

位置大数据的分析处理研究进展

刘经南¹ 方媛¹ 郭迟¹ 高柯夫¹

¹ 武汉大学卫星导航定位技术研究中心,湖北 武汉,430079

摘要:大数据时代的到来,使得社会、科学和经济都发生了巨大的变革。当前,基于各种测量传感网络来感知人类个体和群体与自然环境和社会环境关系的泛在测绘所产生的位置大数据是大数据研究的重要组成部分。位置大数据已经成为当前用来感知人类社群活动规律、分析地理国情和构建智慧城市的重要战略资源。通过对位置大数据的处理分析,可从单纯的定位数据引申出人的社会属性以及与环境的关系,这极大地促进了计算机科学技术、数据科学技术与测绘科学技术的联系,形成了一种智能化、社会化的泛在测绘计算。在定义位置大数据概念的基础上,主要从以下三个方面对位置大数据进行了介绍和分析:1)位置大数据的分类、特征、作用与意义,以及涉及的研究方法体系;2)从社会感知、群体智能系统的建设和地理国情分析三个方面对位置大数据的应用分析进行了具体阐述;3)针对位置大数据的混杂型、复杂性、稀疏性以及研究需求,介绍了常用的位置大数据处理方法,包括地图、轨迹数据的预处理以及降维分析、协同挖掘等方面。

关键词:泛在测绘;位置大数据;位置服务;智慧城市;社会感知;众包

中图法分类号:P23; P208 **文献标志码:**A

大数据(big data)是指大型复杂数据集的聚合。这些数据集的规模和复杂程度常超出目前数据库管理软件和传统数据处理技术在可接受时间下的获取、管理、检索、分析、挖掘和可视化能力。2012年奥巴马政府公布了“大数据研发计划”(Big Data Research and Development Initiative),以国家战略应对大数据时代。中国政府、学术界和产业界也积极应对,目前已经成立了多个大数据的产学研联盟,并在973等科研计划中设立了相关项目。其共同的目标是改进现有从海量和复杂的数据中获取知识的能力,加速人类在科学与工程领域发明的步伐,增强国家安全等。从高德纳(Gartner)公司2013年7月发布的技术曲线成熟度报告分析,大数据研究正处在期望膨胀期,距离成熟期还需要5~10a的时间。如何抓住这一时机,是学术界和工业界面临的机遇与挑战。

大数据时代的到来使得科学、社会和经济产生了巨大的变革。图灵奖获得者吉姆·格雷(Jim Gray)提出:大数据是科学研究的第四范式(the fourth paradigm),是从计算密集型科研发展到以

大数据为基础的数据密集型科研方法。随着大数据的发展,将会形成专门针对数据科学的专业,同时产生一批新的就业岗位,如数据分析师、数据科学家、数据工程师等。大数据将在政府公共服务、医疗服务、零售业、制造业以及涉及个人位置服务等领域得到广泛应用,并产生巨大的社会价值和产业空间。TMR(透明度市场研究)最新发布的报告《大数据市场:2012~2018年全球形势、发展趋势预测》指出,2012年全球大数据市场产值为63亿美元,预计2018年该产值将达483亿美元。

大数据中的一个重要组成部分就是位置大数据(location big data, LBD)。随着位置服务和车联网应用的不断普及,由地理数据、车辆轨迹和应用记录等构成的位置大数据已经成为当前用来感知人类社群活动规律,分析地理国情和构建智慧城市的重要战略资源。

通过对位置大数据的处理分析,使得传统测绘强调的物理世界的测量结果(即位置)可引申到对人类社会的某些动态情况量测中去,这极大地促进了当代计算机科学技术、数据科学技术与测

收稿日期:2014-01-14

项目来源:国家863计划资助项目(2013AA12A206,2013AA12A204);国家自然科学基金资助项目(41104010);高等学校学科创新引智计划资助项目(B07037)。

第一作者:刘经南,教授,博士生导师,中国工程院院士。主要研究方向为大地测量理论及应用,大地测量坐标系理论、导航位置服务和GNSS技术应用。E-mail: jnliu@whu.edu.cn

