随机采样移动轨迹时空热点区域发现及模式挖掘

王 亮[1,2,3],胡琨元[1],库 涛[1],吴俊伟[1,2]

(1.中国科学院 沈阳自动化研究所,沈阳 110016;2.中国科学院大学,北京 100039;3.西安科技大学 电气与控制工程学院,西安 710054)

摘 要:针对随机采样条件下移动轨迹在时间轴分布疏密不均的特点,在将三维时空轨迹转换为一维时间投影数据的基础上,提出一种基于密集时间区间自动检测的时空热点区域发现与移动模式挖掘方法。通过自底向上的动态聚类方式以探测密集时间区间,进而在密集时间区间内进行移动轨迹的时空热点区域发现。最后,采用深度优先的序列模式挖掘算法挖掘频繁移动模式集合。基于合成数据的仿真试验,验证了算法在有效性及可扩展性方面均具有较好的性能。

关键词:人工智能;数据挖掘;随机采样移动轨迹;密集时间区间;热点区域

中图分类号:TP18 **文献标志码**:A **文章编号**:1671-5497(2015)03-0913-08

DOI:10.13229/j.cnki.jdxbgxb201503033

Discovering spatiotemporal hot spot region and mining patterns from moving trajectory random sampling

WANG Liang[1,2,3], HU Kun-yuan[1], KU Tao[1], WU Jun-wei[1,2]

(1. *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang* 110016, *China*; 2. *University of Chinese Academy of Sciences, Beijing* 100039, *China*; 3. *School of Electrical and Control Engineering, Xi′an University of Science and Technology, Xi′an University of Science and Technology, Xi′an* 710054, *China*)

Abstract: The moving trajectory by random sampling distributes unevenly in time dimension. After projecting the three-dimensional spatiotemporal trajectory data into one-dimensional time domain, a spatiotemporal hot spot region discovery and moving pattern mining methods are proposed based on automatic detection of intensive time intervals. Through detecting intensive time intervals dynamically with a bottom-up clustering strategy, the spatiotemporal hot spot regions are discovered in corresponding time intervals. A depth-first algorithm is designed to mine the set of frequency moving patterns. Finally, based on synthetic moving trajectory dataset, the effectiveness and scalability of the proposed algorithms are verified.

Key words: artificial intelligence; data mining; moving trajectory by random sampling; intensive time interval; hot spot region

收稿日期:2013-08-30.

基金项目:国家自然科学基金项目(61402360,61203161,61174164).

作者简介:王亮(1984-),男,讲师,博士.研究方向:数据挖掘,智能信息处理.E-mail:liangwang0123@gmail.com

通信作者:库涛(1979-),男,副研究员.研究方向:感应网络技术,智能信息处理,社会计算.E-mail:kutao@sia.cn

0 引 言

随着GPS、RFID,传感器网络和无线通信等技术的不断发展和普及,越来越多的移动轨迹数据被采集和存储在应用服务器中。对移动轨迹数据进行分析,从中发现所蕴含的有效知识信息,对于交通管理、气象监测、商业推广等领域都具有非常重要的作用[1-2]。在实际的应用中,由于存储空间、网络传输以及计算能力等因素的限制,对于移动对象定位数据的采集往往具有随机采样的特征,即前后轨迹位置数据之间的时间间隔不是固定值,而是某一范围内的随机数值。这种随机采样移动轨迹数据在时间维度分布上不均匀,存在数据分布的密集区间与稀疏区间。密集区间所包含的信息足以刻画移动对象在相应时间段内的移动行为特征,但可能存在一定数量的冗余数据;而稀疏区间的信息由于信息缺失往往难以准确描述移动对象的移动历史轨迹。这种数据空间分布的不均匀特性给移动轨迹的特征聚类及模式挖掘带来了新的难题。

目前对于移动轨迹模式挖掘问题已有较多的方法,大致可分为基于移动轨迹形状的方法[3-6]和基于符号序列的方法[7-10]。基于移动轨迹形状的分析方法通常利用线性插值将移动轨迹在二维平面或三维时空内连接成完整的曲线,进而利用不同移动轨迹曲线之间的距离计算对移动轨迹进行聚类分析和模式发现。基于符号序列的方法将移动轨迹表示为包含移动位置点的空间区域符号序列,通过将原始三元轨迹数据转换成二元数据表示的序列数据,从而利于传统的序列模式挖掘算法对移动轨迹模式进行发现与分析。然而,上述两种移动模式挖掘方法均无法有效解决随机采样移动轨迹数据所带来的问题。同时,现有的文献对于随机采样移动轨迹的模式挖掘问题研究较少,大多集中在数据丢失或是低数据采样率而引起的移动路径重构和地图匹配等方面的问题[11-13]。

