



时空轨迹序列模式挖掘方法综述

康军^{1,2*}, 黄山¹, 段宗涛^{1,2}, 李宜修¹

(1. 长安大学 信息工程学院, 西安 710064;

2. 陕西省道路交通智能检测与装备工程技术研究中心, 西安 710064)

(*通信作者电子邮箱 junkang@chd.edu.cn)

摘要: 全球定位、移动通信技术迅速发展的背景下涌现出了海量的时空轨迹数据, 这些数据是对移动对象在时空环境下的移动模式和行为特征的真实写照, 蕴含了丰富的信息, 这些信息对于城市规划、交通管理、服务推荐、位置预测等领域具有重要的应用价值, 而这些过程通常需要通过时空轨迹数据进行序列模式挖掘才能得以实现。时空轨迹序列模式挖掘旨在从时空轨迹数据集中找出频繁出现的序列模式, 挖掘时空数据中隐藏的信息, 例如: 位置模式(频繁轨迹、热点区域)、活动周期模式、语义行为模式。综述近年来时空轨迹序列模式挖掘的研究进展, 先介绍时空轨迹序列的数据特点及应用, 再描述时空轨迹模式的挖掘过程: 从基于时空轨迹序列来挖掘位置模式、周期模式、语义模式三个方面来介绍该领域的研究情况, 最后阐述现有时空轨迹序列模式挖掘方法存在的问题, 展望其未来的发展趋势。

关键词: 时空轨迹数据; 轨迹序列模式挖掘; 位置模式; 周期模式; 语义模式

中图分类号: TP311

文献标志码: A

Review of spatio-temporal trajectory sequence pattern mining methods

KANG Jun^{1,2*}, HUANG Shan¹, DUAN Zongtao^{1,2}, LI Yixiu¹

(1. School of Information Engineering, Chang'an University, Xi'an Shaanxi 710064, China;

2. Shanxi Road Traffic Detection and Equipment Engineering Technology Research Center, Xi'an Shaanxi 710064, China)

Abstract: Coincident with the rapid development of global positioning technology and mobile communication technology has been the emergence of huge amounts of trajectory data. These data are true portrayals of the moving patterns and behavior characteristics of moving objects in the spatio-temporal environment, and they contain a wealth of information which carries important application implication in the fields of urban planning, traffic management, service recommendation, location prediction, etc. These processes usually need to be achieved by sequence pattern mining of spatio-temporal trajectory data. Spatio-temporal trajectory sequence pattern mining aims to find frequently occurring sequence patterns from the spatio-temporal trajectory data set, and to mine hidden information in the spatio-temporal data, such as location patterns (frequent trajectories, hot spots), activity cycle patterns, and semantic behavior patterns. Summarize the research progress of spatial-temporal trajectory sequence pattern mining in recent years. The research progress of spatial-temporal trajectory sequence pattern mining in recent years is summarized. First, the data characteristics and application of spatial-temporal trajectory sequence are introduced. Then, the mining process of spatial-temporal trajectory pattern is described: The research situation in this field from the perspective of mining position pattern, periodic pattern and semantic pattern based on spatial-temporal trajectory sequence is introduced. Finally, the problems existing in the current spatio-temporal trajectory sequential pattern mining methods are elaborated, and its future development trends are also prospected.

Keywords: spatio-temporal trajectory data; trajectory sequence pattern mining; location pattern; periodic pattern; semantic pattern

0 引言

在全球定位、位置感知、移动通信和遥感等技术高速发展的背景下, 时空轨迹数据, 如: 车辆行驶轨迹、人群移动轨迹、动物迁徙轨迹等, 开始大量涌现。时空轨迹数据含位置、时间、速度等基本信息, 其具有时空序列性、异频采样性、数据质量差的特点^[1], 是对移动对象在时空环境下的个体移动和行为历史的真实写照。它不仅着群体对象泛在的移动行为模式与规律,

例如人群的出行活动特征、车辆移动轨迹特征, 还着城市交通流演化的内在规律。通过从海量的时空轨迹数据中发现这些模式规律, 能够为城市规划、交通管理、交通指挥、服务推荐、位置预测等领域提供新的解决思路和建设性帮助, 而这些过程通常需要通过时空轨迹数据进行序列模式挖掘来得以实现。

时空轨迹序列模式挖掘作为序列模式挖掘的一个重要研

收稿日期: 2020-10-12;

修回日期: 2020-12-03;

录用日期: 2020-12-14。

基金项目: 陕西省重点研发计划项目(2019ZDLGY03-09-01, 2019ZDLGY17-08, 2020ZDLGY09-02)

作者简介: 康军(1975-), 男, 陕西咸阳人, 副教授, 博士, 主要研究方向: 智能交通系统、交通信息工程、机器学习; 黄山(1995-), 男, 湖北孝感人, 硕士研究生, 主要研究方向: 轨迹大数据分析、频繁轨迹模式挖掘。



究分支,其可以定义为从海量的、异构的、含噪声的移动全球定位系统 (Global Positioning System, GPS)轨迹序列中提取潜在的、频繁出现的、具有价值的轨迹序列的过程。随着数据采集和存储等技术的飞速发展,时空轨迹数据呈指数级增长,出现了“数据过剩,信息匮乏”的局面。在这样的背景下,很多领域内的学者,都致力于研究和开发相关时空轨迹序列模式挖掘算法和应用程序,揭示移动轨迹数据中有价值的信息。然而在大数据的时代下,各种时空轨迹数据规模持续增长,应用需求日趋提升,现有的时空轨迹数据序列模式挖掘方法在对海量轨迹数据进行处理时存在着诸多局限性,如挖掘效率低、适用范围有限等。该领域亟待突破分布式挖掘算法,针对实时数据流的处理算法等关键技术。

本文将从时空轨迹序列模式挖掘的需求入手,从时空轨迹序列的数据特征及其应用、时空轨迹序列模式的挖掘过程及研究现状、时空轨迹序列模式挖掘的现有问题及未来发展趋势 3 个方面来阐述该领域的现状和发展。在第一方面介绍时空序列数据的基本特点及其典型的应用场景;在第二方面介绍时空轨迹序列模式的挖掘过程,并阐述近几年该领域的研究情况;最后阐述现有时空轨迹序列模式挖掘方法存在的问题,提出其未来的发展趋势。

