

轨迹数据挖掘城市应用研究综述

牟乃夏^{1,2}, 张恒才², 陈洁², 张灵先¹, 戴洪磊¹

(1. 山东科技大学 山东省基础地理信息与数字化技术重点实验室, 青岛 266590;

2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101)

摘要 轨迹数据作为泛在地理信息环境中社会遥感数据的主要表现形式之一, 为从个体的视角研究群体的空间移动规律, 提供了新的数据支撑和研究思路。特别是在当前的大数据背景下, 通过轨迹数据发掘人类的移动规律和活动模式, 进而探求蕴含的深层次知识, 是解决城市问题的重要途径, 轨迹数据挖掘也由此成为地理信息科学及相关学科的研究热点。本文首先阐述了人类移动规律研究常用的轨迹数据集及在该数据集上开展的相关研究和典型应用; 然后从城市空间结构功能单元的识别及城市韵律分析、人类活动模式的发现与空间移动行为预测、智能交通的时间估算与异常探测、城市计算的其他4个方面, 综述了轨迹数据挖掘在城市中的应用; 最后, 指出了轨迹数据挖掘面临的挑战和进一步的发展方向。

关键词 轨迹; 数据挖掘; 城市计算; 人类移动; 人类活动模式

DOI:10.3724/SP.J.1047.2015.01136

1 引言

人地关系一直是地理学研究的核心问题, 人本化的趋势使研究视角从宏观走向微观, 更加侧重于从个体角度来研究空间问题。随着传感器网络、移动定位、无线通讯、移动互联网、高性能计算与存储技术的快速发展, 数据采集与计算单元的外延不断延伸, 加之地球电子皮肤的逐渐形成、人人都是传感器概念的付诸实施, 特别是3G/4G技术的广泛使用, 以手机为代表的智能移动终端的普及, 使时空数据获取的粒度不断细化, 加速了地理信息的泛化过程^[1]。长期以来, 个体数据依靠传统的活动日志调查等手段获取, 精度低且缺少对时空活动轨迹的连续和完整的描述, 难以大规模、长时间地观测和记录人的空间移动行为^[2]。泛在网络环境下智能移动设备的普及, 使得为用户提供无处不在的位置服务成为可能。同时, 这些智能设备也详细记录了个体在真实世界中的活动轨迹, 如实反映了人们的生活与行为模式。触手可及的轨迹数据给地理学中基于个体视角的研究提

供了新的数据源和研究契机。

轨迹数据不仅记录了人在时间序列上的位置, 也隐喻了人与社会的交互、人在地域上的活动, 乃至人与人之间的关系等社会属性。单一对象的活动反映了个体自身的行为特征, 群体的活动反映了该群体共同的行为特征, 而同一城市大量移动对象的活动, 则反映了该城市总体的社会活动特征^[3]。所以, 轨迹中蕴含的知识也是目前智慧城市、城市计算^[4]、社会遥感等认识城市行为、优化城市决策不可或缺的因素。通过轨迹数据挖掘发现隐含的知识, 探求深层次的城市动力学机制, 也是解决城市交通、城市环境、突发事件应急等重大社会问题的有效手段。

2 人类移动轨迹研究常用数据集

轨迹数据是带有时间戳标记的一系列位置的集合。目前, 已有多个数据集应用到人类移动规律

收稿日期 2015-04-29; 修回日期: 2015-08-06.

基金项目 : 山东省“泰山学者”建设工程专项经费; 资源与环境信息系统国家重点实验室开放基金; 国家自然科学基金项目(41271408, 41101149); 国家“863”计划项目(2013AA120305)。

作者简介 牟乃夏(1973-), 男, 博士, 副教授, 山东平度人, 研究方向为时空信息挖掘与推荐。E-mail: mounaixia@163.com

的研究探索中,其中应用较广泛的是出租车、手机和志愿者数据集等。这些数据集在时间、空间的尺度和粒度上差别较大,分别适用于不同的研究。

人类移动规律研究的数据集及其典型应用如表1所示。

3 轨迹数据挖掘的城市应用

轨迹数据体现个体的活动规律,蕴含个体的行为模式、活动方式、活动范围及社会网络关系等特征。个体群组的群体移动轨迹可推断出城市的人群聚集规律,体现着城市的韵律与动力机制,这是解决城市问题的核心所在。所以,轨迹数据在城市规划、城市交通、城市动态、城市计算、公共安全等领域得到广泛的应用。