对于随机采样移动轨迹数据而言,如何有效检测其密集数据分布的时间区间,尤其是群体轨迹的普遍密集数据分布时间区间,进而以相应时间区间内的移动轨迹特征来发现群体意义上的热点区域是当前所面临的主要问题。Giannotti等[7]基于密度度量的方式定义了空间兴趣区域(Region of interesting,ROI),分别以静态和动态的方式对ROI区域进行了探测与发现,进而对以ROI表示的带时间注释的移动序列数据进行了频繁模式的挖掘。Wang等[9]对固定采样间隔的移动轨迹数据进行了闭合频繁模式的挖掘,基于等间隔的离散时间切片,以等规则空间网格划分的方法对不同时间切片层的热点时空区域进行了探测。

在上述研究的基础上,本文首先将随机采样移动轨迹数据在一维时间轴上进行投影转换得到对应的时间投影数据集;然后采用自底向上和滑动时间窗口相结合的策略以提取密集时间区间;进而基于密集时间区间以发现相应时段上的时空热点区域,利用时空热点区域集合对原始轨迹数据集合进行轨迹简约和数据序列转换;最后基于序列模式挖掘的方法对移动轨迹频繁模式进行挖掘和发现。

1 定义及问题描述

定义1 移动轨迹是由移动对象在空间上的移动而产生的连续运动轨迹。在三维时空域中,移动轨迹 $Tr=\{(x_i,y_i,t_i),1\leqslant i\leqslant N\}$ 是由空间元素和时间元素所表示的三元数据组,其中 (x_i,y_i) 为移动对象对应于瞬时时间 t_i 的空间位置;N 为移动轨迹的长度。

定义2 移动时间间隔是移动轨迹中前后相邻的两个数据 (x_i,y_i,t_i) 与 $(x_{i+1},y_{i+1},t_{i+1})$ 之间的时间差值,其公式为:$TimeInterval=t_{i+1}-t_i$。时间间隔长度表征的是移动轨迹的数据采集周期,在固定采样间隔条件下,相邻时间间隔为一个恒等值;而在随机采用情况下其为某一取值范围内的任意值。

定义3 移动频繁模式是移动轨迹数据库中其支持度不小于用户所给定的最小支持度阈值 min_sup 的移动模式,其具有如下的形式:$(R_i,TI_{ki})\rightarrow(R_j,TI_{kj})$,其中 R_i 为空间中的某一区域;TI_{ki} 为移动时间周期中的某一段时间区间。

随机采样移动轨迹数据可以用图1进行描述:Tr_1 和 Tr_2 分别为两个随机采样的移动轨迹,其中箭头所标注的位置点为其采样时刻所在的空间位置;*Constant Sample* 为固定采样率下等时间间隔移动轨迹;R_a 和 R_b 分别为空间区域。由图可见,在固定采样率下 Tr_1 和 Tr_2 有着相近的移动轨迹 *Constant Sample*。而在随机采样条件下,由于二者密集数据区间与稀疏数据区间互不

重合，轨迹 Tr_1 和 Tr_2 具有完全不同的形状。

由上可知在随机采样条件下，具有相近行为特征的移动轨迹其近似度在很大程度上被弱化。同时，密集数据区间信息冗余与稀疏数据区间信息缺失并存，这对于数据特征的抽取以及移动模式的挖掘都带来了困难和挑战。因此，本文所研究的问题就是在上述随机采样条件下，对于群体移动轨迹下的热点时空区域（Spatiotemporal hot spot region，STHSR）进行检测和发现，同时挖掘其中所蕴含的频繁移动行为模式。

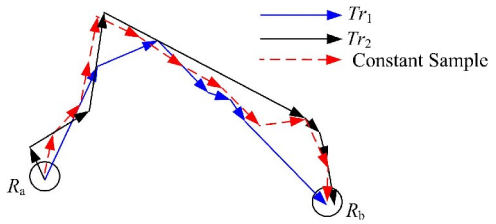


图 1　随机采样移动轨迹比较图示

Fig. 1　Comparison of moving trajectory by random sampling

2　随机采样移动轨迹时空热点区域发现

首先对移动轨迹在时间轴做一维投影，基于滑动时间窗对一维时间投影数据进行自动聚类，以实现对移动轨迹密集时间区间的提取和发现，进而采用密度统计的方法对滑动时间窗内的移动时空轨迹点进行二维平面投影，以发现相应时间段上的时空热点区域 $STHSR$ 。

