

Beijing City Lab

Wang J Y, Li C, Xiong Z, Shan G Z, 2013, A survey of data-centric Smart City technologies. Beijing City Lab. Working paper # 20

以数据为中心的智慧城市研究综述

王静远¹ 李超^{1,2} 熊璋¹ 单志广³

¹(北京航空航天大学计算机学院 北京 100191)

²(北航深圳研究院数据活化(智慧城市)深圳市重点实验室 深圳 518057)

³(国家信息中心 北京 100045)

(jywang@buaa.edu.cn)

A survey of data-centric Smart City technologies

Wang Jingyuan¹, Li Chao^{1,2}, Xiong Zhang¹, and Shan Zhiguang³

¹(School of Computer Science and Engineering, Beihang University, Beijing 100191)

²(Data Vitalization/Smart City Key Laboratory, Research Institute of Beihang University in Shenzhen, Shenzhen 518057)

³(State Information Center, Beijing 100045)

Abstract After twenty years development of urban informatization and digital city, the concept of Smart city is becoming the next stage of urbanization on a global scale. The data science and technologies, such as big data, data vitalization, and data mining, play pivotal roles in building of Smart City. In this paper, we present a survey about data-centric Smart City technologies from an informatics perspective. This paper first summarizes the variety of urban data used in existing Smart City applications, and analyzes their features. Then, the state-of-the-art progresses in the research of data-centric Smart City are surveyed from three aspects: general architectures, application solutions, and core technologies. Finally, the paper concludes the characteristics of the research on data-centric Smart City and raises some directions for future works.

Key words smart city; urban computing; big data; data vitalization

摘要 在城市信息化浪潮与数据科学崛起的共同推动下,智慧城市开始在全球范围内成为未来城市发展的新理念与新实践。大数据、数据活化、数据挖掘等数据管理、应用与分析技术在智慧城市建设当中具有核心作用。本文站在信息科学的视角之上,围绕以数据为中心这一主题,对当前智慧城市研究工作的最新动态进行了综述。文章梳理了当前智慧城市相关研究所广泛采用的城市数据类型及其特点,并从技术体系、相关研究和核心技术与理论三个方面对领域中相关研究现状进行了介绍,最后对现有研究工作的特点以及未来可能的研究方向进行了总结和展望。

关键词 智慧城市; 城市计算; 大数据; 数据活化

在信息技术革命的推动下,以计算机、网络通信等信息通信技术为动力的信息化浪潮席卷全球。经过近十年的信息化与数字化建设,现代城市的运行方式与城市居民的生活环境已经发生了根本性的改变。城市的经济、文化、交通、娱乐等方方面面都已经和信息化的数字空间紧密融合,网络空间(Cyberspace)成为城市居民生活的组成部分。完善的信息基础设施以及丰富的数字化应用成为了现代数字化城市的基本特征之一。丰硕的城市信息化建设成果在为人类生活带来极大便利的同时,也为现代城市形态的进一步

演进奠定了技术与数据的基础。

进入21世纪,以大数据、数据活化为代表的数据科学与技术开始受到人们的广泛关注。以数据为中心的研究方法与技术理念在信息、生物、能源、医药、社会学等不同的学科领域都得到了广泛应用与认可,并促成了大量科研成果的诞生。以信息技术为支撑的数据分析与研究方法,正深刻地改变着传统科学探索的工作方式,成为了人类科技发展与知识获取的一种新兴模式。

在城市信息化浪潮与数据科学崛起的共同推动

基金项目: 国家“八六三”高技术研究发展计划项目(2011AA010502), 国家自然科学基金项目(61272350, 61202426, 61370122 资助)

表 1. 现有研究中常用的城市数据类型与特性
Table 1 Types and features of urban data in existing works

数据类型	数据特性	隐私敏感性	应用范围	数据维度	数据粒度
地图与兴趣点数据	城市基本构架	易获取	应用广泛	低	时间粗、空间细
GPS 数据	单一交通工具	较易获取	应用广泛	中	细
客流数据	多交通工具	隐私敏感	专用为主	中	细
手机数据	个人行为追踪	隐私敏感	专用为主	中	细
位置服务数据	特定兴趣点	较易获取	专用为主	中	细
视频监控数据	固定地点	条块分割	专用为主	高	细
环境与气象数据	空间粒度往往过粗	数据公开	专用为主	底	粗
城市社会活动数据	行业特性	条块分割	专用为主	中/底为主	-

下，智慧城市在全球范围内成为了下一代城市化发展的新理念和 new 实践。智慧城市是一种以新一代信息技术为基础，通过对城市各部分数据进行动态监测、分析、整合和利用，实现对城市生活环境的透彻感知，城市资源的全面调控，城市中各个部分协调配合、城市方方面面便捷运作、人和城市之间和谐共赢等目标的新型城市形态。世界各国，尤其是欧、美、日、韩等发达国家和地区，都在积极开展相关的理论研究与技术探索，发掘城市的数据资源，研发城市智慧应用系统，开展相应的城市试点。在我国，从中央到地方也都在积极探讨发展和建设智慧城市。

在智慧城市建设的过 程中，数据科学与技术发挥着不可替代的重要作用。为了使信息领域的科研人员尽快了解数据科学在智慧城市领域当中的最新应用与研究进展，本文对以数据为中心的智慧城市技术研究现状进行了综述和分析。文章首先对现有智慧城市研究所采用的城市数据类型与特性进行了归纳和梳理，然后从技术体系、相关研究、核心技术与理论三个方面对现有工作进行了综述介绍，并在此基础上对现有工作的特点进行了总结，最后对未来可能的研究方向进行了展望。

智慧城市是一个多学科交叉、多领域融合的综合 性学科，相关工作涉及的内涵和外延都非常广。本文 将综述视角放在信息科学领域，内容紧密围绕以数据 为中心这一主题，分析相关研究工作的整体结构与脉 络。对于非信息学科视角、数据研究以外的相关工作， 由于篇幅问题，不做太多涉及。

1 城市数据

1.1 城市数据类型

在信息城市与数字城市的建设过程当中，城市的

信息基础设施在提供信息服务功能的同时，也积累了 海量的城市动态数据，这些数据种类繁多，难以尽述。 在这里，我们对在现有智慧城市研究工作中较为常用 的城市数据类型进行简要的介绍。

● 地图与兴趣点数据

街道与建筑是城市的基本构架，地图数据是对城 市构架进行描述的基本方式，而兴趣点（Point of Interest, POI）数据则是介绍城市各功能单元的基本 信息。因此，城市地图和兴趣点数据是进行以数据为 中心的智慧城市研究的最基本原料，也是在对其它类 型城市数据进行融合时的空间锚点数据。地图和兴趣 点数据包含了城市布局、结构、行业分布等丰富信息。 如果能够结合大时间跨度的历史数据，还可以从地图 和兴趣点数据中获得城市的发展演进、行业兴衰、布 局迁移等人类学与经济学知识。

● GPS数据

安装有GPS接收芯片的移动设备不但可以为城 市的车辆提供导航与定位服务，同时也可以收集城市 中人、车等流动物体活动信息。例如目前应用比较广 泛的浮动车技术^[1]就是将出租车、公交车等公共交通 工具上安装GPS设备，将其作为传感器对于城市的交 通情况进行采样。浮动车GPS数据对于感知城市中的 交通状况和人类活动行为都有很大的价值。安装有 GPS接收芯片的智能手机也可以当做个人行为轨迹 的收集设备。手机GPS数据可以对某一活动个体的行 为进行全天候追踪采样，因此具有非常大的应用价值。 但是由于隐私、安全等诸多问题，手机GPS数据的很 难大规模收集应用，目前只能依靠志愿者的方式进行 小范围收集和研究。

● 客流数据

城市中市民采用不同交通工具进行日常通勤的 数据称为客流数据。出租车的客流数据可以使用浮动

车GPS数据配合出租车计费表的乘客状态获得。公交车与地铁的客流数据则可以使用市政交通一卡通的刷卡记录进行收集。市政交通一卡通可以同时记录持卡人的刷卡时间、刷卡地点与乘坐的交通工具类型，而且市民通常不会频繁地更换一卡通，因此一卡通数据还兼具身份标示的功能。客流数据包含的城市活动信息非常丰富，可以被用于城区功能分析、人口流动监测、城市交通系统评估、多交通工具人类行为研究、城市交通经济学研究等领域。但是，客流数据在一定程度上也存在有隐私敏感性问题，可以在进行加密或匿名处理之后进行收集和使用。

● 手机数据

手机是人们日常生活必不可少的通讯工具，其所能提供的数据类型很多，包括：通讯录、通话记录、GPS定位信息、与基站间的信令记录、上网记录和APP使用记录等。这些数据可以反映城市中居民活动的兴趣偏好、活动范围、规模频率、社交关系等内容，因此具有非常巨大的应用潜力。但是由于手机数据涉及到详细的个人隐私，不论是政府、运营商还是手机用户对使用这类数据进行科学研究或是商业应用都十分地敏感。同时由于手机数据所包含的个人信息过于丰富和集中，因此简单的匿名处理往往不能完全消除用户隐私泄露的危险。这些原因都导致目前工业界和学术界对手机类数据的利用程度均十分有限。

● LBS位置服务数据

LBS位置服务（Location Based Service）是移动互联网时代一种新兴的网络服务方式，其特点是根据用户所在的地理位置信息有针对性的提供社交、推荐、广告、导引等服务内容。通过LBS应用所收集到的数据具有明确的地理位置坐标并兼具传统Web服务的语义特性。签到数据、商家点评数据、带有位置信息的社交网络数据等均属于LBS数据的范畴。本质上，LBS数据是对POI数据的一种深度的描述和补充，但与地图和POI等简单的城市地理数据相比，LBS数据包含大量的语义信息，可以帮助人们更加深刻地理解城市功能结构与人类活动之间的相互作用方式。

● 视频监控数据

随着视频传感器技术的普及，视频监控技术已经被广泛的应用在了交通管理、社区安保、室内安防、娱乐通讯等城市生活的各个方面。视频监控设备的安装者也包括政府、企业、机构和个人等不同类型。视频监控设备所采集的海量视频数据记录着城市中居民生活的分分秒秒，在数字空间中形成了对物理城市的虚拟“镜像”。充分利用这些视频数据可以从某种程度上再现城市生活的历史，其理论研究与应用价值是不

言而喻的。然而，视频数据同时也存在有维度高、非结构化、语义鸿沟、可理解性差等问题，加之视频数据庞大的生成规模，使得分析和利用城市视频数据变得异常困难。如何开发利用好城市视频这个内容丰富的城市数据宝藏，成为了长期以来困扰工业界和学术界的一大难题。

● 环境与气象数据

气象数据很早便受到城市科学研究的充分关注，典型的研究如天气预报等已经形成了成熟的理论和应用并且得到了广泛的推广。精准的天气预报能够对城市市民的出行规划起到非常大的帮助作用。同时，随着人们对于环境与健康问题的日渐重视，以空气质量为代表的城市环境数据近两年也开始成为社会关注的焦点。雾霾、PM2.5浓度等问题，对于城市市民生活的影响日渐显露。城市环境和气象数据可以在市民出行、工业规划、交通限行、环境治理等诸多领域发挥重要作用。城市的环境与气象数据的一个重要的特点是其地理与时间采样密度低，常规的地理采样密度是以区县为单位，而时间采样密度一般大于一小时一更新。如何实现细粒度、高精度的环境与天气数据分析是该应用类数据的一个重要挑战。

