u,e} - \Delta tp_{v,e}|$$

[0018] 方向相似性 OrientSIM 为：

$$[0019] \quad \Delta \theta_{u,v} = |\theta_u - \theta_v| ;$$

[0020] 速度相似性 velocitySIM 为：

$$[0021] \quad \Delta \bar{v}_{u,v} = |\bar{v}_u - \bar{v}_v| ;$$

[0022] 采用 LCSS 的倒数获得轨迹段间各采样点的相似性 ( $\Delta L_{u,v}$ ) ；

[0023] 轨迹 U 和 V 间的时空距离公式为：

$$[0024] \quad \text{DIS}(U, V) = \frac{\exp\{\frac{I[U]+I[V]}{2}\}}{\min(|U|, |V|)} \times \sum_{\langle u,v \rangle, u \in U, v \in V} \frac{\Delta d_{u,v} \Delta L_{u,v} \Delta \theta_{u,v} \Delta t_{u,v}}{\min(|u|, |v|)}$$

[0025] 进一步, 利用轨迹 U 和 V 间的时空距离公式, 基于 OPTICS 算法对轨迹聚类, 须过滤掉轨迹数量较少的聚类簇, 减少大量破碎的聚类, 让最终轨迹聚类结果能反映出具有重要全局意义的主干路径分布, 具体方法为：

[0026] 设聚类簇 C 中包含的轨迹数目为簇基数  $n_c$ , 给定阈值  $\tau$ , 定义簇基数  $n_b$  高于  $\tau$  聚类簇称为显著簇, 其余的聚类结果为非显著簇, 最终被裁剪掉。

[0027] 本发明将轨迹转换为兴趣点序列, 综合考虑连续兴趣点间形成的轨迹段的速度、方向和时空特性, 进行轨迹聚类, 并挑选出显著簇以获取能反映全局重要性的轨迹聚类形态。实验结果表明, 该方法在聚类中保留了轨迹原有的时空和移动属性特性, 能更全面地反映移动对象的运动和行为模式, 聚类准确性高。

## 附图说明

[0028] 图 1 是本发明实施例提供的多时空特征融合下的兴趣点轨迹聚类方法流程图；

[0029] 图 2 是本发明实施例提供的 MSFF 的聚类后的显著簇；

[0030] 图 3 是本发明实施例提供的 LUCC 方法聚类结果图；

[0031] 图 4 是本发明实施例提供的 LUCC 方法聚类后的显著簇；

[0032] 图 5 是本发明实施例提供的聚类准确率比较图。

## 具体实施方式

[0033] 为了使本发明的目的、技术方案及优点更加清楚明白, 以下结合实施例, 对本发明进行进一步详细说明。应当理解, 此处所描述的具体实施例仅仅用以解释本发明, 并不用于限定本发明。

[0034] 下面结合附图及具体实施例对本发明的应用原理作进一步描述。

[0035] 图 1 示出了本发明的多时空特征融合下的兴趣点轨迹聚类方法流程, 如图所示, 本发明是这样实现的, 一种多时空特征融合下的兴趣点轨迹聚类方法包括：

[0036] S101、根据轨迹连续坐标点间的时间和空间跨度获取用户停留的兴趣点, 通过兴

趣点将轨迹分割为多条轨迹段；

[0037] S102、通过计算轨迹段在空间、时间、速度和方向上的差异来综合判断两条轨迹间的相似度；

[0038] S103、基于 OPTICS 方法对轨迹进行密度聚类，并剪裁掉轨迹数量稀疏的聚类簇。

[0039] 进一步，计算轨迹段时空相似度的具体方法为：

[0040] 步骤一、基于空间四叉树索引，以轨迹 S 中各轨迹段为查询对象，快速获取轨迹 R 中与 S 各轨迹段空间距离最接近的相应的轨迹段，如果两轨迹段间的空间距离小于阈值，则它们构成轨迹段对 SP；

[0041] 步骤二、在同一 SP 内的轨迹段进行后续相似性计算，轨迹段之间的相似性计算包括 4 方面：空间相似性 spatialSIM、时间相似性 tempoSIM、方向相似性 OrientSIM 和速度相似性 velocitySIM。

