

基于多尺度空间划分与路网建模的城市移动轨迹模式挖掘

王亮^{1,2,3} 胡琨元¹ 库涛¹ 吴俊伟^{1,3}

摘要 针对城市移动轨迹模式挖掘问题展开研究, 提出移动全局模式与移动过程模式相结合的挖掘方法, 即通过移动轨迹的起始位置点-终点位置点 (Origin-destination, OD 点) 与移动过程序列分别进行移动全局模式与过程模式的发现. 在移动全局模式发现中, 提出了弹性多尺度空间划分方法, 避免了硬性等尺度网格划分对密集区域边缘的破坏, 同时增强了密集区域与稀疏区域的区分能力. 在移动过程模式发现中, 提出了基于移动轨迹的路网拓扑关系模型构建方法, 通过路网关键位置点的探测抽取拓扑关系模型. 最后基于空间划分集合与路网拓扑模型对原始移动轨迹数据进行序列数据转换与频繁模式挖掘. 通过深圳市出租车历史 GPS 轨迹数据的实验结果表明, 该方法与现有方法相比在区域划分、数据转换等方面具有更好的性能, 同时挖掘结果语义更为丰富, 可解释性更强.

关键词 数据挖掘, 移动轨迹, 多尺度划分, 路网模型

引用格式 王亮, 胡琨元, 库涛, 吴俊伟. 基于多尺度空间划分与路网建模的城市移动轨迹模式挖掘. 自动化学报, 2015, 41(1): 47-58

DOI 10.16383/j.aas.2015.c130804

Mining Urban Moving Trajectory Patterns Based on Multi-scale Space Partition and Road Network Modeling

WANG Liang^{1,2,3} HU Kun-Yuan¹ KU Tao¹ WU Jun-Wei^{1,3}

Abstract In this paper, the problem of discovering moving trajectory patterns in urban environment is studied and the method of integration of moving global pattern and moving local pattern is proposed. Through moving trajectory origin-destination (OD) and moving sequence features, the global patterns and local patterns are mined. In the process of moving global pattern mining, a flexible multi-scale space partition is devised to avoid damage of the dense region edges by hard regular grid division and enhance the ability to distinguish the dense regions and sparse regions. In the process of moving local pattern mining, the modeling method of road network based on moving trajectory is devised to extract the feature of topological relation by key road network nodes. Finally, the raw moving trajectory dataset is converted by partitioned discrete regions and road network model, and the frequent moving trajectory patterns are discovered by a modified sequence pattern mining algorithm. A comprehensive experimental evaluation on Shenzhen taxicabs GPS trajectory dataset is presented, and the evaluation shows that the proposed method outperforms the existing methods in space division, data transform, and interpretability of mined patterns.

Key words Data mining, moving trajectory, multi-scale partition, road network modeling

Citation Wang Liang, Hu Kun-Yuan, Ku Tao, Wu Jun-Wei. Mining urban moving trajectory patterns based on multi-scale space partition and road network modeling. *Acta Automatica Sinica*, 2015, 41(1): 47-58

收稿日期 2013-08-19 录用日期 2014-03-25
Manuscript received August 19, 2013; accepted March 25, 2014
国家自然科学基金 (61003208, 61203161, 61174164, 61105067, 61402360) 资助

Supported by National Natural Science Foundation of China (61003208, 61203161, 61174164, 61105067, 61402360)

本文责任编辑 董海荣

Recommended by Associate Editor DONG Hai-Rong

1. 中国科学院沈阳自动化研究所信息服务与智能控制研究室 沈阳 110016 2. 西安科技大学电气与控制工程学院 西安 710054 3. 中国科学院大学 北京 100049

1. Laboratory of Information Service and Intelligent Control, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016 2. College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an, 710054 3. University of Chinese Academy of Sciences, Beijing 100049

近年来, 随着 GPS、射频识别 (Radio frequency identification, RFID) 等定位追踪技术的不断发展, 对移动物体的轨迹数据进行采集、存储、分析与挖掘成为了现实. 从轨迹数据中发现移动个体及群体的行为模式与规律, 进而为基于位置的服务 (Location-based server, LBS)、人群管理、城市规划等相关应用领域提供可参考和可利用的知识, 已成为当前数据挖掘领域的一个研究热点, 引起了学术界与产业界的广泛关注^[1-14].

1 相关研究工作

目前, 在移动数据挖掘领域, 针对城市移动轨

迹模式挖掘问题的研究已经取得了许多成果^[4-8]. Castro 等基于大规模出租车历史轨迹数据构建了城市交通密度模型,用以预测未来时段的交通状况及车流规模^[4]. Gong 等提出了一种地理信息系统 (Geographic information system, GIS) 算法,对纽约市民的 GPS 出行轨迹数据进行处理并自动检测出 5 种不同的移动出行模式^[5]. Yue 等对基于城市出租车 GPS 轨迹数据构建空间交互模式的可行性问题进行了分析,同时研究了城市不同商业区之间的移动交互行为特征与规律^[6]. Zhan 等利用出租车轨迹数据对城市路网中不同位置之间的平均行程时间进行了预估分析^[7]. Brouwers 等通过 GPS、Wi-Fi 以及手机定位数据研究了城市移动行为中停留空间区域的探测方法^[8].

考虑到移动出行行为的目标导向性,有学者提出了基于起始位置点-终点位置点 (Origin-destination, OD 点) 的移动轨迹分析方法^[9-12]. Yue 等基于 OD 点矩阵分析了武汉出租车移动轨迹,揭示了不同时间段内,市民在城市不同区域之间的移动规律^[9]. Zhang 等分析了出租车 OD flow 下群体移动模式的时间与空间分布特征^[10]. Guo 等基于 GPS 轨迹点的聚类划分,研究了城市移动轨迹 OD pairs 的空间模式特征^[11]. Pan 等基于出租车 OD 点动态模式分析了城市区域功能定位与使用的问题^[12]. 在上述研究中,OD 点侧重于分析移动轨迹的起始-终止全局移动模式,忽略了轨迹的移动过程模式特征,因而无法对研究对象的移动过程数据进行挖掘与分析.

由于移动轨迹在空间维度上的连续性分布,移动模式分析有必要对其覆盖的空间区域进行离散化处理. Yue 等的研究工作或是以有限个城市商业区域为研究对象^[6],或是以城市行政区域边界为基础对移动区域进行离散划分^[9],这属于已知空间领域知识条件下的划分方法.均未对空间区域做进一步细粒度划分处理,因而相应的空间粒度较为粗糙.此外,有相关研究工作^[13-17]采用等空间尺度的规则网格对空间区域进行离散划分,将原始移动轨迹点映射为所属的空间网格 ID 号,进而对转换后的以网格 ID 号表示的移动序列数据进行相关的分析与挖掘.在基于手机定位轨迹数据的分析方面, Sohn 等以及 Caceres 等以移动通信蜂窝网格为空间划分集合,对城市群体移动行为模式与移动规律进行了分析与研究^[18-19].

依据路网数据是否已知,现有的研究可分为已知路网条件下的城市移动轨迹模式挖掘与未知路网条件下的移动轨迹模式挖掘.在已知路网条件下,通

过原始轨迹数据与路网数据的转换与融合实现移动轨迹聚类和移动模式挖掘的目的^[20-21].在未知路网条件下,直接利用原始移动轨迹的空间分布特征,通过热点空间区域 (Region of interesting, ROI) 提取及序列模式挖掘实现对移动轨迹模式的发现^[22-24]. Giannotti 等通过构建空间邻域函数,分别以静态及动态方式发现热点空间区域集合,进而对以热点空间区域表示的移动序列数据进行频繁模式的挖掘^[22]. Lee 等采用复合分层的方法以轨迹的形状与分布特征为基础研究了移动轨迹的分类问题^[23].此外也有相关研究工作基于规则网格单元对覆盖的空间区域进行等粒度划分,从而将原始轨迹数据转换为以网格单元表示的序列数据^[24-25].

