

城市空气质量感知方法综述

王鹏跃^{1,2} 郭茂祖^{1,2} 赵玲玲³ 张 昱^{1,4}

(北京建筑大学电气与信息工程学院 北京 100044)¹

(建筑大数据智能处理方法研究北京市重点实验室 北京 100044)²

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)³

(深部岩土力学与地下工程国家重点实验室 北京 100083)⁴

摘 要 城市空气质量信息对于控制空气污染和保护大众健康都是尤为重要的。城市空气质量感知方法按传感器位置是否发生改变可划分为静态感知方法和动态感知方法两种。其中静态感知方法的数据是基于空气质量监测站、卫星遥感和固定位置的传感器进行采集的。再按成本高低可进一步划分为低成本静态感知和高成本静态感知。动态感知方法按是否以参与者为感知中心划分为参与式方法和非参与式方法。随着感知技术和计算能力的发展,将多源异构的城市数据,如气象数据、交通数据等进行融合,可进一步提高感知的准确性。文中首先对当前空气质量感知方法进行综述,然后分类介绍了各种方法的感知框架和数据处理方法,最后讨论了其面临的问题和挑战。

关键词 空气污染,城市感知,数据采集,机器学习

中图分类号 TP181 文献标识码 A

Review on Urban Air Quality Perception Methods

WANG Peng-yue^{1,2} GUO Mao-zu^{1,2} ZHAO Ling-ling³ ZHANG Yu^{1,4}

(School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)¹

(Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing 100044, China)²

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)³

(State Key Laboratory in China for GeoMechanics and Deep Underground Engineering, Beijing 100083, China)⁴

Abstract Urban air quality information is especially important for controlling air pollution and protecting public health. According to whether the sensor position changes, urban air quality sensing methods can be divided into two methods: static perception methods and dynamic perception methods. The data acquisition of the static sensing method is based on air quality monitoring stations, satellite remote sensing and fixed position sensors. Then, the static sensing method is further divided into low-cost static sensing method and high-cost static sensing method. The dynamic sensing method can be divided into participatory method and non-participatory method according to whether the participant is the perceptual center. With the development of sensing technology and computing ability, the fusion of multi-source heterogeneous urban data, such as meteorological data and traffic data, can further improve the accuracy of perception. This paper firstly summarized current air quality sensing methods, then classified the sensing framework and data processing methods of various methods, and finally discussed the problems and challenges.

Keywords Air pollution, Urban sensing, Data collection, Machine learning

1 引言

空气是人类赖以生存的重要资源。近年来,空气质量在不断恶化。世界卫生组织(WHO)提供的最新数据显示,每10人中就有9人呼吸含有高浓度污染物的空气,每年因空气污染造成的死亡数达到惊人的700万。空气污染问题已经成为一个亟需解决且极具挑战的全球性问题。在空气污染的众

多维度中,细粒度颗粒尤其是PM_{2.5}越来越受到关注,PM_{2.5}的直径过小,可渗透到肺部和心血管系统导致中风、肺癌等疾病,严重影响人们的身体健康^[1]。因此,公布空气质量信息可以让人们时刻了解自身所处环境的空气质量并采取相应的防护措施,如减少外出或佩戴口罩等;还可以提高人们保护环境、减少污染的意识;最为重要的是,可以帮助政府和机构研究污染源,有效控制空气污染。因此,感知城市空气质量具有

本文受国家自然科学基金(61502117),北京市教委科技计划重点项目(KZ201810016019),国家重点研发计划(2016YFC0600901),教育部产学研协同育人项目(201801113001),北京建筑大学市属高校基本科研业务费专项资金(X18197, X18198, X18203, X18018),北京建筑大学双塔计划优秀主讲教师(21082718041)资助。

王鹏跃(1996—),女,硕士生,CCF会员,主要研究方向为机器学习、智慧城市;郭茂祖(1966—),男,博士,教授,博士生导师,主要研究方向为机器学习、智慧城市、生物信息学, E-mail: guomaozu@bucea.edu.cn(通信作者);赵玲玲(1980—),女,博士,讲师,主要研究方向为城市计算与智能信息处理;张昱(1979—),男,博士,副教授,硕士生导师,主要研究方向为大数据、机器学习、智慧城市。

十分重要的意义。

学者们针对这一问题建立了不同的空气质量感知系统。传统的感知空气质量的方法基于空气质量监测站和卫星遥感。随着无线传感网络的发展,各种低成本传感器涌现,将低成本传感器进行大规模部署,可以降低感知的成本并提高感知的细粒度。同时,大数据和人工智能的兴起,将多源的城市数据如气象信息、交通信息等进行融合,再应用机器学习算法对数据进行分析,进一步提高感知的准确度^[2],并为感知空气质量提供了新的思路。图 1 给出了当前城市空气质量感知的基本框架。