2.1　基于滑动时间窗的移动轨迹密集时间区间提取

对于随机采样移动轨迹而言，其数据点在时间维度上的分布随机而无规律，其前后相邻轨迹点之间的时间间隔长度长短不一。由于采样时间间隔过大或数据丢失将使得部分时间区间内的数据分布稀疏；另一方面，由于频繁采样或是传输延时等原因，部分时间区间内的数据可能分布密集。稀疏时段的数据可能存在信息缺失的问题，而密集时段的数据可能存在数据冗余的问题。这种数据空间分布状态的不均衡将对移动行为特征提取和移动轨迹的聚类带来困难。

另一方面，随机采样移动轨迹数据在时间维度上的连续性分布特征，使得其无法像固定采样间隔的轨迹数据一样内在地实现对时间域的离散化划分，因而在时间域上将存在近似化表示的问题。基于此，考虑对移动轨迹数据在一维时间轴上进行投影转换，以形成一维时间投影数据。进而对该时间投影数据进行基于滑动时间窗的聚类划分，以动态区分和发现移动轨迹密集时间区间和稀疏时间区间，最终实现依据移动轨迹时间域上的分布特征对其进行近似化划分的目的。

将三维移动时空轨迹在时间轴上进行投影，对应地轨迹点 (x_i, y_i, t_i) 在时间轴上的投影值为其时间戳 t_i 。如图 2 所示，从移动轨迹点 (x_i, y_i, t_i) 出发的平行于 X-Y 平面的虚线与时间 T 轴相交于 t_i 点位置，也即实现了从三维时空数据向一维时间投影数据的转换：$(x_i, y_i, t_i) \rightarrow (t_i)$。

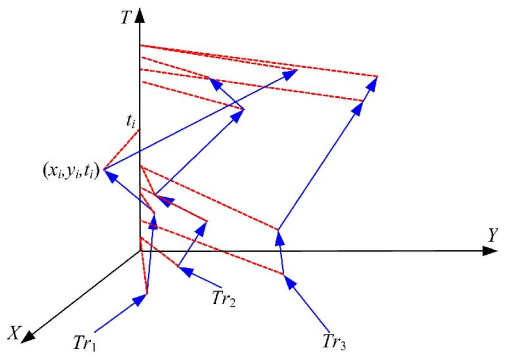


图 2　三维时空移动轨迹的一维时间轴投影

Fig. 2　Projection of spatiotemporal trajectory to timeline

构建了移动轨迹的时间轴投影之后，可得到相应的一维时间投影数据。投影数据在连续时间域上呈不规则分布，密集分布时间区间包含较多的移动轨迹点，稀疏分布时间区间则包含较少的移动轨迹点。轨迹点密集的区间包含有丰富的移动行为信息，而稀疏区间则由于数据丢失或是采样间隔过大而信息量很少。显而易见，轨迹点密集的时间区间对于移动行为模式识别、移动热点空间区域发现、移动轨迹聚类等具有重要的作用。因此，如何从移动轨迹的一维时间投影数据中辨识出分布密集的区域，从而为后续的热点区域发现提供支持是需要解决的首要问题。

考虑采用滑动时间窗对一维时间投影数据进行分割聚类，以实现对轨迹点分布密集时间区间的提取和发现。

定义 4　滑动时间窗 SW（Sliding window）为时间域上的一个宽度大小可变的有限区间，宽度用 ΔT 表示，$\Delta T = T_{\text{end}} - T_{\text{start}}$，$T_{\text{start}}$ 为 SW 的起始时间，T_{end} 为终止时间。当 $\Delta T = 0$ 时，意味着起始时间与终止时间一致，则滑动时间窗退化

为移动轨迹在瞬时时间上的移动信息。

移动轨迹 Tr 在滑动时间窗 SW 内的轨迹点的个数记为 $Num(Tr,SW)$。需要注意的是,对于随机采样间隔数据而言,在滑动时间窗 SW 内可能包含同一个移动轨迹 Tr 的多个轨迹点,也可能不包含 Tr 的任一轨迹点。对于多移动轨迹而言,滑动时间窗 SW 内的轨迹点数记为 $Num(SW)$。

定义 5 密集时间区间为满足条件 $Num(SW)\geqslant\varphi\cdot Mean_Num(SW)$ 的滑动时间窗 SW,其中 φ 为自定义系数,$Mean_Num(SW)$ 为宽度大小为 SW 的滑动窗内的平均移动轨迹点的个数。