1 时空轨迹序列的数据特征及其应用

1.1 数据特征

轨迹序列就是在时空环境下,通过对一个或多个移动对象运动过程的采样所获得的数据信息,包括采样点位置、时间、方向、速度等,这些采样点数据信息根据采样先后顺序排列构成的数据序列。时空轨迹数据可以通过带定位功能的移动智能设备、车载 GPS 设备、通信基站、社交媒体、地图 APP 等不同方式来获取,其来源多样复杂。时空轨迹数据用于分析研究移动对象随着时间特性和空间特性的变化而产生的演化规律,

是对移动对象在时空环境下的历史行为活动的真实写照。

表 1 汇总了不同方式下获取到的几种典型的时空轨迹序列数据。从表 1 可以看出,时空轨迹序列数据具有“3V”特性——量大(Volume)、实时(Velocity)、多样(Variety)。此外,移动对象的时空轨迹数据还具备一下特点:

①时空序列性。轨迹序列是一系列具有位置、时间信息的采样点序列,其体现了移动对象的时空动态性。轨迹序列的最基本特性就是时空序列性;

②异频采样性。不同的移动对象其运动轨迹的采样间隔也呈现出显著的差异。像车辆轨迹、用户移动轨迹、动物迁徙轨迹,其更新频率一般为秒/分钟级;而诸如气温、气流、降水的气象数据其更新频率为由分钟、小时到天不等。不同轨迹序列的差异采样极大影响了轨迹序列模式挖掘的过程;

③数据质量差。移动对象的时空运动轨迹是连续的,轨迹序列则是时空运动轨迹的离散化表示形式,但是其会受到 GPS 设备采样精度的影响,当某对象运动到拐角或突然停止运动时,GPS 设备采集到的轨迹点会出现位置不确定、位置偏离等问题,这就给时空轨迹序列模式挖掘的过程增加了困难。

综上所述,时空轨迹序列数据广泛存在于生活的诸多领域且蕴藏着大量的可用信息,但其数据特性给时空轨迹序列模式挖掘带来极大的挑战。

1.2 时空轨迹序列模式挖掘的应用

时空轨迹序列数据蕴含着丰富的运动信息,通过对其进行序列模式挖掘分析,可从中揭示出有价值的模式信息来服务于社会。城市规划^[2-4]、交通管理^[4]、出行服务推荐^[7]、旅游推荐^[9]、位置预测^[12]和异常轨迹行为检测^[14]等领域已经有诸多应用,如图 1 所示,并深入影响了人们的生活方式。

①城市规划。群体移动对象的运动轨迹蕴含着其在城市时空环境下的行为活动模式与规律,通过挖掘这种行为模式规律,

表 1 代表性时空轨迹序列数据

Tab.1 Representative Spatio-temporal Trajectory Sequence Data

时空轨迹序列数据种类	获取方式	更新频率	数据总量
气温,气流,降水等气象数据	气象数据采集网络	分钟/小时/天	TB 级
车辆轨迹	车载 GPS 设备	秒/分钟级	TB 级
用户移动轨迹	地图 APP、智能移动设备	秒/分钟级	TB/PB 级
动物迁徙轨迹	在动物身上佩戴 GPS 设备、遥感设备	秒/分钟级	TB 级

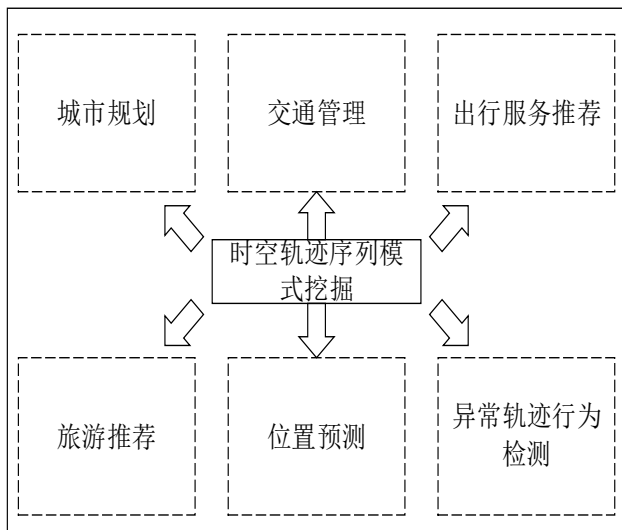


图 1 时空轨迹序列模式挖掘应用

Fig.1 Application of spatio-temporal trajectory sequence pattern mining

有利于揭示城市内部的交通流演变规律、城市不同区域的社会功能和热度特征以及不同移动对象与城市空间的占用关系,这些信息的发现能够为城市公共空间布局、土地利用、基础设施建设等规划提供辅助决策,使得城市规划更符合城市的地理环境、人文条件和经济发展状况。

②交通管理。人群和各种车辆是城市交通中最主要的活动对象,其移动轨迹数据中记录了他们日常出行的频繁历史路线,这些频繁轨迹模式的发现能够深入了解城市交通流的变化特点,还有助于发现人群和交通车辆活动范围的交集区域。这为缓解交通拥堵状况提供了机会,使交通管理人员更加合理地设置交通标志、道路交通标线及十字路口的红绿灯等候时间,为司机和行人提供更好的交通出行环境。

③出行服务推荐。出租车是人们户外出行的重要交通工具,而很多时候人们都面临着打不到车或者候车时间较长的情况。通过对海量的出租车轨迹数据进行分析,挖掘出出租车司机的经验行驶轨迹,发现出租车的频繁活动区域,这些信息能够用于帮助乘客推荐周围有效的打车地点,从而有效减少乘客的打车等候时间,提高打车概率。

④旅游服务推荐。随着在线旅游业的快速发展,旅游推荐也成为其热点服务。海量的用户移动轨迹数据反映了他们的历史活动路线。从大量用户的历史轨迹数据中发现他们的频繁移

动模式,可以揭示出用户旅游出行的频繁热点区域,其代表了大多用户对旅游地标和路线的偏好,这些信息能够为那些没有明确出行目的的旅客提供参考意见,为其进行个性化的推荐服务。

