

移动群智感知质量度量与保障

Quality Measuring and Assurance for Mobile Crowd Sensing

赵东/ZHAO Dong
马华东/MA Huadong
刘亮/LIU Liang

(北京邮电大学 智能通信软件与多媒体
北京市重点实验室, 北京 100876)
(Beijing Key Lab of Intelligent
Telecommunications Software and
Multimedia, Beijing University of Posts and
Telecommunications, Beijing 100876,
China)

目前, 物联网已经进入深度发展阶段, 对物理环境更大规模、更复杂、更全面的感知需求越来越强烈。在过去十多年内, 人们主要关注以无线传感网为代表的固定部署感知网络, 用来对森林、海洋、火山等自然环境进行监测。然而, 这种传统感知模式的网络部署和维护成本很高, 不适宜进行大规模的城市感知。近几年来, 人们开始关注一种新型物联网感知模式, 即“移动群智感知^[1-3]”, 或者叫“以人为中心的感知^[4]”、“参与感知^[5]”、“机会感知^[6]”等。这种感知模式的产生一方面是由于现实世界中存在着大量的移动感知节点, 例如, 具有多达十几种传感器的智能手机、各种可穿戴设备(如智能手环、智能手表、智能眼镜等)、车载感知设备(如全球定位系统(GPS)、第2代车载自动诊断系统(OBD-II)、车载二氧化碳传感器等)或其他便携式电子设备(如Intel的空气质量传感器)。这些

收稿日期: 2015-09-06
网络出版时间: 2015-10-27
基金项目: 国家自然科学基金
(61332005); 中国博士后科学基金
(2015M570059)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2015) 06-0002-004

摘要: 认为移动群智感知网络的感知质量包含时空覆盖质量和数据质量两个层面, 前者关注是否能采集到足够多的数据, 而后者关注数据是否足够准确和可信。分别从这两个层面讨论了感知质量度量与保障的方法, 对移动群智感知网络的部署和应用具有一定的指导意义和实用价值。

关键词: 移动群智感知; 感知质量度量; 感知质量保障

Abstract: The sensing quality of a mobile crowd sensing (MCS) network can be assessed in terms of time-space coverage and data quality. Time-space coverage focuses on whether enough data can be collected, and the data quality focuses on whether the data is sufficiently accurate and credible. We discuss sensing quality measurement and assurance from the two aspects that are significant for guiding the deployment and application of an MCS network.

Keywords: mobile crowd sensing; sensing quality measuring; sensing quality assurance

感知节点通常由移动的人或车携带, 因而更容易实现对整个城市的覆盖, 我们将其称之为“移动感知”。另一方面, 最近学术界和工业界流行一种“众包”思想, 它是一种新的分布式的问题解决模式, 就是将一个复杂的问题分解成很多个简单的问题, 然后外包给大量的普通用户来协同完成。于是, 移动感知与众包思想的结合, 就产生了这种新型物联网感知模式——移动群智感知, 它将普通用户的移动设备作为基本感知单元, 通过移动互联网进行有意识或无意识的协作, 实现感知任务分发与感知数据收集, 完成大规模的、复杂的社会感知任务^[3]。

与传统的固定部署感知模式相比, 移动群智感知有三大优点: 网络部署成本更低、网络维护更容易、系统更具有可扩展性^[7], 因此更适合完成一些大规模的、复杂的感知任务,

可应用于城市环境监测、智能交通、城市管理、公共安全等领域。

然而, 很少有人关注移动群智感知网络的感知质量问题。我们认为, 移动群智感知网络的感知质量包含时空覆盖质量和数据质量两个层面, 前者关注是否能采集到足够多的数据, 而后者关注数据是否足够准确和可信。

然而, 在移动群智感知模式下, 用户的属性、位置、情境等方面的动态变化性使得我们很难对时空覆盖质量进行度量和保障; 而用户感知设备、感知方式、主观认知能力、参与态度等方面的异构性也使得我们很难对感知数据的质量进行相关的度量和保障。