● 社会活动数据

城市社会活动数据包括了城市中的人口户籍、金融物价、医疗卫生、能源消耗等各种社会动态数据。每一类社会活动数据都可以透彻的反映城市居民社会生活的一个方面，是深入理解和分析城市社会化行为的必备原料。这些数据都已经在各自所对应的行业当中发挥了巨大的作用。然而，由于城市社会活动数据行业性较强，容易受到行业条块分割的影响，往往彼此分割孤立，难以统一获取。同时，由于数据的来源与获取渠道十分分散，因此数据之间的时空尺度与采样粒度差别非常大，很难找到合适的同步锚定点。这些特点导致城市社会活动数据之间的相关性作用很难被充分挖掘出来。打破行业条块分割、实现多源异构城市数据的融合，是深度开发利用城市社会活动数据所面临的首要任务。

1.2 城市数据的特性

● 大数据特性

城市是人类活动最为密集的区域，海量的人类活动与社会运行数据不可避免地在城市当中爆发。以北京为例，每天有超过千万的市民出行，每天我们的交通卡产生五千万条刷卡记录，有九百万的车辆在运营，每天仅出租车GPS数据就会产生八千万条左右，同时有近万个交通固定检测器在采集车的瞬时车速，有近

亿万条的手机通讯数据，还有千亿级的我们交通影视影像数据。城市数据种类繁多的（Variety），规模庞大（Volume），对数据的输入和处理速度要求也很高（Velocity）。城市数据包含的信息与知识极为丰富，对于推动人类认识的扩展与科学技术的进步，有着巨大的价值，同时由于数据稀疏性的影响，城市数据的价值密度也很低（Veracity）。综上可以看出，城市数据完全符合大数据（Big Data）所具有的4V特性^[2]。城市数据可谓是大数据概念范畴其中的一个极具代表性的典型样本。

● 时空多维特性

以地图为基础的空间结构是城市数据的一种基本组织方式，而城市快节奏的生活方式，也使得城市数据对于时间维度的变化也非常敏感。因此，时空多特性成为了城市数据的另一个重要特点。在空间上，根据城市地理规模的不同，城市数据具有不同尺度的空间跨度。在时间上，根据产生的时间不同，城市数据具有时间相关的变化和分布。因此在进行城市数据分析和应用时，一方面需要考虑时间和空间两个维度的数据演化特性，另一方面还需要充分利用时间和空间不同维度之间的数据关联关系。这对城市数据的利用技术提出了很高的要求。

● 多尺度与多粒度

研究和利用城市数据除了要考虑时间和空间等多个维度之外，还需要考虑数据尺度和数据粒度对于数据特性的影响。在规模的尺度上，城市可以分为小型城市、中型城市、大型城市、超大型城市等；在地理尺度上，对于城市数据的描述可以小到几个街区或大到数千平方公里；在时间尺度上，城市数据的覆盖时间可以短到一些事件的监控，长到上百年的城市变迁。在地理采样粒度上，可以像遥感测绘数据一样精确到数米，也可以向气象环境数据一样以区县、地区、甚至省市为单位；在时间粒度上，更是根据数据采样设备的时钟、存储与传输能力、计算速度等因素产生不同的时间粒度。在时空多维度的条件下，高效处理多尺度与多粒度的海量数据，是有效利用城市数据所必须解决的核心技术问题之一。

● 多元与异构

如前文所述，城市数据具有非常多的类型与来源，即数据的多元性。这些不同来源的城市数据不论是从结构上、组织方式上、维度尺度与粒度上，都会存在巨大差异，即数据的异构性。智慧城市的应用需求要求我们必须将这些多元异构数据进行有机的融合，通过挖掘活化数据之间的相关性与相互作用方式来获取新知识。如何在一个统一的构架上分析异构性极强

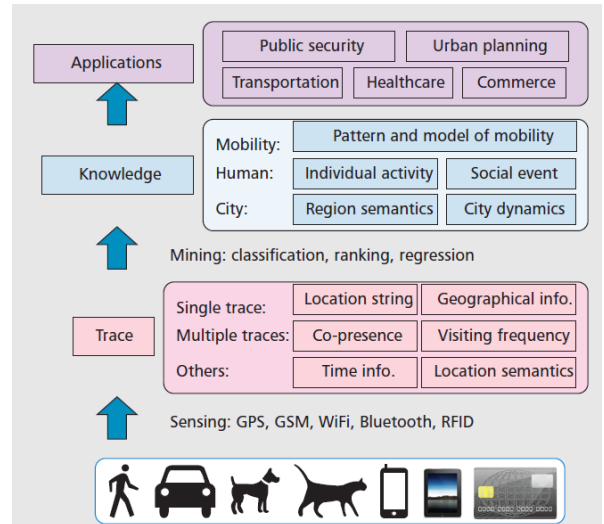


图 1. 基于轨迹挖掘的智慧城市技术体系框架^[3]

Fig 1. Framework of trace mining for smart cities.^[3]

的多元数据，是学术界和工业界在进行智慧城市探索方面所面临的一项共同挑战。

表1对现有研究中常用的城市数据类型与特性进行了简要的总结分析，从表1可以看出现有工作对于城市数据的应用类型还非常有限。城市作为人类生活活动最为密集的场所，所产生的数据类型和数量都是难以尽数的，仅就现有的研究工作来看，我们对于城市数据的开发还远远不足，城市这座数据宝库中积累了无尽的数据资源，等待着科研人员进行深入的开发。

2 技术体系

智慧城市技术体系研究作为对整个智慧城市技术研究工作的顶层设计，对于指导领域技术的发展方向，明确研究工作的内涵与外延，优化现有研究资源的配置与分布等均具有非常重要的意义。相对于工业界琳琅满目的智慧城市解决方案，学术界对于智慧城市的技术体系框架则具有相对较好的共识。

如图1所示，浙江大学的潘纲教授在2013 IEEE Communication Magazine, Smart Cities专题中阐述了基于轨迹数据分析与挖掘的智慧城市技术体系框架^[3]。该框架将基于轨迹数据的智慧城市技术体系分为了轨迹感知(Trace)、知识发现(Knowledge)和具体应用(Applications)三个层次。该框架从概略的层面上系统描述了以数据为中心的智慧城市技术的整体技术路线。微软亚洲研究院的郑宇团队在2013《CCF通讯》城市计算专题当中提出了一种“四层反馈”结构的城市计算技术体系框架^[4, 5]。如图2所示，该技术体系将城市计算的技术框架细分为了“城市感知与数据捕

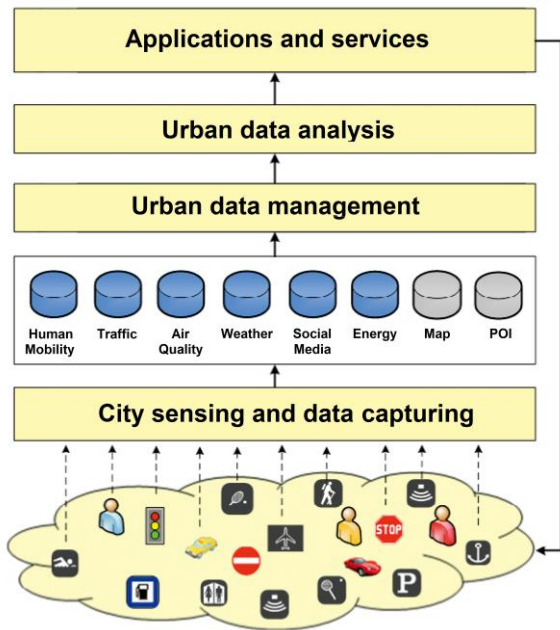
图 3. 微软亚洲研究院城市计算技术框架^[4]

Fig 2. MSRA urban computing technical framework [4]

获”、“城市数据管理”、“城市数据分析”、“服务提供”等四个层次。该框架的一个特色在于其引入了“服务提供”层对于真实物理世界的反馈回路，更加完善的考虑了智慧城市技术对于城市生活的影响。在科技部863“智慧城市（一期）”项目的支持下，863智慧城市项目（一期）总体组提出了“六横两纵”的智慧城市技术框架^[6]。如图3所示，该框架将智慧城市技术体系按照依赖关系划分为了“城市感知层”、“数据传输层”、“数据活化层”、“支撑服务层”、“应用服务层”以及“行业应用层”等六个横向的层次，即“六横”；同时还引入了贯穿六个层次的两大保障体系：“标准与评估体系”以及“安全保障体系”，即“两纵”。该技术框架的一大特色在于其十分具体的细化了各个层次所应包含的技术集合以及其之间的相互依赖关系，为智慧城市技术的整体发展提供了详细的布局性指导。我国未来智慧城市的技术体系建设很有可能将会参照该体系进行丰富和完善。

综合看来，上述几种智慧城市的技术体系框架均有一个共同的特性，即以数据的获取与管理技术作为底层的支撑，以数据挖掘、处理和分析技术作为整个框架的核心构成，在此基础之上对城市用户提供多样化的服务应用。换言之，未来智慧城市的技术的发展方必将是以数据为中心，在这一点上学术界已经从顶层设计的角度达成了基本共识。

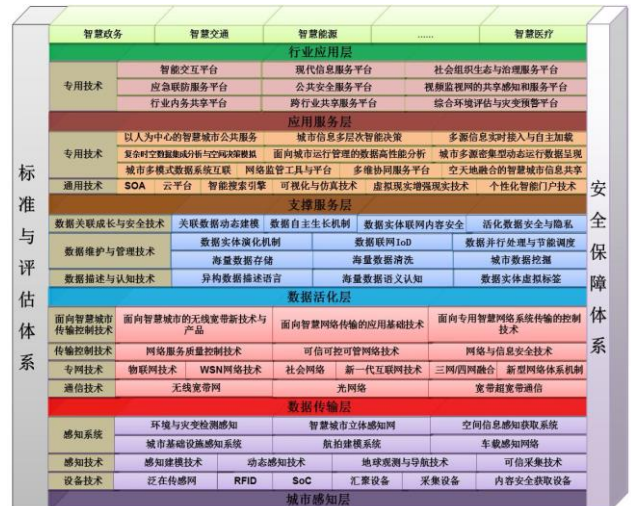
图 2. 863 智慧城市总体技术体系架构^[6]

Fig 3. The Smart City technical framework of the "863" project.