通过原始三维移动轨迹数据的一维时间投影,发现变长滑动窗密集时间区间问题可简化为一维数据的聚类划分问题,如图 3 所示,其中左侧为三维立体图,右侧为一维时间轴图。由图 3 可知,移动轨迹点 p_1 与 p_2 落在宽度为 ΔT 的滑动时间窗 SW 内,因此,$Num(SW)$ 的取值为 2。

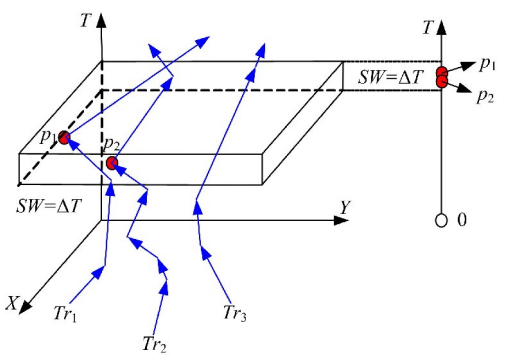


图 3 移动轨迹密集时间区间提取

Fig. 3 Extraction of intensive time interval

由于最终聚类的个数等相关信息未知,因此必须设计出适应于此类问题的聚类算法。考虑采用自底向上的聚类思想,通过最近邻数据的窗体合并和阈值判定相结合的策略,实现对一维时间投影数据的动态聚类。其中阈值判定参数包含时间窗宽度上限 ε 与滑动窗体 SW 的平均移动轨迹点数 $Mean_Num(SW)$。设定时间窗宽度上限 ε 的目的在于:过大的时间间隔上即使其具有足够多的移动轨迹点数量分布,但其所定义热点区域的群体特征不明显,也难以有效刻画有限时间窗体上的移动区域热点程度。而平均移动轨迹点数 $Mean_Num(SW)$ 参数的目的在于保证时间窗体 SW 内有足够多的移动轨迹点,因为稀疏时间区间内移动轨迹点数量有限,无法利用较少的移动轨迹点进行热点时空区域的发现。所得到的密集时间窗体其宽度可以为 0,即为一个瞬时时间切片 $\Delta T=0$,也可为一个小于宽度上限 ε 的滑动时间窗体 $0<\Delta T\leqslant\varepsilon$。最终基于该密集时间窗构造高度为 ΔT,且覆盖整个移动平面区域的时空箱体,进而探测时空热点区域 $STHSR$ 集合。

基于以上描述,移动轨迹密集时间区间提取算法的基本步骤如下:

算法 1 DTIE(Density time interval extracting algorithm from moving trajectory)

输入:$D=\{Tr_1,Tr_2,\cdots,Tr_n\}$,最大时间窗体宽度阈值 ε;自定义系数 φ。

输出:密集时间区间集合 $DSW=\{SW_1,SW_2,\cdots,SW_m\}$。

步骤 1 初始化所有的轨迹;

步骤 2 将移动轨迹在时间 T 轴上进行投影,得到时间分布投影数据集 TD;

步骤 3 对 TD 中的数据进行去重计算,得到无重复数据集 TD';

步骤 4 将 TD' 中的每一个元素进行初始化,令 $TD'=DSW$;

步骤 5 对于集合 DSW 中的元素 SW_i 计算与最近邻元素 SW_j 的距离,若满足条件 $Dist(SW_i,SW_j)+|SW_i|+|SW_j|<\varepsilon$,其中 $|SW_i|$ 为密集时间区间 SW 的宽度,则合并 SW_i 与 SW_j 以形成新的密集时间区间;同时更新信息表 $List$;

步骤 6 重复执行步骤 5 直到集合 DSW 不再进行更新操作为止;

步骤 7 逐次计算集合 DSW 中的每一个元素所包含的轨迹点数,对于满足阈值条件 $Num(SW)\geqslant\varphi\cdot Mean_Num(SW)$ 的元素保留,反之,则删除。

2.2 基于密度统计的时空热点区域发现

定义 6 时空热点区域 $R=(Cell_m,W_n)$ 是某一时间段 $[T_{start},T_{end}]$ 内基于移动轨迹分布特征的热点空间区域。

由于空间区域的不同功能定位和移动群体的行为周期性特点,移动群体意义上的热点空间区域具有显著的时变特征。也就是说,在不同的时间段上热点区域具有不同的空间分布特征。因此,对不同时间内的时空热点区域进行发现和提取,可以分析连续时间周期内热点区域的变化和迁移规律,从而间接得出移动群体在相应时间上