⑤位置预测。移动对象的历史轨迹数据包含了丰富的运动信息,通过挖掘出移动对象的历史频繁轨迹模式以找出其行为潜在的时空规律性及其行为偏好,再与移动对象当前的运动状况相结合,来分析确定对象在未来某个时刻或时间段的运动趋势,这样做可以有助于趋利避害、合理分配道路资源等。

⑥异常轨迹行为检测。一个异常轨迹是在一个距离度量,例如形状和旅行时间上与语料库中其他的轨迹有明显不同的一个轨迹,或者是轨迹的一部分。这种离群轨迹可能是驾驶司机的恶意绕道或意外的道路变化(由于交通事故或施工)。通过频繁轨迹模式挖掘方法,能够从轨迹数据集中检测出轨迹的异常段:如果某一轨迹(或子轨迹)不频繁,则它可能是一个异常轨迹。这种异常轨迹的发现有利于及时发现隐患并辅助决策以确保社会公共环境的财产和人身安全。时空轨迹序列模式挖掘的应用并不仅仅局限于上述几个领域,随着位置感知和定位技术的不断发展,人们对于轨迹模式挖掘的需求越来越大,应用领域也在不断扩大。

2 时空轨迹序列模式的挖掘过程及研究现状

2.1 挖掘过程

时空轨迹序列模式挖掘旨在从时空轨迹数据集中找出频繁出现的序列模式,如普遍性规律或公共性频繁路径等。而从轨迹数据中挖掘频繁的序列模式是一个多步骤相互链接、反复交互过程,如图 2 所示,指从收集原始轨迹点序列到提供可理解的轨迹模式的所有处理步骤,包括:

- (1)轨迹预处理:轨迹数据清洗,轨迹分割,轨迹压缩;
- (2)轨迹模式挖掘:从轨迹数据中发现频繁序列模式,挖掘时空数据中隐藏的信息,例如:位置模式(频繁轨迹、热点区域)、活动周期模式、语义行为模式。

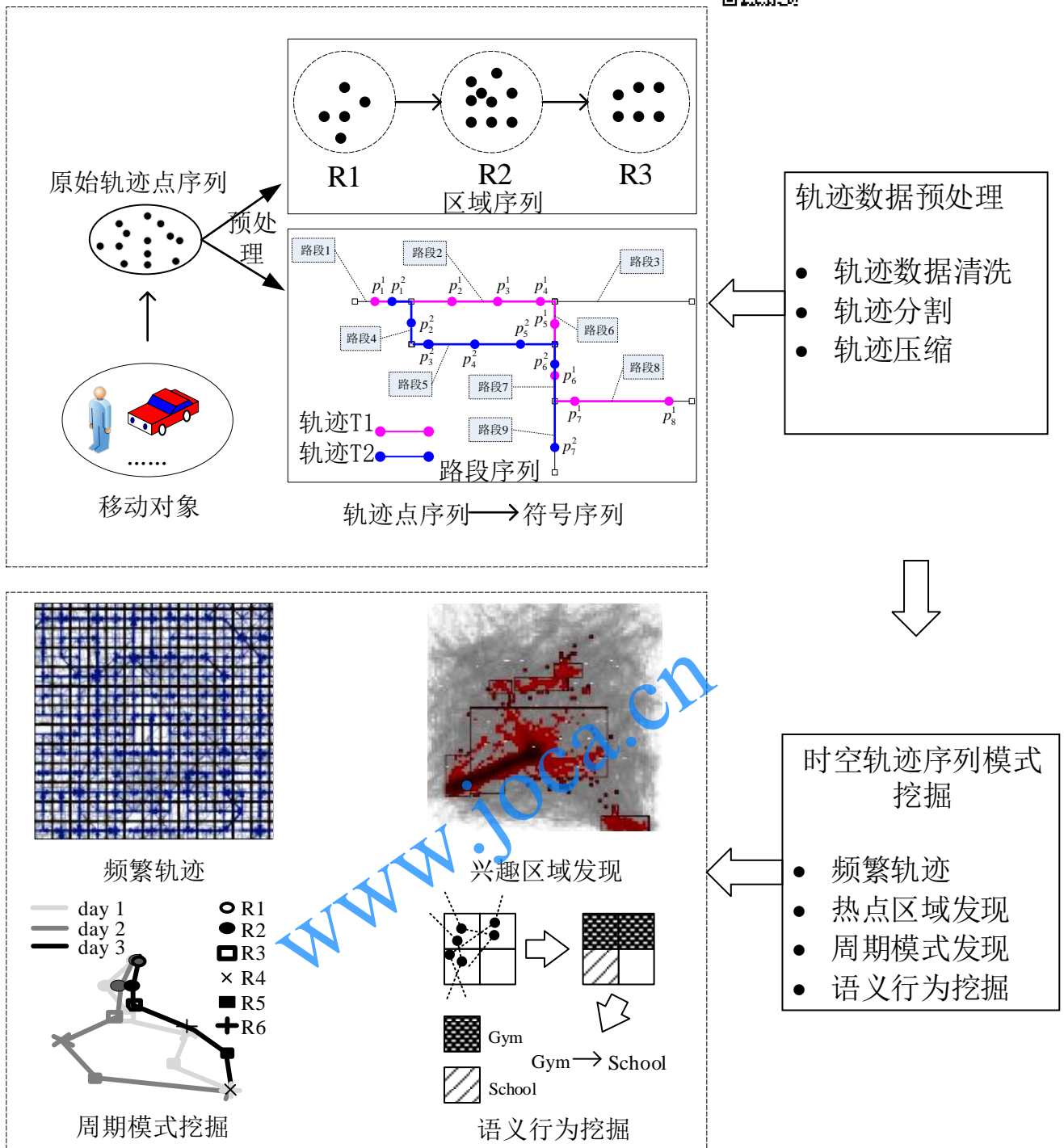


图 2 时空轨迹序列模式挖掘过程

Fig. 2 Process of spatio-temporal trajectory sequence pattern mining

其中最重要的预处理步骤是轨迹分割和轨迹压缩。因为频繁轨迹模式不仅可能出现在整个轨迹之间,而且还可能出现在子轨迹之间^[16]。并且通常采集得到的轨迹数据都是以天为单位的、按周期性间隔采样的连续点序列,其过于冗长无法直接用于挖掘分析。所以有必要基于特定标准(如时隙等)将原始轨迹分割为若干短的子轨迹。另一方面,由于移动轨迹在空间上的连续性分布,需要对其覆盖的空间区域进行离散化处理,如网格划分,聚类,以实现轨迹压缩。经过轨迹分割和空区间域离散化处理可以将原始的细粒度轨迹点序列转换为诸如网格、区域或其他形式的粗粒度符号序列,从而基于序列模式挖掘算法来发现频繁序列模