通讯作者:郭迟,博士。E-mail: guochi@whu.edu.cn

绘科学技术的融合,形成了一种智能化、社会化的泛在测绘计算。泛在测绘(ubiquitous mapping)是指用户在任何地点、任何时间为认知自然和社会环境与人的关系而创建和使用地图的活动。它强调人、环境等地理信息的社会属性^[1],强调人的活动与数据融合。通过对位置大数据进行适当的处理以及社会计算,提供个性化的、实时的、动态的位置信息。每个人都可以是位置数据的提供者,同时也可以成为位置服务的受众。因此,泛在测绘既是位置大数据的主要来源,又能作为提供位置大数据服务的技术平台。

本文针对位置大数据,对其基本内容、应用领域以及相应领域已有的研究成果进行了详细的介绍,最后,针对位置大数据的特征和研究需求,介绍了常用的大数据处理方法。

1 位置大数据

含有空间位置和时间标识的地理和人类社会信息数据即为位置数据。这里,时空标识既可以是严格定义下的时空坐标,也可以是约定俗成、含义明确的位置和时间概念,如地名、方位、上午、下午等等。

1.1 位置大数据分类

位置大数据主要分为地理数据、轨迹数据和空间媒体数据,如表1所示。

1) 地理数据:直接或间接关联着相对于地球上某个地点的数据,包括自然地理数据(土地覆盖类型数据、地貌数据、土壤数据、水文数据、植被数据、居民地数据、河流数据、行政境界)和社会经济数据等。其来源主要是传统测绘和泛在测绘,包括各种遥感影像和大地基准测量数据。其特点是数据体量大,较为规则化,变化较慢。

2) 轨迹数据:通过GNSS等测量手段以及网络签到等方法获得的用户活动数据(个人轨迹数据、群体轨迹数据、车辆轨迹数据),可用于反映用户的位置和用户的社会偏好。来源主要有各类导航数据、智能手机数据、物流数据等等。其特点是数据体量大,信息碎片化,准确性较低,半结构化等。

3) 空间媒体数据:包含位置因素的数字化文字、图形、图像、声音、视频影像和动画等媒体数据,主要来源于移动社交网络、微博等新型互联网应用。特点是数据来源混杂、数据异构性大、数据价值密度低,实时性强但时空标识定义欠严格或欠精确。

表1 位置大数据实例

Tab. 1 Instances of Location Big Data

	实例	体量举例	特点
地理数据	数字矢量线画地图(DLG)	全国1:5万DLG有250 GB,DOM	数据体量大,较为规则化,变化较慢
	数字栅格地图(DRG)	有10 TB,1:1万DLG约5.3 TB,	
	数字正射影像地图(DOM)	DOM约350 TB	
	数字高程模型(DEM)		
轨迹数据	出租车轨迹数据	北京市12 000辆出租车110 d产生577 000 000条轨迹记录 ^[2]	数据体量大,信息碎片化,单记录价值低,准确性低,半结构化
	微博用户位置签到数据	Twitter 63 261用户30 d产生15 944 084条位置签到记录 ^[3]	
空间媒体数据	互联网图像视频数据	2011年全球被创建和被复制的数据总量为1.8 ZB	数据来源混杂,数据异构性大,数据价值密度低,实时性强
	社交网络数据	Facebook每天生成300 TB以上与位置有关的日志数据	
	城市监控摄像头数据	上海平安城市监控摄像头超过60万只,每天产生的位置监控数据达PB级	
	在线电子商务数据	淘宝网每天交易数千万笔,产生约20 TB数据,均含有物流位置信息	

1.2 位置大数据的特点

位置大数据具有典型的体量大、更新(velocity)速度快、多元性(variety)和价值(value)密度低等大数据特性。因此过去那些性能、容量、扩展性等受限,且管理复杂的集中式数据计算模式已不再适用于对具有这样一些特性的大数据进行处

理,需采用去中心化/自组织的分布式计算模式。而位置大数据的“大”更多意义在于:人类可以处理、分析并使用的的数据在大量增加,通过这些数据的处理、整合和分析,可以创造新价值。通过群体智能、社会计算、认知计算对位置数据价值的发酵和提炼,会有越来越多靠“位置数据的体量+

简单的逻辑”的方法来解决复杂问题。

位置大数据的一个鲜明的共性特征是具有时空标识,有时使用坐标、语言文本来描述。另外,位置数据中往往对时空标识附有精确度、可靠度的指标性参数。要求提供位置信息的精确度、可靠度和可信度往往是位置数据处理分析的特殊要求。