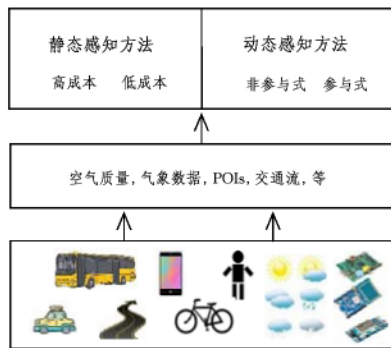


图 1 城市空气质量感知的基本框架

本文对当前的城市空气质量感知方法进行综述,根据传感器位置是否固定将各感知方法划分为静态感知方法和动态感知方法两大类。在静态感知方法中,传统的基于空气质量监测站和卫星遥感的感知方法因成本过高由国家政府来设置,属于高成本静态感知方法。与之相对的低成本静态感知方法,即将低成本传感器进行大规模固定位置部署,以大范围、细粒度地感知空气质量。在动态感知方法中,根据是否以人为感知中心将感知方法划分为参与式感知方法和非参与式感知方法。非参与式感知方法通常将传感器部署在移动车辆上,通过传感器在空间上的移动,达到更低成本、更大范围感知空气质量的目的。参与式感知方法需要公众的参与,将便携式传感器与手机结合或利用手机拍照功能识别图像进行雾霾分析,实时获取数据、上传数据。在参与式感知方法中最重要的问题是数据质量问题,这也带来了相应的问题和挑战。本文详细介绍了各类空气质量感知方法,并讨论其中存在的问题和面临的挑战。

2 静态感知方法

在静态感知系统中,基于空气质量监测站和卫星遥感的方法是自上而下的,因成本过高,通常由国家来部署,属于高成本静态感知方法。低成本静态感知方法是通过选择合适的低成本传感器,将其进行大规模部署来提高感知城市空气质量的细粒度。在低成本静态感知方法中,最关键的问题是如何权衡传感器的成本与准确度。因低成本传感器的成本过低、准确度不高,所以在部署前需对传感器进行校正,在后端感知算法上也需进行改进。

2.1 高成本静态感知方法

2.1.1 基于监测站的感知方法

现今,越来越多的国家开始监测空气质量并公布数据,其中最常用的感知方法是基于空气质量监测站。空气质量监测站的测量精度高,但存在成本过高(约 200 000 美元)、占地面

积大,且需要定期维护等问题^[3],这也就限制了其大规模部署。因此,对空气质量监测站进行合理部署就更为重要。从二十世纪六七十年代开始,美国、欧洲等发达国家的研究人员对如何设计出效率高、代表性好、目的性强的空气质量监测网络进行了大量研究。研究包括从国家级到城市区域各种规模的监测网络。经过长时间的研究、应用及不断改进,美国和欧盟设计了一套适合自身情况的空气质量监测网络。随着城镇化推进和人们对城市环境要求的不断提高,我国的环境空气质量监测系统也在不断发展并完善。2007 年,我国出台了试行版的环境空气质量监测规范,其中对监测网络的设计也做出了相关的规定。目前,我国已建立了 1 497 个监测站点,空气质量监测网络主要有两个功能:1)评价当前城市的空气污染水平;2)监测城市空气质量的历史变化,但这不能为环境污染防治提供决策依据。我国监测站位置较稀疏,不能实时、准确地掌握各个地点的空气质量状况,不能满足公众需求,难以利用监测数据进行分析,更不能为污染源及污染演变的数据挖掘提供足够的数据支持。如北京只有 35 个空气质量监测站,是十分稀疏的。将北京与其他国家的城市设置的空气质量监测站数量进行对比(如表 1 所列),可以看出,在平均覆盖区域中,每个监测站覆盖的平均区域过大,因此北京空气质量监测站的细粒度极低。

表 1 不同城市的空气质量监测站数量

城市	监测站数量	覆盖区域 / km ²	平均覆盖区域 / (km ² / per station)
中国, 北京	35	16 000	457
德国, 柏林	12	890	74
英国, 伦敦	83	1 600	19
美国, 洛杉矶	37	1 290	35
印度, 孟买	13	610	47

在环境科学上感知空气质量主要有两种传统的方法。一种是根据监测站之间的距离来推理未设监测站处的空气质量,如插值、SPA 技术(Single Point Areal Estimation)等方法^[4-5]。由于空气污染浓度在城市空间范围内是非线性的且受很多其他因素如天气、交通流量等影响,因此精度非常低。另一种是经典的色散模型,例如高斯羽流模型和计算流体力学^[6]等。在大多数情况下,这些模型是气象学、交通量和排放因子的函数,基于许多经验假设且不适用于所有城市环境的参数^[6]。如高斯羽流模型假设车辆尾气以高斯方式在垂直和水平方向上分散;有些模型甚至需要建筑物之间的间隙及建筑物表面粗糙系数。由于极难精确地获得这些参数,因此由这种模型产生的结果也不准确。

随着感知技术和计算能力的发展,获取城市信息更加便捷,利用城市大数据进行数据挖掘成为一个热点问题^[2,7]。与这些模型相比,利用城市大数据进行数据驱动的方法不需要经验假设和参数,易于实施并可适用于不同的城市环境。

Hsieh 等^[8]基于城市大数据来确定最佳空气质量监测站的选取位置。Zheng 等^[3]基于数据集识别 5 组特征(F_m, F_t, F_h, F_r, F_p),其中 F_m 表示天气数据, F_t 表示交通相关特征, F_h 表示人流量, F_r 表示道路网络相关特征, F_p 表示 POIs,其提出一种基于协同训练的半监督学习方法,由两个独立的时空分类器组成,其中空间分类器基于人工神经网络(Artificial Neural Network, ANN),将空间相关的特征,如 POIs(point of interests)密度、高速公路长度等作为输入,模拟不同位置的空

气质量之间的空间相关性。时间分类器基于线性链条件随机场(linear-chain conditional random field,CRF),利用涉及时间相关特征,如天气数据、交通信息等模拟空气质量在个体位置上的时间依赖性。这两种分类器在协同训练的框架中互相加强来感知城市空气质量。这种方法也适用于不同的城市环境。但是由于公共监测站部署的稀疏性以及公共监测站数据的时效性,使得该方法存在推断粒度较大、精度一般、不能及时反映紧急情况等问题。Zheng 等^[9]使用时间预测器、空间预测器、结合时间与空间预测器的动态聚合器和拐点预测器共同来预测空气质量未来 48 h 的变化。