3 相关研究

3.1 数据驱动的智能交通

城市交通系统是城市中信息化程度较高的部分，浮动车、一卡通、微波线圈、摄像头等交通传感与信息化设备可以有意或无意的将城市中交通参与者的交通行为记录下来，从而为数据驱动的科学研究提供研究样本。同时，由于城市交通领域自身的数据富集优势，又使得以数据为中心的智慧城市技术能够率先在智能交通领域中发挥重要作用，我们称这类技术为数据驱动的智能交通技术。以数据为驱动的智能交通技术研究所采用的城市数据主要包括地图与POI数据、GPS数据、客流数据、道路微波测量数据等^[7, 8, 9]，并通过多种手段对采集到的数据进行分析理解，实现感知城市的交通运行状况，为市民提供交通引导、导航、推荐等智能服务。

数据驱动的智能交通技术可以被细分为支撑层面和应用两个层面两个层次的研究。支撑层面的研究工作集中在全城交通的感知与分析方面，其目的是感知城市的总体交通状况、分析全市交通的统计行为特征，建立分析模型，为具体的智能交通应用提供数据分析与交通状态评估支撑。例如，基于城市交通监控数据的实时路况报告作为一种成熟的技术已经得到了非常广泛的应用^[10, 11]，对于城市中具有特殊特性路段的检测和查询^[12]、平均通勤时间评估^[13]、交通异常与事故的检测等研究可以极大地提高城市道路交通的管理效率。另外，通过对于GPS数据的深入挖掘和分析，可以进一步理解城市中交通运行的具体模式^[7]。

^[14, 15, 16]，提供交通流量的评估、预测和管理等应用服务^[17, 18, 19]。

应用层面的研究则集中在城市交通管理与运行的各项服务应用当中。基于GPS的路径导航服务是最为典型的应用之一，丰富详尽的地图数据配合实时的路况分析结果可以为用户提供非常优质的行驶路径导航服务^[20]。包含有人类行为信息车辆GPS数据则可以提供汽车司机的驾驶知识来进一步优化导航路径的选择^[21, 22, 23]，微软亚洲研究院开发的T-Driver^[22]车辆导航系统就采用了这样的设计理念。T-Driver系统统计了北京市城区出租车的GPS行驶数据，然后将不同地标之间驾驶技术最娴熟的出租车司机的驾驶路径用图的方式组织起来，就形成了一张包含了出租车司机驾驶知识的地标图。用该地标图来进行路径导航，可以有效的提高车辆在拥堵时段的行驶效率。该研究的主要特色在于将数据统计中获得的人类智慧应用到了传统的信息化交通服务当中，将车辆导航应用由传统的“以计算为中心”变为了“以数据为中心”，所采用的核心技术也由传统的规划技术变为了以数据为驱动力的统计技术，其意义非常深刻。

另一类被广泛研究的应用是出租车与乘客之间的推荐服务。该类研究最重要的任务就是理解出租车与乘客的行为模式以及两者之间的交互关系^[18, 24]。法国电信研究院的D. Zhang团队通过对5350辆出租车驾驶行为的研究，分析比较了不同出租车寻客策略的有效性^[25]。文献[26, 27]对如何通过优化出租车的载客策略来提高出租车司机收入进行了深入研究。北京航空航天大学K. Xue团队^[28]通过对北京1200辆出租车的乘客数据分析，研究了优化的乘客打车策略，并基于该研究开发了名为TaxiWaiter的打车辅助系统。T-Finder系统^[29, 30]则同时考虑了出租车与乘客两个方面的需求，通过对街道打车概率的统计和分析，来进行出租车寻客路线和乘客打车路线的推荐。

路径导航（行车路线推荐）、出租车寻客路线推荐和乘客打车地点推荐都属于有资源约束的分配推荐问题，其本质是一个带约束的多方博弈，现有的研究工作提供的均是局部优化的解决方案。博弈论的相关知识告诉我们，分布式的局部最优并不能保证带来最终的全局最优解^[31]。如果所有的出租车均按照推荐的行驶路线到乘客较多的地点争夺乘客，那么一方面会导致这些地点成为较难寻找顾客的地点，另一方还会导致其它地点的乘客由于缺乏出租车而打不到车。解决该问题的一个途径是采用集中调度的手段实现全局的车辆负载均衡，这种方案虽有较好的理论性能，但是实现起来非常困难。另一个途径是在推荐算法的

设计上引入博弈惩罚机制，以多轮博弈的方式实现分布式的全局最优。相关的理论与应用研究还需要进一步地深入探讨。

出租车行驶的异常轨迹检测也是学术界非常关心的问题^[26, 32]。如何区分出租车司机为躲避拥堵而进行的适当绕行和恶意的“宰客”绕行是该类问题要解决的核心问题。法国国立电信学院设计提出的iBAT/iBOAT^[33, 34]算法可以有效地对绕行出租车进行在线识别或是轨迹识别，其分析结果显示，偏爱绕路的司机所对应的月收入并不比不绕路司机的平均值高，一些司机想通过恶意绕行来增加收入往往只是一种心理安慰行为^[35]，这对于设计合理的出租车收费政策，避免司机恶意绕行有着非常重要的价值。

以数据为驱动的智能交通技术还可以在优化城市公共交通系统方面发挥巨大的作用。B-Planner^[36]系统使用出租车GPS数据所提供的城市通勤需求信息，重新设计了杭州市夜间公交车的行车路线，满足了不同时段人们对公交线路的不同需求。T-Share^[37]出租车拼车系统通过综合考虑打车人的位置、目的地以及出租车的行驶路径等因素，对出租车的拼车路线进行了合理规划，在充分利用出租车自由灵活特性的同时提高了搭载乘客的通勤效率。Flexi^[38]则使用GPS数据设计了一种灵活性介于公交车和出租车之间的小型绿色公交系统。随着轨道交通系统在各个城市的发展，乘坐地铁出行成为城市市民们越来越多的选择，针对地铁轨道交通的智慧城市交通数据研究也受到了越来越多的重视。北京航空航天大学使用北京地铁系统的客流数据，对北京市轨道交通的负载流量进行了预测分析，其研究成果对于保障轨道交用运营安全，提高轨道交通运营效率有着非常重要的意义^[39]。L. Capra等对伦敦地下铁系统(tube)的客流拥堵模式进行了系统的分析，并以此为依据提出了避免地铁拥堵的技术方案^[40]。除此之外，综合利用多种交通工具的客流数据，还可以实现对用户全出行路径的系统规划与通勤时间估计^[41, 42]，以此为基础开发的城市公共服务系统，对于优化城市的整体通勤效率、改善市民的出行交通体验等，都有着非常巨大的帮助。

从上述工作可以看出，科研人员们在数据驱动的智能交通领域均取得了丰硕的研究成果。值得注意的是，现有的系统与成果大多是依靠浮动车GPS、一卡通、微波探测线圈等结构化较好的数据源。对于包含了丰富语义信息但结构化程度低、信息维度高的城市交通视频监控数据，现有研究使用的还非常少。有关监控视频的应用研究依然停留在视频处理、语义提取、事件理解等阶段^[43]，在智慧城市技术体系中扮演着数

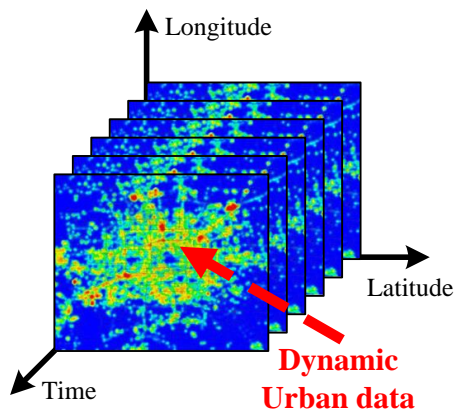


图 4. 城市动态特性的数据感知

Fig 4. Data sensing of urban dynamic

据准备的角色,尚不足以完全支撑以知识发现为目标的集成应用。这一方面是因为高维视频语义分析理解难度大,相关技术尚不成熟;另一方面也是因为视频数据的体量过大,很难按照城市的规模进行协同组织与处理。解决城市交通监控视频上述两个方面的挑战,将会是以数据为中心的智慧城市研究在智能交通领域所要面临的一项重要任务。

3.2 城市计算技术

城市计算是计算机科学以城市为背景,与城市规划、交通、能源、环境、社会学和经济学等学科融合的新兴领域,城市计算技术从城市的全局着眼,通过整合、分析和应用跨行业、跨领域的城市数据,来设计更加高效的城市服务与应用系统,并希望通过这些技术的应用来实现“人—环境—城市”的三者的和谐发展^[4, 44, 45]。城市计算技术的核心是数据挖掘、分析与应用技术,以数据为中心的城市计算技术是构建智慧城市所必不可少的一项支撑技术。在本小节,我们从城市动态特性研究、城市功能识别与规划、附加行业应用等三个方面,介绍在城市计算领域围绕城市数据应用所取得的一些最新研究进展。

3.2.1 城市动态特征的检测与分析

如图4所示,在地理结构上一座城市可以近似的看做一个基于经纬度坐标系的二维平面空间,通过城市传感设施所获得的城数据则可以看做对于二维城市平面空间的动态特性的时间采样。这样,按照经纬度地理位组织起来的城数据采集设备就形成了一个随着时间变化的二维时变信号源。信号源的采样对象是城市的内在动态特性,城市的人口、交通、经济、环境等各种深层次的动态行为都会在该信号源上有所体现,不同的数据收集方式相当于采用了不同类

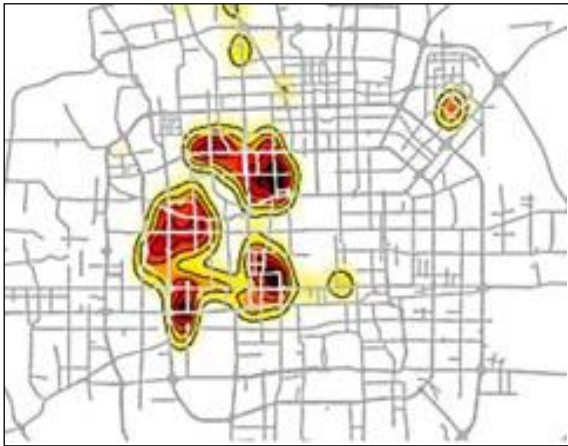
型的传感器,任何一种城市数据都是城市的内在动态特性在某一特殊传感信号空间上的映射。例如,使用安装有GPS设备的出租车与公交车辆收集到的浮动车数据可以被理解使用浮动车作为传感器对于整个城市的动态特性进行采样^[46],所收集到的GPS位置信息实际上是城市深层动态特性在浮动车GPS信号空间上映射。尽管浮动车的GPS数据直接表现出来的是车辆的位置、速度分布等道路交通信息,但事实上整个城市的市民生活方式、经济文化分布、城市地理结构等深层次的城市特性都会在GPS信号空间上有所反映。如果我们能够综合有效的利用信号分析、统计学习、数据挖掘等技术对浮动车GPS数据等城市感知数据进行挖掘,就可以有效的检测发现城市动态特性的内在模式和运行规律,从而改善和优化城市环境,实现城市发展与运行的智慧化。