3.1 城市功能单元识别与城市韵律分析

城市人群移动的时空规律反映了城市人口在时间和空间上的动态变化,体现城市的韵律,隐含城市的功能分区,表征城市活动的动力学机制。城市功能区划分的目的是识别出工业、商业、住宅等不同的城市活动单元。传统上一般采用遥感影像识别的方法,这对于识别土地利用类型是比较有效的,但却无法识别出人类活动产生的功能单元^[22],如职住区的区别等。轨迹隐藏着丰富的人群活动模式,不同城市功能区的人流变化模式不同^[8],依据人群在特定空间上的活动强度和时空序列模式,通过聚类或者分类方法,能有效地识别出不同的活动单元。但这种识别方法精度不高,因为:(1)同一个地

理单元会存在多种不同类型的功能分区,如商住区就是商业和居住区的混合等;(2)城市功能分区本身就存在一定的模糊性和不确定性。但可通过POI语义化信息来精化土地利用类型的识别结果^[23-24],以提高城市功能分区识别的精度。

不仅在城市尺度上通过轨迹数据可进行有效的功能区划分。在城域尺度上,刘瑜等^[25-28]利用手机通话数据,通过社区分割等算法拟合出符合某种条件的边界,这种边界和行政区域具有高度的相关性,并证明了这种相关性除了空间近邻关联外,还与文化、经济和历史等各种因素有关。

轨迹数据时空序列的规律性波动,表达了城市的动力与韵律,职住分离与通勤是城市韵律的表现形式,轨迹数据为城市职住分析与城市通勤研究提供了新的思路。特别是手机数据由于数据量大、普及面广,在统计规律上比交通出行行为的问卷调查更具有优势,因而得到广泛的应用^[29]。基于手机数据进行职住地分析,一是根据特定时间段内的通话频度来识别;二是依据通话的位置分布特征并结合活动时间来判断。Ratti等^[30]使用手机的位置信息研究了米兰的城市活动强度及其时空分布特征,Calabrese等^[31]利用手机网络数据进行了罗马城市活动的实时监测并研究了城市的活动强度^[10, 32-33]。手机数据尽管总体上数据量很大,但是数据的时间和空间粒度较粗,难以刻画城市动态的精细特征。

除使用GPS和手机定位的轨迹数据外,公交智能卡的刷卡数据具有连续性好、信息全面且动态更新等优点,也开始用于城市空间结构方面的研究。龙瀛等^[19]依据北京市850万张公交卡的刷卡记录结

表1 人类移动研究中常用的轨迹数据集
Tab. 1 The trajectory data sets on human mobility study

轨迹数据类型	数据集情况	主要研究人员	典型工作
出租车轨迹	包含出租车位置、速度、载客状态	Lu ^[5]	智能交通
		Zheng ^[6-7]	环境监测、旅游推荐
		Pan ^[8]	城市功能区识别
手机数据	另包含出租车收费数据等	Balan ^[9]	出租车信息服务
	手机所在的基站、手机通信状态	Pei ^[10]	城市功能区识别
		Trasarti ^[11]	城市动态
		Phithakkitnukoon ^[12]	旅游行为分析
志愿者数据集	Geolife(GPS)	Zheng ^[13]	信息推荐、智能交通
	MDC ^[14] (手机)	Etter ^[15] 、Trinh ^[16]	行为分析与预测
	其他自己采集的数据集	张治华 ^[17] 、吕明琪 ^[18]	轨迹语义化、出行信息提取
公交卡	刷卡地点、次数	龙瀛 ^[19] 、Joh ^[20] 等	城市空间结构、功能区识别与通勤关系
签到	社交网络的签到信息	Liu ^[21]	经济地理分析和区域联系强度

合北京市居民出行调查和土地利用类型数据,分析了北京市职住关系和通勤出行特征;Joh等^[20]使用首尔大都市区的1000万次的刷卡记录,分析了出行轨迹与该地区的土地利用特征之间的关系;Roth等^[34]使用伦敦的Oyster卡记录地铁乘客的移动数据分析了城市多中心的空间结构特征。公交卡数据尽管能有效地分析持卡者的移动行为,分析城市的职住区分布和通勤特征。但是,公交卡的覆盖范围有限,对北京等出差人数较多的城市,人群流向分析的精度尚待验证。