Yi 等^[10]提出了一种基于神经网络的方法(Deep-Air),它由空间变换组件和深度分布式融合网络组成。考虑到空气污染物的空间相关性,前者将空间稀疏的空气质量数据转换为一致的输入来模拟污染源。后者采用深度分布式结构来融合异构城市数据,同时得出影响空气质量的因素。与文献^[9]相比,该方法在短期、长期以及突变预测上的相对准确度分别提高了 2.4%,12.2%,63.2%。

目前利用公共监测站的数据来进行空气质量的研究,其优点在于数据精准,但空间上的稀疏分布使得很难在细粒度上进一步分析并得到更深层次的数据挖掘结果,未能进一步探究 PM_{2.5} 污染指数与各种因素,如地理位置因素、环境因素、气象因素等之间的关系,而这些关系正是后续环境治理所必要的。近年来城市计算的发展给基于空气质量监测站的空气质量监测方法带来了新思路,即不需要经验假设。因此,它易于实施并适用于不同的城市环境。

2.1.2 基于卫星遥感的感知方法

与地面空气质量监测方法相比,卫星遥感具有范围大、可获取连续动态数据等优势,能弥补地面监测站在空间上的不足,实现对未设置监测站区域的监测。基于卫星遥感的感知方法主要是围绕气溶胶光学厚度(Aerosol OpticalDepth, AOD)展开,近地层颗粒物浓度与 AOD 数据具有相关性,通过 AOD 数据来研究地表颗粒物浓度是有意义的。建立相关性的方法主要分为两种:1)利用遥感 AOD 数据直接与近地面 PM_{2.5} 建立相关性;2)先对遥感 AOD 数据进行修正,在修正后的数据和近地面 PM_{2.5} 数据之间建立相关性。

利用卫星 AOD 反演地面颗粒物浓度的相关工作中,Engle-Cox 等^[12]将 MODIS 标准 AOD 与美国地面空气质量监测站的 PM_{2.5} 进行相关分析,利用二者的相关系数,估算近地面 PM_{2.5} 情况,验证了卫星遥感估算近地面颗粒物的适用性。Wang 等^[13]研究发现 AOD 数据与 PM_{2.5} 值之间呈线性相关,具有一致性。还有一些学者将所建立的不同模型进行了对比分析^[14]。因此基于遥感技术获取 AOD 数据直接反演得到近地面颗粒物浓度来感知城市空气质量的方法是可行的。

在对 AOD 进行修正后与近地面 PM_{2.5} 建立相关性的研究中,VAN 等^[15]将从 MODIS 推断的 PM_{2.5} 值与加拿大和美国的实际地面测得的 PM_{2.5} 值进行对比。为获得更高的准确性,学者们不断提出新的算法,郑卓云等^[16]利用改进后的反演算法反演得到高分辨率的 AOD 数据,Wang 等^[17]计算了 AOD 与 PM₁₀ 和 PM_{2.5} 的相关性。

目前,AOD 与近地面颗粒物浓度之间仍很难建立一个可靠、稳定、可大范围应用的模型,且该方法极易受到各种环境条件的影响,如云层情况、温度、湿度等,这些影响因素可能会

使得最终的监测结果出现较大的误差^[15]。将气溶胶垂直分布和大气水汽等影响考虑在内,可以提高相关程度,但仍存在局限性。同时,基于卫星遥感的空气质量感知方法所推断出的结果只反映了大气的空气质量,而不是人们更关心的地面空气质量。

2.2 低成本静态感知方法

传统的感知城市空气质量方法都是基于空气质量监测站和卫星遥感,但这两种方法成本昂贵。尽管空气质量监测站十分可靠、精度高,但其占地大、成本昂贵导致监测站在部署上具有稀疏性,不能细粒度地感知城市范围的空气质量。同样,卫星遥感也存在局限性。而人们更希望可以直接监测其当前所处位置的空气质量,这就对传感器的密集部署提出了要求。因此,低成本静态感知方法被提出,通过低成本传感器的大量部署来弥补监测站在空间上的稀疏性。低成本静态感知方法的核心是在保证成本和精度的条件下最大范围地监测空气质量。这就需要权衡传感器的硬件成本和精度。一般的硬件校正是出厂后硬件依靠自身写入的校正曲线来校准。在低成本静态感知方法中可以将出厂标定与实时云端联合对设备进行校准,在后端也可利用算法提升传感器的准确度,这使得使廉价的传感器也可获得较为精准的读数,为低成本静态感知方法提供了精度保证。表 2 对低成本感知方法中常用的低成本传感器的性能进行了对比。

表 2 常用的低成本传感器的性能

传感器	感知颗粒大小	感知颗粒类型	工作温度 /℃	工作湿度 /%	价格 /元
GP2Y1010	<100 μm	TSP	-10~65	0~99	16
PPD42NS	>1 μm & >2.5 μm	Dust	0~45	0~95	26
DN7C3CA006	<2.5 μm	PM _{2.5}	-10~60	10~90	120
PMS3003	<1 μm & <2.5 μm & <10 μm	PM ₁ , PM _{2.5} , PM ₁₀	-5~50	0~99	70