的行为模式。需要注意的是，本文所定义的时空热点区域是具有明确时间区间标注的，即某一特定时间段之内的空间热点区域，这与传统的在整个时间周期内发现热点区域的方法相比具有更为丰富的时间语义信息，同时可动态揭示移动群体的行为模式。

当提取出移动轨迹时间轴分布密集区间 $DSW=\{W_1,W_2,\cdots,W_m\}$ 之后，则可以分别对 $W_i\in DSW$ 滑动时间窗体内的轨迹点进行聚类划分，以发现时空热点区域集合 $STHSR$。需要注意的是在密集时间区间 W_i，若移动轨迹 Tr 在同一个网格单元 $Cell_j$ 内存在多个轨迹点，则只累计计算一次，这样做的目的是为了有效剔除密集时间区间内的冗余轨迹数据。

移动轨迹时空热点区域发现算法的基本步骤如下：

算法 2　STHSRD（Spatio-tempaoral hot spot region discovering algorithm）

输入：移动轨迹数据集 $D=\{Tr_1,Tr_2,\cdots,Tr_n\}$，密集时间区间集合 $DSW=\{W_1,W_2,\cdots,W_m\}$，最小支持度阈值 min_sup。

输出：时空热点区域集合 $STHSR=\{R_1,R_2,\cdots,R_l\}$。

步骤 1 初始化；

步骤 2 对每个 $W_i\in DSW$ 提取时间戳位于此区间的所有轨迹点 TrP_i；

步骤 3 对 TrP_i 中属于同一条轨迹的空间点做去重操作；

步骤 4 将处理后的 TrP_i 轨迹点平面投影，同时对平面做规则网格粒度划分，生成规模为 G 的离散空间网格单元 $Grid=\{Cell_i,i=G\}$；

步骤 5 对每个网格单元内的轨迹位置点进行统计计算，若有 $Cell_i_Num\geqslant min_sup$，即 $Cell_i$ 内的轨迹点数大于或等于最小支持度阈值 min_sup，则 $STHSR=STHSR\cup Cell_i$。

这样，就可得到已知时间区间段内的时空热点区域集合 $STHSR$。

3　基于时空热点区域的移动模式挖掘

在得到了移动轨迹的时空热点区域之后，即可通过时空热点区域集合 $STHSR$ 对原始移动轨迹进行轨迹简约。具体过程为将原始轨迹数据中的时空轨迹点 (x_i,y_i,t_i) 与 $STHSR$ 集合中的时空热点区域元素 R_i 进行匹配计算，只有当同时满足如下两个条件：①时间戳 t_i 属于 R_i 所对应的密集时间窗 $SW_i=[T'_{\text{start}},T'_{\text{end}}]$，即 $T'_{\text{start}}\leqslant t_i\leqslant T'_{\text{end}}$；②空间轨迹位置点 (x_i,y_i) 位于 R_i 所属网格空间单元内，即 $(x_i,y_i)\in R_i$，则轨迹点 (x_i,y_i,t_i) 可转换为时空热点区域 R_i 元素。反之，则认为时空轨迹点 (x_i,y_i,t_i) 或是位于移动轨迹稀疏时间区间或是位于非时空热点区域而予以剔除。

通过上述移动轨迹数据向时空热点区域的转换过程，可得到以时空热点区域 $STHSR$ 为元素的序列数据集，一方面实现了对原始移动轨迹的简约，另一方面也实现了轨迹的时空语义近似化表示。对转换后的 $STHSR$ 序列数据进行频繁模式挖掘，所得到的移动模式形式如：$R_i\rightarrow R_j$，其表示移动对象顺序通过时空热点区域 R_i 与 R_j。需要注意的是，$R_i\rightarrow R_j$ 的模式表示不仅具有空间域上的移动迁移信息，同时具有时间域上的对应区间信息。本文采用深度优先搜索策略，基于经典的序列模式挖掘算法 PrefixSpan 进行相关改进，以对 $STHSR$ 序列数据集进行频繁模式的挖掘。

基于时空热点区域的移动轨迹频繁模式挖掘算法表述如下：

算法 3　FMTPM（Frequent moving trajectory pattern mining algorithm based on STHSR）

输入：移动轨迹数据集 $D=\{Tr_1,Tr_2,\cdots,Tr_n\}$，最小支持度 min_sup，时空热点区域集合 $STHSR=\{R_1,R_2,\cdots,R_l\}$。