在频繁模式挖掘方面,自从 Agrawal 等于 1993 年首次提出“频繁模式”的概念以来^[26],陆续有相关学者提出了适应于不同问题特征的频繁模式挖掘算法^[27-32]. Agrawal 等提出了针对项目集数据的采用广度优先搜索策略的 Apriori 算法^[27]. Srikant 等提出了类 Apriori 算法的频繁序列模式挖掘 GSP 算法^[28]. Pei 等提出了基于投影数据库的频繁序列模式挖掘 PrefixSpan 算法^[29]. Han 等对项目集数据、序列数据、图形数据的频繁模式挖掘相关方法进行了综述分析^[30]. 吴枫等提出了基于时间衰减模型的流数据频繁模式挖掘方法^[31]. 潘云鹤等对面向数据流的频繁模式挖掘的研究进展进行了综述分析^[32].

通过对所述研究工作的分析,我们发现现有的城市移动轨迹模式挖掘问题,尚存在以下需要改进的方面:

- 1) 缺乏对移动轨迹全局模式与移动轨迹过程模式的综合考虑. 现有的研究工作或是只考虑移动轨迹全局模式,基于轨迹 OD 点分析不同区域之间的移动出行规律,忽略了移动过程行为模式^[9, 13-14];或是只考虑移动轨迹过程模式,基于已知路网数据或自定义热点区域 ROI 集合,对原始轨迹进行数据转化与序列模式挖掘,忽略了移动轨迹的全局模式^[22-25]. 前者无法对移动过程中的路径选择、交通状况等信息进行提取与发现,后者无法得出移动过程模式对应的 OD 点目标导向信息,难以在起始点或终止点已知的条件下对移动行为进行分类、解释及预测.

- 2) 缺乏对移动轨迹空间划分伸缩性的考虑. 已有的面向移动轨迹模式分析的空间划分方法大多采用单一尺度的规则网格划分方法 (Regular grid cell, RGC)^[13-17], 尽管该方法具有简单、易操作等优点,但不可避免地带来了尺度参数选取的问题,同时等尺度划分网格这种硬划分方法其空间语义伸缩性较

差, 容易破坏密集区域的边缘部分, 降低最终结果的精确性.

3) 缺乏基于移动车辆轨迹空间分布的路网拓扑关系建模方面的研究. 由于移动车辆轨迹本身就蕴含着城市路网的空间拓扑关系, 因此在未知路网条件下, 如何从移动车辆轨迹空间分布特征中抽取出城市路网的拓扑关系模型, 进而基于路网拓扑模型实现移动过程模式的分析与挖掘, 具有更为现实的意义.

基于上述问题, 本文展开了基于多尺度空间划分与路网拓扑关系建模的城市移动轨迹模式挖掘工作, 主要包括:

1) 着眼于移动轨迹 OD 点的空间分布特征, 提出一种具有多空间尺度的离散区域弹性划分方法. 该方法以 OD 点空间分布密度与划分区域面积为控制参数, 通过空间区域规则网格划分与相近邻网格单元合并两个阶段, 实现对具有相似 OD 点语义分布特征的空间区域多尺度离散划分.

2) 针对未知路网条件下的移动轨迹过程模式挖掘问题, 提出一种基于移动轨迹空间分布特征的城市路网拓扑关系建模方法. 该方法以移动轨迹运动形态的变化为依据, 以路网中交通流量较大的交叉路口位置点为目标, 通过二次点筛选过程构建路网关键位置节点集合, 实现对城市路网拓扑关系的建模.

3) 提出一种移动轨迹全局模式与过程模式相结合的城市移动轨迹模式挖掘与分析的新的框架, 基于多尺度空间划分集合与城市路网模型分别对移动轨迹的 OD 点与过程点进行数据映射转换, 进而对转换后的移动轨迹序列数据进行模式挖掘, 设计并实现了适用于上述转换数据模式挖掘的改进 PrefixSpan 算法.

4) 以深圳市出租车 GPS 移动轨迹数据为实验数据, 对本文提出的方法进行验证. 实验结果表明, 本文提出的城市移动轨迹模式挖掘方法较现有的规则网格空间划分方法及热点区域 ROI 序列模式挖掘方法具有较为明显的优势, 同时挖掘的模式在结果上具有更为丰富的语义知识与更强的可解释性.

2 基于移动轨迹 OD 点位置对的弹性多尺度空间划分

为实现融合语义的移动轨迹表示与地理空间离散化划分, 本节提出一种基于移动轨迹 OD 点位置对的弹性多尺度空间划分方法, 包括等粒度规则网格划分和基于密度的近邻网格合并两个阶段. 下面首先介绍相关的基本概念.

2.1 OD 点位置对相关定义

定义 1 (移动轨迹). 移动轨迹 Tr 是具有时间戳的空间位置序列数据, 可表示为 n 元组, $Tr = (\langle x_1, y_1, t_1 \rangle, \dots, \langle x_i, y_i, t_i \rangle, \dots, \langle x_n, y_n, t_n \rangle)$, $1 \leq i \leq n$, 其中 $\langle x_i, y_i \rangle$ 表示二维地理空间中的位置点信息, 在 GPS 系统中 x_i 和 y_i 分别对应于地理空间经度与纬度数据; t_i 为 $\langle x_i, y_i \rangle$ 位置的时间戳信息, 且 $\forall 1 < i < n, t_i < t_{i+1}$.

通常情况下, 个体的移动行为是目的驱动的, 其在移动轨迹上表现为起始性与目的性的指向, 即从起点出发经由一定的路径最终到达终点. 在路网环境下, 移动个体正是从起始点出发, 通过判断并选择不同的可达路径, 最终到达终点位置.

定义 2 (OD 点位置对). 移动轨迹 OD 点位置对是个体移动出行一段完整轨迹的起始位置点 $\langle x_1, y_1 \rangle$ 与终止位置点 $\langle x_n, y_n \rangle$. 在城市出租车 GPS 移动轨迹场景中, OD 点位置对对应出租车某一段完整载客轨迹中的上客点 (Pick-up) 位置与下客点 (Drop-off) 位置.

以 OD 点位置对来表征移动出行的全局语义信息具有明显的优势: 1) 起始性与目的性指向更为明确; 2) 空间语义下区域间的内在移动关系更为清晰; 3) 描述方式易于理解, 可解释性强. 因此, 本文基于 OD 点位置对, 描述移动轨迹的全局模式. 需要注意, 在以 OD 点位置对表示的移动轨迹全局模式中, 起始点与终止点的描述并非原始轨迹数据中细空间粒度的经纬度位置, 而是粗空间粒度上的区域语义描述. 因此考虑以“区域”这种描述空间范围的对象表示 OD 点位置对, 原因在于经纬度位置点数据不具有代表性、重复性以及丰富的空间语义特征, 且其不利于连续空间维度的离散化与近似化表示.

2.2 基于移动轨迹 OD 点位置对的弹性多尺度空间划分

移动轨迹全局模式的发现需要对覆盖的区域进行离散划分, 将空间平面 S 划分为若干个互不重合的区域集合 s_i , 即有 $S = s_1 \cup s_2 \cup \dots \cup s_n$, 且 $s_i \cap s_j = \emptyset, 1 \leq i, j \leq n$. 基于该离散区域划分集合, 将原始移动轨迹的 OD 点经纬度数据转换为离散区域表示的 OD 点位置对数据, 进而分析移动对象在相应区域之间的移动规律.

通过上述分析, 本文提出一种基于移动轨迹 OD 点空间分布密度的弹性多尺度空间划分方法, 该方法采用划分-合并的两阶段策略, 首先借鉴网格空间区域离散划分的思想^[15-17], 以细空间粒度对整个空间平面 S 进行等尺度划分; 其次以 OD 点密度值

(Density) 与最大划分区域面积 λ 作为阈值参数控制最近邻空间网格单元的合并过程, 最终生成与 OD 点空间分布密度特征相一致的空间划分离散区域集合.