上述城市功能单元识别的相关研究,总体上还处于对其他地理学方法已发现规律基础上的验证层次,就是通过轨迹挖掘再现已知规律。通过轨迹数据发现新规律、新结论的研究相对偏少。下一步的研究重点应考虑细粒度的POI信息及其语义特征,结合轨迹体现的社会属性,提高城市功能单元识别的时空粒度,使相关研究从宏观走向微观,以进行精细化的城市韵律研究,为城市规划及相关社会学的研究提供更为精准的科学支持。

3.2 人类活动模式与移动行为预测

人类移动行为预测,一直是行为地理学和时间地理学关注的热点。长期以来,相关研究一直认为人的行为是不可预测的,但是,Song^[35]、Gonzalez^[36]等的研究发现,个体行为看似随机无序的背后具有高度的时空规律性,甚至推断出人类活动行为具有93%的可预测性,此后人类活动预测的相关研究陆续开展起来。

轨迹中的锚点、出行范围和形状,体现着用户的活动区域、工作性质和生活模式等信息^[22],在了解活动模式的基础上,可进行个体移动行为的预测并实现位置推荐^[37]。

出租车轨迹数据是研究人类活动模式的一个重要数据集,通过轨迹中上下客位置的密度分布与时间间隔,不仅能发现上下车的热点区域,还能发现OD流的移动方向,从而分析人群的移动规律,进行社区结构的划分^[38]。在了解人类活动模式的基础上可进行移动行为的预测,如Veloso等^[39]利用出租车轨迹数据预测了给定条件下的上客点的位置;Li等^[40]根据出租车轨迹预测出一个区域每个时间段的乘客数量;齐观德等^[41]使用出租车轨迹数据通过计算乘客等候时间的概率分布预测出某时某地等候出租车的时间,具有较高的精度。

手机数据也开始用于人类移动行为预测的研

究,Phithakkitnukoon等^[12]使用手机数据研究了日本的手机用户的旅行规律;Csáji等^[42]使用葡萄牙10万手机用户的数据发掘了用户的活动模式;Chen等^[43]根据手机数据猜测了手机用户的活动位置;Ahas等^[44]通过手机数据研究了爱沙尼亚的季节性旅游的游客活动模式。被动式的手机数据尽管为研究人类活动模式提供了有力的数据支撑,但是,该数据集往往只记录位置信息或者通话数据,缺乏进一步的数据描述。为此,国内外研究者主动采集了多个数据集以进行更精细的研究。

志愿者数据集比较知名的有Nokia的MDC数据集和微软中国研究院的GeoLife数据集。Nokia的LDCC计划产生的MDC数据集,不仅记录了手机用户所处的基站数据,还记录了通话与短信、WiFi与蓝牙设备及各种App的使用情况。Trinh等^[16]依据该数据集进行了下一个位置与将使用的App的预测研究,Etter等^[15]在该数据集上开始了从一个位置预测将要去的下一个位置的研究,这2项研究均取得了较好的效果。郑宇等^[45]在Geolife数据集上,也开展了移动行为预测的研究,并根据轨迹的相似性计算用户的相似性,进行了个性化的位置推荐试验,取得了较好的研究效果。

在活动模式与行为预测上,出租车数据体现的是某一辆车的规律,无法定位至特定的人,故难以使用该数据进行人的行为预测,实际上是车的行为预测。手机数据记录的是基站的位置,空间覆盖范围较大,也无法在POI层次上研究人的移动轨迹,基于该数据的活动模式判别和行为预测也是一种粗粒度的、概略性的探测。志愿者数据集特别是GPS采集的数据集(如Geolife)的时间和空间密度很大,能满足精细化的研究需求,活动模式判定和行为预测也具有较高的精度,但是其缺乏长时间序列的数据存储,难以表征人的行为在时间上的不同和演变。

人的活动模式研究的直接应用,是进行移动行为的预测。现有的研究是先从历史轨迹中推断出规律性的活动,然后依据静态的模式进行下一个位置的判断。人的行为的复杂性决定了人的移动是一个受多种因素综合影响的自适应过程。进一步的研究除应考虑动态的上下文环境、人的社会关系等对人的活动趋向的影响,更应考虑个人的偏好和兴趣特征等因素,发展新的模型和算法以提高预测的精度。