[0042] 进一步，所述的计算轨迹段时空相似度的中，spatialSIM 为两轨迹段 U、V 始末兴趣点中心位置的差异值为：

$$[0043] \quad \Delta d_{u,v} = |p_{u,s} - p_{v,s}| + |p_{u,e} - p_{v,e}|$$

[0044] 时间相似性 tempoSIM 包括两轨迹段始末坐标点的时间差异及轨迹段间隔时间差异为：

$$[0045] \quad \Delta t_{u,v} = |t_{u,s} - t_{v,s}| + |t_{u,e} - t_{v,e}| + |\Delta t_u - \Delta t_v|$$

[0046] ；

$$[0047] \quad + |\Delta tp_{u,s} - \Delta tp_{v,s}| + |\Delta tp_{u,e} - \Delta tp_{v,e}|$$

[0048] 方向相似性 OrientSIM 为：

$$[0049] \quad \Delta \theta_{u,v} = |\theta_u - \theta_v| ;$$

[0050] 速度相似性 velocitySIM 为：

$$[0051] \quad \Delta \vec{v}_{u,v} = |\vec{v}_u - \vec{v}_v| ;$$

[0052] 采用 LCSS 的倒数获得轨迹段间各采样点的相似性 ( $\Delta L_{u,v}$ )；

[0053] 轨迹 U 和 V 间的时空距离公式为：

$$[0054] \quad DIS(U, V) = \frac{\exp\{\frac{I(U)+I(V)}{2}\}}{\min(|U|, |V|)} \times \sum_{\langle u,v \rangle, u \in U, v \in V} \frac{\Delta d_{u,v} \Delta L_{u,v} \Delta \theta_{u,v} \Delta t_{u,v}}{\min(|u|, |v|)}$$

[0055] 进一步，利用轨迹 U 和 V 间的时空距离公式，基于 OPTICS 算法对轨迹聚类，须过滤掉轨迹数量较少的聚类簇，减少大量破碎的聚类，让最终轨迹聚类结果能反映出具有重要全局意义的主干路径分布，具体方法为：

[0056] 设聚类簇 C 中包含的轨迹数目为簇基数  $n_{cb}$ ，给定阈值  $\tau$ ，定义簇基数  $n_b$  高于  $\tau$  聚类簇称为显著簇，其余的聚类结果为非显著簇，最终被裁剪掉。

[0057] 1、概念和定义

[0058] 轨迹：平面空间中的随时间推移而连续采用的点序列称为轨迹。轨迹  $TR_i$  定义： $TR_i = \{p_1, p_2, \dots, p_k\}$ ，其中  $p_k = \{x_k, y_k, t_k\}$ ，他们分别代表该点的二维空间坐标和采用时间。

[0059] 兴趣点：假设存在一连续采样点集，该点集中任何两点间的距离不超过阈值  $\varepsilon_p$ ，且采样点数  $s$  大于阈值  $\varepsilon_t$ ，该集合内所有采样点构建的最小外接矩形，称为兴趣点，其表示用户在该位置上车、下车或等红绿灯等行为，隐含了用户行为语义或交通语义信息。因此，

轨迹间的时空相似性度量可近似转化为两条轨迹兴趣点间的邻近度：构成两轨迹的兴趣点间在空间上越接近，空间重叠度越高，重叠的兴趣点越多，则它们越相似。然而，除了考虑兴趣点间的空间邻近性，轨迹间相似性计算还要考虑方向、速度及其他采样点间的相似性等因素。

[0060] 轨迹段：由连续两个兴趣点及它们之间的采样点构成的子轨迹，表示为：SubTrajectory<sub>s</sub> = {p<sub>1</sub>, ..., p<sub>k</sub>} (1 ≤ s ≤ k), k 为该轨迹段所属轨迹的采样点总数。

[0061] 轨迹段速度：轨迹段 u 的速度为该轨迹段所有采样点中的最小速度、最大速度和平均速度来衡量：

$$[0062] \quad \bar{V}_u = (1 - \omega_m - \omega_a)v_{\min} + \omega_m v_{\max} + \omega_a \frac{v_i + v_{i+1} + \dots + v_j}{j - i + 1} \quad (1)$$

[0063] 其中， $\omega_m + \omega_a \leq 1$ ,  $v_{\min}$  为轨迹段中速度最低值， $v_{\max}$  为轨迹段中速度最高值，i 和 j 分别为该轨迹段采样点的下标。

