

# 城市计算概述

郑宇<sup>1,2,3</sup>

1 微软亚洲研究院,北京,100080

2 上海交通大学致远学院,上海,200240

3 西南交通大学信息科学与技术学院,四川 成都,610031

**摘要:**城市计算是计算机科学中以城市为背景,与城市规划、交通、能源、环境、社会学和经济等学科融合的新兴领域。城市计算通过不断获取、整合和分析城市中多种异构大数据来解决城市所面临的挑战。本文介绍了城市计算的定义、框架和主要研究问题,概述了城市计算的典型应用和所需要的技术。

**关键词:**城市计算;大数据;时空数据分析;异构数据融合;城市感知

中图法分类号:P208

文献标志码:A

城市化进程赋予我们现代化的生活,但同时也带来了很多问题和挑战,如空气污染、交通拥堵、能耗增加和规划落后等。由于城市环境极其复杂,要解决这些问题在很多年前看来几乎不太可能。近年来,随着感知技术和计算环境的成熟,各种大数据在城市里悄然而生,如交通流、气象数据、道路网、兴趣点、移动轨迹和社交媒体等。如果使用得当,这些大数据不仅可以及时反映出城市中存在的问题,也可以用来解决城市所面临的挑战,城市计算就是要用城市中的大数据来解决城市本身所面临的挑战,通过对多种异构数据的整合、分析和挖掘,来提取知识和智能,并用这些智能来创造“人-环境-城市”三赢的结果。

## 1 城市计算的概念和框架

### 1.1 城市计算的定义

城市计算是一个交叉学科,是计算机科学中以城市为背景,与城市规划、交通、能源、环境、社会学和经济等学科融合的新兴领域。更具体地说,城市计算是一个通过不断获取、整合和分析城市中多种异构大数据来解决城市所面临的挑战(如环境恶化、交通拥堵、能耗增加、规划落后等)的过程。城市计算将无处不在的感知技术、高效

的数据管理和分析算法,以及新颖的可视化技术结合,致力于提高人们的生活品质、保护环境和促进城市运转效率。城市计算帮助我们理解各种城市现象的本质,甚至预测城市的未来<sup>[1]</sup>。

### 1.2 城市计算的框架

图1给出了城市计算的基本框架,包括城市感知及数据捕获、城市数据管理、城市数据分析和提供服务4个环节。与自然语言分析和图像处理等“单数据、单任务”系统相比,城市计算是一个“多数据、多任务”的系统。城市计算中的任务涵盖改进城市规划、缓解交通拥堵、保护自然环境、减少能源消耗等,而在一个任务中又需要同时用到多种数据。比如,在城市规划的设计过程中,需要同时参考道路结构、兴趣点分布、交通流等多种数据源。

### 1.3 城市计算的核心问题

城市计算是一个新兴的交叉领域,涵盖面较广。从计算机科学的角度来看,其核心问题主要涵盖以下4个方面。

1)城市感知(计算)。如何利用城市中现有的资源(如手机、传感器、车辆和人等),在不干扰人们生活的前提下,不断地自动感知城市的韵律,是一个重要的研究课题。如何从大量的传感器和设备中高效而可靠地收集、传送数据将挑战现有的

收稿日期:2014-09-24

第一作者:郑宇,博士,微软亚洲研究院主管研究员,主要研究方向为城市大数据、时空数据挖掘和普适计算等。他被MIT科技评论评为全球杰出青年创新者(MIT TR35),作为现代创新者代表登上了美国的《时代》周刊。2014年,由于其主导的城市计算具有巨大的商业前景和改变行业格局的潜力,他被《财富》评选为中国40位40岁以下商界精英。E-mail:yuzheng@microsoft.com

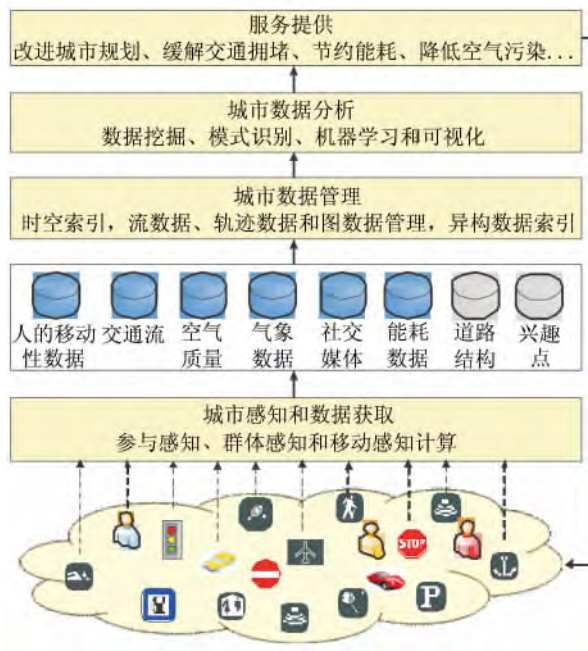


图1 城市计算的基本框架

Fig. 1 Framework of Urban Computing

传感器网络技术。此外,人作为传感器参与到城市感知过程是一个新概念(通常也称为群体感知)。例如,当一起灾难发生后,有些用户会在社交网络上发布消息或上传照片,这些用户其实就在帮助人们感知发生在他们身边的事物;用户出入地铁站时的刷卡行为,也间接帮助人们感知了地铁系统的拥挤程度和人们的出行规律;人们的手机信号,可帮助感知城市中发生的异常和紧急事件。人赋予了传统传感器强大的感知能力和前所未有的灵活性,但此类群体感知技术也面临着以下挑战:①其产生的数据更为自由无序、间接隐晦。比如,用户的开车轨迹数据并不能直接像专业传感器那样告诉我们每条道路上的交通流量、汽车油耗和尾气排放;用户的手机信号也不能直接表达一个地区的交通拥堵情况。②数据的产生时间也变得难以预测和不可控,数据的空间分布也会随着人群的移动而变得很不均匀。这必然会导致数据的缺失和稀疏,即在某些时候、某些地方将没有数据可用。③不能期望所有的用户都参与到感知过程中来贡献数据,因此,能拿到的数据仅仅只是数据全集中一个样本。这个样本可能会跟数据全集存在偏差。比如,出租车的轨迹数据跟整个城市中所有车辆的行驶轨迹就会有偏差。在某条道路上观测到很多出租车,并不一定代表这条路上就会有很多其他车辆。因此,不能直接从出租车的数量来估计其他车辆的数量。以上这三类问题给数据的收集和解析带来新的挑战。城

市感知不再只是感知的过程,也包含能够从复杂、隐晦、缺失和非均匀分布的数据中获取有效知识的计算环节。

2)海量异构数据的管理。城市产生的数据属性差别很大。例如,气象是时序数据,兴趣点是空间点数据,道路是空间图数据,人的移动是轨迹数据(时间+空间),交通流量是流数据,社交网上用户发布的信息是文本或图像数据。如何管理和整合大规模的异构数据将是一个新的挑战。尤其是在一个应用中使用多种数据时,只有提前建立起不同数据之间的关联,才能让后面的分析和挖掘过程变得高效、可行。

3)异构数据的协同计算。①如何从不同的数据源中获取相互增强的知识是一个新的课题。传统的机器学习问题往往基于单一数据,如自然语言处理主要分析文本数据,图像视觉主要基于图像数据。同等对待不同性质的数据,在城市计算的很多应用中效果并不理想。②在保证知识提取深度的同时,如何提高对大数据的分析效率,从而满足城市计算中众多实时性要求较高的应用(如空气质量预测、异常事件监测等),也是一个难题。③数据维度的增加也容易导致数据稀疏性问题,如何应对大数据的数据稀疏性问题,也很重要。当数据规模达到一定程度时,简单的矩阵分解算法都变得很难执行。需要强调的是,大数据和数据稀疏性并不矛盾(后文有更为具体的描述)。