3.3 智能交通时间估算与异常探测

轨迹数据特别是出租车轨迹,能直接反映城市的交通状态,交通时间估算、交通异常探测,是轨迹数据在智能交通上最直接的应用。Chen等^[46]使用轨迹数据进行的短时交通预测取得了较好的效果;Wang等^[47]进行的通行时间估算也具有较高的时间精度。通过轨迹探测交通流中的一些异常行为也取得了较好的实验效果。Pang等^[48]使用LRT模型进行了交通异常快速检测方法的研究;Kanoulas等^[49]使用轨迹数据进行了交通拥堵路段监测和交通异常现象发现的研究。Liu等^[50]探测了交通流中交通异常的时空因果关系;Chawla等^[51]推断了交通流量异常的根本原因;Pan等^[52]根据人群移动和社会网络进行了交通异常中的拥堵感知方面的研究。这些研究为交通异常的快速识别与处置提供了有效的方法支持。目前,轨迹数据上传与处理的速度难以满足实时性的需求,上述研究的时间估算与异常探测是基于历史数据得出的。因此,这些方法的处理效率,能否满足实时性的需求还待进一步研究。

轨迹数据还可用于交通路线的优化与设计。出租车作为城市人群运移的重要工具,出租车司机对道路情况往往比较熟悉,他们的行车路线可被认为是两点之间的最优路径(经典路径),因此,可通过出租车轨迹的统计规律发现最快的行车路线。Yuan等^[53]在T-Drive项目中根据出租车轨迹数据,解算出个性化的最快线路设计,不仅为每30 min驾车路程节约5 min时间,还可通过为不同用户选择不同的道路来缓解可能出现的拥堵。同时,通过最优的路线对比也能发现不合理的道路规划,为城市交通路线设计提供依据^[54]。

现有的交通时间估算和异常探测研究,是在历史数据集上截取不同的时间片段,开展模型与算法的测试与验证。尽管轨迹数据集体量巨大,但是,具体到单项研究可用的数据往往非常稀疏,导致交通时间估算特别是短时交通时间估算总体上不甚理想,而这是动态导航的基础。因此,下一步的研究应侧重于提高高速动态数据的获取、处理与分析的效率,实现动态导航路径的实时及面向个体的可视化表达,并进行合理的交通诱导,以提高现有路网的通行调配能力。

3.4 城市计算的其他应用

智能交通、城市功能区识别和活动模式与行为预测是轨迹数据挖掘的主要研究方向。作为城市

计算的主要数据源,轨迹数据已开始应用于城市环境监测、能源消耗及疾病健康等领域的研究,并取得了初步的成果。Zheng等使用有限个地面监测站的空气质量数据,结合交通流、道路结构、兴趣点分布、气象条件和人口流动规律等大数据,推断出整个城市细粒度的空气质量,取得了较好的实验效果^[6];同时,对于噪声污染精细粒度的监测也在监测点数据,并结合签到数据等的基础上取得了预期的效果^[7]。更有一项创造性的研究是通过出租车轨迹数据,计算在加油站点的排队等候时间,估算能耗及汽车尾气在PM 2.5排放量中的占比^[55]。城市公共安全也是城市管理的一项重要职能,Pang等^[56]通过轨迹研究发现交通流的紊乱从而判定城市突发事件;Song等^[57]研究了160万人在日本大地震和福岛核事故以后的移动规律;Lu等^[58]研究了海地大地震的灾民移动轨迹,后2项研究旨在发现灾后大规模人群的移动规律,以期为灾后应急救援提供理论支持。

除此之外,轨迹数据挖掘在智慧旅游、商业定投、信息推荐等方面的研究也取得了一定的成果。轨迹数据挖掘能探求轨迹背后深层次的知识,已成为大数据背景下研究社会、城市等复杂问题的重要数据支撑和解决方法。

4 结论

以出租车数据、手机数据、志愿者数据等为代表的社会遥感数据,践行着人人都是传感器的理念,感知着人类的社会活动,体现着城市的韵律。作为位置大数据的典型形式,轨迹数据不仅在于表现空间定位信息,更在于引申出人的社会属性、社交网络及所处的社会环境信息。通过轨迹数据能有效地发现个体的活动模式、群体的移动规律,使人们以前所未有的深度认知自己的生活轨迹、社交行为,乃至是环境的变动和城市的演进等。泛在网络下个体活动信息,得到了前所未有的广度和深度的记录,时空粒度的不断细化为社会科学的定量分析提供了可能,同时也给地理信息科学提出了新的挑战。