Wang 等^[18]对常用的各低成本传感器性能进行了分析,为如何选择合适的传感器提供依据。Gao 等^[19]研究了低成本传感器在高 PM_{2.5} 浓度地区的表现,将 PUWP(Portable University of Washington Particle)监测器部署在西安的 8 个站点。研究表明,PUWP 监测器作为一个可行的低成本监测器,可用于加强现有的 PM_{2.5} 采样网络,提高 PM_{2.5} 数据的时空分辨率,但是该研究并没有探究季节变化和气象条件对 PUWP 传感器的影响,虽然研究发现五阶多项式是对西安空气质量的最佳拟合,但同样的模型不一定适用于其他地方,具有局限性。

Khadem 等^[20]设计了一个分布式的节点网络,通过分布式感知和信息处理来监测大型工厂的灰尘,但其着重于网络方面,并没有包含感知方法的详细信息。CitySee 项目^[21]在中国无锡市部署了 1196 个传感器,覆盖约 100 平方千米的区域,实时采集多维数据,包括二氧化碳、温度、湿度、光照、位置等,在传感器供电上采用太阳能驱动。该项目中,虽然每一个传感器的成本很低,但若想感知整个城市范围的空气质量就需要部署大量的传感器,因此总成本仍十分昂贵,但若仅针对某一区域进行监测还是十分可行。

Cheng 等^[22]提出了一种低成本、可大范围部署的空气质量监测系统 AirCloud,其在前端设计了两种监测仪 AQM 和 miniAQM。AQM 成本低廉,可监测 PM_{2.5}、PM₁₀、温度、湿

度、GPS 等多个参数,采集数据频率为 1min,可进行大批量的部署,在部署前需进行校正。在其云端进行空气质量分析,将传感器获得的多源数据进行融合,来学习和建立空气质量模型,并且基于 ANN 的校准模型,提高 AQM 和 miniAQM 的实时精度,在通过一个基于高斯过程的在线推理模型,利用所获得的多源数据来推算未设置传感器地区的空气质量浓度。AirCloud 确实可以采用低成本的传感器来感知空气质量,但因传感器位置固定,存在覆盖域有限的问题,若要实现更高的覆盖率和准确度,就需要部署更多的传感器,这也大大增加了总成本和管理难度。在保证低成本特性不受影响的情况下,需要新的设计来实现更高的覆盖率。

3 动态感知方法

在静态感知中,虽然每个传感器成本很低,但若想大范围监测城市空气质量,就需大量地部署传感器来覆盖整个城市,以中国的一个中等城市杭州为例,其总占地面积约 16 000 平方千米,若想通过部署静态传感器监测整个城市空气质量,那么总成本仍十分昂贵且很难管理这个系统。因此有学者提出可以通过传感器的移动来大范围感知城市空气质量。动态感知是一个相对较新的概念,它通过适量传感器的移动便可实现对一个区域的覆盖。从本质上讲,移动节点的时间分辨率与空间分辨率并不相关,因而不需要成千上万的固定传感器便可实现大范围感知。这既减少了传感器的成本,又达到了大范围感知的效果。在动态感知方法中,将传感器部署在移动车辆上的方法称为非参与式动态感知方法;将与手机相关,需要人参与完成的动态感知方法称为参与式动态感知方法。

3.1 非参与式动态感知

非参与式动态感知方法通常是将传感器安置在移动的车辆上,如自行车、清洁车、公交车等。利用车的移动性来大范围感知整个城市的空气质量,弥补现有空气质量监测站的粗粒度分布,提高感知的细粒度。在动态感知框架中,传感器节点是动态空气质量监测网络的基础模块,许多其他组件如负责处理、过滤、存储和重建信息的服务器是实现该架构所必需的。除了这些组件之外,系统还可使用外部信息源来增加其准确性和稳健性,如区域地图、当地天气和交通信息,以及来自空气质量监测站的数据。

Tudose 等^[23]提出了城市环境的空气质量监测的移动系统,但不是普遍适用的,只监测人们居住地区的污染(大多数时候这些地区也是污染最严重的地区,是人们更为关心的地区)。由于采用模块化设计,该系统可以做更多的扩展和应用,如,确定城市内的主要污染源、预测空气污染模式等。

Devarakonda 等^[24]设计了两种动态感知模型,一种基于公共交通;另一种基于私家车。其设计了网络协议去处理不同传感器产生的数据形式差异,但测量结果受传感器的安置方式以及车辆行驶速度的影响,需进一步讨论。污染程度易受天气影响,但该方法未将天气因素考虑进来。

Hedgecock 等^[25]设计了车载装置来实时感知空气质量,提高感知细粒度。系统基于 Web 可视化进行数据检索,用户不仅能够下载多个移动传感设备的原始传感器数据,而且能够实时地获取吞吐量信息,将数据可视化易于理解的污染图。此系统在互连的分布式服务系统上运行,增加了系统的稳健性和容错性。其将输入的数据分阶段处理,存储每个阶

段的输入、输出,以防某一阶段出现错误。只有成功完成所有阶段的数据任务才会被存储,避免了删除故障数据的繁琐操作并提高了生成数据的整体质量。

Hasenfratz 等^[26]将传感器(MiniDiSCs^[27])安装在公共交通工具上组成移动感知系统;利用采集到的 2 500 万次 UFP 测量数据,设计了土地使用回归(Land-use Regression Models, LUR)模型,生成具有高时空分辨率的精确污染图(Ultrafine Particles, UFP)。LUR 模型使用解释变量(土地使用和交通数据)对节点未覆盖位置的污染浓度进行建模。系统中每个传感器节点的成本都在三千到五万美元之间,虽然获得较高的精度,但成本过高。因此,学者们致力于寻找一个更低成本的空气质量感知系统。