同时,受到数据采集技术等方面的客观制约,位置大数据往往不能全面和正确地反映观察对象的整体全貌,因而具有“复杂但稀疏(complex yet sparse)”的特点。如何从位置大数据中获得价值,进而发现人类社群活动规律,是值得探讨的问题。

1.3 位置大数据的研究框架

位置大数据研究涉及到数据采集、数据处理、计算和存储以及可视化等一套完整的技术方法体系,如图 1 所示。

1) 数据采集方法

建立完整的、满足多种不同精度需求的、室内外一体的位置大数据传感网络是实现位置大数据获取的基本途径。综合利用自然语言处理、图像处理、信息检索等方法,提取互联网多媒体中的位置信息,建立位置信息与互联网媒体的内在关联;利用用户主动分享的移动社会网络数据;通过已经建立的诸如路桥 ETC 系统、地铁收费系统、公交刷卡系统和视频监控系统等,大规模被动收集用户位置数据。



图 1 位置大数据分析框架

Fig. 1 Location Big Data Analyzing Framework

2) 数据处理分析方法

位置大数据处理分析方法分为标准和非标准两种处理分析模式。服务于国民经济各行业和企业业务需求、来源于地理空间信息的位置数据大多采用国标、行标和企业标准进行分析处理;来源于轨迹和空间媒体的位置数据多采用非标准的新方法处理。

位置大数据包含多种来源不确定的数据,使得数据集中往往含有各种错误和误差,体现为数据不正确、不精确、不完全、陈旧或者冗余等。因此需要对数据进行适当的预处理,包括数据补全、

降维、协同挖掘等(下文介绍)。数据挖掘和机器学习技术仍是大数据研究中主要使用的方法。

此外,位置大数据的使用分析更注重数据间的关联关系,不追求数据的精确性和数据间的因果联系,重数据轻模型。

3) 计算和存储方法

利用 Hadoop 等计算框架,建立流媒体、轨迹、地图数据的高效时空索引和分布式分析技术,尤其应注重诸如 Hbase、BigSQL、MongoDB 等非关系时空数据库存储技术。

4) 可视化方法

位置大数据因其体量大、数据繁杂的特征,常规的统计图表无法准确反映其总体趋势,需要借助一些特殊的数据可视化方法进行数据表达,将复杂的交通数据及其分析结果通过可视化方式直观地展现出来,并支持对结果的交互式选取和浏览。目前,常用的方法有热点地图、泛在地图、虚拟现实和增强现实等等。

2 位置大数据在智慧城市中的运用

2.1 位置大数据与泛在地理国情

实现智慧城市的关键之一是建立一种泛在的城市计算(urban computing)体制,其中涉及到泛在测绘、位置大数据分析和提供服务三个重要层面。因此位置大数据的分析处理和辅助决策已成为智慧城市实现与城市地理国情分析中的关键问题。通过对位置大数据的分析应用,主要可形成智慧城市三个方面的服务支持:为城市运行服务,包括城市规划、疾病控制、智能交通、节能减排、环境保护、应急响应等;为个人生活服务,包括社会交流、个性化信息推送、驾驶安全、智能驾驶等;为企业经济服务,包括企业的调度、门店选址、广告推送、位置营销等。

在大数据科学研究中,一些与数据本身来源和主体看似无关的对象,在经过数据价值提取和协同挖掘后往往会表现出密切关联性。同样地,位置大数据的应用分析,不只是被用来进行交通问题的分析,还能够提升我们对更为广泛的人类社会经济活动和自然环境的认知,从而体现其内在价值。从目前的研究和应用来看,位置大数据在社会感知、群体智能系统建设、地理国情分析等方面的应用效果显著。

2.2 基于位置的社会感知

位置社会感知是指通过人类社会生活空间部署的大规模位置传感设备,感知识别社会个体的

行为,分析挖掘群体社会交互特征和规律,引导个体社会行为,支持社群的互动、沟通和协作的一种计算技术^[2,4]。随着移动互联网等技术的发展,位置服务已从单纯的定位服务转变为具有社会化、本地化和移动性特征的新型业态。用户的服务需求已经从获取位置扩展为获取位置背后更为丰富的社会信息与群体智能,从位置搜索、路径规划等普遍化需求扩展为符合自身社会属性的个性化、智能化需求。通过位置大数据的社会感知,可以更好地认识“人的世界”。在不同城市 and 不同人群中,位置数据特征上的差异就可以直接映射到人们在休闲/娱乐方式、群体性格、情感分布、生活压力、城市宜居性等方面上去。再者,通过大规模位置轨迹的分布学习,可以为社群移动行为进行指导,从而参与诸如疾控预防、防灾减灾等工作。