4)虚实结合的混合式系统。城市计算常常催生混合系统,比如云加端模式,即信息产生在物理世界,通过终端设备被收集到云端(虚拟世界)分析和处理,最后云再将提取的知识作为服务提供给物理世界的终端用户。数据在物理和虚拟世界中来回穿行,从分散到集中,再到分散。设计的系统既需要应对上百万的用户,也需要跟成千上万的传感器通信。对不同数据源的同步、存储和更新对系统的设计和搭建提出了更大的挑战。后面要提到的基于浮动车数据的快速行车路线设计<sup>[6]</sup>以及城市异常事件的监测<sup>[48]</sup>都是典型的混合式系统。

#### 1.4 城市计算的应用分类

图2给出了城市计算主要涉及(但不限于)的7类应用:城市规划、智能交通、城市环境、城市能耗、城市经济、社交和娱乐、城市安全。每个类别可再细分为子类,如智能交通包括对自驾车、公交和出租车系统的改进。虽然这7大应用早已存在,城市计算将用大数据的方法来重新演绎这些领域。

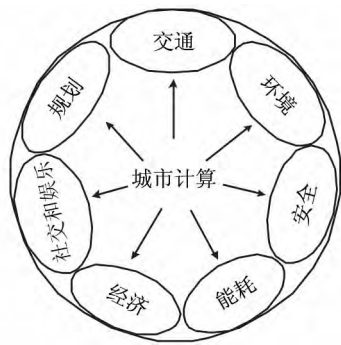


Fig. 2 Categories of Applications in Urban Computing

## 2 城市计算的典型应用

### 2.1 城市规划

城市拥堵的凸显一定程度上是由于现有道路网的设计已经不能满足不断发展的城市交通流的需求。如图 3(a)所示,利用高速和环路等主干道将城市分割成区域,然后分析大规模车流轨迹数据在不同区域之间行驶的一些特征,便可找到连通性较差的区域,从而发掘现有城市道路网的不足之处<sup>[2]</sup>。图 3(b)图给出了根据北京市 3 万多辆出租车 3 个月轨迹数据的一个分析结果。这些结果可以作为制定下一版规划的参考建议。同时,通过对比连续两年的检测结果,可以验证一些已经实施的规划(如新建道路和地铁)是否真的有效。

城市的不断发展催生了不同的功能区域,如文教、商业和住宅区等。准确掌握这些区域的分布对制定合理的城市规划有着极其重要的意义。但是一个区域的功能并不是单一的,如在科学文教区里仍然有饭店和商业设施的存在。因此,一个区域需要由一个功能的分布来表达(如 70% 的功能为商业,20% 的功能为住宅,剩余的为教育)。由于一个区域杂糅了很多不同类别的兴趣点,而且每个兴趣点的作用大小和被访问频度都很难预知,同样都是餐馆,一个小区里的小店和全聚德之类的大饭店所反映的区域功能是完全不一样的,这也给城市规划人员提出了很大的挑战。

结合兴趣点数据和人们的移动模式,文献[3]分析了城市中不同的功能区域。如图 4(a)所示,相同颜色的区域其具有相同功能分布(如红色区域主要为科学文教区)。此处人的移动性数据是从出租车的轨迹数据中提取的(该轨迹数据包含了乘客上车和下车地点的信息)。人的移动性数据很好地区分了相同类别的兴趣点的热度,也



图 3 搜寻城市道路网中不合理的规划

Fig. 3 Glean Problematic Design from Urban Road Networks

揭示了一个区域的功能。如一个区域,大部分人都是早 8 点左右离开、晚上 7 点返回,则这个区域很可能是住宅区。另一方面,一个区域的主要功能是文教,但也不代表该区域的任何一个地点都服务于文教。因此,给定一种功能,我们希望知道它的核心区域所在。图 4(b)显示了成熟商业区的核心区域,颜色越深该区域是成熟商业区的概率越大。

### 2.2 智能交通

估计路网上一条路线的通行时间是很多交通相关应用中常常面临的问题。由于不是所有的道路上都装有线圈传感器,利用浮动车(如出租车)的 GPS 轨迹数据来计算道路的通行时间正受到越来越多的关注。此类方法简单、灵活(如可以根据车辆 GPS 轨迹中两点间的距离和时间来计算速度),但也面临以下挑战:①数据稀疏。不是所有车辆都会贡献 GPS 数据,在一个时间段,只有很少的道路上有贡献数据的车辆(如出租车)通过。如何估计没有 GPS 数据覆盖的道路的通行时间是一大难点。②数据的拼接。对于那些有 GPS 轨迹数据覆盖的路段,如何组合来自不同车辆的轨迹数据也是一个问题。

如图 5 所示,我们希望估计路线 $r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_4$ 的通行时间。首先, $r_4$ 上没有任何数据(数据稀疏);其次,对 $r_1 \rightarrow r_2 \rightarrow r_3$ 的通行时间存在多种选择。如可以先利用轨迹 $T_{r_1}$ 和 $T_{r_2}$ 估计 $r_1 \rightarrow r_2$ 的时间,再用轨迹 $T_{r_2}$ , $T_{r_3}$ 和 $T_{r_4}$ 估计通过 $r_3$ 的时间,最后再求和;也可以用相应轨迹分别估计 $r_1$ 、 $r_2$ 和 $r_3$ 的时间,再求和。由于通过一条线路的轨迹数量跟该线路的长度成反比,而通行时间估计的精度与可用轨迹数据量成正比,需要找到一种最佳的组合来平衡可用轨迹的数据量和被估计线路的长度,从而减少时间估计的误差。

文献[4]根据历史和实时的出租车轨迹数据,并结合兴趣点和路网结构等地理特性,利用矩阵的联合分解和协同过滤机制来实时计算任意单个

路段的通行时间,填补缺失的数据。然后,再利用动态规划算法来找到不同轨迹的最佳组合方式,得到最优时间估计。

T-Drive 系统<sup>[5]</sup>利用装有 GPS 传感器的出租车来感知交通流量,并为普通用户设计真正意义上的最快驾车线路。如图 6 所示,T-Drive 提出了一个基于地标图的路由算法。其中,一个红点表示被出租车司机走过最频繁的路段之一(称为一个地标);红点之间的连线表示连接两个地标点的一条虚拟边,代表连续经过这两个地标点的出租车轨迹的聚合。根据出租车轨迹数据,可以学习出任何一条虚拟边的通行时间。T-Drive 的改进版<sup>[6-7]</sup>进一步考虑了天气以及个人驾车习惯、技能和道路熟悉程度等因素,提出了针对个人的个性化最快线路设计。这个系统不仅可以为每 30 min 驾车路程节约 5 min 时间,也可以通过让不同用户选择不同的道路来缓解可能出现的拥堵。

打车难是很多大城市都面临的一个问题。通过分析出租车乘客的上下车记录,T-Finder<sup>[8-9]</sup>提供了一个司机和乘客的双向推荐服务。一方面,这个系统向出租司机建议一些趴活地点。只要朝着这些地点开,司机将在最短的时间内(在路上或者推荐地点)拉到乘客,并且最大化收入。另一方面,如图 7(a)所示,该系统向乘客推荐一些周边的路段,在这些路段上有更高的概率寻找空车(不同颜色表示不同概率,蓝色最高,红色最低)。同时,T-Finder 还会预测周边的一些出租车停靠站在未来半小时内将驶入的空车数目。T-Finder 通过推荐能够缓解非高峰时段的打车难问题,但对于高峰时段,该系统并不能真正解决问题。