式。

2.2 研究现状

近年来,不少专家学者致力于时空轨迹序列模式挖掘方法的研究,主要包括三类,即基于时空轨迹序列的位置模式挖掘、基于时空轨迹序列的周期模式挖掘和基于时空轨迹序列的语义模式挖掘,如图 3 所示。

2.2.1 基于时空轨迹序列的位置模式挖掘

基于时空轨迹序列的位置模式挖掘是从时空轨迹数据集中发现频繁的含位置信息的序列模式,如网格序列、地理区域序列、



路段序列,故基于时空轨迹序列的位置模式挖掘又可分为基于网格划分的位置模式挖掘、基于聚类划分的位置模式挖掘、基于路网匹配的位置模式挖掘:

①基于网格划分的位置模式挖掘是将轨迹数据覆盖的区域划分为若干个规则的网格,以轨迹点所处网格标识轨迹,将冗长的轨迹点序列转化成简洁的网格序列,再基于改进的频繁模式挖掘算法来发现反复出现的网格序列。

Zhang 等^[17]提出了一种基于时间、经度、纬度三重因素的轨迹定界符将连续轨迹划分为不同子轨迹,再将目标区域进行网格划分,把子轨迹转换为网格序列,通过 Apriori 算法来挖掘频繁轨迹;Yu^[18]首先基于网格划分将原始轨迹点分配到不同网格中,在此基础上将网格建模为节点,将相邻网格间的交通流建模为具有权值的边,构建一个加权有向图。然后用一种扩展标签传播算法来识别相邻时段内有向图中的移动路径,将每条路

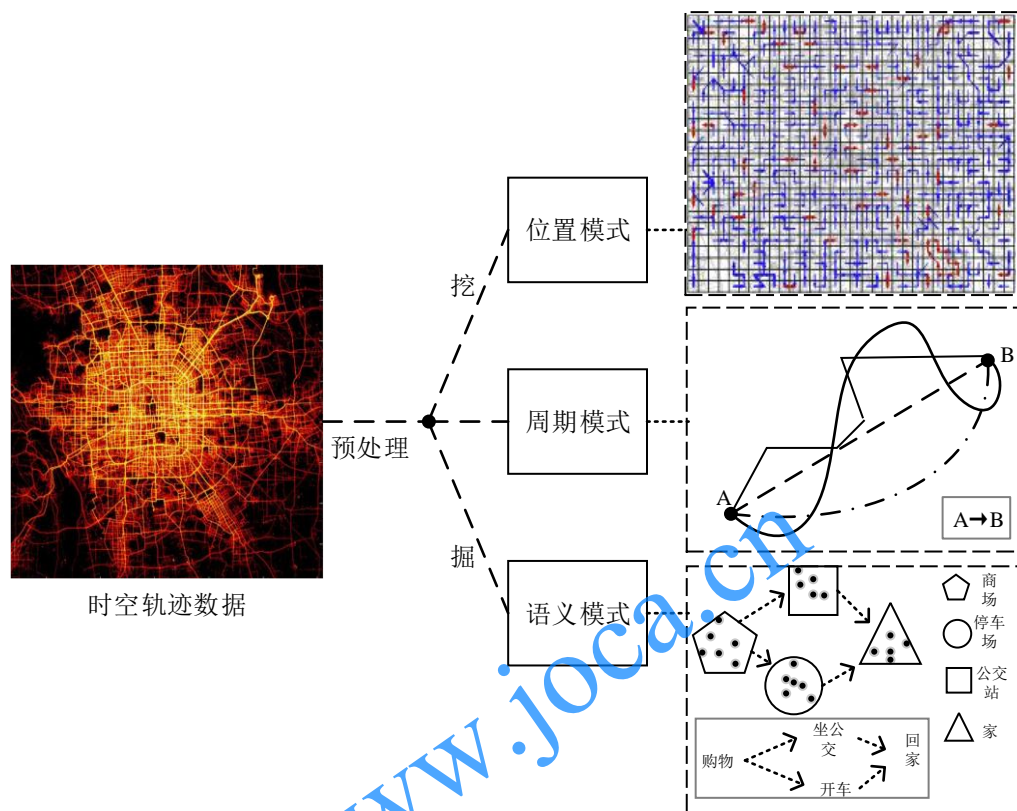


图3 时空轨迹序列模式挖掘的方法分类

Fig. 3 Method classification of spatio-temporal trajectory sequence pattern mining

径当作一条 transaction 使用 Apriori 算法来挖掘一天内不同时间段的频繁移动路径;Enami 等^[19]在对时空轨迹等时段划分的基础上,将经纬度覆盖的地理区域划分成若干网格,时空轨迹转换成具有时间间隔约束的网格序列,最后分别基于改进的 Prefixspan 和 BIDE(BI-Directional Extension paradigm)算法来挖掘频繁时空轨迹,并将结果用于预测未来的移动性。虽然文献[17]以轨迹点所处的网格离散化轨迹达到了简化轨迹的目的,但是对于网格大小的划分没有明确标准,并且网格划分过程中还存在轨迹点位于网格边界上而无法判断具体归属于哪个网格的问题,这些最终都容易导致轨迹模式丢失的情况。为了规避此问题,Wang 等^[20]提出了一种模糊空间划分方法 VSP(Vague Space Partition)将研究区域划分一组模糊网格单元,并通过基于距离的隶属函数把原始轨迹位置可以转换为具有不设计了两种挖掘算法,即 VTPM-PrefixSpan(Vague space partition Trajectory Pattern Mining-PrefixSpan)算法和 VTPM-GSP(Vague space partition Trajectory Pattern Mining-Generalized Sequential Patterns)算法,以发现具有严格时间约束