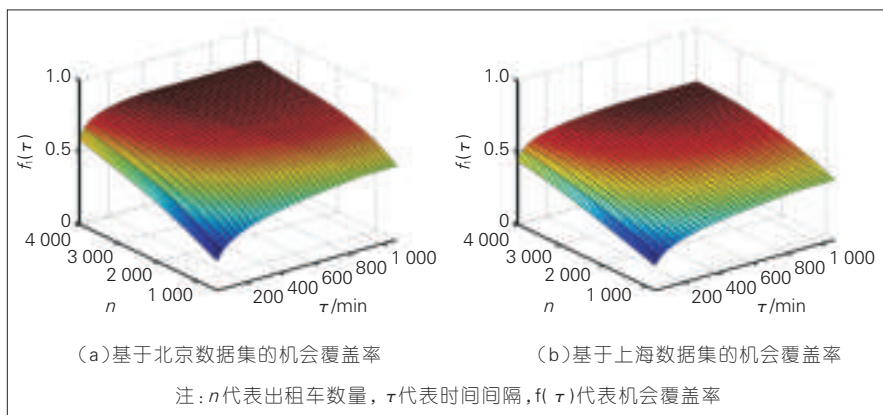
因此, 文章分别从时空覆盖质量和数据两个层面讨论感知质量度量与保障的问题以及对应的解决方法, 对移动群智感知网络的部署和应用

具有一定的指导意义和实用价值。

1 时空覆盖质量度量与保障

随着移动感知设备的持有者随机地到达城市的各个地方,这些节点即可随时随地进行感知。这种移动性对许多应用的感知质量起着重要的作用。以城市空气质量监测为例,假定我们计划使用大量的出租车携带空气质量传感器,对北京的五环内区域进行监测,构建每天早上6点到晚上12点时间段的空气质量感知地图。事实上,有两个基本问题有待解决:怎样度量这些出租车提供的感知机会以及它们能达到的感知质量?需要部署多少辆出租车能达到所需的感知质量?

首先,我们从时间维度来考虑。在传统的固定部署的传感网中,研究者常常使用覆盖率来度量感知质量,通常需要监测区域内每个点总是被至少一个传感器节点覆盖,这种网络的覆盖质量一般不会随着时间而改变。然而,由于人的移动性,移动群智感知网络的覆盖质量是动态变化的。考虑到感知覆盖的时空变化因素,我们将整个监测区域划分为多个网格单元,将每个网格单元被连续覆盖两次的间隔时间作为一个新的度量指标,称之为覆盖间隔时间,用来描述每个网格单元被覆盖的机会。通过对北京和上海两个城市的出租车移动轨迹数据集进行分析,我们发现覆盖间隔时间服从截断的帕累托分布。进一步地,我们提出一个称作“机会覆盖率”的度量指标来表示城市监测区域的整体感知质量与节点个数之间的关系,其定义为在特定时间间隔内能被覆盖的网格单元占所有网格单元的比例的期望值,可以表示成关于覆盖间隔时间分布的函数。图1显示基于北京和上海两个出租车移动轨迹数据集的机会覆盖率与节点个数和时间间隔呈单调递增关系。于是,我们可以推导出至少需要多少节点能使在特定的时间间

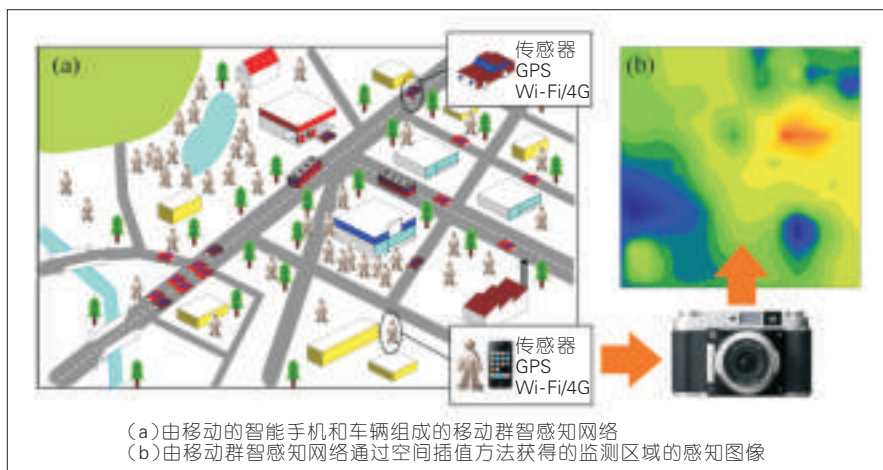


▲ 图1 基于北京和上海两地出租车移动轨迹数据集的机会覆盖率

隔内机会覆盖率不小于指定的阈值。例如,根据对两个出租车移动轨迹数据集的分析,我们需要分别在北京和上海900平方千米的区域内至少部署1700辆和1900辆出租车,才能保证其在1个小时的时间间隔内机会覆盖率不小于50%。更详细的分析方法和结果请见文献[8]。尽管不同城市可能需要不同的节点个数满足所需的机会覆盖率,我们提出的模型和方法可以对网络规划问题提供一般性的指导。