在使用城市感知数据对城市动态特性进行研究的现有工作可以分为城市动态检测和城市动态分析两个大类。接下来,我们对这两个大类的代表性工作进行简要介绍。

● 城市动态特征检测

城市动态检测研究的目的是指对城市感知数据中包含的城市空间特征和时间特征进行检测发现。如图5所示,空间特征是指在时间上频繁或规律出现而在地理上具有独特分布的城市数据特征。城市的空间信号特征往往对应于城市中具有特殊功能的热点地区,例如商业区、交通枢纽站、事故多发点等^[49, 50]。早期的热点检测方法以城市数据信号的强度和密度特性作为衡量指标,不同的强度和密度对应不同的空间特征。香港科大的Lionel M. Ni教授团队最早提出了一种基于移动特性而非密度特性的热点检测方法^[47]。微软亚洲研究院的Y. Zheng团队通过将GPS数据同POI数据相结合的方式,计算了北京各个城市区域中属于某一个功能区划的概率,并以此为依据实现城市功能区的划分^[48],该工作的底层支撑技术就是城市空间信号的特征检测^[49, 50]。类似的工作还有使用推特文本数据流的时空动态检测^[51],根据大规模出租车的运动行为对城市的社交功能分布进行检测^[52],以及对城市人口的移动特性检测^[53]等。

如图6所示,城市时间特征是指在时间上随机出现的异常于该地理区域通常时间分布的异常信号特征,该类特征往往对应于城市中一些特殊事件的发生,例如演出、大型活动、交通管制等。文献[54]使用出租车浮动车的GPS数据设计了分析城市交通流量异常的检测算法,文献[55]在前述算法的基础上进一步实现了对于交通异常模式的早期检测,利用该算法提

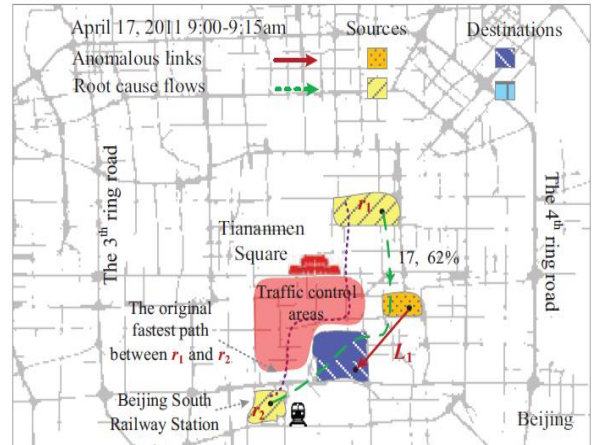
图 5. 城市空间动态特征（城市热点）^[5,48]Fig 5. Spatial dynamic patterns of city (Hot points)^[5,48]

供的服务，可以更早的对道路交通中的一些拥堵事件进行具有前瞻性的预防处置。文献[56]则使用群体感知的方式收集了城市中市民的活动数据和社交网络的媒体数据，并利用这些数据进行城市交通事件的异常检测。由于社交网络中的媒体数据具有更加丰富语义特性，因此可以获得更加准确检测效果。

● 城市动态特征分析

城市动态特性的分析工作更加侧重于对城市时间和空间模式的成因及内在联系进行分析和理解。例如，研究工作[57]提供了对于道路交通中异常事件根本成因的分析挖掘的方法。研究工作[63]使用关联数据挖掘技术，设计了分析城市交通异常事件在时间和空间上的因果关联的实现算法，并采用该算法对北京市的一些路段的热点时间之间的因果关联进行了研究分析。文献[58]将研究重点放在了城市空间动态特性的分析之上，其利用浮动车的GPS的轨迹数据建立了城市不同空间特征之间的交互作用模型。文献[59]进一步分析了城市空间特征与时间特征背后的成因联系，将城市动态特性分解为一些基本模式，并将复杂的城市空间与时间特征用这些基本模式的线性组合表示，研究的实验结果具有很好的可解释性。研究工作[60]则使用可视化的方式，对于北京市城市交通拥堵事件的关联关系进行了分析。

比起面向具体城市应用的研究工作，城市动态特性的检测与分析的内涵和外延都更加抽象，研究工作范围边界很难明确界定，很多研究都作为具体应用的支撑技术在相关文献的某一章节中进行介绍，系统性不强。但是作为应用研究的支撑基础，该类研究工作的价值不容忽视，构建和梳理系统的城市动态特性检测分析技术与理论体系是深入开展该项研究工作的当务之急。

图 6. 城市时间动态特征（异常事件）^[5,57]Fig 5. Temporal dynamic patterns of city (Events)^[5,57]

3.2.2 城区功能识别与规划

识别城市不同区域的主要功能作用是理解城市特点的首要任务之一，开展城市规划工作是指导城市建设发展的主要手段，以数据为中心的智慧城市技术在上述领域同样发挥了非常重要的作用。

城区的功能识别是指利用城市运行数据和动态特征对于城市的各个不同区域按照其在整个城市中所承担的功能进行识别和标注。例如划分出商业区、住宅区、工业区、行政区等。该类研究的难点一方面在于城市的功能区划往往具有非常大的重叠性和混合性，即便是采用人工标注的方法也很难准确定义某一个城区的所承担的具体功能；另一方面，由于城市的功能区划属于城市动态特性的深层次语义，并没有非常适当的城市监控指标可以直接反映城市区域的功能，科研人员需要从不同城市数据空间中将城市功能区划的投影从众多深层次特性的投影当中提取出来，这具有非常大的难度；此外，由于城区功能是城市的深层次特性不能直接观测，也就必然难以找到合适ground truth对分类结果进行验证，这也造成了对于研究工作评价的困难。尽管具有诸多挑战，但是该类研究的价值依然是显而易见的，对于城市的设计者来说，城市的功能结构对应了城市的内在发展规律，通过了解城市的内在的发展规律，可以指导未来城市的建设方向，纠正已有不合理的城市规划方案等。

早期的城市功能区域识别工作只能对城区进行粗粒度的功能区分，例如研究工作[61, 62, 63, 64, 65]仅能区分住宅区、非住宅区两种类型的城市功能区划，研究工作[66]将分区的精度提到为了三种，即住宅区、工商业区、公共机构区。研究工作[67]则将城市的功能区进一步细化为了四种，即商业区、工业区、服务

区、住宅区。这些粗粒度的城市功能区域划分在一些结构单一、形成时间短、规划效果好的城市当中能够表现出较好的性能。但对于我国城市普遍存在的一些问题，例如城区功能复杂、混合程度高、先建设后规划等问题，上述算法很难满足实际的应用需求，还需要有更加细粒度的城区识别技术。针对上述需求，研究工作[59]将城市的日常通勤模式划分成了住宅区模式、工作区模式和其它模式，每一个城区的功能可以被建模为三种不同模式根据自身功能需求而形成的线性组合，线性组合的系数可以被用来当做该城区功能成分的度量。例如，如果某一城区在住宅区通勤模式上的系数比较大并有中等的工作区通勤模式系数，那么该区域就可以被认为包含有较多的住宅，同时还兼有一部分工作区的功能。研究工作[49]对于一个城市商业区的密度水平进行了评估，每一个城区都可以使用该文献提出的算法来估算商业功能在该个城区中的概率。文献[50]提出了对于城市的住宅区密度分布进行测算的方法，这些方法可以用来度量某个城区属于住宅区的概率。文献[6]使用OD语义流来探究不同出发-到达城区之间的相互关系。浙江大学的G. Pang团队使用出租车的GPS数据对于杭州市不同城区的功能进行了细粒度的划分^[错误!未找到引用源。]，该研究通过细致的城市区域功能特征设计，使用基本的SVM算法获得了高达95%的城区功能识别率。微软亚洲研究院的Y. Zheng团队借鉴了自然语言处理领域中文章主题分类的方法对北京市的城区域功能进行了识别和分类^[48]。该研究的核心思想是将不同的城市区域建模为文章主题分类模型中的文章，城市的区域功能类型建模为文章的主题，将城区的通勤模式分布建模为文章的词频分布，并配合POI数据使用改进的LDA对城市区域进行分类，取得了非常好的分类效果。

除了使用交通通勤数据外，文献[7]使用手机数据对城市的区域功能进行了识别和划分。类似的技术还可以应用到对于道路和特殊兴趣点的类型划分，例如，研究工作[7, 7]对于城市的街道类型进行了识别和划分。研究工作[7]发现城市轨道交通的客流数据在反映乘客出行情况的同时，也揭示了城市区域间的功能联系和演进规律，因此使用北京市的地铁客流进出站数据对北京地铁站周边的城市区域进行了功能识别与划分，为北京城市地铁站周边区域的规划与设计提供了指导信息。研究工作[7]则使用纽约地铁的客流数据对纽约地铁站的功能类型进行了划分，为纽约地铁的建设规划提供了指导帮助。

使用城市数据对于建设规划的合理性进行评估也是近两年新兴的一个研究方向。研究工作[46]使用

城市出租车GPS轨迹数据测算了北京市不同城区之间的连通性，并设计了三种指标对于城区之间的交通性能进行评估。该研究所取得的成果可以用来检测城市的设计失误和评价城市的设计优劣。

上述研究工作可以看做是对城市动态特征检测技术的一种应用，他们的一个共同特点是其并不关心城市数据与某一具体城市动态特性之间的相关关系，而是使用城市数据所包含的整体信息来对城市的结构特性进行理解。研究方法则是直接对采集到的城市数据进行模式提取与分类，并用分类后的结果近似城市的功能区划。这些研究之所以能够这样做是因为城市的功能与规划本身就是多种城市动态特性的综合体现，我们可以不用具体理解某一特定的城市动态特征对城区功能的直接作用方式，而只需要比较其之间的差异性即可。在该类研究中，数据对城市动态特征的综合程度越高，越有利于问题的求解。

3.2.3 附加行业应用

以数据为中心的智慧城市技术也能够为不同的城市行业应用提供巨大的帮助。我们知道海量的城市数据的收集过程本身就是为了支持与之相对应的行业应用。例如，GPS浮动车数据是为了监测道路的拥堵状况、手机数据是为了提供手机通讯服务、一卡通客流数据是为了提供方便的公共交通服务等。除了这些数据本身所对应的专门应用之外，城市数据还可以用于提供与最初数据收集过程无关的行业应用，我们称之为附加行业应用。附加行业应用的一个重要特点是，人们无法获得充分的目标行业数据，而只能采用相关的外围城市数据，建立行业信息同外围数据之间的关联模型，再利用模型和综合数据反推行业应用所需的信息。如何从包含了城市综合特性的外围城市数据当中提取某一特定附加行业应用所需要的信息，是附加行业应用研究所面临的主要挑战。一些具有代表性的研究工作包括能源消耗、空气质量、住房价格和地图测绘等。接下来我们对这些工作进行简要介绍。

● 能源消耗

能源是维系城市运转的动力所在，随着全球能源的日益枯竭，降低城市能源消耗、构建绿色城市成为了智慧城市建设的核心目标之一。然而，城市作为一个复杂的能量代谢系统，即便是弄清楚城市对某一种特定形式能源的消耗量也是非常困难的。为解决这一问题，微软亚洲研究院的Y. Zheng团队利用出租车GPS数据和城市加油站的POI数据对北京市机动车辆的每日汽油消耗量进行了估算^[7]。该研究所要解决的挑战一方面在于出租车并不能完全代表城市中全部