的频繁轨迹模式;Yi 等^[21]在此基础上又提出了一种基于垂直投影距离的网格划分方法 VGS(Vague Grid Sequence method based on vertical projection distance),根据垂直投影距离将每个网格划分为显式区域和模糊区域;然后将轨迹变换为模糊网格序列;最后提出 VGS-PrefixSpan(Vague Grid Sequence method based on vertical projection distance-PrefixSpan)算法,在 PrefixSpan 算法的基础上,通过记录轨迹序列中项的位置来减少构造投影数据库的时间。

②基于聚类划分的位置模式挖掘即把目标研究区域聚类为若干不相重叠的区域,如热点区域^[22]、停留点区域^[23],以区域序列来表示轨迹,再通过序列模式挖掘从中挖掘频繁轨迹模式。文献[24]先对原始轨迹数据所处跨度区域划分为若干子区域,通过 k-means 算法将所有子区域聚类为多个簇群,并根据各簇所含轨迹数以降序排列来提取热点簇群,最后从这些热点簇群中挖掘频繁轨迹模式;文献[25]中提出一种基于时空热点区域的 FMTMP(Frequent Moving Trajectory Pattern Mining)移动模式挖掘算法,通过动态聚类方式提取移动轨迹密集时间区间,在



不同时间区间对轨迹点进行聚类划分以识别时空热点区域,实现移动轨迹序列向时空热点区域序列的转换;文献[26]中提出一种基于 3D 网格的时空频繁轨迹挖掘方法。识别出轨迹数据覆盖范围内的停留区域,用密度聚类算法将轨迹数据转化为停留区域序列并将其映射到水平坐标为经纬度空间信息、垂直坐标为单一粒度时间信息的 3D 网格中,以挖掘频繁轨迹模式;文献[27]中提出基于从社交网站提取的用户轨迹数据利用 Apriori-like 算法发现频繁轨迹模式的方法,其主要是通过 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)算法识别出用户日常活动区域中的频繁访问区域,将原始数据转换为时序排列的频繁区域序列来实现的。

虽然文献[25-28]通过不同的聚类方式来简化原始轨迹序列的表达形式,缩减轨迹序列的长度,以降低了挖掘过程的复杂度,但是聚类算法也存在时间复杂度较高的问题,同时对于各聚类簇内部的序列模式并不能很好地挖掘。

③基于路网匹配的位置模式挖掘是通过假设研究对象被限制在目标区域的相关道路网络上移动,通过路网匹配算法将轨迹点映射到路网中,使其转化为具有路网含义的路段序列,以便于进行频繁轨迹模式挖掘。

文献[28]通过提取道路网络中的关键位置、特征点以表示原始轨迹。王亮等^[28]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。Wang 等^[29]研究了道路网络中的个人轨迹频繁模式挖掘。利用拐角检测技术来识别嵌入在个人 GPS 轨迹中的物理道路拓扑信息,从 GPS 轨迹中提取特征点,将轨迹变换为道路网络中有序的拐角序列,来挖掘频繁轨迹。Bermingham 等^[30]提出了一种 DC-SPAN(Distinct Contiguous Sequential Patterns)算法从车辆轨迹中来挖掘轨迹模式。通过假设车辆被限制在研究区域的相关道路网络上行驶,使用地图匹配将轨迹映射到路网中,将原始轨迹转换为道路节点访问序列,以挖掘连续轨迹模式。

2.2.2 基于时空轨迹序列的周期模式挖掘

基于时空轨迹序列的位置模式挖掘更多侧重于发现移动对象频繁经过的轨迹路线,然而移动对象的行为在很多时候还呈现出一定的周期性规律,即移动对象总是会在特定时间间隔内或多或少遵循相同的运动轨迹,例如,鸟类定期的迁徙行为,人们每天按固定的路线上下班,商业客机按固定的时间表从一个地方到另一个地方等。对于这种周期性轨迹模式的发现同样很有研究价值,其可用于压缩运动数据,预测对象未来的运动趋势。

文献[31]中提出了一种基于 Traclus 的融合轨迹点方向、速度、时间等时空语义信息的层次轨迹聚类方法用于时空周期性模式挖掘,克服了传统周期模式挖掘方法其忽略了序列和固有层次结构的缺陷,但是其未能考虑轨迹点的语义背景信息,同时也无法处理不规则轨迹;文献[32]针对[31]算法的缺陷,在从不规则采样的时空轨迹中检测参考点时,综合考虑所有空间性、时间性、语义和层次结构,提出一种分层语义周期性模式挖掘方

法,以找到分层周期性模式。

Yang 等提出一种两阶段方法:阶段一,使用运动参数对周期周期性行为建模。阶段二,基于循环周期模式特点,提出了轨迹距离矩阵搜索算法 TDMS(Trajectory Distance Matrix Search)用于识别周期性轨迹和提取周期轨迹。利用 Delaunay 三角剖分和逆向地理编码方法,从收集的周期轨迹中提取运动位置信息^[33];Shi 等^[34]提出了一种 GPMine(Group Periodic moving pattern Mining)算法用于挖掘群体对象的周期性运动模式;文献[36]则研究了周期模式挖掘中面临不确定性时空数据采样频率及数据稀疏性的问题。

上述的周期模式挖掘方法大都只专注于发现特定时空范围内具有不同周期长度的模式,与固定时空粒度的周期模式相比,不同时空粒度下的周期模式则具有更大的实用价值,这样的模式可以更准确地描述个人行为。文献[37]引入多粒度模型来描述对象活动的时空信息,并通过空间优先和时间优先的多粒度活动发现算法,来识别出不同的粒度下对象的活动行为。最后提出一种多粒度周期性活动发现算法 MPAD(Multi-granularity Periodic Activity Discovery)来挖掘物体活动的周期性模式;文献[39]中提出了一个发现具有不同时空粒度下周期模式的框架:首先通过轨迹抽象将轨迹点序列转换为具有位置信息的时间序列。然后,根据时空信息定义了一个多粒度行为模型。最后,提出了 SBPM(Single Behavior Period Mining)和 MBPM(Multi-Behavior Period Mining)算法,SBPM 可以在不知道周期长度的情况下发现单个行为周期模式,MBPM 可以从单个行为模式中识别多个行为的周期模式。