输出：移动轨迹频繁模式集合。

步骤 1 将原始移动轨迹数据库 D 转换为 $STHSR$ 序列数据库；

步骤 2 对 $STHSR$ 序列数据库进行扫描，得到长度为 1 的频繁序列模式 L_1；

步骤 3 对每个频繁序列模式 L_1 构建投影数据库，探测所有频繁项，生成长度为 2 的频繁序列模式 L_2；

步骤 4 对所产生的长度为 k 的序列模式 L_k 构建投影库，探测所有频繁项，生成长度为 $(k+1)$ 的频繁序列模式 L_{k+1}；

步骤 5 循环步骤 4 以产生频繁移动轨迹模式集 $PatternSet$。

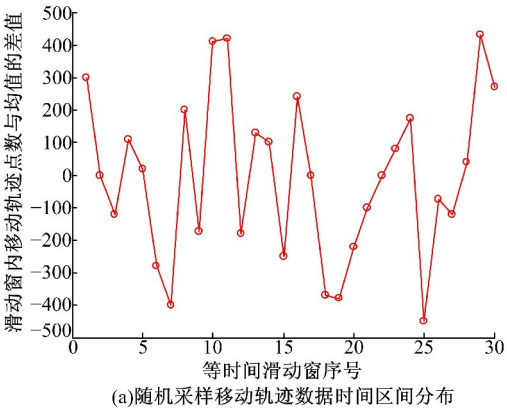
4　试验及结果分析

为了验证本文所提出方法的有效性，采用

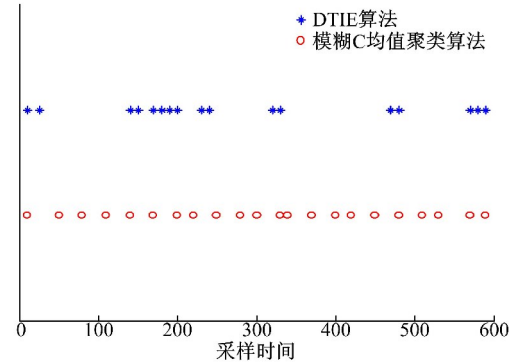
GSTD 时空轨迹数据产生算法[14-15]生成仿真数据,其可虚拟生成时空三维特征属性表示下的移动轨迹数据集合。本文在 GSTD 算法的基础之上,针对随机采样移动轨迹的条件要求,通过算法的相关调整和修改产生了试验部分所需要的仿真测试移动轨迹数据。其中数据规模为 2000～5000 条,数据采样周期为 600～1200 s,对应的数据平均长度为 86～172,自定义系数 φ 设置为 1.1。

所用试验设备为个人 PC,中央处理器为 Pentium(R) Dual-Core CPU T4400 @ 2.20 GHz,内存为 2.00 GB(1.99 GB usable),操作系统为 Windows XP 32 b。所用编程语言为 Matlab 2010 b。

图 4 为随机采样移动轨迹数据密集时间区间比较结果。其中图 4(a)为在等时间滑动窗 $\Delta T = 20$ 上统计得出的分段移动轨迹数据点与整个采样周期平均移动轨迹点的差值,由图可见,以等宽度划分的时间间隔内的移动轨迹点分布极不均匀,部分区间与平均点数相近,而部分区间与平均点数相差较大。对应的差值为正数的时间区间为密集数据区间;差值为负数的时间区间为稀疏时间区间。



(a)随机采样移动轨迹数据时间区间分布



(b)随机采样移动轨迹数据密集时间窗提取

图 4 移动轨迹密集时间区间提取试验结果

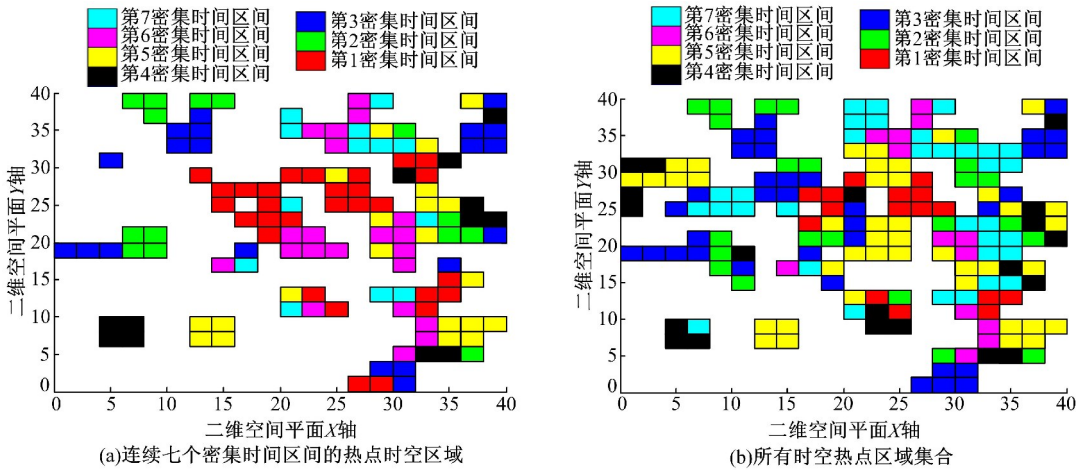
Fig. 4 Experiment of extracting intensive time interval