Dong 等^[28]设计了一种低成本的可感知城市规模的移动传感器 Mosaic,整个感知系统由 3 层组成:传感器网络层、云层和应用层。在传感器网络层,传感器数据(PM_{2.5}、温度、湿度)和轨迹信息(GPS 提供)都被 GPRS 传输到云端。在云层采用阿里云提供的云服务器,对原始传感数据存储和处理。在应用程序层,通过使用 HTTP 协议与云层交互,可实时查询数据并进行可视化,为终端用户提供服务。他们还提出了一种新的校正方法,将传统的人工神经网络(ANN)和支持向量机(SVMs)进行结合,并与 ANN, Multi-SVMs 对比,验证其在精度和 recall 方面的优越性。结果表明, Mosaic 是一种可行的城市空气质量感知方法,可以产生高精度的 PM_{2.5} 测量值。

Gao 等^[29]利用 POI 来筛选公交线路以期获得更高的覆盖率,并对文献^[28]中将传统的人工神经网络(ANN)和支持向量机(SVMs)进行结合的方法进行改进,将采集的数据进行过滤,通过平滑处理和一个 3 层 ANN 模型^[22],再加入温度和湿度等天气因素,得出最终的空气质量数据。利用 POI 筛选公交线路,就可以使用少量的传感器感知更大范围的空气质量,但是采集的数据并不具有时效性,并不能实时感知空气质量。

非参与式动态感知方法可以通过部署适量的传感器,借助传感器在空间上的移动达到大范围感知的效果。但有些问题需要进一步讨论,在动态感知中,传感器进风口朝向不同会影响监测数据结果,如何确定最佳进风口朝向;因采样结果会受到车辆移动速度和污染物空间梯度影响,如何确定最佳的采样时间;需要进一步调查普通司机在获得细粒度空气数据时的反应,有多少司机会选择空气更清洁但需花费更久时间的路线。

3.2 参与式动态感知

目前有很多手机应用直接将公共监测站的数据记录提供给用户,但这并不是用户当前所在位置的空气质量。而用户所处位置的 PM_{2.5} 值才是用户更为关注的,与其健康有直接关联的。随着移动通信和传感技术的发展,可将一个复杂的问题分解成众多小任务,并将这些小任务分配给用户网络,从而形成能够解决复杂问题的集体知识^[30-31]。由此,参与式动态感知将便携传感器与手机结合使用,可解决无线传感网络大规模部署所带来的高成本和难管理的问题,成为感知城市空气质量的一种新途径^[32]。便携式传感器可以提升现存感知网络的密度,参与用户越多,就可以覆盖更多的区域获得更好的细粒度,从而更好地感知城市空气质量。此外,设备之间

的共享信息可以帮助协同位置的设备校正测量误差。将实际数据的色散模型、校准时常、设备类型等综合考虑进来校正数值,能够进一步提高测量准确性。参与式感知方法使得大众和专业人员都可以共享并分析数据^[33],提高了空间分辨率,达到对城市范围的准确估计。参与式感知系统的结构如图2所示。

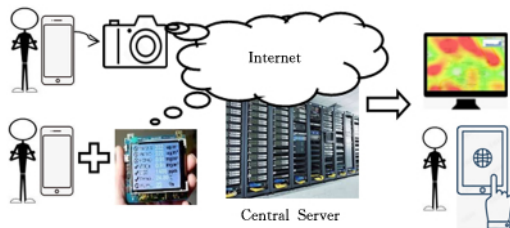


图2 参与式感知系统的结构

分布式的、可移动的空气质量管理传感器应该满足以下要求:1)便携性,传感器体积小,便于携带;2)低成本,参与式动态感知需要吸引大众参与进来,因此需要低成本;3)可用性,避免频繁的维护,如频繁充电、专家校准等;4)响应性,识别来源并启用,确保数据的及时性;5)准确性,读数应有意义。

Zhuang等^[34]展示了一个便携、低成本的个人空气质量监测系统 AirSense,将与使用者所关联的信息如环境、行为等因素考虑进来,即加入位置信息、行为活动、温湿度等,让使用者获得更准确的空气质量信息,但是这种监测方法并不适用于所有的场景,具有局限性。Budde等^[35]调查了将廉价的、商品粉尘传感器集成到手机中进行空气质量测量的可行性,并首次讨论了应用这种低成本方法测量不同场景中空气质量的情况。Budde等^[36]提出了一种用于参与式感知场景的可移动、低成本的颗粒物传感方法。将高精度的测量设备性能与廉价的商用现货(COTS)传感器进行对比,验证了廉价的COTS传感器可以用于分布式移动的个人测量装置,不仅成本低而且可以达到有意义的准确性,在实际应用中效果很好。Chen等^[37]设计了一个PM_{2.5}监测的参与式城市感知系统,并在30个国家中共部署了超过2500台设备,这是目前世界上最大规模的PM_{2.5}监测项目之一。此系统的关键特征是其开放式系统架构,该架构基于开源硬件、开源软件和开源数据。为了促进框架的部署,其对低成本颗粒传感器的准确性进行了全面的评估,来确定最可靠的传感器,最终选择Sharp-D和G3传感器,构建一个有效的参与式感知PM_{2.5}的系统。该系统可以为城市提供数据服务和决策支持,且具有高度的可扩展性。但是开放架构可能会面临数据质量和数据安全的问题。例如,当人们可以很便捷地访问数据时,就需要防止恶意用户上传虚假数据;而对于数据质量问题,要确保用户上传数据的准确性。这些都是开放系统结构所需要考虑和解决的问题。