2.2.3 基于时空轨迹序列的语义模式挖掘。

从时空轨迹序列中挖掘的位置模式和周期性模式往往只能反映移动对象频繁经过的位置信息或在某些特定时间内反复遵循的路线,却无法捕获轨迹的语义信息。然而轨迹数据所蕴含的丰富语义信息,如其所在的场景地理知识、在某处的行为模式与状态等,能更好地反映与体现真实环境,通过从时空轨迹数据集中挖掘出的语义轨迹模式更具有实际应用性。Chen 等^[39]中定义了一种语义轨迹模式,即具有空间、时间和语义属性的移动模式。通过从原始轨迹数据中提取出这三种属性,将其转换为语义轨迹序列,再采用基于匹配的 MB 算法和基于 Prefixspan 的 PS 算法来发现语义模式,以预测用户的移动行为。Comito 等^[40]通过基于密度的聚类算法将 GPS 轨迹点聚类到具有实际语义的位置区域中,实现轨迹序列到语义位置区域序列的转换,再运用改进的 Prefixspan 算法来挖掘转换后具有时间间隔约束的序列,使其只获得最大频繁模式,并将所得结论用于分析人们的移动行为。

文献[42]中提出了一种基于停留时间的语义行为模式挖掘方法。先将原始轨迹转化为停留点区域序列并赋予其实际的区域背景信息,构建停留语义行为集,再结合 Prefixspan 算法挖掘频繁语义行为模式,通过定义语义行为模式之间的相似性度量方法以找出具有相似行为模式的群体移动对象;文献[43]中提出了 SMOPAT(Semantic MOBility PATterns)算法来从私家车轨迹数据中挖掘时空语义移动模式,以理解司机的行为目的;文



献[44]中提出了一种基于语义轨迹模式挖掘的路线推荐系统,该系统通过从带有地理标签的照片中提取轨迹,用所需的语义去扩展轨迹来构建语义轨迹集并挖掘语义轨迹模式,这些语义轨迹模式将连同用户的各种需求来生成一组满足用户指定约束和目标的语义轨迹路线;文献[45]中使用丰富的上下文语义注释扩展了从社交媒体获取的用户时空轨迹数据,并提出一种语义轨迹模式挖掘算法来挖掘基本和多维语义轨迹模式以深入理解用户的移动行为模式。

3 时空轨迹序列模式挖掘的现有问题及发展趋势

关于移动对象的时空轨迹序列模式挖掘已经得到许多专家学者的广泛研究,并且开发出许多时空轨迹序列模式挖掘的方法,然而大部分算法只是从理论上进行分析,没有投入到实际应用中,适用性也差。现有的时空轨迹序列模式挖掘技术仍然存在诸多局限性,具体包括以下几个方面:

1)对于移动时空轨迹序列模式挖掘的研究没有形成一个完整的、适应性强的理论框架和模型,同时缺乏必要的相关模式定义、分类方法以及处理流程规范,没有一个统一的开发平台。

2)在实际的挖掘过程中,大多数算法仍然沿用着在单一时空粒度下挖掘频繁轨迹模式的方式,这样得到的结论往往不能够充分体现与反映移动对象的活动规律及模式。因此,今后更多的研究工作应基于多维时空粒度来挖掘时空频繁轨迹模式中蕴含的群体活动规律语义信息。

3)信息时代下时空轨迹数据无时无刻不在增加,而现有的研究方法仅仅只是基于移动对象的历史轨迹数据来进行模式挖掘,一旦轨迹数据集发生改变,整个挖掘过程将要重新开始,这样会造成巨大的计算开销。

4)大数据时代下,时空轨迹数据日益激增,大多现有的轨迹模式挖掘算法都是基于单机环境的,在处理海量的时空轨迹数据时,效率低下且算法各方面性能遭遇瓶颈,难以满足大量时空轨迹数据的挖掘需求。

5)由于GPS设备的采样频率基本都是秒级的,因此获取的时空轨迹数据具有实时更新的特性,从而使其呈现出数据流式的结构。现有的对于实时更新的轨迹数据流进行序列模式挖掘的技术方法还不够成熟,基于时空轨迹数据流的序列模式挖掘方法还有待深入研究。

综上所述,时空轨迹序列模式挖掘的研究工作今后应更多地聚焦于以下几个发面:

1)基于多维时空粒度的轨迹序列模式挖掘方法。由于时空信息是移动对象行为活动的两个最基本属性,该方法通过将移动对象的轨迹所对应的时空信息进行层次划分,形成多维的时空粒度,再挖掘出移动对象在所有时空粒度下的频繁位置模式,从而得到移动对象的活动规律,为位置感知和个性化位置等服

务提供相应的技术支持。

2)基于增量式的时空轨迹序列模式挖掘方法。该类方法是专门用于规避一旦数据库原始数据发生改变,对改变后的整个数据集重新挖掘而造成的时空消耗,其只需要通过对变化部分的数据挖掘,并在原有所得轨迹序列模式的基础上进行更新就能够挖掘得到每次数据集改变后的轨迹序列模式,从而节省大量计算开销,更具灵活性和适应性。

3)基于分布式的时空轨迹序列模式挖掘方法。这类方法旨在通过将规模巨大的轨迹数据集分散在不同的服务器上存储,然后每台机器对分布在本机上的数据进行挖掘,最后将所有机器的挖掘结果汇总整合,具体可以借助于并行编程模型如Spark、MapReduce等来实现,以适应对于海量轨迹数据的挖掘需求。由于整体的工作量是不可预测的,因此如何在服务器之间均匀地分配任务是非常棘手的。同时,并行挖掘需要各个机器之间的信息交互,通信IO就无可避免。这些问题都将是分布式的轨迹序列模式挖掘方法即将面临的难题。