下面对提出的弹性多尺度空间划分方法进行详细阐述. 考虑一个空间平面 S 以及移动 OD 点集合 P , 首先利用细空间粒度网格 $G = \{g_1, g_2, \dots, g_m\}$ 对 S 进行等尺度划分以生成 m 个规则网格单元. 在每个划分的网格单元 g_i 内分布有数量不等的 OD 点, 采用基于密度的方法依次对最近邻的网格单元进行合并计算. 若合并后的多网格 $mulg$ 区域对应的 OD 点密度大于设定的阈值 ε , 则停止合并过程, 多网格 $mulg$ 区域即为新生成的离散划分区域; 否则继续寻找最近邻且未被合并的网格单元 g_j 进行合并过程.

考虑到过大的区域划分面积不利于空间语义的表示与发现, 因为大面积区域内的移动过程无法检测到, 同时过大的区域面积将导致转化后的轨迹数据非常粗糙. 因此, 本文在上述划分算法中引入了最大区域面积阈值参数 λ , 限制稀疏区域面积的过大增长. 在网格合并的过程中判断新生成多网格 $mulg$ 区域的面积与阈值参数 λ , 如果该区域面积大于 λ 则停止此轮合并. 实际上, 若我们所使用的初始网格空间粒度足够小, 则该方法可以对任意形状的空间平面进行划分.

基于 OD 点密度的弹性多尺度空间划分算法表述如下:

算法 1. 基于 OD 点密度的弹性多尺度空间划分算法

输入. 空间平面 S , 细粒度网格单元集合 $G = \{g_1, g_2, \dots, g_m\}$, OD 点集合 P .

输出. 空间划分离散区域集合 $S = \cup_{l=1}^L s_l$.

参数. 区域划分数量期望值 ∂ , 密度阈值 ε , 区域密度上限 φ , 最大区域面积阈值 λ .

步骤 1. 构建细粒度空间网格单元 $G = \{g_1, g_2, \dots, g_m\}$, 并建立其与空间平面 S 的映射对应关系表 $List$;

步骤 2. 网格单元 OD 点分布密度计算, 对于每一个网格单元 g_i 计算其对应的 OD 点密度值 $Density(g_i)$.

步骤 3. 最近邻网格单元合并.

步骤 3.1. 对网格单元 g_i 进行密度阈值判断, 若 $Density(g_i) < \varepsilon$, 则选择与 g_i 最邻近的网格单元 g_j 进行合并操作以生成多网格 $mulg$ 区域; 否则 g_i 独立成为新的区域 s_l , 更新 $List$ 并返回步骤 2;

步骤 3.2. 判断步骤 3.1 中合并生成的多网格

$mulg$ 区域的 OD 点密度 $Density(mulg)$, 判断条件: 1) $Density(mulg) \geq \varepsilon$; 2) $|mulg| \geq \lambda$. 其中 $|mulg|$ 表示多网格 $mulg$ 区域的空间面积. 若满足条件 1) 或条件 2), 则 $mulg$ 区域即为新生成的离散划分区域 s_l , 更新关系表 $List$ 并返回步骤 2; 反之若条件 1) 和条件 2) 都不满足, 则返回步骤 3.1, 重新选择次近邻网格单元 g_{j+1} 进行新一轮合并过程.

在弹性多尺度空间划分算法的参数部分, 区域离散划分数量期望值 ∂ 表示对空间平面 S 进行多尺度划分前设定的期望值, 该参数类似于模糊 C 均值聚类算法中预先设定的聚类数 ψ . 需要注意的是, 模糊 C 均值聚类算法中用户设定的聚类数 ψ 是严格约束的, 因此最终产生的聚类数量恒等于预先设定的 ψ 值. 而本文提出的多尺度空间划分方法中的区域划分数量期望值 ∂ 并非严格约束关系, 因此算法最终产生的多尺度空间划分离散区域集合的规模并非恒等于期望值 ∂ , 而是近似于期望值 ∂ .

在本文的弹性多尺度空间划分实验部分, 细粒度空间网格单元数量为 3900 个, 区域离散划分期望值 $\partial = 100$, 在不限定单个划分区域面积时最终产生的空间划分离散区域规模为 82 个; 在限定单个划分区域最大面积 $\lambda = 50$ 时, 最终产生的空间划分离散区域规模为 136 个.

密度阈值 ε 、区域密度上限 φ 和最大区域面积阈值 λ 的取值与最终产生的空间划分离散区域之间没有严格的数学对应关系, 且在不同的应用场景 (包括实验数据、应用需求、空间分布特征等) 下对于最终的空间划分离散区域规模有不同的数量或粒度要求. 因此, 我们给出一般性的参数取值指导性说明. 经过算法及实验测试结果分析, 密度阈值 ε 取值应为 $|P|/\partial$ 取值的 1~1.2 倍, 其中 $|P|$ 为实验数据中 OD 点集合 P 的规模; 区域密度上限 φ 取值应为密度阈值 ε 取值的 1.5~1.8 倍; 最大区域面积阈值 λ 取值应为 $|G|/\partial$ 取值的 1.2~1.5 倍, 其中 $|G|$ 为实验中细粒度空间网格单元的数量, $|G|$ 的取值在覆盖的空间区域面积一定的情况下, 与采用的空间粒度大小成反比关系, 空间粒度取值取决于不同的应用场景及用户的应用需求.

通过对算法 1 进行时间复杂度分析, 可得出其值为 $O((|G| - 1)!)$. 由此可见, 算法 1 的时间复杂度与空间初始划分阶段所选定的网格数量紧密相关.

3 基于移动轨迹的城市路网拓扑关系建模

移动轨迹不仅具有出发点与目的点的全局属性, 且在移动过程中具有过程属性, 即移动过程点的时间、空间分布特征. 探究轨迹的过程属性是研究移

动对象在移动过程中不同时间段内的路径选择以及相应的交通状况等信息. 例如, 相同 OD 点位置对对应的不同过程数据中可能含有出租车司机的恶意绕远行为信息, 也可能含有不同路段在不同时间段内的交通流状态信息. 另外, 研究相同移动过程中的行为模式特征对于目前推广的绿色拼车、节能减排具有一定的应用价值.

显而易见, 城市车辆的移动轨迹分布在路网之上, 路网的存 在不仅对移动行为的有序性起到规范作用, 同时也是一种空间约束. 而对于城市交通路网进行整体建模不仅耗时, 而且不利于移动轨迹在空间上的近似化分析. 因此, 本文考虑基于移动轨迹的路网分布特征, 实现对城市路网拓扑关系模型的构建. 我们定义城市路网拓扑关系模型为路网上关键空间位置点的集合. 路网上的关键位置点以交叉路口最具有代表性, 如果确定了移动轨迹所途经的若干个关键交叉路口, 则其大致的移动路径即可以基本确定. 同时路网交叉口交通流量大, 以其表示的移动过程模式对于交通诱导、路线规划等具有直接的指导作用. 故考虑通过探测城市路网关键交叉口位置点构建路网的拓扑关系模型.

一般而言, 在城市路网交叉口会有大量的移动车辆轨迹发生方向偏转. 基于此移动特征, 本文以移动轨迹的方向变化为依据提出一种移动路网交叉口发现方法, 该方法包括两个阶段: 1) 通过计算移动轨迹转向角提取移动轨迹转向点集合 TP ; 2) 基于转向点集合 TP 进行二次提取, 通过计算设定区域内的轨迹转向点分布特征, 得到路网交叉口位置集合 $Intersection$, 即城市路网的拓扑关系模型.

定义 3 (移动轨迹角). 移动轨迹角 α 是移动个体在移动过程中其前后相邻的两个轨迹位置点连线与水平轴的夹角.

定义 4 (移动轨迹转向角). 移动轨迹转向角 $\Delta\alpha$ 是前后相邻的两个移动轨迹角差值的绝对值, 即 $\Delta\alpha = |\alpha_{i+1} - \alpha_i|$.