车辆的行为,每一个加油站中正在加油的车辆中只有一小部分是出租车,也并非每个加油站每时每刻都有出租车辆在加油;另一方面GPS数据只包含出租车的行驶轨迹与运营状态信息,没有明确的车辆行驶意图信息,一辆出租车在加油站附近出现并不能说明其正在加油,需要有专门的算法对出租车的加油行为进行检测和判断。针对上述两个方面的问题,Y. Zheng团队设计实现了从GPS轨迹数据中发现加油事件的检测方法,提出了一种能够在稀疏张量当中分析汽车在加油站中加油所消耗时间的评估算法,并实现了能够通过加油时间推断加油站车辆到达频率的排队计算方法。该研究可以为普通用户提供加油站的推荐服务,也可以为石油公司的加油站建设规划提供意见,同时还可以让政府了解和掌握整个城市的能源消耗情况,从而制定更为合理的能源管理政策。

除了Y. Zheng团队的研究工作之外,研究工作[7]还通过分析城市人口、车辆GPS数据以及POI等数据来规划电动汽车充电站的建设。文献[7]则通过分析车内总线传感器的数据来设计更加节能环保的汽车驾驶方式。

● 空气质量

大气污染问题是我国北方主要大型城市所面临的一个巨大的环境问题。尤其是近几年,北京的空气质量问题受到了从政府到公众的一致关注,PM2.5、雾霾、空气指数等不断成为新闻媒体所热议的关键词。为解决我国城市近两年持续出现的空气污染问题,国务院于2013年9月10日印发了《大气污染防治行动计划》^[7],提出要从十个方面采用综合手段防治大气污染。然而,空气污染也具有非常复杂的成因,人们在对于大气污染防治的认识依然存在着许多空白。比如大气中的首要污染物就是竟是由什么原因造成的,交通、工厂、气候、天气、人口、植被,究竟哪一个才是对空气质量影响最大的因素,现有的空气监控系统能否满足大气污染防治的实际需要等。只有回答了这些问题,才能够真正实现对于大气污染防治的有效防治。

基于城市计算技术的U-Air系统在该领域取得了初步的研究进展^{错误!未找到引用源。1}。U-Air系统对北京市22个空气监测站的PM2.5数据读数进行了分析研究,发现即便是在相同的天气条件下,距离非常接近的两个空气监测站的PM2.5数据读数依然会有数倍的差距。这意味着如果采用线性差值的方法,目前较为稀疏的监测站分布并不能完全反映整个北京市的空气质量情况。市民在得知空气质量较好的情况下,很可能会外出进入一个空气质量非常差的市区,从而引发健康问题。为了解决这一问题,U-Air系统利用

机器学习技术,使用城市的气象数据、交通数据、城市结构数据、POI数据等训练获得了城市空气质量的时空模型,并使用该模型实现了以一平方公里为单位的细粒度城市空气质量报告。此外,该模型还能够很好的度量不同城市动态因素对空气质量的影响情况。该研究成果为进一步的空气质量预报和空气污染治理等大气污染防治工作奠定了初步的基础。

● 城市经济

城市经济学是经济学的一个重要分支^[8],其研究对象是城市中各要素在社会经济系统中的相互作用关系和运行方式。以数据为中心的智慧城市技术采用全新的视角来分析城市经济学问题,与传统的经济学分析模型不同,数据挖掘、机器学习等人工智能技术在这里扮演了核心角色。

英国科研人员设计开发的Geo-Spotting系统使用机器学习的方法,利用Foursquare应用提供的LBS数据,对纽约城区的店铺地理位置与营业收益的关系进行了分析,并以此为依据帮助商户进行店铺选址^[8]。研究工作[8, 8]使用POI数据对城市的经济活动分布进行研究。此外,一些关于城市商业区分布的城市功能区域识别工作也可以列入城市经济学的范畴当中^[48]。

● 地图测绘

城市交通数据在地图的测绘方面也能够发挥非常大的作用,对于城市当中一些新修建但并未进行地图测绘的街道,可以使用车辆行驶的轨迹数据进行测量,这样的测绘方式可以极大地提升城市地图的测绘效率并有效的降低测绘成本。文献[8]率先研究了使用城市交通的GPS数据进行地图绘制的方法,研究工作[8]提出了一种能够检测双向道路路链的地图生成算法,文献[8]则进一步设计了能够对城市中立交桥交叉点进行检测地图绘制方法。

综上所述,附加行业应用的相关研究专注于智慧城市建设某一领域的特殊需求,核心任务是建立已知城市数据同城市未知特性之间的关系模型。该类研究所面临的关键挑战在于如何降低现有数据同无关特性之间的耦合程度,挖掘数据同目标应用之间的相关联系。在该类应用当中,数据所包含的城市动态信息越单一越有助于问题的求解。

3.3 城市人类活动统计力学

人类行为统计力学是统计物理学的一个重要分支,主要研究内容是使用统计的手段揭示人类行为的内在规律,采用的研究手段以复杂网络、复杂系统等物理学工具为主,并综合融入了信息科学、社会学等

多学科研究工具。该领域的研究早期由物理学家发起,近几年越来越多的受到信息科学等其他领域科研人员的关注。城市环境下的人类行为统计力学研究我们称之为城市人类活动统计力学。该类研究与智能交通、城市计算等信息学科研究的不同之处在于其更加关注揭示数据背后所蕴含的自然规律,应用色彩并不浓重。接下来,本文从城市交通网络分析和市民行为建模两个方面简要介绍相关的研究工作。

3.3.1 城市交通网络分析

与城市交通网络分析相关的统计力学研究来源于复杂网络的相关研究。1998年Watts和Strogatz在Nature杂志上发表文章提出了小世界(Small-World)网络模型^[8],该模型描述了从完全规则的网络到完全随机网络的网络转变。小世界网络既具有与规则网络类似的聚类特性,又具有与随机网络类似的较小直径。随后,1999年Barabási和Albert在Science杂志上发表文章指出许多实际的复杂网络的连接度分布都具有幂律形式,由于幂律分布没有明显的特征长度,该类网络又被称为无标度(Scale-Free)网络^{错误!未找到引用源。}¹。在两篇经典网络研究论文的推动之下,复杂网络理论开始在各个学科显现出巨大的能量,并逐渐成为了交叉学科研究的热点之一。

在城市研究领域,道路网络特别是轨道交通网络成为复杂网络理论应用的主要领域。文献[9]对美国波士顿市的交通网络进行了研究,将波士顿的地铁网络和地铁及公交复合网络各抽象成为一个无向含权图,并对其进行了分析对比。研究结果显示,单纯地铁网络的局部有效性非常低,因此不是小世界网络。然而,包含地铁和公交的网络则同时具有较高的局部有效性和全局有效应,因此属于小世界网络。

文献[9]对印度的铁路网络进行了研究,结果显示印度铁路网络具有明显的小世界效应。文献[9]对印度铁路网络中边的定义与其它研究中的定义略有不同。在其它研究中,两节点间有边存在表明这两节点被一条或多条线路连接且两点间的线路上无其它节点。而在对印度铁路网络研究中,两点相连表示从一点到另一点有一条铁路路线直达,即同一线路上各站均可视为相邻节点。两点间的最短路线即为从一点到另一点所需最少换乘次数。印度铁路网络的度分布特性表明,该网络聚集系数随网络总结点数的变化趋势较无标度网络略缓,因此不具有无标度网络的特性。文献[9]对连通的定义还被用在了随后对波士顿和维也纳轨道交通网络的研究中^[9],据此得出这两个城市交通网络均具有小世界网络特征。

文献[9]采用公共交通数据和私人交通数据对韩国的高速公路网络进行了研究。研究结果显示,公共高速公路网络为无标度网络。但与私人交通网络合并后,网络不再具备无标度网络特点,而是符合重力模型。文献[9]对韩国首尔的地铁网络进行了研究,研究显示首尔地铁边的权重分布为幂律分布,对于客流数据通过生成树模型构造出的图,其强度为对数正态分布,度分布为幂律分布。

文献[9]对新加坡公共交通网络(包含轨道交通网络和公交网络)进行了分析。结果显示,就网络拓扑特性而言,轨道交通网络更接近于随机网络。但在考虑客流网络后,网络整体呈现出无标度网络特性。此外,该研究在对网络拓扑性质研究的基础上,还利用特征向量分析方法对客流网络不同时间段表现出的特征进行了研究。研究中,客流数据被分为工作日和周末两组,即分别对工作日和周末两个客流网络进行分析,发现其均为小世界网络。

文献[9]对加拿大多伦多市的地铁网络进行了研究,作者使用图论的方法分别对多伦多地铁2010年、2025年和2035年的网络拓扑进行了分析。研究将地铁网络抽象为一个由换乘站、终点站和地铁线路组成的图。利用覆盖率,直达性和连通性这三个图论中的指标,作者对多伦多地铁网络进行了纵向比较。比较结果显示,经过15到25年之后的建设,多伦多地铁网络的规模将与其他国际城市地铁网络的规模相当,年人均乘坐数将增加2到3倍。由于年人均乘坐数通常被当作衡量地铁网络优劣的指标,因此上述数据说明多伦多地铁网络在之后25年将更加优化。

由于我国城市轨道交通发展的相对较晚,国内城市交通网络的研究目前主要集中于公交网络^[9,9] ^{错误!未找到引用源。}¹,涉及到地铁网络的只有对北京和上海地铁网络的研究。但这些研究大多局限于利用仿真模拟方法讨论网络抗毁性方面的特性等^{错误!未找到引用源。}^{10,10]}。

3.3.2 城市居民行为建模

城市居民的行为建模是人类时空动力学研究的一个子集^[10]。人类行为的时空动力学研究开始于2005年Barabási教授在Nature杂志上发表了一篇有关人类行为时间特性的论文^[10]。论文认为人类行为在时间上分布并不符合泊松过程,并提出了一种基于优先级的排队论模型,该模型显示人类活动的时间分布更加接近于幂律分布(Power Law)。随后,Brockmann等在Nature上发表了关于人类行为空间标度律的研究^[10],该研究显示人类长程旅行中表现出来空间行为的

既不同于随机游走过程又不同于Lévy飞行过程的独特性质。在文献[10, 10]的推动下,人类行为的时空动力学很快成为了统计物理学和复杂性科学研究的热点领域^[10]。

城市作为人类活动最为密集的区域之一,在该研究领域必然不会缺席。在众多城市数据当中,最先被用来分析人类时空行为的是手机数据。文献[10]使用志愿者3~6个月的短消息通信记录进行了人类活动的时间分析,发现其行为符合幂指数在1.2~1.7之间的幂律分布。文献[10]利用2006年新年期间6百万用户的3千万条短信数据进行统计分析,发送时间间隔和回复时间间隔均符合截尾幂律分布。文献[10]对来自3家公司14万用户的150万条短信记录进行了分析,并发现其符合双模分布,即前段为幂律分布,后段为指数分布。