4)基于实时轨迹数据流的模式挖掘方法。轨迹数据流是指那些数据量非常巨大的,无法全部存放在存储介质上进行分析 and 计算的数据。其特点是数据持续到达,且速度快、多变化、规模宏大;基于实时轨迹数据流的模式挖掘方法其核心是设计高效的单遍数据集扫描算法,在一个远小于数据规模的内存空间里不断更新一个代表数据集的结构,使得在任何时候都能根据这个结构迅速挖掘出轨迹序列模式。目前已经有了一些流式处理框架如Spark Streaming、Flink、Storm等,所以可以以这些框架为基础来设计基于轨迹数据流的挖掘方法。

4 结语

时空轨迹序列模式挖掘是在全球定位技术与位置感知技术快速进步的背景下催生出的热门研究领域之一,已经成为分析移动对象行为模式规律的重要手段。为此,各领域业界人士开展广泛的研究,相继提出针对移动对象时空轨迹的序列模式挖掘方法,其主要包括从时空轨迹数据中挖掘位置模式、周期模式和语义模式三类方法,并在城市规划、交通管理、服务推荐等领域取得诸多应用。但是随着GPS采集技术、位置感知技术等的不断完善,时空轨迹数据规模呈现出爆炸式增长,现有的技术方法在算法效率、适应性等方面已经无法满足实际的应用需求。未来,该领域的研究工作重点可能主要在以下方面:①基于多维时空粒度的时空轨迹序列模式挖掘;②基于增量式的时空轨迹序列模式挖掘;③基于分布式的时空轨迹序列模式挖掘;④基于时空轨迹数据流的序列模式挖掘。

参考文献(References)

- [1] XU J J, ZHENG K, CHI M M, et al. Trajectory big data: Data, applications and techniques [J]. Journal on Communications, 2015,36(12): 97-105



- [2] LIU X, GONG L, GONG Y X, et al. Revealing travel patterns and city structure with taxi trip data [J]. *Journal of Transport Geography*, 2015, 43: 78-90.
- [3] MAO F, JI M H, LIU T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data [J]. *Frontiers of Earth Science*, 2016, 10(2): 205-221.
- [4] HUANG W, LI S N. An approach for understanding human activity patterns with the motivations behind [J]. *International Journal of Geographical Information Science*, 2019,33(2): 385-407.
- [5] CUI J X, LIU F, JANSSENS D, et al. Detecting urban road network accessibility problems using taxi GPS data [J]. *Journal of Transport Geography*, 2016, 51: 147-157.
- [6] ANDRIENKO N, ANDRIENKO G, RINZIVILLO S. Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics [J]. *Journal of Intelligent Information Systems*, 2016, 57(1): 172-194.
- [7] TU W, CAO R, YUE Y, et al. Spatial variations in urban public ridership derived from GPS trajectories and smart card data[J]. *Journal of Transport Geography*, 2018, 69: 45-57.
- [8] KONG X J, XIA F, WANG Ji Z, et al. Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data [J]. *IEEE Transactions on Industrial Informatics*, 2017, 13(3): 1202-1212.
- [9] ZHANG S, TANG J J, WANG H X, et al. Revealing intra-urban travel patterns and service ranges from taxi trajectories [J]. *Journal of Transport Geography*, 2017, 61: 72-86.
- [10] SUN Y R, FAN H C, BAKILLAH M, et al. Road-based travel recommendation using geo-tagged images [J]. *Computers, Environment and Urban Systems*, 2015, 53: 110-122.
- [11] TSAI C Y, LAI B H. A Location-Item-Time sequential pattern mining algorithm for route recommendation [J]. *Knowledge-Based Systems*, 2015, 73: 97-110.
- [12] CHEN C, MA J T, SUSILO Y, et al. The promises of big data and small data for travel behavior(aka human mobility)analysis [J]. *Transportation Research Part C: Emerging Technologies*, 2016, 68: 285-299.
- [13] CHEN M, YU X H, LIU Y. Mining moving patterns for predicting next location [J]. *Information Systems*, 2015, 54: 156-168.
- [14] WU R Z, LUO G C, SHAO J M, et al. Location prediction on trajectory data: A review [J]. *Big Data Mining and Analytics*, 2018, 1(2): 108-127.
- [15] LEI P R. A framework for anomaly detection in maritime trajectory behavior [J]. *Knowledge and Information Systems*, 2016, 47(1): 189-214
- [16] JONITZ D, BUCHER D. Continuous Trajectory Pattern Mining for Mobility Behaviour Change Detection [C] // *LBS 2018: Proceedings of 14th International Conference on Location Based Services*. Cham: Springer, 2018: 211-230.
- [17] KARSOUM S, GRUENWALD L, LEAL E. Impact of Trajectory Segmentation on Discovering Trajectory Sequential Patterns [C] // *Proceedings of the 2018 IEEE International Conference on Big Data*. Washington, DC: IEEE Computer Society, 2018: 3432-3441.
- [18] ZHANG Y L, PENG P F, LIU J S, et al. AIS Data Oriented Ships' Trajectory Mining and Forecasting Based on Trajectory Delimiter [C] // *Proceedings of the 10th International Conference on Intelligent Human-Machine Systems and Cybernetics*. Piscataway: IEEE, 2018: 269-273.
- [19] YU W H. Discovering Frequent Movement Paths From Taxi Trajectory Data Using Spatially Embedded Networks and Association Rules [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(3): 855-866.
- [20] ENAMI S, SHIOMOTO K. Spatio-temporal human mobility prediction based on trajectory data mining for resource management in mobile communication networks [C] // *Proceedings of the IEEE 20th International Conference on High Performance Switching and Routing*. Piscataway: IEEE, 2019: 1-6.
- [21] WANG L, HU K Y, KU T, et al. Mining frequent trajectory pattern based on vague space partition [J]. *Knowledge-Based Systems*, 2013,50: 100-111.
- [22] YI C, YUAN P S, QIU M, et al. An Indoor Trajectory Frequent Pattern Mining algorithm Based on Vague Grid Sequence [J]. *Expert Systems with Applications*, 2019,118: 614-624.
- [23] ZHAO P X, QIN K, YE X Y, et al. A trajectory clustering approach based on decision graph and data field for detecting hotspots [J]. *International Journal of Geographical Information Science*, 2017, 31(6): 1101-1127.
- [24] LUO T, ZHENG Y W, XU G L, et al. An Improved DBSCAN Algorithm to Detect Stops in Individual Trajectories [J]. *International Journal of Geoinformation*, 2017, 6(3): 46-63.
- [25] SHAW A A, GOPALAN N P. Finding frequent trajectories by clustering and sequential pattern mining [J]. *Journal of Traffic and Transportation Engineering*, 2014, 1(6): 393-403.
- [26] 王亮, 胡琨元, 库涛, 等. 随机采样移动轨迹时空热点区域发现及模式挖掘[J]. *吉林大学学报(工学版)*, 2015, 45(3): 913-920.(WANG L, HU K Y, KU T, et al. Discovering spatiotemporal hot spot region and mining patterns from moving trajectory random sampling [J]. *Journal of Jilin University(Engineering and Technology Edition)*, 2015, 45(3): 913-920.)
- [27] YIN F, LI X, YAO C, et al. Mining Frequent Spatio-Temporal Items in Trajectory Data[J]. *International Journal of Database Theory and Application*, 2015, 8(4): 149-156.
- [28] HUANG Q Y, LI Z L, LI J, et al. Mining Frequent Trajectory Patterns from Online Footprints [C] // *Proceedings of 7th ACM SIGSPATIAL International Workshop on GeoStreaming*. New York: ACM, 2016: 1-7.
- [29] 王亮,胡琨元,库涛, 等.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[J].*自动化学报*,2015,41(1): 47-58.(WANG L, HU K Y, KU T, et al. Mining Urban Moving Trajectory Patterns Based on Multi-scale Space Partition and Road Network Modeling [J]. *Journal of Acta Automatica Sinica*, 2015,41(1): 47-58.)
- [30] WANG T B, ZHANG D Q, ZHOU X S, et al. Mining personal frequent routes via road corner detection [J]. *IEEE Transactions on Systems Man and Cybernetics Systems*, 2016, 46(4): 445-458.
- [31] BERMINGHAM L, LEE I. Mining distinct and contiguous sequential patterns from large vehicle trajectories [J]. *Knowledge-Based Systems preprint*, *Knowledge-Based Systems*, 2019, 189: 1-12.