在第 1 个阶段, 我们对移动轨迹样本数据逐次计算移动轨迹角 α , 进而比较相邻移动轨迹角变化量的绝对值, 即移动轨迹转向角 $\Delta\alpha$. 如图 1 所示, 对于一段移动轨迹的空间位置点逐次计算得到移动轨迹角 α_1 与 α_2 , 而 α_1 与 α_2 的差值绝对值 $\Delta\alpha$ 为移动轨迹转向角. 若 $\Delta\alpha$ 近似成直角则认为移动对象在该点处发生了一次转向, 本文中 $\Delta\alpha$ 近似成直角的判断为 $\Delta\alpha$ 属于 $90^\circ \pm 10^\circ$ 或者 $-90^\circ \pm 10^\circ$ 的区间内, 即裕度范围为 20° . 对应的前后两条轨迹直线的交叉点即为轨迹转向点 TP .

在第 2 个阶段, 基于轨迹转向点集合 TP 进行

二次提取, 逐个对 TP 集合中的空间点 tp_i 进行分析. 在以 tp_i 为中心的设定范围 R_p 内计算转向点个数 n . 如果转向点个数 $n \geq N$, 其中 N 为设定的最小点数阈值参数, 则认为该点即是路网交叉口点.

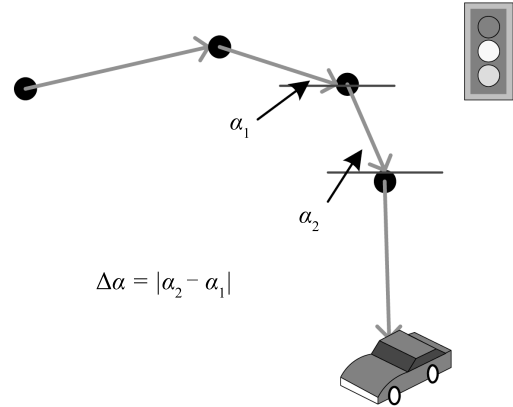


图 1 移动轨迹转向点计算示意图

Fig. 1 Moving trajectory turning point calculation

基于移动轨迹的城市路网交叉口发现算法表述如下:

算法 2. 路网交叉口集合发现算法

输入. 移动轨迹数据集 $MTraj$.

输出. 路网交叉口点集合 $Intersection$.

参数. 区域设定值 ρ , 最小转向点数 N .

步骤 1. 对移动轨迹 $Traj_l$ 逐次计算轨迹位置点 $Traj_l.loc_i$ 的移动轨迹角 α .

步骤 2. 逐次计算轨迹 $Traj_l$ 的移动轨迹转向角 $\Delta\alpha$, 判断 $80^\circ \leq \Delta\alpha \leq 100^\circ$ 或 $-100^\circ \leq \Delta\alpha \leq -80^\circ$, 若满足任一条件则认为 $\Delta\alpha$ 近似直角, 故而 $TP = TP \cup Traj_l.loc_i$.

步骤 3. 对 TP 集合中的每一个点 tp_i 在设定大小 ρ 的区域空间 R_p 内计算转向点数 n .

步骤 4. 满足最小转向点数 n 的 tp_i 点即为路网关键交叉口点.

对算法 2 的时间复杂度进行分析, 可得其复杂度值为 $O(|MTraj| \times (P_{aver} - 1))$, 其中 $|MTraj|$ 为移动轨迹数据集的规模数 $MTraj$ 的规模数, P_{aver} 为移动轨迹的平均轨迹位置点数. 显而易见, 算法 2 的实时性与选取的移动轨迹的规模及特征直接相关.

4 多尺度空间划分与城市路网拓扑关系建模实验

4.1 数据介绍

本文的实验数据为广东省深圳市 3000 辆出租车 15 天的移动轨迹历史记录, 该数据格式为 [ID,

Current time, Longitude, Latitude, Speed, Status]. 其中, ID 表示车辆的标识号, Current time 为数据采样时间, Longitude 和 Latitude 分别为轨迹的地理空间经度与纬度, Speed 为车辆的当前移动速度, Status 为出租车的载客信息, 包括空载与载客两个取值. 本研究用于多尺度空间划分与路网拓扑关系建模实验的数据为 100 辆出租车 2~5 天的有效移动轨迹数据, 经过数据预处理得到的轨迹点数为 825 599 个, 有效载客轨迹为 12 820 条, 对应的有效载客轨迹点为 446 269 个.

图 2 为有效载客轨迹数据在 Google Earth 上的分布图示, 可以看出, 轨迹分布具有较大的不均匀性, 中心城区轨迹分布非常密集, 其他区域则相对稀疏, 这也与城市规划、市民移动出行的一般规律相契合.



图 2 出租车 GPS 移动轨迹 Google Earth 分布图示
Fig. 2 The taxi GPS moving trajectory on Google Earth

4.2 弹性多尺度空间划分实验

本文提出的 OD 点位置对在城市出租车移动轨迹场景中表现为一段有效载客轨迹的第一个点与最后一个点, 即上客 (Pick up) 点与下客 (Drop off) 点. 本文对实验数据中的 Pick up 点与 Drop off 点的空间分布进行分析, 如图 3 所示, 圆形点表示 Pick up 点, 菱形点表示 Drop off 点, 可见大部分的点都集中在中心城区, 少部分的点分布在市郊区, 这一特征与城市功能定位及建设规划相一致.

本文以移动轨迹的 OD 点作为数据集, 对城市区域进行多尺度空间划分. 如第 2.2 节所述, 基于 OD 点密度的弹性多尺度空间划分算法, 对上述 OD 点所覆盖的深圳市地理空间区域进行划分实验, 结果如图 4 所示, 横坐标表示经度, 纵坐标表示纬度. 图 4(a) 为不限制区域划分面积的结果, 可以看出, 划分的区域面积大小差异较大, 最小的划分区域面积为单位 1, 最大的划分区域面积为 583. 面积小的区域表示该区域的 OD 点分布密集, 即为密集区域;

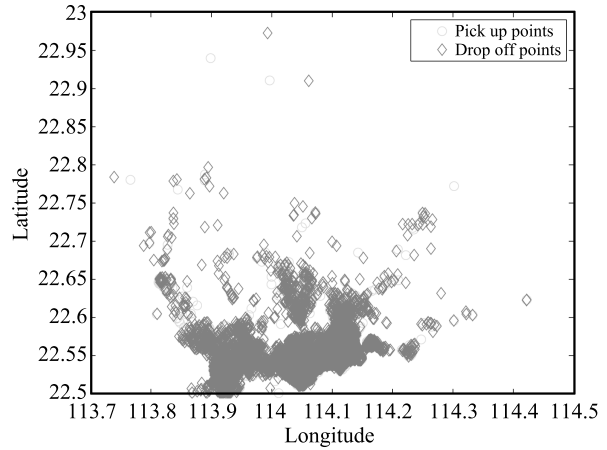
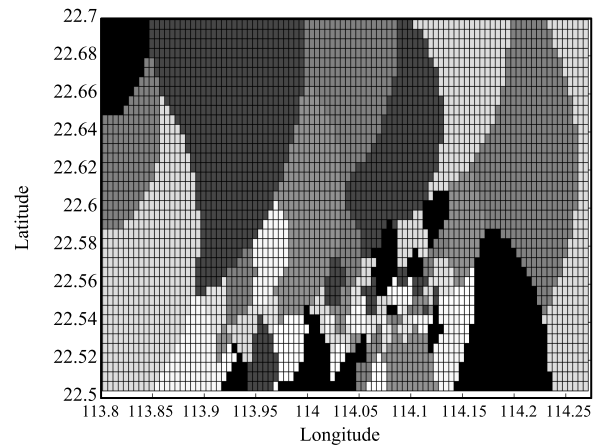
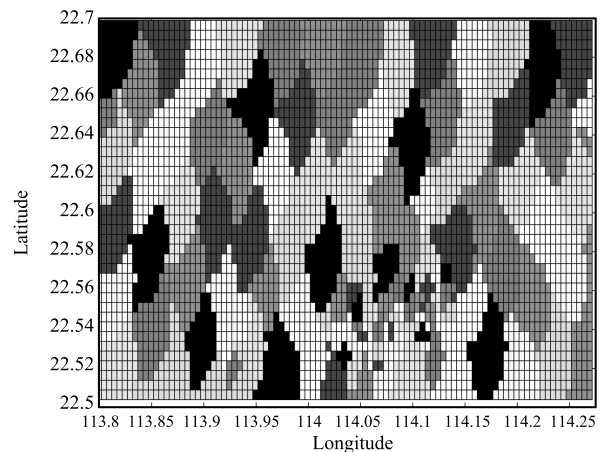