[0064] 轨迹段方向：为轨迹段 u 始末兴趣点之间形成的角度，也称运动方向角：

$$[0065] \quad \theta_u = \arctan\left[\frac{(y_{u,e} - y_{u,s})}{(x_{u,e} - x_{u,s})}\right] \quad (2)$$

[0066] 其中，(x<sub>u,s</sub>, y<sub>u,s</sub>) 为轨迹段起点，(x<sub>u,e</sub>, y<sub>u,e</sub>) 为轨迹段终点。

[0067] 轨迹兴趣点停留时间：为用户在轨迹段始末兴趣点分别驻留的时间，间接反映了移动对象的繁忙程度，可分别表示为  $\Delta t_{p_{u,s}}$  和  $\Delta t_{p_{u,e}}$ 。

[0068] 轨迹段间隔时间：为轨迹段始末兴趣点之间的时间间隔，表达了用户在这两个兴趣点间的交通通勤状况，及移动对象所采用的交通方式。其可表示为：

$$[0069] \quad \Delta t_u = t_{u,s} - t_{u,e} \quad (3)$$

[0070] 其中，t<sub>u,s</sub> 和 t<sub>u,e</sub> 分别表示轨迹段 u 始末采样对应的时间。

[0071] 2、轨迹时空聚类

[0072] 2.1 轨迹段相似性量度

[0073] 最长公共子序列 (Longest common subsequence, LCSS)：是指两个或者多个序列中存在的最长的共同子序列。对于时空轨迹来说，计算其最长公共子序列并转化为 LCSS 距离可以衡量轨迹间的相似程度。LCSS 值通过递归方式获取：

$$[0074] \quad \text{LCSS}(R, S) = \begin{cases} 0, & \text{if}(m = n = 0) \\ \text{LCSS}(\text{Rest}(R), \text{Rest}(S)) + 1, & \text{if } |r_{1,x} - s_{1,x}| \leq \delta \cap |r_{1,y} - s_{1,y}| \leq \psi; \\ \max\{\text{LCSS}(\text{Rest}(R), S), \text{LCSS}(R, \text{Rest}(S))\}, & \text{otherwise.} \end{cases} \quad (3)$$

[0075] 式中，LCSS(R, S) 表示轨迹 R 与 S 间的相似性。假设轨迹在二维空间进行比较， $\delta$  和  $\psi$  分别表示 x 轴和 y 轴上的相似阈值，当横坐标差小于  $\delta$  且纵坐标差小于  $\psi$  时，认为这对采样点相似，LCSS 值加 1，其他各项意义与前述相同。当轨迹记录点数都为 0 时，LCSS(R, S) 为 0。本实施例引用 LCSS 计算轨迹段时空相似性。

[0076] 首先，基于空间四叉树索引，以轨迹 S 中各轨迹段为查询对象，快速获取轨迹 R 中与 S 各轨迹段空间距离（采用轨迹段中心间的空间距离表示）最接近的相应的轨迹段。如果两轨迹段间的空间距离小于阈值，则它们构成轨迹段对 (Sub-Trajectory Pair, SP)。该

方法避免了 LCSS 中通过递归的方法搜寻两轨迹间最近的兴趣点,降低了相似性计算的时间复杂度。其次,在同一 SP 内的轨迹段进行后续相似性计算。轨迹段之间的相似性计算包括 4 方面:空间相似性 spatialSIM、时间相似性 tempoSIM、方向相似性 OrientSIM 和速度相似性 velocitySIM。其中,spatialSIM 为两轨迹段始末兴趣点中心位置的差异值,即:

$$[0077] \quad \Delta d_{u,v} = |p_{u,s} - p_{v,s}| + |p_{u,e} - p_{v,e}| \quad (5)$$

[0078] 其中,  $\Delta d_{u,v}$  为轨迹段  $u$  和  $v$  之间的空间距离,  $p_{u,s}$  为轨迹段  $u$  的起始兴趣点的中心位置,  $p_{u,e}$  为轨迹段  $v$  的终点兴趣点的中心位置。此外,还要进一步考虑轨迹段间各采样点的相似性 ( $\Delta L_{u,v}$ ),这里采用 LCSS 的倒数获得该值。