其次,我们从空间维度考虑。一个监测区域的环境现象(如PM2.5浓度、二氧化碳浓度、噪音等)可以表示为一个二维信号,类似于一个图像。在大部分环境监测应用中,移动感知节点将采集到的感知数据发送到数据中心,然后数据中心汇集现有数

据,并利用空间插值技术估计未知数据,从而得到一个完整的感知地图,相当于一个感知图像。如图2所示,移动群智感知网络就像一个“城市摄像机”,而每个移动感知节点就相当于这个摄像机的每个“像素”。在传统的图像系统中,分辨率是度量图像质量的一个重要指标。从这个概念得到启发,我们利用群智感知分辨率作为指标来度量感知图像的质量。分辨率越高,则代表所部署的移动群智感知网络越能准确地捕获到环境现象的变化。然而,与传统数字图像系统中的分辨率定义不同,我们不能简单地将像素数(即移动感知节点个数)看做群智感知分辨率。这是因为,数字摄像机的像素会形成一个精细的网格,而城市摄像机的像素在城市中则呈现分散化的动态化分布。



▲ 图2 基于移动群智感知网络的城市环境监测

为了解决这个问题,我们首次提出“城市分辨率”这一新的指标来度量城市感知图像的质量。简单地说,我们分别基于移动群智感知网络和 $n \times n$ 网格化部署的感知网络采集到的部分感知数据,利用空间插值技术来估计未知数据,得到两个完整的监测区域感知数据矩阵,然后利用相关系数来评估两个矩阵的相似性,如果它们的相似性足够高,则认为该移动群智感知网络的城市分辨率是 $r=n \times n$ 。我们分别使用3种不同变化度的二维信号,通过蒙特卡罗仿真研究了分辨率 r 与移动感知节点个数 s 之间的关系,发现它们之间存在一个近似的线性关系: $r=\alpha^2 s$,其中,在真实的人或车移动模型下, α 的参考值范围是[0.5,0.6]。一方面,基于该线性关系,我们就可以根据移动群智感知网络的节点个数推出所能达到的感知质量;另一方面,我们也可以反过来推出,需要部署多少移动感知节点能达到所需的分辨率需求。同样以北京和上海两个出租车移动轨迹数据集为例,图3显示了其城市分辨率分布情况。可以看出,人或车密集分布的区域具有更高的城市分辨率,例如北京的中心和东部区域,以及上海的中心和西南区域。更详细的介绍请见文献[9]。

上面提到的方法主要适用于城

市环境监测、交通拥堵状况和道路健康状况监测等需要对整个城市的每个区域进行连续监测的大规模城市感知应用。与此不同,Chon等人则研究了“以地方为中心”的移动群智感知应用的覆盖质量^[10]。所谓以地方为中心的应用,就是自动识别或跟踪用户每天访问的不同地方(如咖啡馆、超市、办公室、家、学校等),来帮助用户认识和分析自己的日常行为模式,或者获取基于位置的搜索和信息推荐等服务。构建和部署这些应用的前提是对用户访问的每个地方采集足够的感知数据(如GPS位置、声音、图像、光照、Wi-Fi信号指纹等)来建立各种模型。这里的覆盖问题就是:多长时间内多少用户采集数据能覆盖到多少人们经常访问的地方?为此,Chon等人在韩国招募了85人并在两个月时间内收集了大约4.8万次用户访问不同地方的感知数据然后进行分析,得到了一些有趣的结果:仅仅利用少量的用户(85人),就能对人们常去的地方提供高覆盖率(最流行的地方的15%);用户访问地方的个数服从幂律分布,基于该模型可估计出需要多少用户能达到所需的地方覆盖率;用户对泄露个人隐私的担心并没有带来太大的影响,例如,用户允许在93%的食物相关的地方收集声音数据,在82%的所有类型

的地方收集声音数据。

以上方法主要用来度量时空覆盖质量,而保障时空覆盖质量则需要综合利用多种方法:(1)可以将固定部署感知模式和移动群智感知模式相结合,在移动用户很难到达的空白区域,通过优化部署固定的感知网络来保障时空覆盖质量;(2)利用空间插值、压缩感知等方法来弥补感知数据的缺失,也可以利用感知现象的时空相关性,或者天气、交通、重大事件等外部因素与感知现象的关联性,采用机器学习的方法来实现未知数据的准确估计;(3)设计合适的激励机制来鼓励更多的用户来参与这些感知活动。