在人类空间行为研究方面,文献[1]分析了10万名手机用户在6个月内的通话地点记录,发现这些用户的移动距离服从带有指数尾的幂律分布。进一步,文献[1]使用同样的手机通话地点数据对人类空间活动模式的可预测性进行了分析。分析结果显示,样本数据中的用户活动在空间上具有93%的可预测性,且不同个体的可预测性差异并不大。

文献[11]使用瑞典4个城市中50辆出租车六个月的GPS轨迹数据,对出租车乘客的出行行为进行了统计分析,研究认为出租车乘客的乘车距离服从双段的幂律分布。北航的K. Xue团队^[11]使用了北京市近两万辆出租车3个月内的GPS轨迹数据,对出租车乘客的乘车距离特征进行了更加精确的统计测算。其结果发现,出租车乘客的乘车距离与时间长度均服从指数分布,而非文献[11]所认为的幂律分布。A. Bazzani教授的研究小组使用意大利罗马、博洛尼亚、西尼加利亚和佛罗伦萨等城市的私家车GPS数据,对私家车主驾车出行的行为特征进行了统计分析^[11, 11, 11, 11]。统计结果显示这三个城市的车辆行驶里程都近似服从指数分布,验证了文献[118]的统计结论。

文献[11]使用伦敦地铁中200万名乘客在一周时间内的IC卡刷卡数据,对于乘客的乘车距离进行了统计,研究发现地铁乘客的乘车距离更加近似于负二项式分布,而非通常认为的幂律分布或是类似于出租车乘客/私家车主出行距离的指数分布。文献[11]对230名志愿者填写的日常出行日志数据进行了研究,该研究的一大亮点在于对个体行为的空间特性进行了分析。研究结果显示,在混合交通情况下,交通费用通常与出行距离通常具有线性关系而出行时间与距离则具有“对数+线性”关系,出行距离分布近似为带有

指数截断的幂函数律;而在单一交通工具下,出行时间和距离是线性关系。

2012年Barabási教授在Nature杂志发表的一篇文章提出了一种关于城市间市民通勤活动范围的“辐射”模型,该模型相比交通领域中传统的引力模型更加准确且没有参数^[1]。该文章很快引起了学术界对城市交通通勤行为的关注。人们对于城市的人类活动行为的研究重点也开始从出行距离的标度律研究扩展到了城市/城区之间的人口移动流量研究。北航K. Xue团队基于多个城市的出租车与出行数据分析,提出了一种新的群体移动模型,能够模拟和预测城市内不同区域间的人口移动的流量。基于这一模型的理论分析进一步表明造成出租车出行距离呈指数分布的根本原因在于城市内人口密度呈指数衰减^[1]。针对城市内通勤模式与城市间通勤模式的不同特性,北京师范大学的W. Wang教授团队提出了一种conduction-like stochastic process模型来进一步准确刻画城市内人口流动的通勤特性^[12]。

综合上述研究可以看出,利用来源丰富的城市多源数据以及强大的统计物理工具,我们可以深刻的理解城市道路、交通结构以及人类活动行为的本质特性。相对于基于机器学习、数据挖掘的等计算机技术的城市计算等研究,该类研究领域的研究结论具有很强的可解释性。在应用方面,该领域的文献虽然很少介绍具体的应用方式,但是我们可以看出,相关的理论研究成果可以很好的为智慧城市中的交通规划建设、定制推荐、传染病防控等应用服务。

4 核心技术与理论

● 数据感知技术

以数据为中心的智慧城市技术从顶层设计到具体的技术细节,均是以海量的城市数据作为核心支撑的。因此城市数据的感知与获取技术就成为了研究工作的整体技术基础。传统的城市传感器技术在为不同的行业应用提供服务的同时,也直接或间接地收集到了大量的城市动态感知数据,这些数据使得现有的研究工作成为了可能。与此同时,随着移动互联网的兴起,人们可以随时随地的记录和分享自己的所见所闻,使得“以人为传感器”对城市进行感知的群体感知技术开始日渐兴起。例如研究工作^[12]使用微博数据对2012年北京市7.21暴雨的积水点进行了检测,其结果与实际积水地点基本相符。目前,群体感知技术在智慧城市应用当中还只是传统数据获取方式的一种补充,相信在不久的将来,群体感知技术将会逐步走向

前台，成为智慧城市应用的一项主要数据来源。

● 数据管理技术

海量城市数据的存储、管理与检索是一项非常具有挑战性的工作，已有的一些研究工作显示，分布式数据库、Hadoop等现有的数据管理方式尚无法完全满足智慧城市应用对于海量数据查询与管理的实时性和灵活性要求^[12]。造成该问题的根本原因在于，用于智慧城市应用的海量城市数据并非专门为建设智慧城市而收集（多为城市信息化和数据化过程中的信息基础设施），数据管理系统的设计者更多考虑的是其初始的数据应用需求，而没有对智慧城市应用做任何数据管理优化。同时城市计算等智慧城市技术在数据利用方面又有着异常广阔的想象空间，即便是智慧城市研究的专家也无法列举所有可能的应用模式。这就导致数据管理系统的设计者在开发系统时无法获知系统应用者的具体应用需求，从而引起数据管理系统与应用系统之间的不匹配，影响系统的整体性能。如何解决这一问题目前学术界和工业界尚并没有非常成熟的技术方案和支撑理论。

● 数据挖掘技术

数据挖掘技术是大数据时代进行数据利用和知识发现的另一项核心关键技术，构建以数据为中心的智慧城市也必然需要数据挖掘技术的大力支持。该领域一个主要的挑战在于如何在海量多源的城市数据当中寻找合适具体应用的数据子集。例如，城市动态检测研究往往需要所使用的城市数据子集尽可能多的涵盖城市动态特性的各个方面，而具体的附加行业应用则要求数据所包含的信息尽可能的单一纯粹。这些互斥性的数据需求给城市数据的收集、管理和挖掘都带来了巨大挑战。解决好数据子集的选择问题与子集间不同数据的融合问题，是在智慧城市领域进行数据挖掘研究的关键所在。

● 数据活化技术

数据活化（Data Vitalization）概念是由北京航空航天大学熊璋教授所提出的一种全新的数据管理与应用模式^[12]。数据活化也就是赋予数据生命的意思，数据活化技术的核心思想在于将真实物理世界中的数据内在联系映射到数据存储与管理的数字空间中，使存储空间中相互隔离的数据变为一个有机的整体，恢复数据在物理世界当中的关联性，并突破信息空间在进行数据利用时的局限性。数据活化体系结构采用“活化细胞”（Vitalized cell）的对数据进行组织。活化细胞作为数据组织和管理的最基本单元，一方面具有对数据的存储能力，另一方面也具有映射物理空间中数据联系与相互作用的计算能力。活化细胞在存储数



图 7. Data revitalization 解决方案架构^[127]

Fig 7. The framework of Data revitalization Solution^[127]

据的同时，会根据实际应用的不同，不断地学习用户应用行为并对细胞中的数据进行重组，使数据可以更好的适应多样化的用户需求。同时，当物理世界中数据描述对象发生变化时，活化细胞还可以通过自主演化的方式来改变数据的储存结构和内容，从而实现存储数据的代谢与演化过程。

数据活化技术非常适合追踪物理世界中的对象实体在数据空间中的映射演进过程，是处理海量城市数据、构建智慧城市的有力工具。例如，研究工作^[12]使用数据活化的思想设计了一种海量城市数据联网的数据组织工具IOD (The Internet of Data)，该工具可以对现有物联网数据的活化组织，非常适合城市中海量物联网与传感器数据的管理。研究工作^[12]将数据活化思想直接应用到了智慧城市当中，设计了一种面向城市数据的数据活化体系架构。研究工作^[12]使用数据活化技术对监控视频数据进行结构化的描述，该研究的成果可以用以解决城市视频监控数据的高维度和非结构化问题。除此之外，数据活化技术也受到了工业界的密切关注，如图7所示，美国Bardess Group Ltd. 公司基于数据活化技术开发了一种 Data Revitalization的数据管理解决方案^[12]。该方案的数据管理架构共有五层组成，并且具有更新、学习和演进的能力，使用该架构可以有效的提高企业数据资产的整体管理质量。

数据活化技术已经在智慧城市、视频数据分析、企业数据资产管理等数据密集性应用领域初步展示了其技术优势。使用数据活化技术对城市数据的组织与管理方式进行重构，从数据的底层结构开始实现数据的智慧化，将会是未来智慧城市技术研究的一个重要的发展方向。

● 数据可视化技术

智慧城市技术是一种与普通民众生活紧密结合的应用技术，友好亲切的服务呈现方式是其中必不可少的环节。数据可视化技术能够将复杂纷繁的城市数据以一种简洁有序的方式呈献给用户，可以很好的填补技术到用户之间的理解鸿沟。在城市数据应用的推动下，数据可视化必然会成为未来智慧城市技术体系

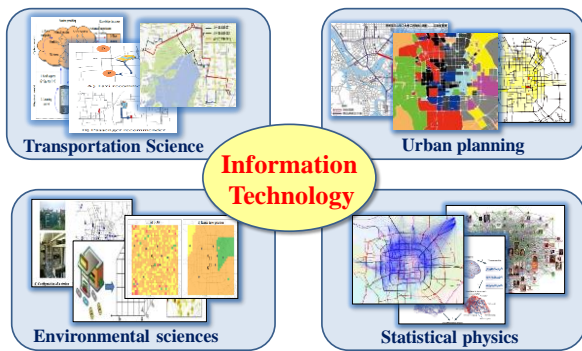


图 9. 以数据为中心的智慧城市当中各学科关系

Fig 8. Data-centric Smart City related disciplines

当中至关重要的一环。

● 统计物理学理论

物理学是研究物质世界最基本结构、最普遍相互作用、最一般运动规律及所使用实验手段和思维方法的自然科学。基础物理理论对于工业应用技术的影响往往是变革性的。17世纪牛顿经典力学引发的工业革命、20世纪量子力学引发的新技术革命都是非常好的证明。进入21世纪之后，人类活动行为的基础理论研究成为了统计物理学的一个新兴研究热点，复杂网络理论、人类活动时空动力学等研究成果不断涌现，这些成果必然会引发城市技术发展的巨大变革。在这样的大趋势下，谁能够更好的利用统计物理学的相关研究成果，谁就能够在日趋激烈的智慧城市技术竞争中占领先机。

5 领域研究特点

5.1 多学科交叉

以数据为中心的智慧城市是一个多学科，多领域相互交融的交叉学科。该领域的工作涉及到信息科学、交通科学、环境科学、规划与建筑学、统计物理学以及社会科学等多个不同领域的基本理论与技术工具。图8给出了不同学科在以数据为中心的智慧城市技术研究中所扮演的角色地位。从图中可以看出，由于数据收集、管理、分析、处理与应用均需要依赖于信息技术，因此信息科学在整个领域当中起到了核心的主导作用。同时，与城市生活运行密切相关的各项专业学科知识作为信息科学的在某一专业领域的知识补充，极大的扩展了信息技术所能够涵盖的应用范围。在对城市数据进行应用时，某一专门学科的专业知识可以帮助信息技术建立所需的数据分析与处理模型。在整个以数据为中心的智慧城市研究领域当中，信息技术即是多学科进行交叉合作的关联核心，也是支撑