T-Share<sup>[10-11]</sup>则通过出租车实时动态拼车的方案来解决这一难题。在 T-Share 系统中,用户通过手机提交打车请求,表明上下车地点、乘客人数和期望到达目的地的时间。后台系统实时维护着所有出租车的状态,在接收到一个用户请求后,搜索出满足新用户条件和车上已有乘客条件的最优车。最优是指出租车去接一个新的用户所增加的里程最小。如图 7(b)所示,该出租车被规划为接 $u_1$ 、接 $u_2$ 、放下 $u_1$ 、接 $u_3$ ,再放 $u_2$ 、放 $u_3$ (+表示上车,-表示下车)。根据仿真结果,T-Share 系统一年可以为北京市节约汽油 8 亿升(可供 1 百万辆车开 10 个月,价值 10 亿人民币,减排  $\text{CO}_2$  16 亿 kg),乘客能打到车的概率提高 3 倍、费用降低 7%,出租车司机的收入增加 10%。

还有一些工作,利用乘客在地铁系统中的刷

卡数据来估计单个地铁站点内的拥挤程度和不同站点间的通行时间,从而优化人们的出行线路、时间和购票方式的选择<sup>[12-13]</sup>。还有人通过分析出租车的轨迹数据来建议公交线路<sup>[14]</sup>,如果有大量的人通过打出租车从一个地点到另一个地点,则说明这两个地点需要公交线路来联通。

## 2.3 环境

### 2.3.1 城市空气

空气质量信息对控制污染和保护人们身体健康有着重要意义。很多城市都开始建地面空气监测站来实时感知地面的空气质量。但是由于建设监测站的成本高昂,一个城市通常只有有限个站点,并不能完全覆盖整个城市的范围。如图 8(a)所示,北京城区仅有 22 个站点(大概 100 多  $\text{km}^2$  一个站点)。不幸的是空气质量受多方面因素的影响(如地表植被、交通流量、楼房密度等),而且随地域不均匀地变化。如果一个区域没有监测站,我们不知道该地区的空气质量如何,更不能用一个笼统的数据来概括整个城市的空气状况。

利用群体感知的方法是解决这个问题的一种方式。比如,“哥本哈根车轮”项目在自行车车轮里安装一些传感器,并通过用户手机将收集的数据发送至后台服务器。这样依靠群体的力量,就可以感知整个城市不同角落的温度、湿度和  $\text{CO}_2$  浓度。由于受传感器大小和感知时间的限制,这种方式只能适用于部分气体,如  $\text{CO}$  和  $\text{CO}_2$ 。像  $\text{PM}_{2.5}$  这样的悬浮物,传感器体积较大,不便于携带,而且要求 2~4 h 的测量时间才能产生较为精确的读数。

U-Air<sup>[15-16]</sup>利用有限个地面监测站的实时和历史空气质量数据,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,利用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。图 8(b)显示了北京某时刻的细粒度空气质量(其中,不同颜色表示不同污染指数,绿色为优)。

### 2.3.2 城市噪音

城市的发展带来了许多噪音源,如汽车鸣笛、酒吧和建筑施工等。这些噪音不仅会影响人的睡眠质量、降低工作效率,还会对人体的精神和健康产生危害。应对城市噪音污染首先要了解整个城市的噪音状况和噪音构成,但对整个城市的噪音建模却非常困难。首先,噪音随时间变化很快(转瞬即逝),随空间变化剧烈(百米外的噪音情况可能跟当前地点的差异很大)。因此,要监控细粒度的城市噪音需要安装数百万声学传感器。由于传



传感器的安装和维护成本巨大,这种做法很不现实。此外,噪音污染的衡量不仅取决于噪音的强弱(分贝数),还取决于人们对噪音的容忍度,而后者会随着时间的变化而变化。例如,人在夜间对噪音的容忍度要远远小于白天,噪音的分贝数虽然较白天低,但仍有可能被人们认为是严重的噪音污染。再者,噪音通常是多种声音源的混合体。单纯的传感器数据只能反应声音的强弱,却不能告诉我们噪音的构成。如一个地方在中午的时候,40%的噪音来自于交通流,20%来自于酒吧,20%来自于建筑工地。因此,即便能安装无数的传感器,也不能真正解决噪音污染的问题。

在大数据时代,其他数据可以间接反映城市的噪音污染。比如,从 2010 年开始,美国政府开设了 311 服务,让市民可以通过打电话或者用手机应用来抱怨这个城市中让他们感到不满意的地方。每一个抱怨都关联时间、地点和类别信息。根据过去 4 a 的数据统计,噪音是被抱怨次数第三的类别。图 9 显示了纽约市在 2012~2013 年 311 数据的分布。其中,不同的颜色表示不同的噪音类别,柱状体的高度表示抱怨的次数。

这个数据其实是人作为传感器以及群体感知的结果。具体来说,一个地方人们对噪音的抱怨次数间接反映了这个地方噪音污染的程度。311 抱怨数据在不同噪音类别上的分布反映了这个地方噪音的构成。但是这个数据很稀疏,因为并不能确保在任何一个地方和时间段都有人提供抱怨数据。为解决数据的稀疏性问题,CityNoise<sup>[17-19]</sup>结合路网数据、兴趣点数据和社交媒体中的签到数据来协同分析各个区域在不同时间段和噪音类别上的污染指数。有了这样细粒度的噪音指数,可以对不同区域在不同时间按照不同的指标来排名。