图 3 Pick up 点与 Drop off 点的空间分布
Fig. 3 The distribution of pick up and drop off points



(a) 不限制区域划分面积

(a) Not limiting the area division



(b) 限制区域划分面积

(b) Limiting the area division

图 4 城市移动区域多尺度空间划分实验结果图示
Fig. 4 Urban mobile regional multi-scale space partition

面积大的区域表示该区域的 OD 点分布稀疏, 即为稀疏区域. 图 4 (b) 为限制区域划分面积的情况, 其中最大区域面积阈值 λ 设为 50, 从图中可以看出, 该划分方法在保证对 OD 点密集分布小区域细粒度划分的同时, 对稀疏区域的面积合并进行了限制, 从而避免区域划分面积无限制时较大面积区域的产生.

4.3 城市路网拓扑关系建模实验

对实验数据中的 12820 条有效载客轨迹中的 420629 个移动过程点进行城市路网拓扑关系建模实验, 经过移动轨迹转向角的计算、比较, 得到轨迹转向点集合 TP , 如图 5 所示.

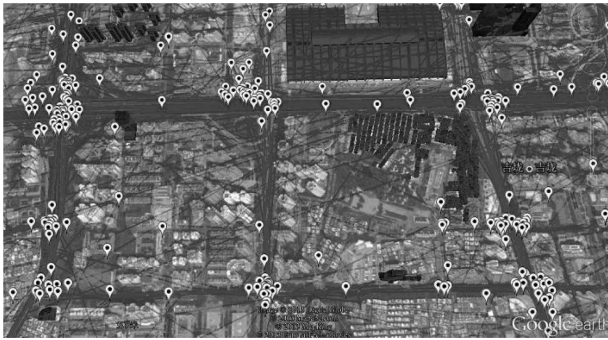


图 5 转向点 TP 集合分布 Google Earth 图

Fig. 5 The Google Earth map of turning point TP set

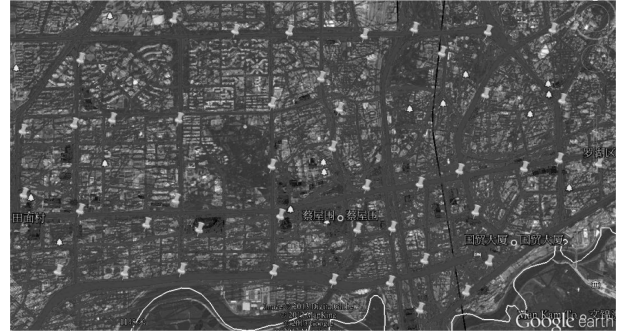
图 5 中, 线段表示出租车移动轨迹, 圆形图标代表转向点, 可以看出, 得到的转向点在实际路网的交叉口分布非常集中, 其他区域的转向点分布原因可能是由于轨迹采集的时间间隔较大所致.

如第 3 节所述, 在 TP 点集合的基础上进行二次提取, 对每一个转向点 tp 在设定大小的区域空间内, 计算包含的转向点数, 若包含的转向点数不小于给定的阈值, 则认为该点为关键路网交叉口点, 否则删除该转向点. 最终结果如图 6 所示, 图 6 (a) 为路网拓扑关系模型局部图示, 其中图钉状小图标代表发现的路网交叉口点; 图 6 (b) 为路网拓扑关系模型整体图示.

在本文实验中, 区域设定值 ρ 的取值与城市道路交通的相关参数取值相关, 考虑城市一级干道中双向八车道宽度为 28 米, 两侧非机动车道宽度共约 11 米, 绿化带宽度 3 米, 此外考虑到 GPS 数据采集之间的时间延迟, 增加 8 米的空间裕度, 区域设定值 ρ 的取值为直径 50 米. 该参数的取值在不同的应用场景中可结合具体情况进行一定范围的调整.

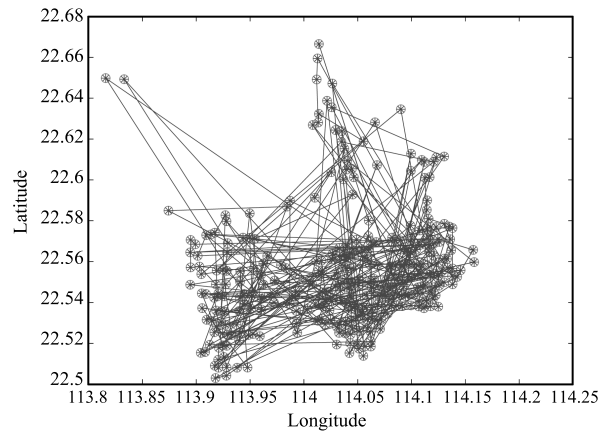
最小转向点数 N 的取值与实验数据的规模成正比, 与最终发现的关键路网交叉口点 *Intersection* 的规模成反比. 最小转向点数 N 的取值, 建议对任

意相邻若干个交叉路口转向点 TP 计算其平均分布值, 通过数次测试、比较选取合适的取值. 本文选取最小转向点数阈值为 10, 最终得到的关键路网交叉路口的规模为 317.



(a) 局部图示

(a) The local graphic



(b) 整体图示

(b) The overall graphic

图 6 城市路网拓扑关系建模实验图示

Fig. 6 The result of urban road network topology relationship modeling

5 城市移动轨迹模式发现

构建了移动空间的多尺度划分与路网拓扑关系模型之后, 可以基于空间划分离散区域集合与路网拓扑关系模型中的交叉口位置点集合, 对原始移动轨迹数据做空间近似化转换处理. 一段移动轨迹 $Tr = \{(p_1, t_1), (p_2, t_2), \dots, (p_n, t_n)\}$ 在相应的全局模式与过程模式下可表示为

$$Tr' = \{(R_i, t_1), \dots, (cross_{k_i}, t_{k_i}), \dots, (R_j, t_n)\}$$

其中, $p_i = (lon_i, lat_i)$ 表示轨迹的经纬度位置信息, t_i 是其对应的的时间戳信息, R_i 表示移动全局模式中所划分的离散区域, $cross_{k_i}$ 为移动过程模式中的路网交叉路口位置点, $0 \leq k_i \leq n - 2$. 通过这样的转

换, 原始移动轨迹数据在保留其 OD 点位置对全局信息的同时, 将移动过程点进行关键路网交叉口位置点序列的映射转换, 在最大程度上保留原始轨迹移动过程信息的同时简化了对应的数据规模与空间分布. 下面对转换后的轨迹序列数据进行模式的挖掘与发现.

1) 基于热点区域 ROI 的模式挖掘方法^[22, 25]及本文提出的移动全局模式与过程模式相结合的方法对原始移动轨迹进行预处理转换实验对比. 为使转换结果尽量在同一标准下进行对比, 采用基于分布密度的方法对原始移动轨迹提取 ROI 热点区域集合, 产生 420 个热点 ROI 空间区域. 本文提出的方法产生多尺度空间划分离散区域的规模为 80, 移动过程模式中路网交叉口集合的规模为 317. 图 7 为不同规模原始移动轨迹数据分别在两种不同方法下的转换序列数据规模. 由图 7 可以看出, 随着原始轨迹数据规模的增大, 对应的转换数据的规模呈线性增长趋势, 而本文提出的方法较 ROI 方法转换后的数据规模增长更快, 其原因在于移动过程模式中的关键路网交叉口集合空间分布较 ROI 集合的空间分布更为均匀.