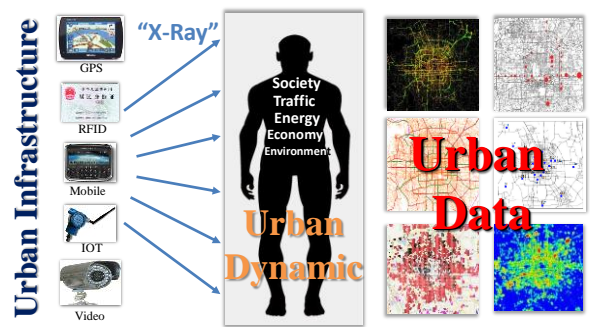


图 8. 智慧城市数据应用模式的 X-Ray 模型

Fig 9. A X-Ray of Smart City applications patterns

各个学科发挥其专业优势的公共技术平台。如果没有信息技术作为衔接，专业学科知识很难在完全不相关的领域当中建立起必要的联系。同时，如果没有专业学科知识的帮助，信息技术所能发挥的作用也只能停留在信息领域当中，无法发挥出其应有的强大作用。

因此在智慧城市领域开展数据研究时，技术思路需要充分考虑其学科交叉的特点，注意吸收和使用多学科的知识、经验与技术。一方面要使用信息技术的视角，将信息科学的知识与核心思想应用到不同的学科领域当中，解决智慧城市建设过程中的不同领域的应用问题。另一方面也需要多多借鉴其他学科的问题分析与解决模式，将其融合到信息技术领域，从而丰富和扩展信息技术的内涵与外延。

5.2 X-Ray 模型

为了理清以数据为中心的智慧城市研究中，城市信息基础设施、城市运行动态以及城市数据所扮演的不同角色及其相关关系，我们用一个“X-Ray模型”来对城市数据应用的总体模式进行说明。如图9所示，我们将想要了解的城市运行动态比作一个需要检查身体的病人，病人身体中的神经、血管、肌肉、骨骼、内脏等器官分别代表了城市当中的交通、人口、环境、能源、经济、物流等深层次的动态韵律。如果想要了解一个城市的健康状况，我们必须对这些深层次的动态进行感知。最直接的方式是安装传感设备对我们想要了解的城市动态行为进行直接监测，但是由于成本、技术、隐私等问题，直接监测往往难以实现，我们只能使用从现有信息基础设施中获得的的外围数据来对城市进行分析。这就好比医生不能总是通过穿刺、手术等有损的方式来了解病人的身体状况一样，我们需要通过一些类似X光的医学影像技术来了解病人。例如，车辆的GPS导航定位数据和手机通信的基站位置数据虽然并非了研究城市中人类活动行为而收集，但是却可以在不侵犯用户隐私和引起个体反感的情

况下，对于人类日常活动的行为方式进行研究。这些城市数据就像X光射线一样，无损的透过城市的肌体将城市内大量有用的信息呈现给了科研人员。

智慧城市研究中进行分析数据的研究人员，需要像医生将X片影像同疾病建立起联系一样，将不同的城市数据同所要分析的具体城市动态特性建立起联系。在这一过程中，丰富的数据来源同精湛的分析技术都是必不可少的因素。充足的数据来源可以极大降低问题求解的难度，但是最终结果的可靠性还是需要依赖于研究人员对分析算法的设计和选取。就像综合采用CT、核磁、超声等复合医学影像手段可以更好的帮助医生诊断病情，但最终的治疗效果依然取决于医生凭借专业知识与个人经验而给出的系统治疗方案。智慧城市领域的研究人员需要综合运用信息科学、物理学、人类学、社会学知识等多学科知识，系统分析所能获得的多种城市动态数据，对智慧城市的建设中所存在的问题做出精准判断，然后因地制宜、对症下药的解决所面临的问题。这就是以数据科学在智慧城市领域进行应用的学科特点，也是以数据为中心的智慧城市技术的魅力所在。

6 研究展望

学术界和工业界围绕着以数据为中心的智慧城市这一主题展开了多层次、多视角、多方位的理论与技术探索，并取得了令人瞩目的研究成果。然而，构建和谐宜居的智慧城市系统，实现人与环境、人与城市和谐发展的智慧城市愿景，依然存在有诸多挑战，这些挑战需要多学科、多领域的研究人员们相互配合、凝聚智慧、共同面对。在这里，我们对未来在以数据为中心的智慧城市领域可能有所突破的研究方向和臻待突破的研究问题进行展望。

● 多源城市数据的紧耦合

城市是一个复杂的、庞大的复杂动态系统，其任何一方面的动态特性都会在不同的数据空间中有所体现。现有的智慧城市研究工作往往只能利用到城市数据空间中的一两种数据，即使是多源数据融合的研究工作也往往是某一类数据为核心，加之地图\POI等城市GIS锚定数据做支撑，例如：出租车GPS数据+地图\POI数据等。多源数据的耦合度并不高，利用模式也相对单一。如何充分挖掘城市核心数据的关联性，实现多源数据之间的紧耦合，将会是未来智慧城市数据应用技术发展的必然趋势。

● 信息世界-物理世界的交互

开展智慧城市技术研究的根本目的是为了服务

城市与市民，即将信息空间的智慧应用到真实的物理世界当中。同时，应用了智慧城市新技术的物理世界也会发生变化，并在信息空间中有所反映，研究人员还需要根据这些反馈进一步的改进技术。这种信息世界同物理世界的交互过程将不断的迭代进行，并且逐步实现自主演化，最终构成一个具有自我改进能力的反馈系统。目前业界还没有太多相关工作的报道，但相信不久的将来，物理世界与信息世界的交互研究将会成为智慧城市研究的一个重要方向。

● 城市深层次数据的利用

城市的运行过程涵盖了交通、经济、社会、物流等多层次、多方面的社会活动。现有研究工作所采用的地图数据、交通数据、通讯数据、环境数据等都只是反映了城市活动的物理与信息接触行为，并且均是以地理位置坐标进行锚定标注的。对于一些深层次的城市逻辑行为，例如资金流动、政策导向、社会心理、流行时尚等，现有研究工作都少有涉及。开发利用城市深层次数据，挖掘城市运行的内在逻辑关系，将会是智慧城市技术发展的未来方向之一。

● 城市行为动力学理论体系

城市行为动力学是理解城市内在行为与市民活动规律的基础理论，其在整个智慧城市研究的理论体系当中扮演者非常重要的角色。目前的城市行为动力学理论研究，还只是依附于人类活动时空动力学的一个统计物理学分支，尚没有形成完整的理论体系，具体内涵与外延边界也尚不清晰。因此，未来完善和发展系统的城市行为动力学理论，构建完整独立的理论体系，将会是智慧城市基础理论研究的核心任务。

7 结束语

构建智慧城市是未来现代化城市的发展方向，以数据为中心的智慧城市技术是支撑智慧城市建设的技術基础，相关研究具有很重要的在理论价值与现实意义。我们可以看出，随着城市信息基础设施建设的日渐成熟，人们对城市智慧化的需求也越来越迫切，智慧城市建设正在受到来自不同学科科研人员的密切关注。不论是在信息科学领域、城市科学领域，还是环境、能源、物理学等学科领域，以数据为中心的智慧城市都在逐步成为人们感兴趣的热点领域。尤其是在数据科学崛起的大背景之下，Data-centric Smart City 很可能会成为数据信息技术从虚拟数字空间走向真实物理世界的第一步阶梯，未来的前景与影响力不可估量。

在信息技术领域的很多方面，我国同欧美等发达

国家还具有较大的差距,大多还处于追踪国际研究前沿热点的水平,很难实现对研究方向和潮流的引领。然而,以数据为中心的智慧城市领域确极有可能成为我国学者引领国际学术潮流的一个突破点。其主要原因是我国学者在该领域当中有着得天独厚的优势。首先,中国具有非常丰富的城市资源。根据最新的统计数字显示,截至2012年我国的城市化水平已经达到52.57%^[1],全球人口超过百万的特大城市21%位于中国^[1]。丰富的城市资源为我国学者提供了充分且熟悉的研究样本。其次,随着我国国民经济的发展,我国城市的信息基础设施的建设水平也有了显著的提高,城市的信息化程度有了长足的进步。物联网、GPS终端、智能手机、市政信息化等城市数据采集设施与渠道日渐普及与畅通,这使得我国学者可以相对较容易的获得丰富的城市数据。除此之外,在提高城市管理水平、改善城市环境、治理拥堵污染等城市病、建设和谐社会、构建智慧城市方面,我国从政府到民间均有着非常迫切的需求和实际投入,这使得我国的科研人员可以得到充足的研究条件支持。上述三个方面优势使得我国科研人员在以数据为中心的智慧城市研究领域有可能率先实现理论和技术上的突破,从而引领世界智慧城市发展的潮流。

参考文献

- [1] Floating car data [OL]. Wikipedia, 2013 [2013-10-11]. http://en.wikipedia.org/wiki/Floating_car_data.
- [2] Big Data [OL]. Wikipedia, 2013 [2013-10-11]. http://en.wikipedia.org/wiki/Big_data.
- [3] Pan Gang, Qi Guande, Zhang Wangsheng, Li Shijian, Wu Zhaohui, and Yang Laurence Tianruo. Trace analysis and mining for smart cities: Issues, methods, and applications [J]. IEEE Communications Magazine, 2013, 51(6):120–126.
- [4] 郑宇. 城市计算 [J]. 中国计算机学会通讯, 2013, 9(8):6–7.
- [5] 郑宇. 城市计算与大数据 [J]. 中国计算机学会通讯, 2013, 9(8):8–18.
- [6] 《智慧城市白皮书(2012)》[R]. 863计划“智慧城市(一期)”项目总组. 2012
- [7] Liu Liang, Andris Clio, and Ratti Carlo. Uncovering cabdrivers' behavior patterns from their digital traces [J]. Computers, Environment and Urban Systems, 2010, 34(6):541–548.
- [8] Calabrese Francesco, Colonna Massimo, Lovisolo Piero, Parata Dario, and Ratti Carlo. Real-time urban monitoring using cell phones: A case study in rome [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1):141–151.
- [9] Barria Javier A and Thajchayapong Suttipong. Detection and classification of traffic anomalies using microscopic traffic variables [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(3):695–704.
- [10] 实时路况查询. [OL] 北京市交通委, 2013 [2013-10-11]. http://eye.bjjtw.gov.cn/Web-T_bjtt_new/Main.html
- [11] Gühnemann Astrid, Schäfer Ralf-Peter, Thiessenhusen Kai-Uwe, and Wagner Peter. Monitoring traffic and emissions by floating car data [R]. Institute of Transport Studies Working Paper, 2004, (ITS-WP-04-07).
- [12] Kanoulas Evangelos, Du Yang, Xia Tian, and Zhang Donghui. Finding fastest paths on a road network with speed patterns [C]. // Proceedings of the 22nd International Conference on Data Engineering. IEEE, 2006, Article10.
- [13] Pfoser Dieter, Brakatsoulas Sotiris, Brosch Petra, Umlauf Martina, Tryfona Nektaria, and Tsironis Giorgos. Dynamic travel time provision for road networks [C]. // Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM:475–478.
- [14] Zheng Yu, Chen Yukun, Li Quannan, Xie Xing, and Ma Wei-Ying. Understanding transportation modes based on GPS data for web applications [J]. ACM Transactions on the Web, 2010, 4(1):1–136.
- [15] Zheng Yu, Li Quannan, Chen Yukun, Xie Xing, and Ma Wei-Ying. Understanding mobility based on gps data [C]. // Proceedings of the 10th International Conference on Ubiquitous Computing. ACM, 2008:312–321.
- [16] Zheng Yu, Liu Like, Wang Longhao, and Xie Xing. Learning transportation mode from raw gps data for geographic applications on the web [C]. // Proceedings of the 17th International Conference on World Wide Web. ACM, 2008:247–256.
- [17] Fabritiis Corrado, Ragona Roberto, and Valenti Gaetano. Traffic estimation and prediction based on real time floating car data [C]. // Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems. IEEE, 2008:97–203.
- [18] Li Xiaolong, Pan Gang, Wu Zhaohui, Qi Guande, Li Shijian, Zhang Daqing, Zhang Wangsheng, and Wang Zonghui. Prediction of urban human mobility using large-scale taxi traces and its applications [J]. Frontiers of Computer Science, 2012, 6(1):111–121.
- [19] Castro Pablo S, Zhang Daqing, Shijian LI, et al. Urban traffic modelling and prediction using large scale taxi gps traces [C]. // Proceedings of the 10th International Conference on Pervasive Computing, volume 7319, 2012:57–72.
- [20] Gonzalez Hector, Han Jiawei, Li Xiaolei, Myslinska Margaret, and Sondag John P. Adaptive fastest path computation on a road network: a traffic mining approach [C]. // Proceedings of the 33rd International Conference on Very Large Data Bases. VLDB Endowment,