尽管利用外置PM_{2.5}传感器与手机结合进行参与式感知是可行的,但这对于普通使用者来说很不方便。随着社交媒体的不断发展,当空气质量变得糟糕时,人们会拍照上传社交媒体并进行评论,这就提供了一种新的参与式感知空气质量的思路。学者们考虑利用手机拍照功能,对所获取的图像进行分析,来进行空气质量的监测。Poduri等^[38]通过空气能见度监测来测量场景中的大气尘埃。图像去雾霾方法主要基

于图像增强和基于物理模型这两大类方法^[39]。He等^[40]发现图像与雾霾模型中的参数之间存在关系,可以通过光学模型(见式(1))表达出来,这展现了利用图像获得PM_{2.5}的可行性。

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (1)$$

其中, I 代表所看到图片的强度, J 表示场景的亮度, A 表示全球大气光成分, t 描述光线通过媒介时没有被散射掉的那部分光线。

Budde等^[41]讨论了用移动设备测量颗粒物的方法,介绍了使用专用传感器设备的解决方案,以及将传感器改装为照相手机的新颖方法,其无需进行电气修改,手机的闪光灯和照相机分别用作光学灰尘传感器的光源和接收器。

Li等^[42]使用计算机视觉技术来分析可从社交媒体上获取的照片。在从照片上直接计算得到的污染水平与所拍摄城市的官方PM_{2.5}记录之间建立相关性。实验表明,这种基于图像的方法可用于监测空气污染。

Liu等^[43]首次提出利用智能手机拍照来实现参与式监测PM_{2.5}的方法,他们提出一种基于学习方法(LB),从拍照获取的图像中提取空间差异、暗通道、HSI色彩差异等3个主要图像特征,建立雾霾模型,来低成本、细粒度地感知城市PM_{2.5}。人们用手机拍摄天空的照片,并把它们上传到中央数据库,根据图像、位置和手机传感器数据(例如方位等)计算出空气污染。但是这种方法在极恶劣的天气环境下存在一定的限制。

与自上而下的部署模式不同,参与式传感系统最为关键的问题是数据质量,因为没有机制来选择接受或拒绝参与者的贡献,所以存在以下数据质量问题:1)参与者所使用的设备可能是准确度较低的低成本PM_{2.5}传感器;2)设备可能部署在不合适的位置(例如室内或接近污染源);3)设备在运行一段时间后可能会出现故障,而对每一位参与者的设备进行定期的检查和校正是有难度的。

由此,可以设计一种机制,根据其历史数据和与附近设备的相关性来评估每个设备的数据质量。此外,还可在数据平台上应用校准算法,无需对现场设备进行任何修改。

参与式动态感知方法将一个大问题划分为若干个小问题,利用大众的力量解决每一个小问题,属于自下而上的方式,通过这种方式可以细粒度、实时感知整个城市的空气质量,满足人们关心的当前所在位置的空气质量问题。参与式动态感知方法中最为关键的是数据质量,只要确保用户上传数据的准确性,便可达到大范围、细粒度的感知。

结束语 本文对当前的空气质量监测方法按传感器位置是否发生移动,划分为静态感知方法和动态感知方法两大类。在静态感知中,传统的基于空气质量监测站和卫星遥感的方式属于高成本静态感知方法,其监测站成本过高,但数据准确,可利用其获得的有限数据,加上所监测到的其他城市信息,如天气数据、道路网络、车流量等,将多源异构数据进行融合,利用大数据和机器学习算法来感知城市空气质量,具有很好的前景。在低成本静态感知方法中,低成本传感器经过校正后可大规模部署,弥补了空气监测站在空间上的稀疏问题,可以更好地感知城市范围的空气质量。但是,若想细粒度地

感知城市空气质量,仍需部署成百上千的传感器。因此,整个系统的总成本仍十分昂贵,并且大规模的传感网络很难管理。

在动态感知中,根据是否需要人的参与来上传数据将方法划分为非参与式动态感知方法和参与式动态感知方法。在非参与式动态感知方法中,大多是通过交通工具来进行感知的,传感器位置会发生移动,因此只需要合理设计移动轨迹,再部署适量的传感器便可达到大范围感知的效果,既降低了成本又扩大了感知的范围。在参与式动态感知方法中,多是将传感器与手机相结合,或利用手机拍照功能,由用户上传数据,属于由下至上的方式。

随着各类方法的不断发展和进步,感知城市空气质量的成本在不断下降,精度不断提高。感知城市范围的空气质量,可将静态感知与动态感知方法进行结合,各取其优势,建立空气质量感知网络。随着大数据时代的到来,对数据的关注度增加,城市数据也日益受到重视,如天气数据、交通数据、道路网、轨迹数据等,为城市空气质量的数据分析和数据挖掘提供了数据支持,对分析空气污染源也有着很大的作用。近年兴起并快速发展的一些技术也对进一步改进空气质量感知方法提供了新的方向,如深度学习、增强学习等,值得重点关注。