文献[3,5]依照全球定位系统数据,分析纽约63万名微博用户的440万条微博,绘制身体不适用户位置“热点”地图,显示流感在纽约的传播情况。根据热点地图和视频,最早可在个体出现流感症状前8d作出预测,准确率为90%。文献[6]通过对微博位置数据进行分析 and 挖掘,研究地震的实时预警。文献[7-8]通过分析160万人1a的GPS移动轨迹,对东日本大地震和福岛核事故发生后的灾民移动、避难行为进行了建模、预测和模拟。

2.3 群体智能系统的建设

随着互联网的发展,各种共享性知识已经成为人们学习的主要途径。很多传统的专家系统都逐渐发展为具有社会化计算特性的众包(crowdsourcing)系统。所谓众包,就是众多个体通过互联网络自主或自发地提供数据或服务协同参与某一任务的完成,其产生的经济或社会效益大大超过任务设计和预想的本身,众包同时也是通过大数据发掘出群体智能的一种方法。众包的这一特点在智能交通等领域尤为突出。位置大数据恰恰是大众智慧的一种历史日志。通过对其进行反向建模,就能够获得有关人活动于物理世界的群体智能。比如通过位置大数据轨迹特征建立的导航引擎就取得了比普通导航引擎好得多的效果^[9],还可实时分析城市交通流量情况,为用户出行服务。

文献[10]基于北京市3万多辆出租车3个月的轨迹数据分析城市交通中的不合理区域。文献[9,11]依照用户历史数据为司机和乘客提供了一个双向最优的出租车招车/候车服务模型。文献[12]同样基于出租车GPS轨迹数据并结合天气及个人驾车习惯、技能和道路熟悉程度等因素,提

出了针对个人的最优导航算法,平均为每30min的路线节约5min。可以预见,基于位置大数据的群体智能平台体系的建设将是智慧城市智能决策支持的关键之一。

2.4 城市地理国情关联分析

传统的地理国情分析偏重一些地理性的静态事务,而融入对位置大数据的分析应用后,将地理国情情况作为隐变量,可以利用全局建模和协同挖掘的手段对城市生活中社会以及环境情况进行推测分析,包括经济运行情况(如CPI指数、绿色GDP指数、房价指数、城市投资与负债情况)、资源与环境情况(如PM_{2.5}值)、土地规划以及政府执政水平等。

文献[13]利用出租车在加油站的等待时间来估计加油站的排队长度,估算出此时加油站内的车辆数目及加油量。通过将全城的加油站数据汇总,可估计出一个城市每天的能源消耗量,为政府制定更为合理的长期城市能源战略。文献[14]利用地面监测站有限的空气质量数据,结合交通流道路结构、兴趣点分布、气象条件和人们流动规律等大数据,基于机器学习算法建立数据和空气质量的映射关系,从而推断出整个城市细粒度的空气质量。文献[15-16]分别通过出租车载客状态、人们的移动行为特征和POI类型来分析区域的功能类型。

3 位置大数据的处理方法

1) 预处理。包括对地图的预处理和位置轨迹数据的预处理。位置大数据分析一般需要基于地图或路网数据展开。因此需要将原本是连续空间的平面地图离散化,划分为多个区域。其主要方法有网格化分区^[17]、依道路网分区^[16]、依位置密度分区^[18-19](DBSCAN)和依参考点分区^[20](泰森多边形)等。在对位置大数据尤其是轨迹数据进行分析时,一般要求数据集具有较高的采样率,因此常需要对轨迹数据进行简单的线性插值^[17]。而将轨迹数据与地图数据进行关联,地图匹配是必不可少的处理步骤,现已存在一些经典的算法,如ST-Matching,IVMM,Passby等^[21-23]。

2) 局部位置数据的特征提取。大数据的分析首要任务是从局部研究对象中提取出价值,即初步的特征提取。针对位置大数据的混杂性,可分析和抽象出三类特征:①区域静态特征。以一定区域范围为观察对象,提取区域内与地图地貌相关的一些指标,包括路网特征和兴趣点变化