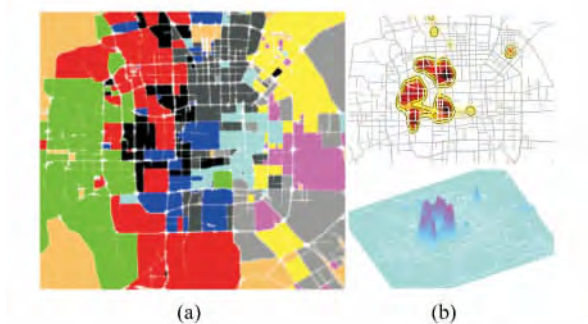


图 4 发现城市的功能区域

Fig. 4 Identify Functional Regions of a City

如图 10(a)显示了工作日夜间 0~5 点(或周末晚上 7~11 点),整个城市的总体噪音情况(颜

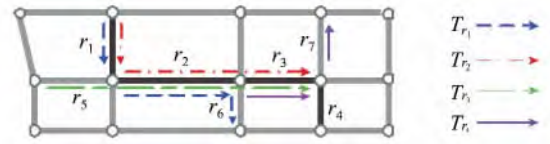


图 5 基于 GPS 轨迹的路线通行时间估计

Fig. 5 Estimate Travel Time of a Path Based on GPS Trajectories

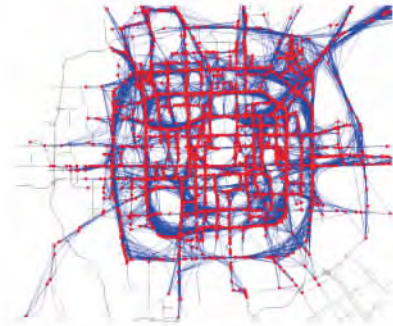


图 6 基于地标图的 T-Drive 系统

Fig. 6 T-Drive: A Landmark-Graph-Based System

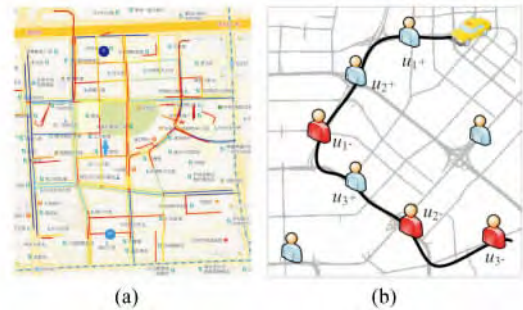


图 7 城市计算中的出租车解决方案

Fig. 7 Solutions for Improving Urban Taxi Services

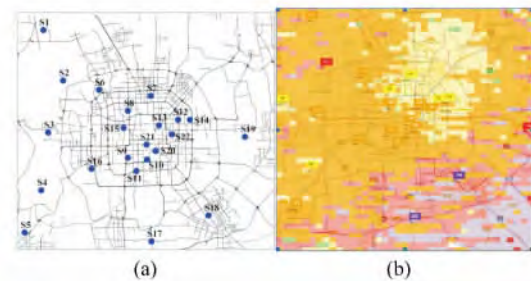


图 8 基于大数据的细粒度空气质量预测

Fig. 8 When Urban Air Quality Meets Big Data

色越深,噪音污染越重)。也可以根据某一个类别(如建筑施工)的噪音来看城市中不同地区的污染情况(见图 10(b))。这样的结果甚至可以帮助我们分析出一个地点(如图 10(c)所示的时代广场)在不同时间段中的噪音污染构成。有了这样的信息,政府才能够做出相应决策来缓解城市噪音污染。



图 9 纽约市市民对噪音的抱怨数据

Fig. 9 311 Complaint Data in New York City

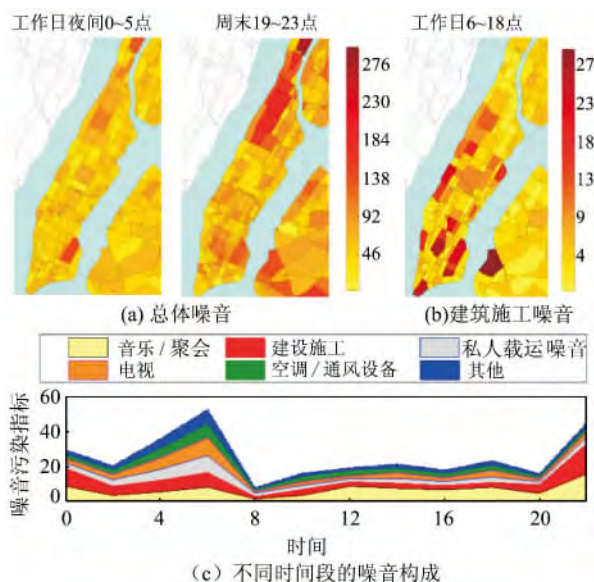


图 10 分析城市不同区域的噪音情况和噪音构成

Fig. 10 Diagnose Urban Noises

## 2.4 能源消耗相关应用

文献[20,24]利用装有 GPS 的出租车在加油站的等待时间去估计加油站的排队长度,从而估算出此时加油站内的车辆数目及加油量。将全城的加油站数据汇总,便可计算出任意时刻有多少燃油被消耗掉(加入到汽车的油箱里)。如图 11 所示,这些数据能实现三方面的应用:①给需要加油的用户提供推荐信息,寻找排队时间最短的加油站。②可让加油站运营商知道各个地区的加油需求,从而考虑增加新的站点或动态调整某些加油站的工作时间。③政府可以实时掌握整个城市的油耗,制定合理的能源战略。

如图 12 所示,文献[22]利用城市中一部分车辆(如出租车)的 GPS 轨迹信息进一步分析了整个路网上行驶汽车的实时油耗和尾气排放情况。这项研究不但可以向用户建议最低油耗路线,还可以做到细粒度的空气污染预警。对长期数据进行分析,还可以发现城市中高能耗路段(建议改进

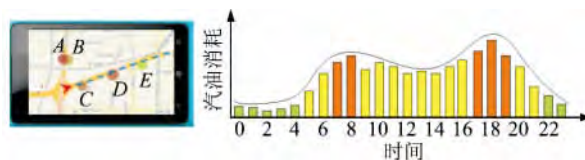


图 11 加油站排队时间及实时油耗估计

Fig. 11 Infer the Queuing Time and Gas Consumption of a Gas Station

城市规划),并帮助分析汽车尾气排放的 PM<sub>2.5</sub> 占空气中总量的比重,从而为政府决策提供参考建议,如限制交通流量是否真的能够减缓污染情况。

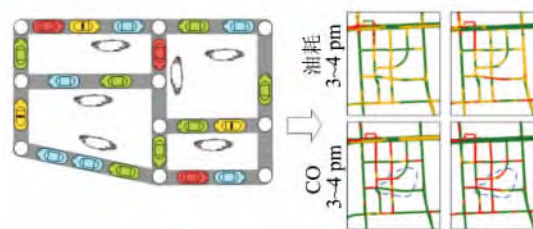


图 12 基于浮动车数据的油耗和尾气排放计算

Fig. 12 Estimate Gas Consumption and Vehicle Emissions on Roads

该研究项目先通过出租车的轨迹数据计算已有数据路段上的车辆行驶速度,再结合兴趣点和路网结构等其他数据源,将有限的速度信息利用协同过滤的机制传播到整个路网。然后,利用图模型将每条道路上的速度信息转化为车流量信息。最后,通过环境学理论,根据车速和流量计算出能耗和尾气排放情况。

文献[23]通过分析人口数据、车辆轨迹数据、各地区能源消耗情况和兴趣点的分布来研究未来新能源汽车的充电站建在何处最优。也有不少工作通过分析汽车内部的传感器数据(如踩油门、刹车的时间和次数等)来建议经济省油的开车方式[24]。

## 2.5 社交和娱乐

社交网络的盛行,尤其是基于位置的社交网络[25-26]的风靡,带来了丰富的媒体数据,如用户关系图、位置信息(签到和轨迹)、照片和视频等[27-30]。这些数据不仅表现了个人的喜好和习惯[31],也反映了整个城市里人们的生活方式和移动规律。基于这些数据,很多推荐系统被提出,包括朋友推荐[32-35]、社区推荐、地点推荐[36-39]、旅行线路推荐[40-45]和行为活动推荐[46-48]。文献[49]综述了基于位置的社交网络中的各种推荐系统。

城市计算中的社交应用更加强调从大量用户的社交媒体数据中提炼出群体智慧。人作为一个



重要的感知和计算单元参与到计算的过程中是城市的重要特点之一。例如,一个用户的签到或者带有地标的照片数据都可被看成是不确定的轨迹,因为用户不会不停地签到或拍照。给出这样一条轨迹数据时,我们无法判断出该用户选择的具体线路(如图 13(a))。但是,当把很多个用户的不确定线路叠加到一起的时候,就能猜测出最有可能的线路(如图 13(b)所示),即“不确定+不确定→确定”<sup>[43-45]</sup>。这样的应用可以帮助人们规划旅行线路。比如,一个用户想在一条线路中去后海、天坛和颐和园三个地方,他把这三个点输入到系统里,便可根据大众的签到数据计算出一条最热门的游玩路线。

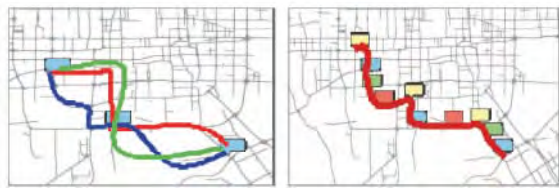


图 13 从不确定的社交媒体中推断路线

Fig. 13 Identify the Most-Likely Route from Uncertain Social Media

社交媒体数据同时也向城市计算的其他方面有所贡献。比如,通过网民在社交网络中发布的信息来预测总统选举结果、疾病蔓延和房价走势,发现异常事件和灾难,分析交通流量,以及设计广告推送和商业选址。通过社交媒体还能分析一个城市的风格以及不同城市之间的相似性。