参 考 文 献

- [1] BOLDO E, MEDINA S, TERTRE A L, et al. Apehis: Health impact assessment of long-term exposure to PM_{2.5} in 23 European cities[J]. *European Journal of Epidemiology*, 2006, 21(6): 449-458.
- [2] ZHENGY, CAPRA L, WOLFSON O, et al. Urban Computing: Concepts, Methodologies, and Applications[J]. *ACM Trans. Intelligent Systems and Technology*, 2014, 5(3): 38-55.
- [3] ZHENG Y, LIU F, HSIEH H-P. U-Air: When urban air quality inference meets big data [C]// *Proceedings of the 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. New York: ACM, 2013: 1436-1444.
- [4] MAEKIEWICZ M. A review of mathematical models for the atmospheric dispersion of heavy gases. Part I. A classification of models [J]. *Ecological Chemistry and Engineering S*, 2012, 19(3): 297-314.
- [5] WANG J F, HU M G, XU C D, et al. Estimation of citywide air pollution in Beijing [J]. *PLOS ONE*, 2013, 8(1): e53400.
- [6] VARDOULAKIS S, FISHER B E. A., PERICLEOUS K A, et al. Modelling air quality in street canyons: A review [J]. *Atmospheric Environment*, 2003, 37(2): 155-182.
- [7] ZHU J Y, ZHANG C, ZHANG H C, et al. pg-Causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data[J]. *IEEE Transactions on Big Data*, 2017, 6(6): 1.
- [8] HSIEH H-P, LIU S D, ZHENG Y. Inferring air quality for station location recommendation based on urban big data [C]// *Proceedings of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining*. New York: ACM, 2015: 437-446.
- [9] ZHENG Y, YI X W, LI M, et al. Forecasting fine-grained air quality based on bigdata [C]// *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2015: 2267-2276.
- [10] YI X, ZHANG J, WANG Z, et al. Deep distributed fusion network for air quality prediction [C]// *24th ACM SIGKDD International Conference*. 2018.
- [11] LU J, CAO X. PM_{2.5} pollution in major cities in China: Pollution status, emission sources and control measures [J]. *Fresenius Environ. Bull*, 2015, 24: 1338-1349.
- [12] ENGEL-C J A, HOLLOMAN C H, COUTANT B W, et al. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality [J]. *Atmospheric Environment*, 2004, 38(16): 2495-2509.
- [13] WANG J, CHRISTOPHER S A. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies [J]. *Geophysical Research Letters*, 2003, 30(21).
- [14] LIU Y, FRANKLIN M, KAHN R, et al. Using aerosol optical thickness to predict ground-level PM_{2.5} concentrations in the St. Louis area: A comparison between MISR and MODIS [J]. *Remote Sensing of Environment*, 2007, 107(1-2): 33-44.
- [15] VAN D A, MARTIN R V, PARK R J. Estimating ground-level PM_{2.5} using aerosol optical depth determined from satellite remote sensing [J]. *Journal of Geophysical Research: Atmospheres*, 2006, 111(D21): 201-210.
- [16] 郑卓云, 陈良富, 郑君瑜, 等. 高分辨率气溶胶光学厚度在珠三角及香港地区区域颗粒物监测中的应用研究 [J]. *中国环境科学*, 2011, 31(6): 1154-1161.
- [17] WANG Z F, CHEN L F, TAN J H, et al. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical- and RH correcting method [J]. *Remote Sensing of Environment*, 2010, 114(1): 50-63.
- [18] WANG Y, LI J Y, HE J, et al. Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement [J]. *Aerosol Science and Technology*, 2015, 49(11): 1063-1077.
- [19] GAO M, CAO J, SETO E. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM_{2.5} in Xi'an, China [J]. *Environmental Pollution*, 2015, 199: 56-65.
- [20] KHADEM M I, SGARCIU V. Dust monitoring systems [C]// *The Sixth International Conference on Systems and Networks Communications (ICSNC 2011)*. 2011: 68-71.
- [21] LIU Y, MAO X, HE Y, et al. CitySee: not only a wireless sensor network [J]. *IEEE Network*, 2013, 27(5): 42-47.
- [22] CHENG Y, LI X, LI Z, et al. AirCloud: a cloud-based air-quality monitoring system for everyone [C]// *Proceedings of the 12th ACM Conf on Embedded Network Sensor Systems*. New York: ACM, 2014: 251-265.
- [23] TUDOSE D S, PATRASCU T A, VOINESCU A, et al. Mobile sensors in air pollution measurement [C]// *Proceedings of the 8th Positioning Navigation and Communication*. Dresden: IEEE, 2011: 166-170.
- [24] DEVARAKONDA S, SEVUSU P, LIU H, et al. Real-time air quality monitoring through mobile sensing in metropolitan areas [C]// *Proceedings of the 2nd ACM SIGKDD Int workshop on urban computing*. New York: ACM, 2013: 15.

(下转第 51 页)