率等静态特征信息;② 区域交通力学特征。抽取区域内群体移动目标的移动行为,如区域交通流动性的时间演化性等动态特征信息;③ 个体移动模式特征,以单个移动对象为观察对象,提取个体在一段时间内的移动行为规律特征,如在一条路径中发生状态转换的概率期望等^[24]。不同于传统轨迹分析,这些特征大都具有二阶统计性,并能够很好地被应用于具体的位置服务计算中。

3) 大数据的降维分析。大数据的一个显著特征是体量大而价值密度低,如何对大数据进行合理的降维分析是大数据处理中的重要环节。对于位置大数据网络在空间尺度上的降维处理,其核心就是减少网络中的节点(即区域)或边(即区域间的关联),通过关键分量的分析获得全局特征,其主要方法包括依超介数的降维^[10]和依主分量的降维^[24]。而在时间尺度上的降维则主要指对时间的离散化,降低各时间段间的相似性。

4) 特征关联和协同挖掘。数据稀疏性是大数据科学研究中一个突出的问题,这种特性会导致结果失真。如文献^[17]发现,由于出租车群体往往比较喜欢在一些特定场所聚集,造成这些地方的位置数据密度过高,从而使得根据城市出租车轨迹密度所绘制的城市交通热点图与真实情况

存在很大的差异。而真正交通热点区域,由于缺失适当的采集手段,无法反映到热点图中。因此,针对数据的“复杂但稀疏”性,需从时间和空间尺度上进行关联和协同构造。其中时间尺度上可以通过建立概率图模型,如隐马尔科夫模型(HMM)、条件马尔科夫模型(conditional Markov model, CMM)或条件随机场模型(conditional random fields, CRF),将从位置大数据中提取的特征与某一特定应用关联。解决空间数据稀疏性的主要手段则包括关联聚类和协同过滤。区域聚类是指通过一个层级聚类方法将特征向量差异小的区域归于一类,从而用数据样本较多的区域去“替代”数据样本较少的区域;协同过滤是一种被广泛应用于推荐系统中的数据项聚类方法,通过矩阵分解方法(UV、SVD)将原始矩阵划分为维度较低的若干矩阵的乘积,通过对分形矩阵中值的猜测试探,估计原始矩阵中的缺失项。

为此,本文提出了一种位置大数据分析平台,将上述方法集成,支撑面向城市运行和大众群体应用的位置服务,如图 2 所示。平台的底层运用了 Hadoop 结构,上层按照前述处理步骤,调用多个数据挖掘工作包,完成对位置大数据价值提取的工作。