2 感知数据质量度量与保障

感知数据质量受很多方面因素的影响,主要包括:

- 用户所使用的感知设备类型。例如,价格高昂的高端手机的传感器一般比那些价格低廉的低端手机的传感器精度要高。

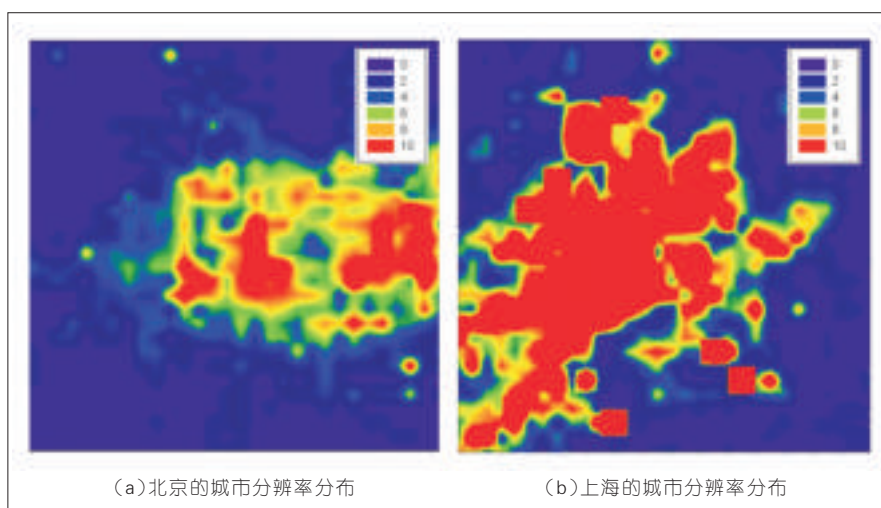
- 用户采集数据的环境和方式。例如,把手机拿在手里采集环境噪声的数据质量比把手机放在衣服口袋或手提包里采集环境噪声的数据质量高。

- 用户的主观认知能力。例如,基于移动群智感知的图像搜索应用依赖用户对图像的识别能力,而不同用户对同一图像的认知可能是不一样的。

- 用户的参与态度。例如,有的用户会严格按照要求来采集数据,而有些用户会比较随意,甚至有些恶意用户会上传虚假伪造的数据。

以上因素都会造成感知数据质量的参差不齐。下面,我们首先根据感知任务或对象的类型来介绍几种典型的感知数据的质量度量和保障的方法:

- 面向二进制型任务的方法。二进制型任务的结果只有两种。事件检测是一种典型的二进制型任务,



▲ 图3 基于北京和上海两地出租车移动轨迹数据集的城市分辨率分布

即判断某种事件是否发生。最简单的方法是投票,即当判定事件发生的用户数量超过特定阈值的时候,才最终确定事件发生。

- 面向多类别型任务的方法。多类别型任务的结果多于两种,例如,用户对某个事物的评价可以打分为1~5的某个分数。投票法虽然也可以用来度量结果的不确定性,但还不够准确。最大期望法是一种常用的更准确的方法,它采用迭代的方式工作,即首先根据用户的感知数据来估计用户的可靠性,然后根据用户的可靠性来估计最终的任务结果,并不断重复上述过程。

- 面向连续信号型任务的方法。对区域环境现象的连续监测属于连续信号型任务。Koutsopoulos针对这类任务提出了一种感知数据质量度量方法,即计算某个用户提交的历史数据与所有用户数据的平均值之间的累积误差作为该用户的感知数据质量指标^[11]。

以上3种方法的基本思想都是发挥集体的智慧来抵御个人数据不准确的影响,从而提高整体数据的可靠性。然而,这些方法并不能充分应对恶意用户的攻击。Mousa等人总结了串谋攻击、女巫攻击、GPS欺骗等11种可能的恶意用户攻击方式^[12]。面对这些攻击,一般有两类方法解决感知数据的可信性问题:

- 可信平台模块。这类方法是在用户的移动感知设备设置专门的硬件模块,保证用户感知和上报到数据中心的数据是由真实的、授权的感知设备所采集,还可以采用签名和硬件加密机制来保护感知数据只能由授权用户访问。与此类似,我们基于“安全数码相机”的思想,利用MD5算法和基于随机数的加密算法设计了一个图像篡改检测方法来保障用户上传图像数据的真实性^[13]。