- [55] AZADEH A, NEGAHBAN A, MOGHADDAM M. A hybrid computer simulation-artificial neural network algorithm for optimisation of dispatching rule selection in stochastic job shop scheduling problems [J]. *International Journal of Production Research*, 2012, 50(2): 551-566.
- [56] 徐迭石. 大数据环境下车间生产异常的智能发现与响应方法研究[D]. 哈尔滨: 哈尔滨理工大学, 2017.
- [57] 张存吉. 智慧制造环境下感知数据驱动的加工作业主动调度方法研究[D]. 广州: 华南理工大学, 2016.
- [58] 海尔集团. 海尔 COSMOplat-工业大数据平台[EB/OL]. (2017-4-28) [2018-4-23]. <https://cosmoplat.com/solution/micro/?solutionId=132>.
- [59] ESMALIFALAK M. A data mining approach for fault diagnosis: An application of anomaly detection algorithm[J]. *Measurement*, 2014, 55(3): 343-352.
- [60] RASHID M M, AMAR M, GONDAL I, et al. A data mining approach for machine fault diagnosis based on associated frequency patterns[J]. *Applied Intelligence*, 2016, 45(3): 1-14.
- [61] KIM J, LEE D, KO D, et al. An autonomic computing based on big data platform for high-reliable smart factory[J]. *Information*, 2017, 20(6): 3947-3956.
- [62] 黄天恩, 孙宏斌, 郭庆来, 等. 基于电网运行仿真大数据的知识管理和超前安全预警[J]. *电网技术*, 2015, 39(11): 3080-3087.
- [63] 钟福磊. 工业大数据环境下的混合故障诊断模型研究[D]. 西安: 西安电子科技大学, 2015.
- [64] 胡亮, 刘洋. 工业大数据在航天制造领域的集成应用研究[J]. *军民两用技术与产品*, 2015(23): 48-51.
- [65] 刘苑红, 张伟, 崔艳妍, 等. 大数据在用户供电可靠性预测评估中的应用[J]. *电力信息与通信技术*, 2016, 14(3): 55-59.
- [66] 武方方. 基于大数据的物流配送中心选址优化研究[D]. 合肥: 合肥工业大学, 2015.
- [67] 王杨珠. 基于物流信息平台的数据挖掘系统的设计与实现[D]. 北京: 北京邮电大学, 2017.
- [68] SAP. What's new in SAP S/4 HANA [EB/OL]. (2017-9-15) [2018-4-23]. https://help.sap.com/doc/b870b6ebcd2e4b5890f16f4b06827064/1709%20000/en-US/WN_OP1709_EN.pdf.
- (上接第 40 页)
- [25] HEDGECK W, VOLGYESI P, LEDECZI A, et al. Dissemination and presentation of high resolution air pollution data from mobile sensor nodes[C]// *Proceedings of the 48th Annual Southeast Regional Conf*. New York: ACM, 2010: 6.
- [26] HASENFRATZ D, SAUKH O, WALSER C, et al. Pushing the spatio-temporal resolution limit of urban air pollution maps[C]// *Proceedings of IEEE PerCom'14*. Budapest: IEEE, 2014: 69-77.
- [27] FIERZ M, HOULE C, STEIGMEIER P, et al. Design, calibration, and field performance of a miniature diffusion size classifier [J]. *Aerosol Science and Technology*, 2011, 45(1): 1-10.
- [28] DONG W, GUAN G Y, CHEN Y, et al. Mosaic: towards city scale sensing with mobile sensor networks[C]// *Proceedings of IEEE ICPADS'21*. Melbourne: IEEE, 2015: 29-36.
- [29] GAO Y, DONG W, GUO K, et al. Mosaic: a low-cost mobile sensing system for urban air quality monitoring [C]// *Proceedings of IEEE INFOCOM'35*. San Francisco: IEEE, 2016: 1-9.
- [30] MUSTHAG M, GANESAN D. Labor dynamics in a mobile micro-task market[C]// *Proceedings of the SIGCHI Conf on Human Factors in Computing Systems*. New York: ACM, 2013: 641-650.
- [31] CHEN X, WU X, LI X Y, et al. Privacy-preserving high-quality map generation with participatory sensing[C]// *Proceedings of IEEE INFOCOM'14*. Toronto: IEEE, 2014: 2310-2318.
- [32] 于瑞云, 王鹏飞, 白志宏, 等. 参与式感知: 以人为中心的智能感知与计算[J]. *计算机研究与发展*, 2017, 54(3): 457-473.
- [33] BURKE J, ESTRIN D, HANSEN M, et al. Participatory sensing [C]// *Proceedings of ACM SenSys Workshop on World-Sensor-Web*, 2006.
- [34] ZHUANG Y, LIN F, YOO E H, et al. AirSense: A portable context-sensing device for personal air quality monitoring [C]// *Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare*. New York: ACM, 2015: 17-22.
- [35] BUDDE M, BUSSE M, BEIGL M. Investigating the use of commodity dust Sensors for the embedded measurement of particulate matter [C]// *Proceedings of IEEE INSS'9*. Antwerp: IEEE, 2012: 1-4.
- [36] BUDDE M, MASRI R E, RIEDEL T, et al. Enabling low-cost particulate matter measurement for participatory sensing scenarios [C]// *Proceedings of the 12th IntConf on MUM'13*. New York: ACM, 2013: 19.
- [37] CHEN L J, HO Y H, LEE H C, et al. An open framework for participatory PM2.5 monitoring in smart cities [J]. *IEEE Access*, 2017, 5: 14441-14454.
- [38] PODURI S, NIMKAR A, SUKHATME G S. Visibility monitoring using mobile phones[J]. *Annual Report: Center for Embedded Networked Sensing*, 2010: 125-127.
- [39] 苗启广, 李宇楠. 图像去雾霾算法的研究现状与展望[J]. *计算机科学*, 2017, 44(11): 1-8.
- [40] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[J]. *IEEE Computer Society*, 2011, 33(12): 2341-2353.
- [41] BUDDE M, BARBERA P, MASRI R E, et al. Retrofitting smart phones to be used as particulate matter dosimeters [C]// *Proceedings of the 2013 IntSymp on Wearable Computers*. New York: ACM, 2013: 139-140.
- [42] LI Y, HUANG J, LUO J. Using user generated online photos to estimate and monitor air pollution in major cities[C]// *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*. ACM, 2015: 79.
- [43] LIU X Y, SONG Z, NGAI E, et al. PM2.5 monitoring using images from smartphones in participatory sensing[C]// *Proceedings of IEEE INFOCOM*. Hong Kong: IEEE, 2015: 630-635.