今后轨迹数据挖掘的研究重点应关注如下4个方面的问题:(1)由于轨迹数据的体量巨大、信息碎片化、结构化程度差,涉及到高动态移动对象的存储与分析,对现有的数据库管理软件及现有的数据处理、存储、分析技术,都提出了更高的要求,关系

数据库的管理模式和时空索引技术,对处理大规模的轨迹数据显得力不从心,需研究和发展新的数据管理与存储、分析的手段。(2)轨迹数据的高动态、高实时性使其更加关注数据感知、管理与分析的速度,对算法的时间和空间复杂度提出了更高的要求,并行计算等高性能计算模式应更深入地融合进轨迹数据的计算中。(3)隐私问题是轨迹数据处理与分析过程中的一个不可回避的问题,其不仅是一个技术问题,也是一个道德与法律问题,应引起进一步的关注。(4)多学科研究方法的交叉和融合将进一步延展,在传统的计算机科学等学科融合的基础上,将进一步和物理学、复杂系统、社会学、管理学,甚至是自然语言处理等学科领域结合,从不同侧面发展新的方法和技术,共同推动轨迹数据的深度应用与创新。

总之,还需在轨迹大数据的智能感知、智能提取、智能服务等几个环节投入更多的研究,以更准确地感知城市脉搏,使城市更加智能,提供更个性化、更精准的位置服务。

参考文献:

- [1] 陆锋,张恒才.大数据与广义GIS[J].武汉大学学报(信息科学版),2014,39(6):645-654.
- [2] 刘瑜,肖昱,高松,等.基于位置感知设备的人类移动研究综述[J].地理与地理信息科学,2011,27(4):8-13.
- [3] 潘纲,李石坚,齐观德,等.移动轨迹数据分析与智慧城市[J].中国计算机学会通讯,2012,8(5):31-37.
- [4] 郑宇.城市计算概述[J].武汉大学学报(信息科学版),2015,40(1):1-13.
- [5] Lu F, Duan Y Y, Zheng N B. A practical route guidance approach based on historical and real-time traffic effects [C]. Proceedings of the 17th International Conference on GeoInformatics, Fairfax, USA, 2009:12-14
- [6] Zheng Y, Liu F, Hsieh H P. U-Air: When urban air quality inference meets big data[C]. KDD, Chicago, USA, 2013.
- [7] Zheng Y, Liu T, Wang Y, *et al.* Diagnosing New York city's noises with ubiquitous data[C]. Ubi-Comp. Seattle, WA, USA, 2014.
- [8] Pan G, Qi G, Wu Z, *et al.* Land-use classification using taxi GPS traces[J]. IEEE Transactions on Intelligent Transportation Systems, 2013,14(1):113-123.
- [9] Balan R K, Nguyen K X, Jiang L. Real-time trip information service for a large taxi fleet[C]. Proceedings of the 9th international conference on Mobile systems, applications, and services, 2011:99-112.
- [10] Pei T, Sobolevsky S, Ratti C. A New insight into land use classification based on aggregated mobile phone data[J]. International Journal of Geographical Information Science, 2014,28(9):1988-2007.
- [11] Trasarti R, Olteanu-Raimond A, Nanni M, *et al.* Discovering urban and country dynamics from mobile phone data with spatial correlation patterns[J]. Telecommunications Policy, 2015,39(3-4):347-362.
- [12] Phithakkitnukoon S, Horanont T, Witayangkurn A, *et al.* Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan[J]. Pervasive and Mobile Computing, 2015,18:18-39.
- [13] Zheng Y, Xie X, Ma W Y. GeoLife: A collaborative social networking service among user, location and trajectory[J]. IEEE Data(base) Engineering Bulletin, 2010,33(2):32-39.
- [14] Laurila J K, Gatica-Perez D, Aad I, *et al.* The mobile data challenge: Big data for mobile computing research[C]. Proc. MDC Workshop, Newcastle, UK, 2012.
- [15] Etter V, Kafsi M, Kazemi E, *et al.* Where to go from here? Mobility prediction from instantaneous information [J]. Pervasive and Mobile Computing, 2013,9(6):784-797.
- [16] Do T M T, Gatica-Perez D. Where and what: Using smart-phones to predict next locations and applications in daily life[J]. Pervasive and Mobile Computing, 2014,12:79-91.
- [17] 张治华.基于GPS轨迹的出行信息提取研究[D].上海:华东师范大学,2010.
- [18] 吕明琪.基于轨迹数据挖掘的语义化位置感知计算研究[D].杭州:浙江大学,2012.
- [19] 龙瀛,张宇,崔承印.利用公交刷卡数据分析北京职住关系和通勤出行[J].地理学报,2012,67(10):1339-1352.
- [20] JOH C, Hwang C. A Time-geographic analysis of trip trajectories and land use characteristics in Seoul metropolitan area by using multidimensional sequence alignment and spatial analysis[C]. AAG Annual Meeting, Washington, DC, 2010.
- [21] Liu Y, Sui Z, Kang C, *et al.* Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data[J]. PloS One, 2014,9(1):e86026.
- [22] 李婷,裴韬,袁烨城,等.人类活动轨迹的分类、模式和应用研究综述[J].地理科学进展,2014,33(7):938-948.
- [23] 陆锋,刘康,陈洁.大数据时代的人类移动性研究[J].地球信息科学学报,2014,16(5):665-672.
- [24] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs[C]. The 18th ACM SIGKDD International conference on knowledge discovery and data mining. Beijing, China, 2012.
- [25] Gao S, Liu Y, Wang C, *et al.* Discovering spatial interaction communities from mobile phone data[J]. Transactions