- [32] ZHANG D Z, LEE K, Lee I. Hierarchical trajectory clustering for spatio-temporal periodic pattern mining [J]. Expert Systems with Applications, 2018, 92: 1-11.
- [33] ZHANG D Z, LEE K, Lee I. Mining hierarchical semantic periodic patterns from GPS-collected spatio-temporal trajectories [J]. Expert Systems with Applications, 2019, 122: 85-101.
- [34] YANG W, LU W, AI T H, et al. Cycle Periodic Behavior Detection and Sports Place Extraction Using Crowdsourced Running Trace Data [C] // Proceedings of the 26th International Conference on Geoinformatics. Piscataway: IEEE, 2018: 1-5.
- [35] SHI T T, JI G L, LIU Y, et al. Mining Group Periodic Moving Pattern from Spatio-Temporal Trajectories [C] // Proceedings of the 7th International Conference on Advanced Cloud and Big Data (CBD). Piscataway: IEEE, 2019: 108-113.
- [36] 李海. 基于 GPS 轨迹的周期模式发现[J]. 电子设计工程, 2015,23(21): 24-27.(LI H. Periodic pattern mining based on GPS trajectories [J]. Journal of Electronic Design Engineering, 2015, 23(21): 24-27.)
- [37] CHEN X P, SHI D X, ZHAO B H, et al. Periodic Pattern Mining Based on GPS Trajectories [C] // Proceedings of the 2016 International Symposium on Advances in Electrical, Electronics and Computer Engineering. Paris: Atlantis Press, 2016: 181-187.
- [38] YUAN G, ZHAO J, XIA S X, et al. Multi-granularity periodic activity discovery for moving objects [J]. International Journal of Geographical Information Science, 2017, 31(3): 435-462.
- [39] ZHOU K C, TIAN Z S, YANG Y W. Periodic Pattern Detection Algorithms for Personal Trajectory Data Based on Spatiotemporal Multi-Granularity [J]. IEEE Access, 2019, 7: 99683-99693.
- [40] CHEN C C, KUO C H, PENG W C. Mining Spatial-Temporal Semantic Trajectory Patterns from Raw Trajectories [C] // Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW). Piscataway: IEEE, 2015: 1019-1024.
- [41] COMITO C, FALCONE D, TALIA D. Mining human mobility patterns from social geo-tagged data [J]. Pervasive and Mobile Computing, 2016, 23: 91-107.
- [42] 郭黎敏, 高需, 武斌, 等. 基于停留时间的语义行为模式挖掘[J]. 计算机研究与发展, 2017, 54(1): 111-122.(GUO L M, GAO X, WU B, et al. Discovering Common Behavior Using Staying Duration on Semantic Trajectory [J]. Journal of Computer Research and Development, 2017, 54(1): 111-122.)
- [43] WAN C C, ZHU Y M, YU J D, et al. SMOPAT: Mining semantic mobility patterns from trajectories of private vehicles [J]. Information Sciences, 2018, 429: 12-25.
- [44] CAI G C, KYUNGMI L, ICKJAI L. Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos [J]. Expert Systems with Applications, 2018, 94: 32-40.
- [45] CAI G C, KYUNGMI L, ICKJAI L. Mining Semantic Trajectory Patterns from Geo-Tagged Data [J]. Journal of Computer Science and Technology, 2018, 33(4): 849-862.
- [46] This work is partially supported by the Funds for Key Research and Development Plan Project of the Shaanxi Province (2019ZDLGY03-09-01, 2019ZDLGY17-08, 2020ZDLGY09-02).
- KANG Jun**, born in 1975, Ph. D., associate professor. His research interests include artificial intelligence and machine learning.
- HUANG Shan**, born in 1995, M. S. candidate. His research interests include trajectory big data analysis and frequent trajectory pattern mining.
- DUAN Zongtao**, born in 1977, Ph. D., professor. His research interest includes context-aware computing in transportation.
- LI Yixiu**, born in 1996, M. S. candidate. His research interests include big data spatiotemporal trajectory analysis and map matching.