随机移动轨迹数据的密集时间窗提取问题,其在一定程度上属于聚类划分问题,即都是将具有相同分布特征的数据划分为同一类别,且不同聚类簇之间的差别越大越好。然而其与聚类问题所不同之处在于,密集时间窗提取只关注数据分布密集的时间区间。由于随机采样移动轨迹数据时间维分布疏密不均的特点,其既不可以通过等尺度划分的方法进行提取,亦也不可以直接通过聚类的方法进行区间提取。本文将所提出的 DTIE 算法与模糊 C 均值聚类算法进行试验仿真比较,以验证 DTIE 算法的有效性。图 4(b)为本文所提出的密集时间窗提取方法与模糊 C 均值聚类方法对上述试验数据处理后的结果,其分别为聚类中心点与密集时间区间中心点。为了在同一标准下对两种方法进行比对,模糊 C 均值聚类的聚类数与 DTIE 算法的密集时间区间数目取相同设置。由图可见,模糊 C 均值聚类所得出的时间区间分布均匀,且包含稀疏数据区间,显然难以有效抽取随机移动轨迹数据中的密集时间区间;而 DTIE 算法所得出的密集时间区间与图 4(a)所示的数据分布具有相近的分布特征,在剔除掉稀疏数据区间的同时可以有效发现密集时间区间。同时 DTIE 算法所得出的密集时间区间平均宽度为 10.23,而模糊 C 均值聚类所得出的时间区间平均宽度为 25.27,这是因为 DTIE 算法剔除了稀疏时间区间,而模糊 C 均值聚类方法无法剔除稀疏时间区间。因而 DTIE 算法在密集时间区间的检测方面具有良好的性能。

图 5 为移动轨迹时空热点区域发现算法 STHSRD 检测到的热点时空区域集合 STHSR,其中图 5(a)为连续 7 个密集时间区间所对应的热点时空区域;图 5(b)为所发现的所有时空热点区域集合。注意由于采用的网格单元颜色数为 7 个,故而后续的密集时间区间所对应的时空热点区域的网格单元颜色按照图 5(c)所示的顺序循环往复。由图可见,随着时间的推移,移动热点区域具有随时间变化的特点,这同时也反映出移动群体在不同的时段具有不同的区域停留与偏好。

图 6 为算法 3 基于时空热点区域移动模式挖掘的结果。图 6(a)为不同移动轨迹规模下,算法所挖掘的频繁模式结果与最小支持度的关系比较,其中移动轨迹规模分别为 3000 条,4000 条及 5000 条。可以看出,随着最小支持度阈值的不断

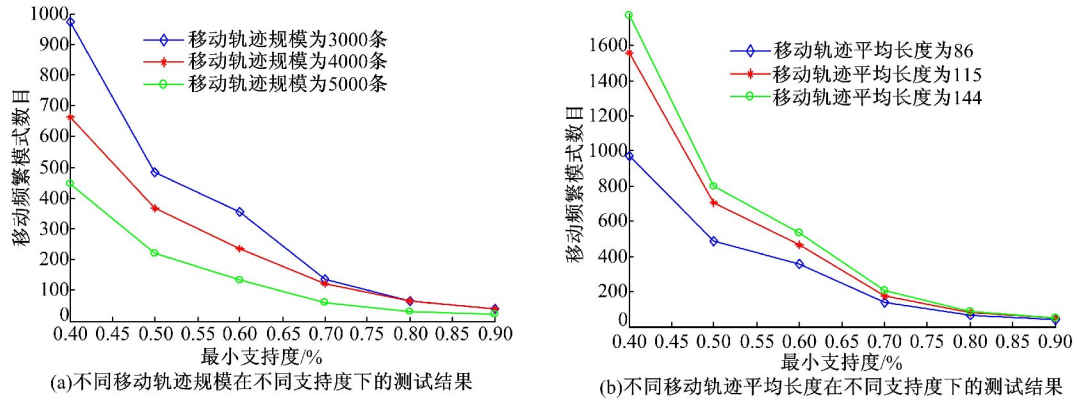
增大，所挖掘的频繁移动轨迹模式数量不断减小。同时随着轨迹规模的增大，所挖掘的移动频繁模式数目不断增加。图 6(b)为不同移动轨迹平均长度下，算法所挖掘的频繁模式结果与最小支持度的关系比较，其中移动轨迹的平均长度分别为 86，115 及 144。可以看出，随着最小支持度阈值的不断增大，所挖掘的频繁移动轨迹模式数量不断减小。同时随着移动轨迹平均长度的增大，所挖掘的移动频繁模式数目也不断增加。