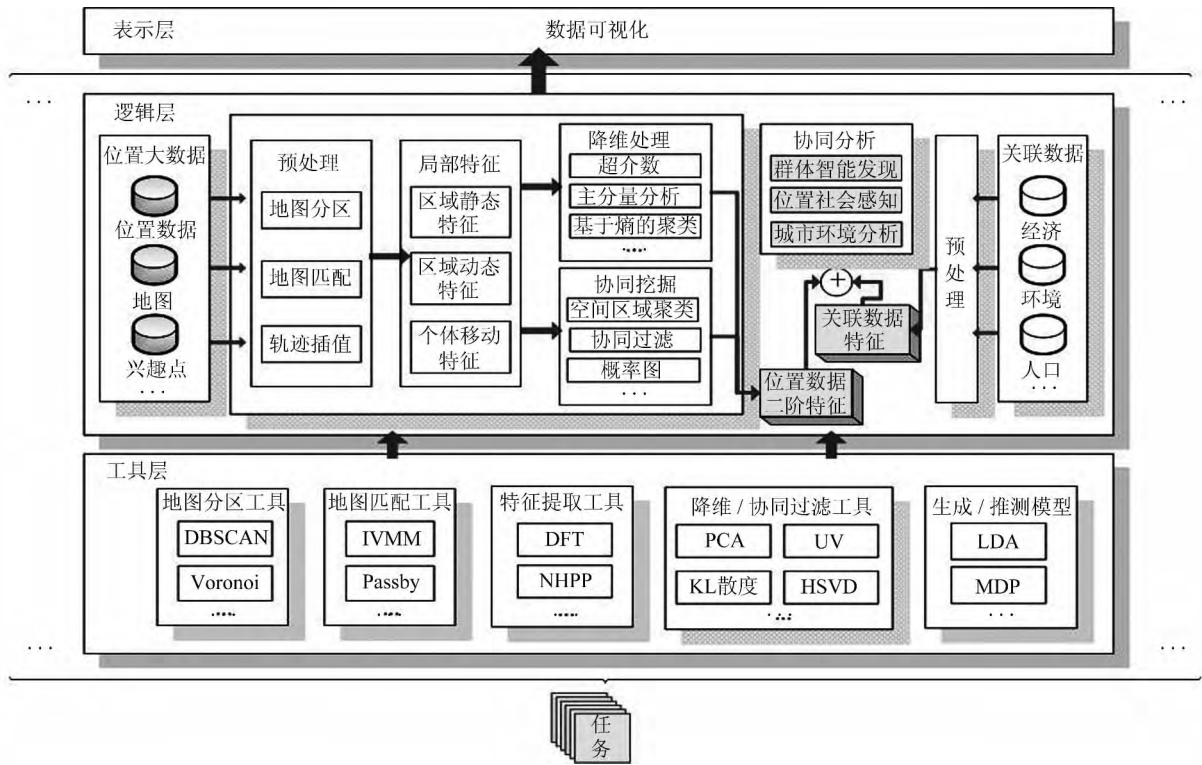


图 2 位置大数据处理分析平台

Fig. 2 Location Big Data Processing and Analyzing Platform

4 结 语

位置大数据的研究需要测绘技术与信息技术高度融合,将单纯的定位数据引申到人的社会属性以及与环境的关系中去,形成一种智能化、社会化的泛在测绘计算。传统测绘技术一直追求的都是数据的准确性(微观),而位置大数据研究恰恰允许数据的混杂性、不确定性(宏观)。传统测绘在微观数据处理的层面积累较深,泛在测绘和位置大数据服务要求研究宏观数据处理层面的新方法、新思维。因此,要使测绘行业应对大数据时代的变革,使位置大数据成为智慧城市建设和地理国情分析监测的关键,就必须改变当前土地、规划、测绘、城管、交通等部门各自掌握各自数据的局面,建立政府、学术界和产业界位置大数据共同分享、共同分析和共同经营的新机制。

位置大数据不只用来进行交通问题的分析,其已经成为当前用来感知人类社群活动规律、分析地理国情和构建智慧城市的重要战略资源,提升了大众对更为广泛的人类社会经济活动和自然环境的认识,充分体现了位置数据的价值。

参 考 文 献

- [1] Liu Jingnan. The Concept and Progress of Ubiquitous Mapping and Ubiquitous Position[J]. *Digital Communication World*, 2011, 4: 28-30 (刘经南. 泛在测绘与泛在定位的概念与发展[J]. 数字通信世界, 2011, 4: 28-30)
- [2] Guo Chi, Fang Yuan, Liu Jingnan, et al. Study on Social Awareness Computation Methods for Location-based Services[J]. *Journal of Computer Research and Development*, 2013, 50(12): 2 531-2 542(郭迟, 方媛, 刘经南, 等. 位置服务中的社会感知计算方法研究[J]. 计算机研究与发展, 2013, 50(12): 2 531-2 542)
- [3] Sadilek A, Kautz H A, Silenzio V. Modeling Spread of Disease from Social Interactions[C]. The 6th International AAAI Conference on Weblogs and Social Media (ICWSM), Dublin, 2012
- [4] Pentland A. Socially Aware, Computation and Communication[J]. *Computer*, 2005, 38(3): 33-40
- [5] Sadilek A, Kautz H, Bigham J P. Finding Your Friends and Following them to Where You Are[C]. The 6th ACM International Conference on Web Search and Data Mining, New York, 2012
- [6] Sakaki T, Okazaki M, Matsuo Y. Tweet Analysis for Real-time Event Detection and Earthquake Reporting System Development [J]. *IEEE Trans. Knowl. Data Eng.*, 2013, 25(4): 919-931
- [7] Song X, Zhang Q, Sekimoto Y, et al. An Intelligent System for Large-scale Disaster Behavior Analysis and Reasoning[J]. *IEEE*, 2013, 28(4): 35-42
- [8] Song X, Zhang Q, Sekimoto Y, et al. Modeling and Probabilistic Reasoning of Population Evacuation During Large-scale Disaster [C]. The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2013
- [9] Yuan J, Zheng Y, Zhang L, et al. T-finder: A Recommender System for Finding Passengers and Vacant Taxis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 25(10): 2 390-2 403
- [10] Zheng Y, Liu Y, Yuan J, et al. Urban Computing with Taxicabs[C]. The 13th International Conference on Ubiquitous Computing, New York, 2011
- [11] Yuan J, Zheng Y, Zhang L, et al. Where to Find my Next Passenger [C]. The 13th International Conference on Ubiquitous Computing, New York, 2011
- [12] Yuan J, Zheng Y, Xie X, et al. T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(1): 220-232
- [13] Zhang F, Wilkie D, Zheng Y, et al. Sensing the Pulse of Urban Refueling Behavior[C]. The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, 2013
- [14] Zheng Y, Liu F, Hsie H P. U-Air: When Urban Air Quality Inference Meets Big Data[C]. The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2013
- [15] Pan G, Qi G, Wu Z, et al. Land-use Classification Using Taxi GPS Traces[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012, 14(1): 113-123
- [16] 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, New York, 2012
- [17] Liu S, Liu Y, Ni L M, et al. Towards Mobility-based Clustering[C]. The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2010
- [18] Pan G, Qi G, Wu Z, et al. Land-use Classification Using Taxi GPS Traces[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012, 14(1): 113-123
- [19] Li Z, Ding B, Han J, et al. Mining Periodic Behavior