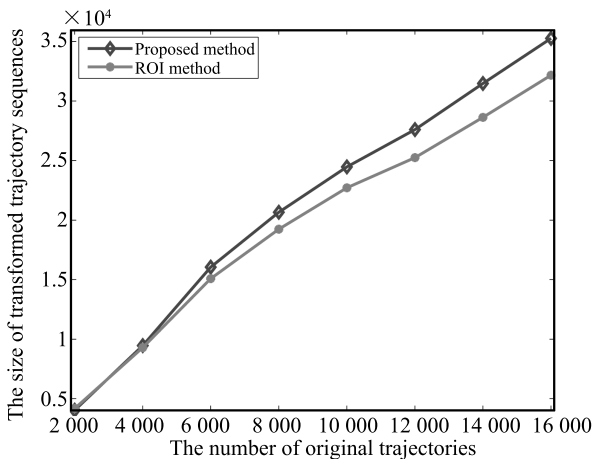


图 7 不同方法下的原始移动轨迹与转换后的轨迹规模比较
Fig. 7 The original mobile trajectory compared with the converted track scale under different methods

2) 对规则网格空间划分方法^[13-17]及本文所提出的多尺度空间划分方法进行 OD 位置对移动统计分析对比, 结果表示城市环境中移动群体在不同离散区域之间的移动规律, 具体结果如图 8 所示. 其中本文提出的多尺度空间划分区域的规模为 80, 规则网格空间划分区域的规模为 100. 图 8 中 FP-total 表示多尺度空间划分方法统计得出的所有频繁移动行程个数, FP-part 为 FP-total 中剔除相同出发-目的区域的移动行程之后的频繁行程个数, 相应的 RP-total 为规则网格空间划分方法统计得出的所有频繁移动行程个数, RP-part 为 RP-total 中剔除相同出发-目的区域的移动行程之后的频繁行程个数.

目的区域的移动行程之后的频繁行程个数, 相应的 RP-total 为规则网格空间划分方法统计得出的所有频繁移动行程个数, RP-part 为 RP-total 中剔除相同出发-目的区域的移动行程之后的频繁行程个数.

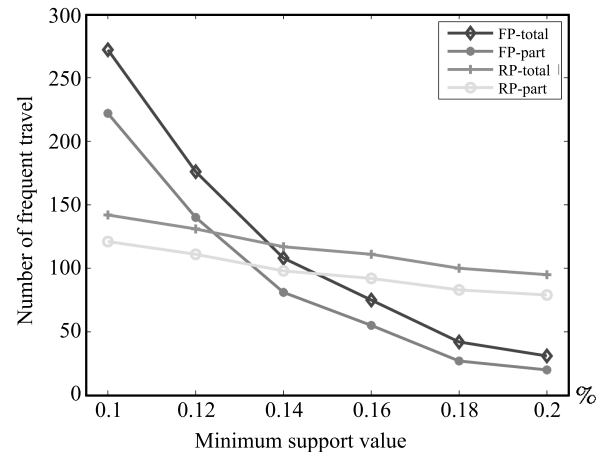


图 8 不同支持度下的频繁 OD 移动行程比较
Fig. 8 Comparing of frequent OD mobile travelling under different support

由图 8 可见, 本文提出的多尺度空间划分方法统计得出的区间频繁移动行程个数, 在较小支持度下, 比规则网格空间划分方法统计得出的区间频繁移动行程个数数量更多, 最大相差 130. 同时, 随着支持度的不断增大, 两种方法得出的频繁区间移动行程数不断减小, 且多尺度空间划分方法其数目变化较大, 平均变化率为 34.94%; 规则网格空间划分方法频繁行程数目变化较小, 平均变化率为 7.69%, 二者相差 4.5 倍, 说明支持度的变化对规则网格空间划分方法的区域移动统计结果影响较小, 故该方法难以有效反映移动群体在不同离散区域之间的移动规律, 而本文提出的方法比规则网格空间划分方法在移动模式空间分布性上有显著改善, 因其较好地克服了规则网格划分容易破坏密集区域边缘的问题, 同时基于相邻区域密度合并的弹性多尺度划分比硬性划分方法可以更好地区分移动语义下的密集区域与稀疏区域.

下面对具有相同移动全局模式下的过程属性进行分析统计, 图 9 所示为具有相同 OD 点位置对全局特征属性的一组移动轨迹, 线段与圆形小图标表示原始移动轨迹, 图钉状小图标表示过程模型中的关键路网交叉点. 其中关键路网交叉点数量为 7 个, 单个移动轨迹平均点数为 21 个 (包括 6 个停止点与 16 个移动点). 由图 9 可见, 本文提取出的过程模型关键路网交叉点基本上可以完全表征其移动过程特

征,同时其轨迹数据规模显著减小.其原因在于基于关键路网交叉点的轨迹描述方法不仅可以有效剔除移动停止点、噪声干扰点等无效数据,而且可以有效减少同一路段相邻空间冗余轨迹点,从而以较小规模数据表征原始移动轨迹数据.因此,本文提出的路网拓扑关系建模方法与 ROI 热点区域方法及规则网格划分方法相对比,可以在细空间粒度上对移动轨迹进行近似化表示与转换,在最大程度保留原始轨迹移动特征的前提下,其数据规模也具有显著优势,可以以较少规模的转换数据完整表征较大规模的原始移动轨迹数据.

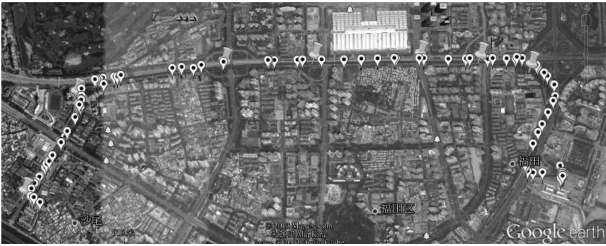


图 9 具有相同 OD 点的移动轨迹与过程轨迹分布
Fig.9 The moving trajectory and process point distribution under same origin-destination

3) 对基于移动全局模式与过程模式表示的移动轨迹数据进行频繁序列模式挖掘分析.挖掘算法采用传统的序列模式挖掘算法 PrefixSpan^[29],为适应本文提出的轨迹序列数据挖掘需要,在原算法的频繁模式全局信息保留、轨迹序号匹配等方面进行相应的改进.移动全局模式与移动过程模式相结合的频繁移动模式算法表述如下:

算法 3. 频繁移动模式挖掘算法

输入. 原始移动轨迹数据的转换序列数据库 $TD = DS_{MP} \cup DS_{OD}$, 最小支持度阈值 min_sup .

输出. 频繁移动模式集合 $FP = \cup_{k=1}^K fp_k$.

参数. 频繁移动序列模式 fp , 序列模式 fp 的长度 k , fp -投影数据库 $DS_{MP}|_{fp}$, 若 $fp = \langle \rangle$, 则 $DS_{MP}|_{fp} = DS_{MP}$.

调用函数.

FP -PrefixSpan($fp, k, DS_{MP}|_{fp}$).

步骤 1. 扫描 $DS_{MP}|_{fp}$ 一次产生频繁关键路网点 Ri 集合;

步骤 2. 对于每一个频繁关键路网点 Ri , 将其与对应的前缀频繁移动序列模式 fp 进行连接操作, 产生新的移动序列模式 fp' , 且输出新模式 fp' ;

步骤 3. 检索与模式 fp' 对应的移动轨迹 ID 集合 Set_ID , 在移动全局数据库 DS_{OD} 中统计计算相应全局移动信息, 包括出发区域、目的区域以及

平均移动速度;

步骤 4. 对于任一个 fp' , 构造以 fp' 为前缀的投影数据库 $DS_{MP}|_{fp'}$, 同时调用函数 FP -PrefixSpan($fp, k, DS_{MP}|_{fp}$).