[0079] 时间相似性 tempoSIM 包括两轨迹段始末坐标点的时间差异及轨迹段间隔时间差异,即:

$$[0080] \quad \Delta t_{u,v} = |t_{u,s} - t_{v,s}| + |t_{u,e} - t_{v,e}| + |\Delta t_u - \Delta t_v|$$

$$[0081] \quad (6)$$

$$[0082] \quad + |\Delta tp_{u,s} - \Delta tp_{v,s}| + |\Delta tp_{u,e} - \Delta tp_{v,e}|$$

[0083] 其中,  $t_{u,s}$  为轨迹段  $u$  的起始点时间,  $t_{u,e}$  为轨迹段  $u$  的终点时间。

[0084] 方向相似性 OrientSIM 为:

$$[0085] \quad \Delta \theta_{u,v} = |\theta_u - \theta_v| \quad (1)$$

[0086] 速度相似性 velocitySIM 为:

$$[0087] \quad \Delta \bar{V}_{u,v} = |\bar{V}_u - \bar{V}_v| \quad (2)$$

[0088] 最终,将以上公式合并,即可获取轨迹  $U$  和  $V$  间的时空距离公式:

$$[0089] \quad DIS(U, V) = \frac{\exp\{\frac{I[U] + I[V]}{2}\}}{\min(|U|, |V|)} \times$$

$$\sum_{\langle u, v \rangle, u \in U, v \in V} \frac{\Delta d_{u,v} \Delta L_{u,v} \Delta \theta_{u,v} \Delta t_{u,v}}{\min(|u|, |v|)} \quad (3)$$

[0090] 其中,  $I[U]$  表示在  $U$  不属于任何 SP 的轨迹段数量,  $\langle u, v \rangle$  构成一个 SP,它们分别为轨迹  $U$  和  $V$  中的轨迹段。

[0091] 2.2 基于 OPTICS 的轨迹聚类

[0092] 利用轨迹相似性公式 (9),基于 OPTICS (Ordering Points To Identify the Clustering Structure) 算法对轨迹聚类。

[0093] 定义  $\xi$  邻域  $N_\xi(R)$ :对于轨迹  $R$ ,给定邻近阈值  $\xi$ ,如果存在轨迹  $S$ ,满足  $N_\xi(R) = \{S \in D | DIS(R, S) \leq \lambda, R \neq S\}$ 。其中,  $D$  为所有轨迹数据集。轨迹邻域用以在 OPTICS 聚类中,判断每条轨迹的当前密度,进而将密度较大的轨迹聚为同一组。

[0094] 须过滤掉那些轨迹数量较少的聚类簇,减少大量破碎的聚类,让最终轨迹聚类结果能反映出具有重要全局意义的主干路径分布。

[0095] 设聚类簇  $C$  中包含的轨迹数目为簇基数  $n_{cb}$ ,给定阈值  $\tau$ ,进行如下定义:

[0096] 显著簇:  $C_{pro} = \{C | C \in O \cap n_{cb} > \tau\}$ ,其中,  $O$  为聚类的结果集。即簇基数  $n_b$  高于  $\tau$  聚类簇称为显著簇。其余的聚类结果为非显著簇,最终被裁剪掉。

[0097] 该方法能较好过滤到大部分不重要的聚类,在全局空间分异上凸显出重要聚类的影响范围,而其他类似的轨迹密度聚类方法无法做到这一点。

[0098] 3、聚类方案实施



[0099] 本方案部署环境中的服务器硬件为 CPU(CORE 2 DUO 2.8GH), 内存 8GB, 采用 Visual Studio 2010 编程, 采用 MySQL 存储和检索轨迹数据。采用武汉市武昌区 2010 年 2 月至 4 月的出租车数据集作为实验数据, 共 10835 条轨迹, 每条轨迹的采样点包括了经纬度坐标、采样时间。聚类方法统一采用 OPTICS。

### [0100] 3.1 聚类有效性比较

[0101] 本文提出的方法简称为 MSFF, 比较方法为采用 LUCC 针对原始轨迹空间坐标信息的轨迹相似性计算方法, 简称为 LUCC。图 2- 图 4 分别显示了 MSFF 方法和 LUCC 方法的轨迹聚类空间分布图, 图中每条线代表 100 条不区分速度、方向和时间的轨迹, 小于 100 条的也用一条线段表示, 而带有颜色的线条为聚类, 灰色线条为非显著簇结果。