## 2.6 经济

城市经济<sup>[50]</sup>是一个相对成熟的研究领域,比如,分析决定土地价格的因素、土地使用限制对经济的影响,公司选址和人们选择的住宅位置对未来经济将产生的影响。

文献[51]通过分析大量用户的签到数据来为商业选址提供位置建议。比如,麦当劳的选址问题。文献[52]结合道路结构、兴趣点分布、人口流动等诸多因素来对房屋的价值排序。即在市场向好时,哪个小区的房价将会涨得更多;相反,市场下行时,哪些小区比较抗跌。与使用传统经济学模型不同,以上这两个例子采用了机器学习算法和数据驱动的方法。

## 2.7 城市安全和应急响应

城市中总是会有一些突发事件,比如自然灾害(地震和洪水等)、大型赛事和商业促销、交通事故和临时管制,以及一些群体性事件。如果能及时感知,甚至预警这些事情,将能极大地帮助城市

管理,提高政府对突发事件的应对能力,保障城市安全,减少悲剧的发生。

文献[53-55]通过分析北京 3 万多辆出租车的轨迹来发现城市中的异常事件,其主要思路是当异常事件发生时,附近的交通流将出现一定程度的紊乱。文献[56]试图用具体的交通线路来进一步解释异常出现的原因。如图 14 所示, $L_1$ 联通的两个区域之间出现了交通流异常,但问题本身可能并不在这两个区域。其原因在于天安门附近因马拉松比赛而导致了交通管制,之前通过紫色虚线出行的车流就不得不绕道到绿色虚线的线路。所以绿色的线路才是产生这次异常的原因。文献[57]根据司机们路线选择方式的改变来捕捉交通异常,并进一步从相关的微博中提取关键词来解释异常的原因,如婚博会、道路坍塌。

文献[58]通过分析 160 万日本人一年的 GPS 移动轨迹数据库来对日本大地震和福岛核事故发生后的灾民移动、避难行为进行建模、预测和模拟。这样,日后再有同样事情发生时,便可从之前的灾难中吸取经验,提前做好准备,例如,为人们推荐合理的撤退线路。

## 3 城市计算的主要技术

城市计算是一个交叉学科,需要与一定的行业背景相结合,可使用的技术手段较多。从计算机学科的角度来看,城市计算的主要技术涉及 5 个方面。

### 3.1 传感器技术

1) 传感器网络技术,实现现有专业传感器(如温度传感器、位置传感器、交通流线圈、空气质量监测仪等)之间的互联互通,完成数据的快速收集。

2) 主动参与式感知技术<sup>[59]</sup>。用户主动分享自己获取的数据来共同完成一个复杂的任务。例如,每个用户都利用手机上的传感器来贡献自己周边的气温和湿度,从而构建出全城细粒度的气象信息。

3) 被动群体感知。城市里各种发达的信息基础设施(如蜂窝移动通信系统和公交卡系统)为城市计算提供了良好的感知平台。这些基础设施可能并不是专门为城市计算设立的,但当用户在使用这些基础设施时会产生大量的数据,将这些数据融合到一起就能很好地反映城市的韵律。例如,通过分析大量用户的地铁刷卡数据就能掌握城市的人口流动规律,通过分析大规模的出租车

轨迹数据就能感知城市路面的交通流。与第二种感知技术不同,被动式群体感知中的用户并不知道自己的数据将作何使用,甚至不知道自己在产生数据。

### 3.2 数据管理技术

1) 流数据管理技术。由于大量的传感器数据都以流的形式输入,高效的流数据库技术是城市计算数据管理层的基石。

2) 轨迹管理技术。交通流、人的移动,以及带位置标签的社交媒体都可以表示为轨迹数据(即带有时间戳并按时间排序的点序列)。对轨迹的处理技术<sup>[60]</sup>在城市计算中经常用到,如地图匹配算法<sup>[61-62]</sup>、轨迹压缩<sup>[63-64]</sup>、轨迹搜索<sup>[65-67]</sup>、轨迹频繁模式挖掘<sup>[68-71]</sup>等。

3) 图数据管理技术。社交网络中人的关系、不同地区之间的人口流动、道路上的交通流等都可表达为图模型。因此,图数据的管理和模式发现技术尤为重要。城市计算的应用中更多会用到带有时空属性的图模型,即每个节点都有空间坐标信息,图中边和点的属性(甚至图结构)会随时间而变化。前面提到的最快行车路线设计<sup>[5-6]</sup>、查找路网中不合理规划<sup>[2]</sup>、发现城市不同的功能区域<sup>[3]</sup>、交通流异常检测<sup>[56]</sup>都以带有时空属性的图作为研究模型。

4) 时空索引技术。有效的索引可以大大提高数据提取的效率。由于空间和时间是城市计算中最常用的两个数据维度,各种空间索引<sup>[72-73]</sup>和时空索引技术<sup>[74]</sup>都常常被使用。更重要的是利用时空索引技术将不同种类的数据(如文本、车流等)关联和组织起来,为之后的高效数据挖掘和分析做好准备。

### 3.3 数据挖掘技术

可用于城市计算的数据挖掘和机器学习算法很广泛,并没有太严格的限定。各种模式发现、统计学习和人工智能方法都可以应用到该领域。但在挑选这些技术时需要重点考虑以下两个因素。

#### 3.3.1 从异构数据中学习相互增强的知识

实现这个目标通常有三种方法。第一种方法是分别从不同的数据中提取特征,然后简单地将这些特征直接拼接并归一化到一个特征向量里,输入到机器学习的模型中。由于不区分不同数据的特性,实践证明这种方法并不是最有效的。其二,在计算模型的不同阶段先后使用不同数据。比如,文献<sup>[2]</sup>先用道路数据将城市分割成很多区域,然后再将轨迹数据映射到这些区域上构建图,最后通过分析图模型来找出不合理的道路规划。

第三种方法是将不同的数据分别输入到同一个计算模型的不同部分。如图15所示,文献<sup>[3]</sup>将人的移动性数据和兴趣点数据分别输入到一个主题模型的两个不同部分来分析城市的不同功能区域。

如图16所示,文献<sup>[15]</sup>将交通流、人的移动性和气象数据等时变信息输入到一个随机条件场去模拟一个地点空气的时序相关性,而将道路结构、兴趣点分布等空间(非时变)信息输入到神经网络中以模拟不同区域间空气质量的相关性。最后,两个模型在半监督学习的框架里相互迭代、增强,共同推断出一个地点的空气质量。如果只是简单将所有数据输入到一个分类器中,由于那些空间数据不随时间变化,将会被忽略,预测效果并不好。

#### 3.3.2 应对数据的稀疏性

大数据与数据的稀疏性并不矛盾。以预测城市的细粒度空气为例,能观测到的交通流、人流、道路和兴趣点数据都是大数据,但由于只有有限个监测站能产生空气质量的读数,因此,训练数据很稀疏(虽然特征数据很大)。另外一个例子是利用出租车来估计城市的燃油消耗<sup>[22]</sup>。总的来说,出租车的GPS轨迹数据巨大,但某些时刻有相当一部分路段上并没有出租车行驶。如何估计这些路段的速度和油耗,也是一个应对数据稀疏性的问题。应对数据的稀疏性通常可采用以下三种方法。

1) 采用矩阵(或张量)分解算法和协同过滤。如图17所示,文献<sup>[17]</sup>用张量分解结合协同过滤来应对分析城市噪音过程中所面临的数据稀疏性问题。中间的三维张量模型表达了不同时间段,任意地理区域在不同噪音类别上的用户抱怨情况。周边的三个矩阵根据其他三个数据源建立,分别表示地区之间因地理特性而造成的相似性、根据人的移动性之间的相似性(以及不同时间之间的相似性)、噪音类别之间的相关性。这种手段,其实也是一种数据融合的方法。在城市油耗估计<sup>[22-23]</sup>和路线通行时间估计<sup>[4]</sup>中也使用了类似的技术。