对算法 3 的时间复杂度进行分析, 可得其值为 $O(|Ri|^q \times |TD|)$, 其中 $|Ri|$ 为频繁关键路网点 Ri 集合的规模, $|TD|$ 为移动轨迹数据转换序列数据库 TD 的规模, q 为所挖掘到的最长频繁移动模式的长度.

需要注意的是, 本文提出的移动模式有别于现有的传统移动轨迹模式^[22, 25]. 传统的移动轨迹模式基于移动过程, 未考虑移动轨迹的全局属性, 而本文提出的移动模式不仅具有频繁移动过程属性, 且包含相应的 OD 点位置对、移动速度等全局属性, 即有效结合了移动全局模式与过程模式.

下面基于算法 3 对具有相同移动过程模式下的全局属性进行挖掘和分析. 图 10 所示为通过算法 3 挖掘得出的频繁轨迹模式, 其中图钉状小图标为关键路网交叉点. 如图 10 所示, 移动车辆频繁经过三个关键路网交叉点 (图钉状小图标) 从右向左行驶, 依次经过若干个交通路口. 同时我们得到与此频繁模式对应的全局移动模式信息, 如表 1 所示. 其中第 1 列表示移动出行的起始区域, 第 2 列表示移动目的的区域, 第 3 列表示对应模式轨迹段内的平均速度值.



图 10 频繁移动轨迹模式图示
Fig.10 The frequent movement patterns

从表 1 中可以看出, 具有该模式的移动轨迹基本上都是从区域 53 出发, 但目的区域各不相同. 另外平均移动速度值变化较大, 其中最大值为 50.61, 最小值为 13.51, 相应的出行时间段分别为夜间 23 点左右与下午 18 点左右, 由此可得该路段相应时间段内的交通状态信息, 进而可以为特定时间段内的出行提供参考. 由此可以得出, 本文提出的移动全局模式与过程模式相结合的移动轨迹模式挖掘方法, 与单纯的 OD 点移动全局模式和 ROI 序列移动过程模式相对比, 不仅可以有效发现移动过程中的频繁

移动序列, 而且可以给出相应移动过程模式的 OD 点全局移动信息与区间内移动平均速度等信息, 其挖掘结果语义信息更为丰富, 结果的可解释性更强.

表 1 相同轨迹模式下的全局移动信息

Table 1 The global mobile information in same trajectory pattern

起始区域	目的区域	平均速度 (km/h)
53	6	36.56
53	15	37.24
53	35	50.61
53	26	13.51
53	18	30.36
53	24	17.05
53	32	31.94
53	5	36.06
60	46	21.69

4) 对本文提出的方法与 ROI 方法进行移动频繁轨迹模式挖掘比较分析. 由图 11 可见, 随着最小支持度的不断增大, 两种方法挖掘的频繁轨迹模式数量呈线性下降趋势. 在相同最小支持度下, 本文提出的方法与 ROI 方法所挖掘的频繁模式相对比数量更大, 可以更多地发现移动轨迹模式特征, 而且随着支持度的增大, 二者的差异越来越大. 其原因在于 ROI 方法较路网拓扑关系建模方法空间粒度更大, 同时其空间分布均匀性较差, 且在移动轨迹空间分布不均衡的情况下该问题将更为明显. 此外, 实验测试结果表明, 对于模式挖掘算法的实时性而言, 其时

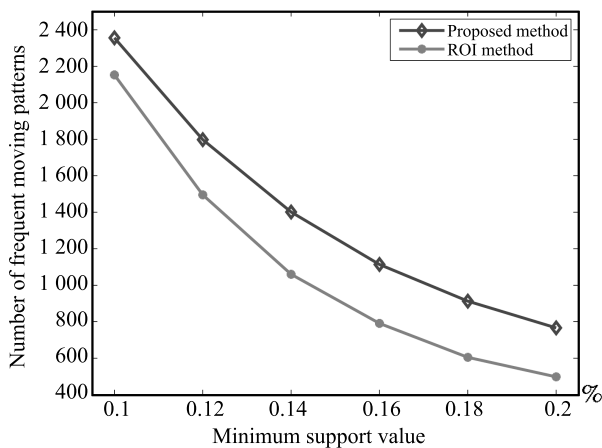


图 11 不同支持度下频繁移动过程模式结果图示

Fig. 11 The frequent movement process patterns under different support values

间开销与实验数据的规模成正比关系, 与最小支持度阈值的大小成反比关系.

6 结论

本文针对城市移动轨迹模式发现问题展开研究, 提出了移动全局模式与移动过程模式相结合的挖掘思路, 即通过移动轨迹 OD 点特征与移动过程序列特征分别进行移动全局模式与过程模式的发现. 在移动全局模式发现中, 针对现有的规则网格划分方法空间语义伸缩性较差这一问题提出了弹性多尺度空间划分方法, 避免了对移动密集区域边缘部分的破坏, 同时可以更好地对密集区域与稀疏区域进行区分. 在移动过程模式发现中, 针对路网未知条件下移动模式发现无法有效结合路网拓扑特征, 提出了基于移动轨迹的路网拓扑关系模型构建方法, 通过路网位置关键节点的探测抽取出路网拓扑关系模型. 最后基于多尺度空间划分区域集合与路网拓扑关系模型对原始移动轨迹进行转换与模式发现. 通过实际 GPS 移动轨迹数据的实验仿真, 与现有的规则网格方法以及 ROI 热点区域方法进行比对, 结果表明本文提出的方法在移动区域划分、移动轨迹数据转换等方面具有更好的性能, 同时发现的移动轨迹模式结果语义更为丰富, 可解释性更强.

References

- Liu Yu, Xiao Yu, Gao Song, Kang Chao-Gui, Wang Yao-Li. A review of human mobility research based on location aware devices. *Geography and Geo-Information Science*, 2011, **27**(4): 8–13
(刘瑜, 肖昱, 高松, 康朝贵, 王瑶莉. 基于位置感知设备的人类移动研究综述. *地理与地理信息科学*, 2011, **27**(4): 8–13)
- Wang Ming-Sheng, Huang Lin, Yan Xiao-Yong. Exploring the mobility patterns of public transport passengers. *Journal of University of Electronic Science and Technology of China*, 2012, **41**(1): 2–7
(王明生, 黄琳, 闫小勇. 探索城市公交客流移动模式. *电子科技大学学报*, 2012, **41**(1): 2–7)
- Zou Yong-Gui, Wan Jian-Bin, Xia Ying. LBSN user movement trajectory clustering mining method based the road network. *Application Research of Computers*, 2013, **30**(8): 2410–2414
(邹永贵, 万建斌, 夏英. 基于路网的 LBSN 用户移动轨迹聚类挖掘方法. *计算机应用研究*, 2013, **30**(8): 2410–2414)
- Castro P S, Zhang D Q, Li S J. Urban traffic modelling and prediction using large scale taxi GPS traces. In: *Proceedings of the 2012 Pervasive Computing Lecture Notes in Computer Science*. Berlin Heidelberg: Springer, 2012. 57–72
- Gong H M, Chen C, Bialostozky E, Lawson C T. A GPS/GIS method for travel mode detection in New York city. *Computers, Environment, and Urban Systems*, 2012, **36**(2): 131–139