[0102] 图 2 为本方法 MSFF 聚类后的显著簇空间分布结果。该图中, 轨迹数据被分为 76 个聚类簇。本实施例将轨迹划分为各兴趣点后, 原来在同一主干道路上速度和方向相似、但兴趣点位置相差大的轨迹, 不会被聚合到同一聚类簇。由于位于枝干道路上移动对象的速度和方向的相似度较低, 大部分的处于主干道之外的轨迹被作为非显著簇处理。

[0103] 图 3 中显示为 LUCC 的聚类结果。OPTICS 经过参数调优, 其处理后的轨迹被分为 282 个类。从总体上看, 尽管能体现城市车流的主要分布, 但由于聚类数量众多, 许多旁支小路与主要公路聚到了一块, 难以体现出城市主要交通干道的分布。

[0104] 图 4 为 LUCC 方法图 3 聚类结果基础上过滤掉非显著簇的空间分布。这时的 OPTICS 聚类结果过滤掉了大量轨迹数量过少的聚类簇, 最终轨迹仅仅被分为 12 个类, 仅能体现出城市部分主要交通干道的分布。

[0105] 可见, 本方法所得轨迹聚类结果的空间分布反映了车流在城市中的最为主要的流动状况, 良好表现出了城市主干要道, 聚类效果显著。

### [0106] 3.2 聚类准确率比较

[0107] 进行 MSFF 和 LUCC 方法比较:

$$[0108] \quad \eta = 1 - \frac{\text{dist}_m(i)}{\sum_{n=1, n \neq m}^K \text{dist}_m(i)}, \text{ 其中, } \text{dist}_n(i) = \frac{\sum_{j \in C_n} \text{DIS}(i, j)}{|C_n|} \quad (4)$$

[0109]  $\text{dist}_n(i)$  为轨迹  $i$  与聚类簇  $n$  之间的距离。  $\eta$  越大, 聚类准确率越高。

[0110] 方案实施结果如图 5 所示。随着 OPTICS 方法中邻近阈值  $\xi$  的增加, MSFF 的聚类准确率都明显优于 LUCC 方法。这是因为 MSFF 综合考虑了兴趣点和采样点间的空间和时间多种特征, 还兼顾了社会语义信息, 因此每个聚类中轨迹的时空信息一致性高, 能表达时空粒度更加细致的相似轨迹; 而 LUCC 方法仅仅考虑全局环境下的空间相似性, 因此每个聚类中的时空信息的一致性相对要低, 同一聚类中的轨迹不能正确表达具有社会移动语义相似性的移动对象, 聚类适用面窄。

### [0111] 4、结论

[0112] 轨迹时空聚类在动物迁徙模式、交通流量演变、地理社交推荐等领域有重要应用价值。本文将轨迹转换为兴趣点序列, 综合考虑连续兴趣点间形成的轨迹段的速度、方向和时空特性, 进行轨迹聚类, 并挑选出显著簇以获取能反映全局重要性的轨迹聚类形态。实验结果表明, 该方法在聚类中保留了轨迹原有的时空和移动属性特性, 能更全面地反映移动对象的运动和行为模式, 聚类准确性高。