- 信誉系统。这类方法是评估和记录用户的历史感知数据的可信性,并将其用在未来的系统交互过程

中,对于信誉度低的用户感知数据采用的可能性也比较低,同时也会采用相应的激励或惩罚措施。贝叶斯系统是一种常用的具体方法^[14]。

3 结束语

作为物联网的新型感知模式,移动群智感知促进了大量创新应用的出现,同时也面临一系列新的问题与挑战。文章分别从时空覆盖质量和数据质量两个层面分析了移动群智感知网络中感知质量度量和保障的各种问题,并提供了一些对应的解决方法。随着相关技术的不断进步和成熟,移动群智感知质量度量与保障将对移动群智感知网络的快速发展和广泛应用提供重要的支撑作用。

参考文献

- [1] GANTI R K, YE F, and LEI H. Mobile Crowdsensing: Current State and Future Challenges [J]. IEEE Communications Magazine, 2011, 49 (11): 32-39
- [2] MA H D, ZHAO D, and YUAN P. Opportunities in Mobile Crowd Sensing [J]. IEEE Communications Magazine, 2014, 52 (8): 29-35
- [3] 刘云浩. 群智感知计算[J]. 中国计算机学会通讯, 2012, 8 (10): 38-41
- [4] CAMPBELL A, EISENMAN S, LANE N, et al. The Rise of People-Centric Sensing [J]. IEEE Internet Computing, 2008, 12 (4): 12-21
- [5] BURKE J, ESTRIN D, HANSEN M, et al. Participatory Sensing [C] // Workshop on World-Sensor-Web, Co-Located with ACM SenSys, Boulder, Colorado, USA, 2006: 1-5
- [6] LANE N, EISENMAN S, MUSOLESI M, et al. Urban Sensing Systems: Opportunistic or Participatory? [C] // in Proceedings of HotMobile, Silverado Resort, Napa Valley, USA, 2008: 11-16
- [7] 赵东,马华东. 群智感知网络的发展及挑战[J]. 信息通信技术, 2014(05): 66-70
- [8] ZHAO D, MA H D, LIU L, and LI X Y. Opportunistic Coverage for Urban Vehicular Sensing [J]. Computer Communications, 2015: 71-85. doi:10.1016/j.comcom.2015.01.018
- [9] LIU L, WEI W, ZHAO D, and MA H D. Urban Resolution: New metric for Measuring the Quality of Urban Sensing [J]. IEEE Transaction on Mobile Computing, 2015(09): 21-25. doi: 10.1109/TMC.2015.2404786
- [10] CHON Y, LANE N, KIM Y, et al. Understanding the Coverage and Scalability of Place-Centric Crowd Sensing [C] // in Proceedings of ACM UbiComp, Zurich, Switzerland, 2013: 3-12
- [11] KOUSOPOULOS I. Optimal Incentive-Driven Design of Participatory Sensing Systems [C] // Proceedings of IEEE

- INFOCOM, Turin, Italy, 2013: 1402-1410
- [12] MOUSA H, MOKHTAR S B, HASAN O, et al. Trust Management and Reputation Systems in Mobile Participatory Sensing Applications: A Survey [J]. Computer Networks, 2015, 90. doi:10.1016/j.comnet.2015.07.011
- [13] QIN T, MA H D, ZHAO D, et al. Crowdsourcing Based Event Reporting System Using Smartphones with Accurate Localization and Photo Tamper Detection [C] // in Proceedings of BIGCOM, Taiyuan, China, 2015: 141-151
- [14] TANAS C, HERRERA- JOANCOMARTI J. When Users Become Sensors: Can We Trust Their Readings? [J]. International Journal of Communication Systems, 2015, 28 (4): 601-614

作者简介



赵东, 北京邮电大学博士后; 主要研究方向为物联网、群智感知、无线传感网、机会网络等; 已发表论文20篇。



马华东, 北京邮电大学教授、博士生导师、计算机学院执行院长, 智能通信软件与多媒体北京市重点实验室主任, 国家杰出青年科学基金获得者, 教育部长江学者特聘教授, 国家“973”计划项目“物联网体系结构的基础研究”首席科学家; 主要研究方向为多媒体系统与网络、物联网与传感网; 主持完成20余项科研项目; 已发表论文200余篇。



刘亮, 北京邮电大学副教授、博士生导师; 主要研究方向为物联网与移动计算; 主持和参加基金项目10余项; 已发表论文60余篇。