2) 使用半监督学习算法或转移学习算法。如文献<sup>[15]</sup>使用半监督学习算法来弥补因空气监测站少而带来的训练样本稀疏性问题。半监督学习从具有不同分布的其他数据源中获取知识,从而解决一个机器学习任务中训练样本不足的问题。比如,可否根据出租车在路网上的分布来学习出其他车辆的分布情况。



3) 基于相似性的聚类算法。假设需要根据埋在地面的线圈传感器来估计行驶在道路上的车辆数,由于并不是所有路面都埋有线圈,很多道路上的流量无法估计。根据道路的拓扑结果、周边的兴趣点分布等信息,我们可以计算不同道路之间的相似性,从而对道路进行聚类。这样,被分在同一个类里的道路很可能具有相同的车流模式。于是,在一个类中,可以将有传感器道路的读数赋给那些没有传感器的道路。

### 3.4 优化技术

城市计算中也经常用到各种优化技术,比如文献[10-11]就是将时空搜索技术和路径优化相结合来寻找能够接送乘客的最佳出租车;文献[56]通过线性规划来分析最可能造成交通异常的车流;文献[8-9]向出租车司机推荐最优的乘客寻找路线;文献[4]利用动态规划算法来优化组合不同轨迹的运算复杂度。

### 3.5 混合数据的可视化技术

可视化以直观的方式帮助我们理解获取的知识和模式。与单一数据可视化不同,城市计算中的可视化技术需要同时考虑多个维度,其中,空间和时间是两个至关重要的维度。

图 18 是(工作日下午 12~14 点之间)乘出租车到达各个区域的人数的热度图(颜色越深,人越多)。将不同时间段的此类热度图连续播放,便可以动态地反映整个城市的人口流动规律。相对而言,北京东部的中央商业区具有更高的人气。

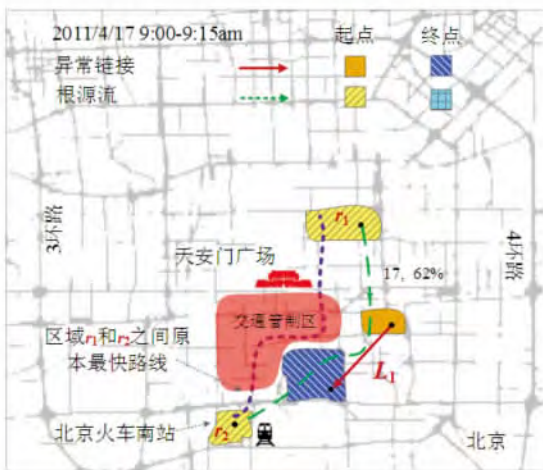


图 14 根据人的移动性来分析交通异常

Fig. 14 Detect Traffic Anomalies from Human Mobility Data

图 4 表达了城市不同功能区域的空间分布;图 8 反映了城市中细粒度的空气质量;图 9 和 10

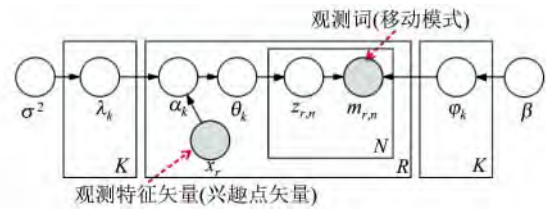


图 15 基于图模型的城市功能区域发现

Fig. 15 Identify Functional Regions Based on a Topic Model

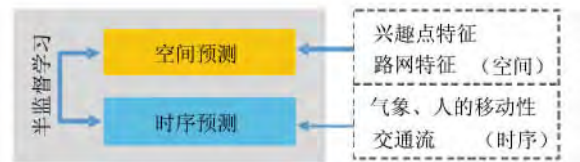


图 16 基于半监督学习的异构数据融合

Fig. 16 Fusion Heterogeneous Data Sources in a Co-training-based Framework

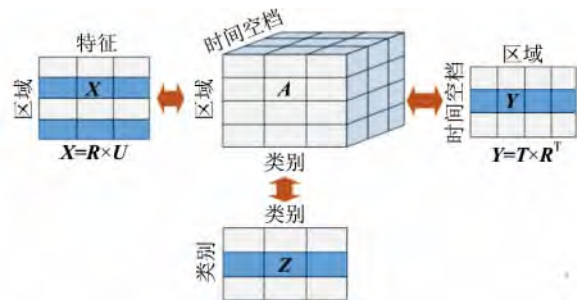


图 17 基于张量分解和协同过滤的城市噪音分析

Fig. 17 Diagnose Urban Noises Through Context-Aware Tensor Decomposition

显示了纽约市的噪音污染的时间和空间分布情况;图 12 可视化了城市道路上的油耗和污染物排放。这些可视化不仅帮助我们展现数据和结果,也有利于进一步发觉更深层次的知识,并对城市面临的问题做出相应的决策。

## 4 结 语

城市计算是一个新兴且非常重要的交叉领域,是计算机学科与传统城市规划、交通、能源、经济、环境和社会学等多个领域在城市空间的交汇。它关系到人类未来的生活质量和可持续性发展。本文提出了城市计算的概念和基本框架,概述了其研究问题、常用技术和应用分类,并介绍了一些典型的案例。大数据时代的到来为城市计算提供了更多的机遇和更广阔的前景。