(a)连续七个密集时间区间的热点时空区域　　(b)所有时空热点区域集合

图 5　时空热点区域发现试验结果图示

Fig. 5　Experiment of discovering STHSRs



(a)不同移动轨迹规模在不同支持度下的测试结果　　(b)不同移动轨迹平均长度在不同支持度下的测试结果

图 6　移动频繁模式挖掘试验结果

Fig. 6　Experiment results of mining frequent moving patterns

5　结束语

对随机采样条件下的移动轨迹模式挖掘问题进行了研究，针对随机采样移动轨迹数据在时间域密集数据区间与稀疏数据区间随机且并存的特点，通过将三维时空轨迹投影转换为一维时间数据，提出了一种基于聚类的密集时间区间检测和时空热点区域发现方法，进而以时空热点区域为基础对原始移动轨迹数据进行映射转换，最后采用深度优先搜索策略对转换序列数据中的移动频繁模式进行挖掘和发现。试验结果证明，本文所提出的方法具有良好的性能。

参考文献：

[1] Lee J W, Paek O H, Ryu K H. Temporal moving pattern mining for location-based service[J]. Journal of Systems and Software, 2004, 73(3): 481-490.

[2] 郑宇，谢幸. 基于用户轨迹挖掘的智能位置服务[J]. 中国计算机学会通讯，2010，6(6)：23-30.
Zheng Yu, Xie Xing. Enable smart location-based services by mining user trajectories[J]. Communications of the CCF, 2010, 6(6): 23-30.

[3] Cao H P, Mamoulis N, Cheung D W. Mining frequent spatio-temporal sequential patterns[C]// Fifth IEEE International Conference on Data Mining,

2005:82-89.

[4] Lee J G, Han J W, Whang K Y. Trajectory clustering: a partition-and-group framework[C]// Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Beijing, China, 2007:593-604.

[5] De Lucca Siqueira F, Bogorny V. Discovering chasing behavior in moving object trajectories[J]. Transactions in GIS, 2011, 15(5):667-688.

[6] Yanagisawa Y, Akahani J, Satoh T. Shape-based similarity query for trajectory of mobile objects[C]// Mobile Data Management, Berlin, Germany, 2003: 63-77.

[7] Giannotti F, Nanni M, Pinelli F, et al. Trajectory pattern mining[C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2007: 330-339.

[8] Kang J, Yong H S. Mining spatio-temporal patterns in trajectory data[J]. JIPS, 2010, 6(4):521-536.

[9] Wang L, Hu K, Ku T, et al. Discovering closed frequent patterns in moving trajectory database[C]// 14th International Conference on Communication Technology, Chengdu, 2012:567-572.

[10] Lee A J T, Chen Y A, Ip W C. Mining frequent trajectory patterns in spatial—temporal databases[J]. Information Sciences, 2009, 179(13): 2218-2231.

[11] Zheng Y, Lou Y, Zhang C, et al. Map-matching for low-sampling-rate GPS Trajectories[P]. U. S. Patent: 12/712, 857, 2010-02-25.

[12] Pfoser D, Jensen C S. Capturing the uncertainty of moving-object representations[C]// Advances in Spatial Databases, Berlin, Germany, 1999:111-131.

[13] Miwa T, Kiuchi D, Yamamoto T, et al. Development of map matching algorithm for low frequency probe data[J]. Transportation Research Part C: Emerging Technologies, 2012, 22:132-145.

[14] Theodoridis Y, Nascimento M A. Generating spatiotemporal datasets on the WWW[J]. ACM SIGMOD Record, 2000, 29(3):39-43.

[15] Theodoridis Y, Silva J R O, Nascimento M A. On the generation of spatiotemporal datasets[C]// In Proceedings of the 6th Int'l Symposium on Large Spatial Databases, Hong Kong, China, 1999: 147-164.