- in GIS, 2013,17(3):463-481.
- [26] Kang C, Sobolevsky S, Liu Y, *et al.* Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages[C]. Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM, 2013.
- [27] Calabrese F, Ratti C, Colonna M. Real-time urban monitoring using cell phones: A case study in Rome[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1):141-151.
- [28] Calabrese F, Di Lorenzo G, Liu L, *et al.* Estimating origin-destination flows using mobile phone location data[J]. IEEE Pervasive Computing, 2011,10(4):36-44.
- [29] 许宁,尹凌,胡金星.从大规模短期采样的手机定位数据中识别居民的职住地[J].武汉大学学报(信息科学版), 2014,39(6):750-756.
- [30] Ratti C, Pulselli R M, Williams S, *et al.* Mobile landscapes: Using location data from cell phones for urban analysis[J]. Environment and Planning B: Planning and Design, 2006,33(5):727-748.
- [31] Calabrese F, Ratti C. Real time Rome[J]. Networks and Communication Studies, 2006(20):247-258.
- [32] Reades J, Calabrese F, Ratti C. Eigenplaces: Analysing cities using the space-time structure of the mobile phone network[J]. Environment and Planning B: Planning and Design, 2009,36(5):824-836.
- [33] Steenbruggen J, Borzacchiello M T, Nijkamp P, *et al.* Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: A review of applications and opportunities[J]. GeoJournal, 2013,78(2):223-243.
- [34] Roth C, Kang M S, Batty M, *et al.* Structure of urban movements: Polycentric activity and entangled hierarchical flows[J]. PLoS ONE, 2011,6(1):e15923.
- [35] Song C, Qu Z, Blumm N, *et al.* Limits of predictability in human mobility[J]. Science, 2010,327(19):1018-1021.
- [36] González M C, Hidalgo C A, Barabási A. Understanding individual human mobility patterns[J]. Nature, 2008,453(7196):779-782.
- [37] 牟乃夏,刘文宝,张灵先,等.空间信息服务的个性化问题[J].测绘科学,2011,36(3):104-106.
- [38] Guo D. Flow mapping and multivariate visualization of large spatial interaction data[J]. IEEE Transactions on Visualization and Computer Graphics, 2009,15(6):1041-1048.
- [39] Veloso M, Phithakkitnukoon S, Bento C. Urban mobility study using taxi traces[C]. Proceedings of the 2011 international workshop on Trajectory data mining and analysis, 2011:23-30.
- [40] Li X, Pan G, Wu Z, *et al.* Prediction of urban human mobility using large-scale taxi traces and its applications[J]. Frontiers of Computer Science, 2012,6(1):111-121.
- [41] 齐观德,潘遥,李石坚,等.基于出租车轨迹数据挖掘的乘客候车时间预测[J].软件学报,2013,24(S2):14-23.
- [42] Csáji B C, Browet A, Traag V A, *et al.* Exploring the mobility of mobile phone users[J]. Physica A: Statistical Mechanics and its Applications, 2013,392(6):1459-1473.
- [43] Chen C, Bian L, Ma J. From traces to trajectories: How well can we guess activity locations from mobile phone traces?[J]. Transportation Research Part C: Emerging Technologies, 2014(46):326-337.
- [44] Ahas R, Aasa A, Mark Ü, *et al.* Seasonal tourism spaces in Estonia: Case study with mobile positioning data[J]. Tourism Management, 2007,28(3):898-910.
- [45] Li Q, Zheng Y, Xie X, *et al.* Mining user similarity based on location history[C]. Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. Irvine: ACM, 2008:1-10.
- [46] Chen B Y, Lam W H K, Li Q Q, *et al.* Shortest path finding problem in stochastic time-dependent road networks with stochastic first-in-first-out property[J]. IEEE Transactions on Intelligent Transportation Systems, 2013,14(4):1907-1917.
- [47] Wang Y, Zheng Y, Xue Y. Travel Time Estimation of a Path using Sparse Trajectories[C]. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014:25-34.
- [48] Pang X L, Chawla S, Liu W, *et al.* On mining anomalous patterns in road traffic streams[C]. The 7th International Conference on Advanced data mining and applications. Berlin, Germany: Springer, 2011:237-251.
- [49] Kanoulas E, Du Y, Xia T, *et al.* Finding fastest paths on a road network with speed patterns[C]. The 22nd international conference on data engineering, 2006:3-7.
- [50] Liu W, Zheng Y, Chawla S, *et al.* Discovering spatio-temporal causal interactions in traffic data streams[C]. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011:1010-1018.
- [51] Chawla S, Zheng Y, Hu J. Inferring the root cause in road traffic anomalies[C]. IEEE International Conference on Data Mining, 2012:141-150.
- [52] Pan B, Zheng Y, Wilkie D, *et al.* Crowd sensing of traffic anomalies based on human mobility and social media[C]. Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013:344-353.
- [53] Yuan J, Zheng Y, Zhang C, *et al.* T-Drive: Driving direc-