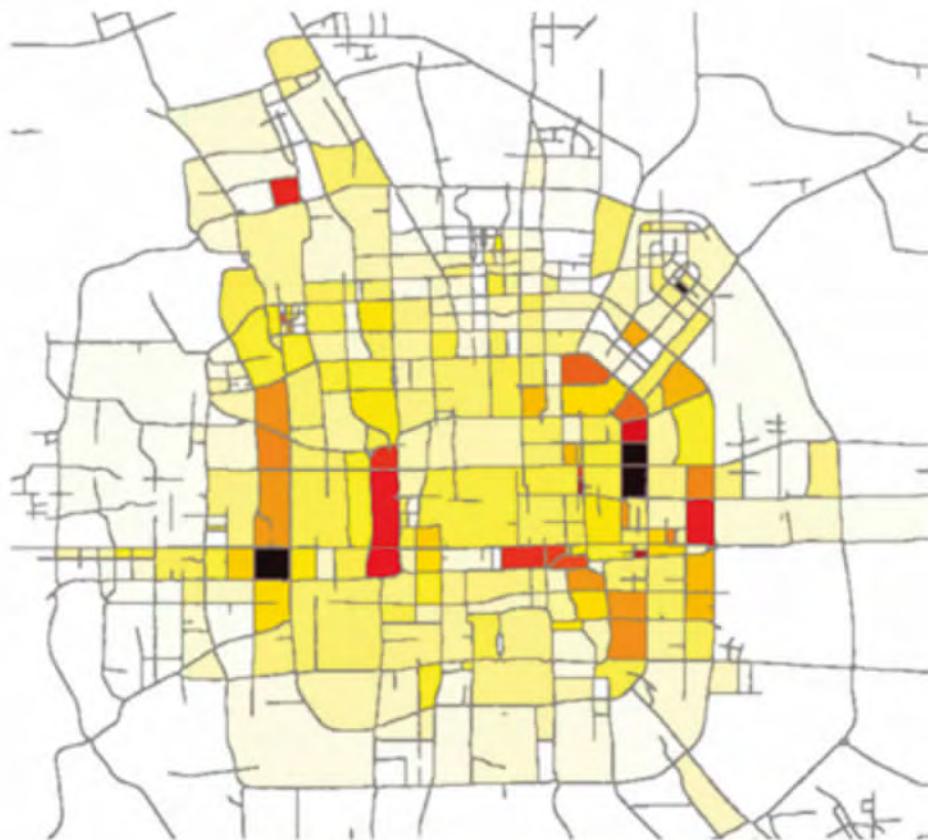


图 18 北京市区域人口到达热度图

Fig. 18 Heat Map of Beijing Urban Areas in Terms of People's Arrivals

## 参 考 文 献

- [1] Zheng Y, Capra L, Wolfson O, et al. Urban Computing: Concepts, Methodologies, and Applications [J]. *ACM Transactions on Intelligent Systems and Technology*, 2014, 3 (5), DOI: 10. 1145/1290002. 1290003
- [2] Zheng Y, Liu Y, Yuan J, et al. Urban Computing with Taxicabs[C]. *UbiComp*, Beijing, 2011
- [3] Yuan J, Zheng Y, Xie X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs[C]. *KDD*, Beijing, 2012
- [4] Wang Y, Zheng Y, Xue Y. Travel Time Estimation of a Path Using Sparse Trajectories [C]. *KDD*, New York, 2014
- [5] Yuan J, Zheng Y, Zhang C, et al. T-Drive: Driving Directions Based on Taxi Trajectories[C]. *SIGSPATIAL GIS*, San Jose, California, 2010
- [6] Yuan J, Zheng Y, Xie X, et al. Driving with Knowledge from the Physical World[C]. *KDD*, San Diego, California, 2011
- [7] 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
- [8] Yuan J, Zheng Y, Zhang L, et al. Where to Find My Next Passenger? [C]. *UbiComp*, Beijing, 2011
- [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*, 2013, 25(10): 2 390-2 403
- [10] Ma S, Zheng Y, Wolfson O. T-Share: A Large-Scale Dynamic Taxi Ridesharing Service[C]. *ICDE*, Brisbane, Australia, 2013
- [11] Ma S, Zheng Y, Wolfson O. Real-Time City-Scale Taxi Ridesharing [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, DOI: 10. 1109/TKDE. 2014. 2334313
- [12] Lathia N, Capra L. Mining Mobility Data to Minimise Travellers' Spending on Public Transport[C]. *KDD*, San Diego, California, 2011
- [13] Ceapa I, Capra C S L. Avoiding the Crowds: Understanding Tube Station Congestion Patterns from Trip Data[C]. *UrbComp*, Beijing, 2012
- [14] Bastani F, Huang Y, Xie X, et al. A Greener

- Transportation Mode: Flexible Routes Discovery from GPS Trajectory Data[C]. ACM SIGSPATIAL GIS, Beijing, 2011
- [15] Zheng Y, Liu F, Hsieh H P. U-Air: When Urban Air Quality Inference Meets Big Data[C]. KDD, Chicago, IL, USA, 2013
- [16] Zheng Y, Chen X, Jin Q, et al. A Cloud-Based Knowledge Discovery System for Monitoring Fine-Grained Air Quality[R]. MSR-TR-2014-40, Beijing, 2014
- [17] Zheng Y, Liu T, Wang Y, et al. Diagnosing New York City's Noises with Ubiquitous Data[C]. UbiComp, Seattle, WA, USA, 2014
- [18] Wang Y, Zheng Y, Liu T. A Noise Map of New York City[C]. UbiComp, Seattle, WA, USA, 2014
- [19] Liu T, Zheng Y, Liu L, et al. Methods for Sensing Urban Noises[R]. MSR-TR-2014-66, USA, 2014
- [20] Zhang F, Wilkie D, Zheng Y, et al. Sensing the Pulse of Urban Refueling Behavior[C]. UbiComp, Zurich, Switzerland, 2013
- [21] 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
- [22] Shang J, Zheng Y, Tong W, et al. Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City[C]. KDD, New York, 2014
- [23] Momtazpour M, Butler P, Hossain M S, et al. Coordinated Clustering Algorithms to Support Charging Infrastructure Design for Electric Vehicles[C]. UrbComp, Beijing, 2012
- [24] Tulusan J, Staake T, Fleisch E. Providing Eco-driving Feedback to Corporate Car Drivers: What Impact does a Smartphone Application have on Their Fuel Efficiency? [C]. UbiComp, Pittsburgh, PA, USA, 2012
- [25] Zheng Y. Tutorial on Location-Based Social Networks[C]. WWW, Pittsburgh, PA, USA, 2012
- [26] Zheng Y. Location-Based Social Networks: Users [M]//Zheng Y, Zhou X. In Computing with Spatial Trajectories. New York: Springer, 2011
- [27] Zheng Y, Chen Y, Xie X. GeoLife2.0: A Location-Based Social Networking Service[C]. IMDM, Taiwan, 2009
- [28] Zheng Y, Wang L, Zhang R, et al. GeoLife: Managing and Understanding Your Past Life over Maps [C]. MDM, Beijing, 2008
- [29] Zheng Y, Xie X, Ma W. GeoLife: A Collaborative Social Networking Service Among User, Location and Trajectory[J]. *IEEE Data Engineering Bulletin*, 2010, 33(2): 32-40
- [30] Zheng Yu, Xie Xing. Enabling Smart Location-Based Services by Mining GPS Traces[J]. *CCF*, 2010(6): 23-30(郑宇, 谢幸. 基于用户轨迹挖掘的智能位置服务[J]. *中国计算机学会通讯*, 2010(6): 23-30)
- [31] Ye Y, Zheng Y, Chen Y, et al. Mining Individual Life Pattern Based on Location History[C]. IEEE MDM, Taiwan, 2009
- [32] Li Q, Zheng Y, Xie X, et al. Mining User Similarity Based on Location History[C]. ACM SIGSPATIAL GIS, Irvine, CA, USA, 2008
- [33] Xiao X, Zheng Y, Luo Q, et al. Inferring Social Ties Between Users with Human Location History [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2014, 5(1): 3-19
- [34] Xiao X, Zheng Y, Luo Q, et al. Finding Similar Users Using Category-Based Location History[C]. ACM SIGSPATIAL GIS, San Jose, California, 2010
- [35] Zheng Y, Zhang L, Ma Z, et al. Recommending Friends and Locations Based on Individual Location History[J]. *ACM Transaction on the Web*, 2011, 5(1): 1-5
- [36] Zheng Y, Zhang L, Xie X, et al. Mining Correlation Between Locations Using Human Location History[C]. ACM SIGSPATIAL GIS, Seattle, Washington, 2009
- [37] Zheng Y, Xie X. Learning Location Correlation from GPS Trajectories[C]. MDM, Kansas City, Missouri, USA, 2010
- [38] Bao J, Zheng Y, Mokbel M F. Location-Based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data [C]. SIGSPATIAL GIS, Redondo Beach, California, 2012
- [39] Zheng Y, Zhang L, Xie X, et al. Mining Interesting Locations and Travel Sequences from GPS Trajectories[C]. WWW, Madrid, Spain, 2009
- [40] Zheng Y, Xie X. Learning Travel Recommendations from User-Generated GPS Traces[J]. *ACM Transaction on Intelligent Systems and Technology*, 2011, 2(1): 2-19
- [41] Yoon H, Zheng Y, Xie X, et al. Smart Itinerary Recommendation Based on User-Generated GPS Trajectories[C]. UIC, Xi'an, 2010
- [42] Yoon H, Zheng Y, Xie X, et al. Social Itinerary Recommendation from User-Generated Digital Trails[J]. *International Journal on Personal and Ubiquitous Computing*, 2012, 16: 469-484
- [43] Wei L, Zheng Y, Peng W C. Constructing Popular Routes from Uncertain Trajectories[C]. KDD, Beijing, 2012