[0113] 以上所述仅为本发明的较佳实施例而已, 并不用以限制本发明, 凡在本发明的精

神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

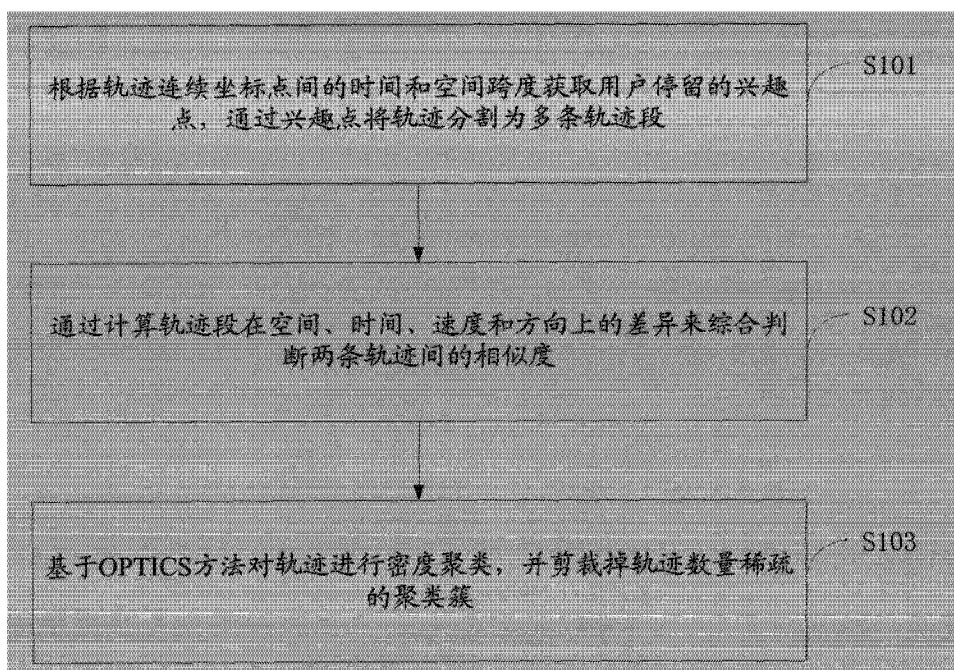


图 1

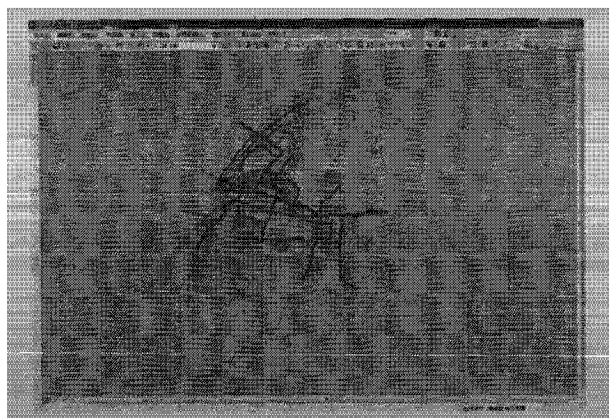


图 2

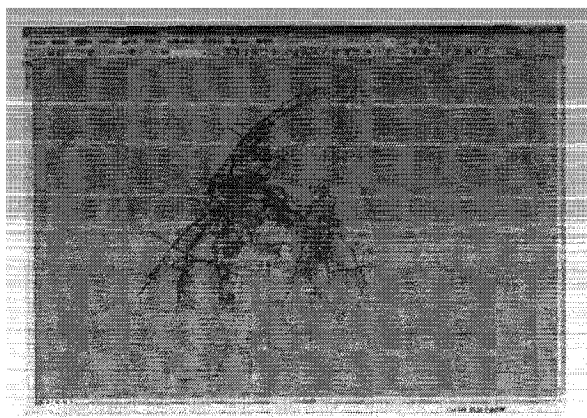


图 3

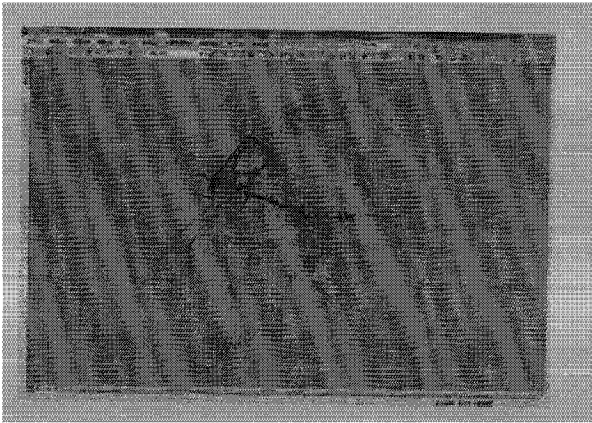


图 4

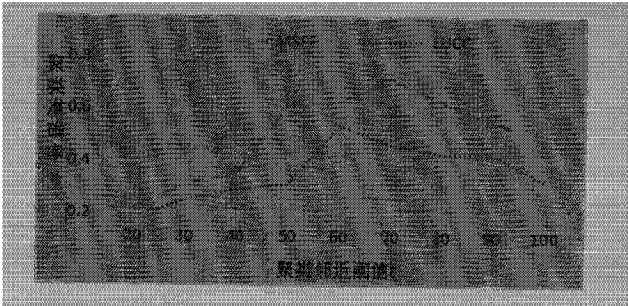


图 5