- tions based on taxi trajectories[C]. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010:99-108.
- [54] Zheng Y, Liu Y, Yuan J, *et al.* Urban Computing with Taxicabs[C]. Proceedings of the 13th international conference on Ubiquitous computing, 2011:89-98.
- [55] Zhang F, Yuan N J, Wilkie D, *et al.* Sensing the pulse of urban refueling behavior:a perspective from taxi mobility [J]. ACM Transaction on Intelligent Systems and Technology, 2013,9(4):1-24.
- [56] Pang X L, Chawla S, Liu W, *et al.* On detection of emerging anomalous traffic patterns using GPS data[J]. Data & Knowledge Engineering, 2013(87):357-373.
- [57] Song X, Zhang Q, Sekimoto Y, *et al.* Modeling and probabilistic reasoning of population evacuation during large-scale disaster[C]. Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, 2013:1231-1239.
- [58] Lu X, Bengtsson L, Holme P. Predictability of population displacement after the 2010 Haiti earthquake[J]. Proceedings of the National Academy of Sciences(PNAS), 2012, 109(29):15576-15581.

A Review on the Application Research of Trajectory Data Mining in Urban Cities

MOU Naixia^{1,2*}, ZHANG Hengcai², CHEN Jie², ZHANG Lingxian¹ and DAI Honglei¹

(1. Shandong Provincial Key Laboratory of Geomatics and Digital Technology of Shandong Province, Shandong University of Science and Technology, Qingdao 266590, China; 2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: The trajectory datasets record a series of position information at different times, so they become the new data sources to study the laws of human mobility. As a main form of social remote sensing data, trajectory datasets also bring a new individual viewpoint to study geographical phenomena. With the emergence of big data, trajectory data mining becomes a hot topic in geographical information science, urban computing and other correlative disciplines. In this paper, we gave a brief review on trajectory data mining and its applications in cities. First, we listed the data sets frequently adopted by human mobility research, gave the classification and their typical applications using FCD data, mobile phone data, smart cards data, check-in data, etc. Then, we summarized its application in solving cities' problems from four aspects: (1) the identification of urban spatial structure and function unit; (2) the patterns recognition of human activity and the behavior prediction of human movement; (3) the traffic time estimation and the anomaly detection of intelligent transportation; (4) other applications in urban computing such as in urban air and noise pollution, disaster prevention and rescue, even in intelligent tourism and information recommendation. At the end, we pointed out the challenges and further research directions of trajectory data mining.

Key words: trajectory data; data mining; urban computing; human mobility; human activity patterns

***Corresponding author:** MOU Naixia, E-mail: mounaixia@163.com