- [44] Liu H, We L, Zheng Y, et al. Route Discovery from Mining Uncertain Trajectories[C]. ICDM, Vancouver, Canada, 2011
- [45] Zheng K, Zheng Y, Xie X, et al. Reducing Uncertainty of Low-Sampling-Rate Trajectories[C]. ICDE, Brussels, Belgium, 2012
- [46] Zheng V W, Zheng Y, Yang Q. Joint Learning User's Activities and Profiles from GPS Data[C]. LBSN, Seattle, WA, USA, 2009
- [47] Zheng V W, Zheng Y, Xie X, et al. Collaborative Location and Activity Recommendations with GPS History Data[C]. WWW, Raleigh, NC, USA, 2010
- [48] Zheng V W, Cao B, Zheng Y, et al. Collaborative Filtering Meets Mobile Recommendation: A User-Centered Approach[C]. IAAAI, Atlanta, Georgia, USA, 2010
- [49] Bao J, Zheng Y, Wilkie D, et al. A Survey on Recommendations in Location-Based Social Networks[J]. *GeoInformatica*, 2015, DOI:10.1007/S10707-014-0220-8
- [50] Arthur O. Urban Economics[M]. Boston, Mass: McGraw-Hill/Irwin, 2003
- [51] Karamshuk D, Noulas A, Scellato S, et al. GeoSpotting: Mining Online Location-Based Services for Optimal Retail Store Placement[C]. IKDD, Chicago, IL, USA, 2013
- [52] Fu Y, Xiong H, Ge Y, et al. Exploiting Geographic Dependencies for Real Estate Appraisal: A Mutual Perspective of Ranking and Clustering[C]. KDD, New York, 2014
- [53] Liu W, Zheng Y, Chawla S, et al. Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams[C]. KDD, San Diego, California, 2011
- [54] Pang L X, Chawla S, Liu W, et al. On Mining Anomalous Patterns in Road Traffic Streams[C]. ADMA, Beijing, China, 2011
- [55] Pang L X, Chawla S, Liu W, et al. On Detection of Emerging Anomalous Traffic Patterns Using GPS Data[J]. *Data & Knowledge Engineering*, 2013, 87: 357-373
- [56] Chawla S, Zheng Y, Hu J. Inferring the Root Cause in Road Traffic Anomalies[C]. ICDM, Brussels, Belgium, 2012
- [57] Pan B, Zheng Y, Wilkie D, et al. Crowd Sensing of Traffic Anomalies based on Human Mobility and Social Media[C]. ACM SIGSPATIAL, GIS, Orlando, Florida, USA, 2013
- [58] Song X, Zhang Q, Sekimoto Y, et al. Modeling and Probabilistic Reasoning of Population Evacuation during Large-Scale Disaster[C]. KDD, Chicago, IL, USA, 2013
- [59] Goldman J, Shilton K, Burke J, et al. Participatory Sensing: A Citizen-Powered Approach to Illuminating the Patterns that Shape our World[EB/OL]. [http://www.mobilizingcs.org/wp-content/uploads/Participatory\\_Sensing.pdf](http://www.mobilizingcs.org/wp-content/uploads/Participatory_Sensing.pdf), 2014
- [60] Zheng Y, Zhou X. Computing with Spatial Trajectories[M]. New York: Springer 2011
- [61] Lou Y, Zhang C, Zheng Y, et al. Map-Matching for Low-Sampling-Rate GPS Trajectories[C]. ACM SIGSPATIAL GIS, Seattle, Washington, 2009
- [62] Yuan J, Zheng Y, Zhang C, et al. An Interactive-Voting Based Map Matching Algorithm[C]. MDM, Kansas City, Missouri, USA, 2010
- [63] Chen Y, Jiang K, Zheng Y, et al. Trajectory Simplification Method for Location-Based Social Networking Services[C]. LBSN, Seattle, WA, USA, 2009
- [64] Song R, Sun W, Zheng B, et al. PRESS: A Novel Framework of Trajectory Compression in Road Networks[C]. VLDB, Hangzhou, China, 2014
- [65] Wang L, Zheng Y, Xie X, et al. A Flexible Spatio-Temporal Indexing Scheme for Large-Scale GPS Track Retrieval[C]. MDM, Beijing, China, 2008
- [66] Chen Z, Shen H T, Zhou X, et al. Searching Trajectories by Locations: An Efficiency Study[C]. SIGMOD, Indianapolis, Indiana, USA, 2010
- [67] Tang L, Zheng Y, Xie X, et al. Retrieving k-Nearest Neighboring Trajectories by a Set of Point Locations[C]. SST, Minneapolis, MN, USA, 2011
- [68] Xue A Y, Zhang R, Zheng Y, et al. Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection against Such Prediction[C]. ICDE, Brisbane, China, 2013
- [69] Tang L, Zheng Y, Yuan J, et al. Discovery of Traveling Companions from Streaming Trajectories[C]. ICDE, Washington DC, USA, 2012
- [70] Tang L, Zheng Y, Yuan J, et al. A Framework of Traveling Companion Discovery on Trajectory Data Streams[J]. ACM TIST, 2013, DOI: 10.1145/2542182.2542185
- [71] Zheng K, Zheng Y, Yuan N J, et al. On Discovery of Gathering Patterns from Trajectories[C]. ICDE, Brisbane, China, 2013
- [72] Sheng C, Zheng Y, Hsu W, et al. Answering Top-k Similar Region Queries[C]. DASFAA, Tsukuba, Japan, 2010
- [73] Shekhar S, Chawla S. Spatial Databases: A Tour[M]. New Jersey: Prentice Hall, 2003
- [74] Güting R H, Schneider M. Moving Objects Databases[M]. Massachusetts: Academic Press, 2005

## Introduction to Urban Computing

ZHENG Yu<sup>1,2,3</sup>

1 Microsoft Research, Beijing 100080, China

2 Zhiyuan College, Shanghai Jiaotong University, Shanghai 200240, China

3 School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China

**Abstract:** Urbanization's rapid progress has led to many big cities, which have modernized people's lives but also engendered big challenges, such as air pollution, increased energy consumption and traffic congestion. Tackling these challenges can seem nearly impossible years ago given the complex and dynamic settings of cities. Nowadays, sensing technologies and large-scale computing infrastructures have produced a variety of big data in urban spaces, e. g. human mobility, air quality, traffic patterns, and geographical data. The big data implies rich knowledge about a city and can help tackle these challenges when used correctly. Urban computing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human, to tackle the major issues that cities face. Urban computing connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytics models, and novel visualization methods, to create win-win-win solutions that improve urban environment, human life quality, and city operation systems. This article introduces the concept of urban computing and its main application scenarios. It also discusses the key research problems and major methodologies of urban computing from the perspective of computer sciences.

**Key words:** urban computing; big data; spatio-temporal data analytics; data fusion; urban sensing

**First author:** ZHENG Yu, PhD, a lead researcher from Microsoft Research. He is also a chair professor at Shanghai Jiaotong University and an affiliate professor at Southwest Jiaotong University. His research interests include urban big data analytics, spatio-temporal data mining, and ubiquitous computing. Currently, he is leading the research on urban computing in Microsoft Research, passionate about using big data to tackle urban challenges. In 2013, he was named one of the Top Innovators under 35 by MIT Technology Review (TR35) and featured by Time Magazine for his research on urban computing. In 2014, he was named one of the top 40 Business Elites under 40 in China by Fortune Magazine, because of the business impact of urban computing he has been advocating since 2008. E-mail: Yuzheng@microsoft.com