- 6 Yue Y, Wang H D, Hu B, Li Q Q, Li Y G, Yeh A G O. Exploratory calibration of a spatial interaction model using taxi GPS trajectories. *Computers, Environment, and Urban Systems*, 2012, **36**(2): 140–153
- 7 Zhan X Y, Hasan S, Ukkusuri S V, Kamga C. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies*, 2013, **33**: 37–49
- 8 Brouwers N, Woehrle M. Dwelling in the canyons: dwelling detection in urban environments using GPS, Wi-Fi, and geolocation. *Pervasive and Mobile Computing*, 2013, **9**(5): 665–680
- 9 Yue Y, Zhuang Y, Li Q Q, Mao Q Z. Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In: Proceedings of the 17th International Conference on Geoinformatics. Fairfax, USA: IEEE, 2009. 1–6
- 10 Zhang W S, Li S J, Pan G. Mining the semantics of origin-destination flows using taxi traces. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. New York, USA: ACM, 2012. 943–949
- 11 Guo D S, Zhu X, Jin H, Gao P, Andris C. Discovering spatial patterns in origin-destination mobility data. *Transactions in GIS*, 2012, **16**(3): 411–429
- 12 Pan G, Qi G D, Wu Z H, Zhang D Q, Li S J. Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems*, 2013, **14**(1): 113–123
- 13 Veloso M, Phithakitnukoon S, Bento C. Urban mobility study using taxi traces. In: Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis. New York: ACM, 2011. 23–30
- 14 Veloso M, Phithakitnukoon S, Bento C. Sensing urban mobility with taxi flow. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. New York, USA: ACM, 2011. 41–44
- 15 Huang Jian-Bin, Zhang Pan-Pan, Huangfu Xue-Jun, Sun He-Li. A trajectory prediction approach for mobile objects by combining semantic features. *Journal of Computer Research and Development*, 2014, **51**(1): 76–87
(黄健斌, 张盼盼, 皇甫学军, 孙鹤立. 融合语义特征的移动对象轨迹预测方法. *计算机研究与发展*, 2014, **51**(1): 76–87)
- 16 Zhao Yue, Liu Yan-Heng, Yu Xue-Gang, Wei Da, Shan Chang-Wei, Zhao Yang. Method for mobile path prediction based on pattern mining and matching. *Journal of Jilin University (Engineering and Technology Edition)*, 2008, **38**(5): 1125–1130
(赵越, 刘衍珩, 余雪岗, 魏达, 单长伟, 赵洋. 基于模式挖掘与匹配的移动轨迹预测方法. *吉林大学学报(工学版)*, 2008, **38**(5): 1125–1130)
- 17 Gid6falvi G, Pedersen T B. Mining long, sharable patterns in trajectories of moving objects. *GeoInformatica*, 2009, **13**(1): 27–55
- 18 Sohn K, Kim D. Dynamic origin-destination flow estimation using cellular communication system. *IEEE Transactions on Vehicular Technology*, 2008, **57**(5): 2703–2713
- 19 Caceres N, Wideberg J P, Benitez F G. Deriving origin destination data from a mobile phone network. *IET Intelligent Transport Systems*, 2007, **1**(1): 15–26
- 20 Won J I, Kim S W, Baek J H, Lee J. Trajectory clustering in road network environment. In: Proceedings of the 2009 Computational Intelligence and Data Mining. Nashville, USA: IEEE, 2009. 299–305
- 21 Li X L, Han J W, Lee J G, Gonzalez H. Traffic density-based discovery of hot routes in road networks. *Advances in Spatial and Temporal Databases*, Berlin Heidelberg: Springer, 2007. 441–459
- 22 Giannotti F, Nanni M, Pinelli F, Pedreschi D. Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2007. 330–339
- 23 Lee J G, Han J W, Li X L, Gonzalez H. TraClass: trajectory classification using hierarchical region based and trajectory based clustering. In: Proceedings of the 2008 VLDB Endowment. Auckland, New Zealand: VLDB, 2008. 1081–1094
- 24 Wang L, Hu K Y, Ku T, Yan X H. Mining frequent trajectory pattern based on vague space partition. *Knowledge-Based Systems*, 2013, **50**: 100–111
- 25 Li H J, Tang C J, Qiao S J, Wang Y, Yang N, Li C. Hotspot district trajectory prediction. In: Proceedings of the 2010 Web-Age Information Management. Berlin Heidelberg: Springer, 2010. 74–84
- 26 Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data. New York, USA: ACM, 1993. 207–216
- 27 Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings of the 1994 International Conference on Very Large Data Bases. Santiago, Chile: VLDB, 1994. 487–499
- 28 Srikant R, Agrawal R. Mining sequential patterns: generalizations and performance improvements. In: Proceedings of the 5th International Conference on Extending Database Technology. Avignon, France: EDBT, 1996. 3–17
- 29 Pei J, Han J W, Mortazavi-Asl B, Wang J Y, Pinto H, Chen Q M. Mining sequential patterns by pattern-growth: the prefixspan approach. *IEEE Transactions on Knowledge and Data Engineering*, 2004, **16**(11): 1424–1440
- 30 Han J W, Cheng H, Xin D, Yan X F. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 2007, **15**(1): 55–86
- 31 Wu Feng, Zhong Yan, Wu Quan-Yuan. Mining frequent patterns over data stream under the time decaying model. *Acta Automatica Sinica*, 2010, **36**(5): 674–684
(吴枫, 仲妍, 吴泉源. 基于时间衰减模型的数据流频繁模式挖掘. *自动化学报*, 2010, **36**(5): 674–684)
- 32 Pan Yun-He, Wang Jin-Long, Xu Cong-Fu. State-of-the-art on frequent pattern mining in data streams. *Acta Automatica Sinica*, 2006, **32**(4): 594–602
(潘云鹤, 王金龙, 徐从富. 数据流频繁模式挖掘研究进展. *自动化学报*, 2006, **32**(4): 594–602)



王 亮 西安科技大学讲师, 博士. 2009 年获东北大学控制理论与控制工程专业工学硕士学位, 2014 年获中国科学院沈阳自动化研究所工学博士学位. 主要研究方向为移动计算, 智能信息处理, 复杂系统建模与优化理论. 本文通信作者.

E-mail: liangwang0123@gmail.com

(WANG Liang Lecturer at Xi'an University of Science and Technology. He received his master degree in control theory and control engineering from Northeast University in 2009, and his Ph. D. degree in mechatronic engineering from Shenyang Institute of Automation, Chinese Academy of Sciences in 2014. His research interest covers mobile computing, intelligent information processing, and complex system modeling and optimization theory. Corresponding author of this paper.)



胡琨元 中国科学院沈阳自动化研究所研究员. 1994 年获西安电子科技大学检测技术与仪器专业工学学士学位, 2003 年获东北大学系统工程专业工学博士学位. 主要研究方向为智能信息处理技术, 移动商务与现代物流, 企业信息化.

E-mail: hukunyuan@sia.cn

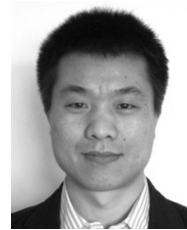
(HU Kun-Yuan Professor at Shenyang Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree in measurement technique and instrument from Xidian University in 1994, and his Ph. D. degree in systems engineering from Northeast University in 2003. His research interest covers RFID information integration,

mobile commerce and modern logistics, and enterprise informatization.)



库 涛 中国科学院沈阳自动化研究所副研究员. 2003 年获山西大学计算机科学与技术专业学士学位, 2009 年获中国科学院沈阳自动化研究所机械电子工程专业工学博士学位. 主要研究方向为感应网络技术, 现实挖掘与社会计算, 智能信息处理. E-mail: kutao@sia.cn

(KU Tao Associate professor at Shenyang Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree in computer science and technology from Shanxi University in 2003, and his Ph. D. degree in mechatronic engineering from Shenyang Institute of Automation, Chinese Academy of Sciences in 2009. His research interest covers sensor network, reality mining and social computing, and intelligent information processing.)



吴俊伟 中国科学院沈阳自动化研究所博士研究生. 2007 年获河南科技大学计算机应用专业工学学士学位. 主要研究方向为移动计算, 机器学习与模式识别.

E-mail: wujunwei@sia.cn

(WU Jun-Wei Ph. D. candidate at Shenyang Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree in computer application from Henan University of Science and Technology in 2007. His research interest covers mobile computing, machine learning, and pattern recognition.)