- iors for Moving Objects[C]. The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, 2010
- [20] de Montjoye Y A, Hidalgo C A, Verleysen M, et al. Unique in the Crowd: The Privacy Bounds of Human Mobility[J]. *Scientific Reports*, 2013, 3: 1-5
- [21] Lou Y, Zhang C, Zheng Y, et al. Map-matching for Low-sampling-rate GPS Trajectories[C]. The 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, 2009
- [22] Yuan J, Zheng Y, Zhang C, et al. An Interactive-voting Based Map Matching Algorithm[C]. The 11th International Conference on Mobile Data Management (MDM), Kansas City, 2010
- [23] Liu K, Li Y, He F, et al. Effective Map-matching on the Most Simplified Road Network[C]. The 20th International Conference on Advances in Geographic Information Systems. New York, 2012
- [24] Chawla S, Zheng Y, Hu J. Inferring the Root Cause in Road Traffic Anomalies[C]. IEEE 12th International Conference on Data Mining (ICDM), Piscataway, 2012

Research Progress in Location Big Data Analysis and Processing

LIU Jingnan¹ FANG Yuan¹ GUO Chi¹ GAO Kefu¹

¹ Global Navigation Satellite System Research Center, Wuhan University, Wuhan 430079, China

Abstract: As the age of big data is coming, the social, science and economic have been undergoing a great revolution. At present the location big data mainly produced by ubiquitous mapping is an important part of the big data research. Ubiquitous mapping is to analyze the relation of human, including the individual and the group, to the natural and social environments. Location big data has become the important strategic resource for analyzing people's behavior patterns and geographic situation, and building smart city. Through the location big data processing and analyzing, the positioning data can be simply extended to the relationship of human social attributes and environments. This has greatly promoted the relations among computer, data and mapping technologies, and formed the ubiquitous mapping calculation with intelligence and socialization. This paper introduces the location big data mainly from three aspects: Firstly, the classification, feature, effect and significance of location big data; secondly, the connection applications of location big data in social awareness, swarm intelligence system development and geographic situation analysis; lastly, the common processing methods of location big data, including the maps and trajectory data preprocessing, dimension reduction analysis and collaborative mining.

Key words: ubiquitous mapping; location big data; location-based service; smart city; social awareness; crowdsourcing

First author: LIU Jingnan, Academician of Chinese Academy of Engineering, professor, PhD supervisor. He has long been dedicated in teaching and researching geodetic surveying theories and application, location service theories and software development, and GNSS technology application. E-mail: jnliu@whu.edu.cn

Corresponding author: GUO Chi, PhD. E-mail: guochi@whu.edu.cn

Foundation support: The National High-tech R&D Program of China(863 Program), Nos. 2013AA12A206, 2013AA12A204; the National Natural Science Foundation of China, No. 41104010; the Program of Introducing Talents of Discipline to Universities, No. B07037.