- 2007:94–805.
- [21] Ziebart Brian D, Maas Andrew L, Dey Anind K, and Bagnell Andrew. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior [C]. // Proceedings of the 10th International Conference on Ubiquitous Computing. ACM, 2008:322–331.
- [22] Yuan Jing, Zheng Yu, Xie Xing, and Sun Guangzhong. Driving with knowledge from the physical world [C]. // Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:316–324.
- [23] Yuan Jing, Zheng Yu, Xie Xing, and Sun Guangzhong. T-drive: Enhancing driving directions with taxi drivers' intelligence [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(1):220–232.
- [24] Zheng Kai, Zheng Yu, Yuan Nicholas Jing, and Shang Shuo. On discovery of gathering patterns from trajectories[M]. IEEE Computer Society Press, 2013.
- [25] Li Bin, Zhang Daqing, Sun Lin, Chen Chao, Li Shijian, Qi Guande, and Yang Qiang. Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset [C]. // Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops. IEEE, 2011:63–68.
- [26] Sun Lin, Zhang Daqing, Chen Chao, Castro Pablo Samuel, Li Shijian, and Wang Zonghui. Real time anomalous trajectory detection and analysis [J]. Mobile Networks and Applications, 2013, 18(3):1–16.
- [27] Ge Yong, Liu Chuanren, Xiong Hui, and Chen Jian. A taxi business intelligence system [C]. // Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:735–738.
- [28] Zheng Xudong, Liang Xiao, and Xu Ke. Where to wait for a taxi? [C]. // Proceedings of the ACM SIGKDD International Workshop on Urban Computing. ACM, 2012:149–156.
- [29] Yuan Jing, Zheng Yu, Zhang Liuhang, Xie Xing, and Sun Guangzhong. Where to find my next passenger [C]. // Proceedings of the 13th International Conference on Ubiquitous Computing, ACM, 2011: 109–118.
- [30] Yuan Nicholas Jing, Xie Xing, Zheng Yu, and Zhang Liuhang. T-Finder: A recommender system for finding passengers and vacant taxis [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 99(1):1–1.
- [31] Harsanyi John C. and Selten Reinhard. A General Theory of Equilibrium Selection in Games, volume 1 [M]. The MIT Press, 1st edition, 1988.
- [32] Chen Chao, Zhang Daqing, Samuel Castro P, Li Nan, Sun Lin, and Li Shijian. Real-time detection of anomalous taxi trajectories from GPS traces [J]. Mobile and Ubiquitous Systems: Computing, Networking, and Services, 2012:63–74.
- [33] Chen Chao, Zhang Daqing, Castro Pablo S, Li Nan, Sun Lin, Li Shijian, Wang Zonghui, et al. IBOAT: isolation-based online anomalous trajectory detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2):806–818.
- [34] Zhang Daqing, Li Nan, Zhou Zhi-Hua, Chen Chao, Sun Lin, and Li Shijian. ibat: detecting anomalous taxi trajectories from gps traces [C]. // Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, 2011:99–108.
- [35] 张大庆, 陈超, 杨丁奇, and 熊昊一. 从数字脚印到城市计算 [J]. 中国计算机学会通讯, 2013, 9(8):17 – 24.
- [36] Chen Chao, Zhang Daqing, Zhou Zhi-Hua, Li Nan, Atmaca Tülin, and Li Shijian. B-planner: Night bus route planning using large-scale taxi gps traces [C]. // IEEE International Conference on Pervasive Computing and Communications, volume 18, 2013:225–233.
- [37] Ma Shuo, Zheng Yu, and Wolfson Ouri. T-Share: A large-scale dynamic taxi ridesharing service[C]. // Proceedings of the 29th IEEE International Conference on Data Engineering , IEEE, 2013:1-12.
- [38] Bastani Favyen, Huang Yan, Xie Xing, and Powell Jason W. A greener transportation mode: flexible routes discovery from gps trajectory data [C]. // Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2011:405–408.
- [39] Leng Biao, Zeng Jiabei, Xiong Zhang, Lv Weifeng, and Wan Yueliang. Probability tree based passenger flow prediction and its application to the beijing subway system [J]. Frontiers of Computer Science, 2013, 7(2):1–9.
- [40] Ceapa Irina, Smith Chris, and Capra Licia. Avoiding the crowds: understanding tube station congestion patterns from trip data [C]. // Proceedings of the ACM SIGKDD International Workshop on Urban Computing. ACM, 2012:134–141.
- [41] Balan Rajesh K, Nguyen Khoa X, and Jiang Lingxiao. Real-time trip information service for a large taxi fleet [C]. // Proceedings of the 9th International Conference on Mobile systems, Applications, and Services. ACM, 2011:99–112.
- [42] Lathia Neal and Capra Licia. Mining mobility data to minimise travellers' spending on public transport [C]. // Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:1181–1189.
- [43] Hao Jiuyue, Li Chao, Kim Zuwhan, and Xiong Zhang. Spatio-temporal traffic scene modeling for object motion detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1):295–302.
- [44] Kindberg Tim, Chalmers Matthew, and Paulos Eric. Guest editors' introduction: Urban computing [J]. IEEE Pervasive Computing, 2007, 6(3):18–20.
- [45] Shklovski Irina and Chang Michele F. Guest editors' introduction: Urban computing–navigating space and context [J]. Computer, 2006,

- 39(9):36–37.
- [46] Zheng Yu, Liu Yanchi, Yuan Jing, and Xie Xing. Urban computing with taxicabs [C]. // Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, 2011:89–98.
- [47] Liu Siyuan, Liu Yunhuai, Ni Lionel M, Fan Jianping, and Li Minglu. Towards mobility-based clustering [C]. // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:919–928.
- [48] Yuan Jing, Zheng Yu, and Xie Xing. Discovering regions of different functions in a city using human mobility and pois [C]. // Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:186–194.
- [49] Stefanov William L, Ramsey Michael S, and Christensen Philip R. Monitoring urban land cover change: an expert system approach to land cover classification of semiarid to arid urban centers [J]. Remote Sensing of Environment, 2001, 77(2):173–185.
- [50] Chen D, Stow DA, and Gong P. Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case [J]. International Journal of Remote Sensing, 2004, 25(11):2177–2192.
- [51] Pozdnoukhov Alexei and Kaiser Christian. Space-time dynamics of topics in streaming text [C]. // Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. ACM, 2011 8:1–8:8.
- [52] Qi Guande, Li Xiaolong, Li Shijian, Pan Gang, Wang Zonghui, and Zhang Daqing. Measuring social functions of city regions from large-scale taxi behaviors [C]. // Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops. IEEE, 2011:384–388.
- [53] Veloso Marco, Phithakkitnukoon Santi, and Bento Carlos. Urban mobility study using taxi traces [C]. // Proceedings of the 2011 international workshop on Trajectory data mining and analysis. ACM, 2011:23–30.
- [54] Pang Linsey Xiaolin, Chawla Sanjay, Liu Wei, and Zheng Yu. On mining anomalous patterns in road traffic streams [J]. Lecture Notes in Computer Science, 2011, 7121:237–251.
- [55] Pang Linsey Xiaolin, Chawla Sanjay, Liu Wei, and Zheng Yu. On detection of emerging anomalous traffic patterns using GPS data [J]. Data & Knowledge Engineering, 2013, 87(9):357–373.
- [56] Pan Bei, Zheng Yu, Wilkie David, and Shahabi Cyrus. Crowd sensing of traffic anomalies based on human mobility and social media [C]. // 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2013:1–12.
- [57] Chawla Sanjay, Zheng Yu, and Hu Jiafeng. Inferring the root cause in road traffic anomalies [C]. // Proceedings of the 12th IEEE International Conference on Data Mining. IEEE, 2012:141–150.
- [58] Yue Yang, Wang Han-dong, Hu Bo, Li Qing-quan, Li Yu-guang, and Yeh Anthony GO. Exploratory calibration of a spatial interaction model using taxi gps trajectories [J]. Computers, Environment and Urban Systems, 2012, 36(2):140–153.
- [59] Peng Chengbin, Jin Xiaogang, Wong Ka-Chun, Shi Meixia, and Li òPietro. Collective human mobility pattern from taxi trips in urban area [J]. PloS one, 2012, 7(4):e34487.
- [60] Wang Zuchao, Lu Min, Yuan Xiaoru, Zhang Junping, and Wetering Huub van de. Visual traffic jam analysis based on trajectory data [J]. Visualization and Computer Graphics, IEEE Transactions on, 2013, 19(12):2159–2168.
- [61] Puissant Anne, Hirsch Jacky, and Weber Christiane. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery [J]. International Journal of Remote Sensing, 2005, 26(4):733–745.
- [62] Seto Karen C and Fragkias Michail. Quantifying spatiotemporal patterns of urban land-use change in four cities of china with time series landscape metrics [J]. Landscape ecology, 2005, 20(7):871–888.
- [63] Carleer AP and Wolff El ònore. Urban land cover multi-level region-based classification of vhr data by selecting relevant features [J]. International Journal of Remote Sensing, 2006, 27(6):1035–1051.
- [64] Deng Jin S, Wang Ke, Hong Yang, and Qi Jia G. Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization [J]. Landscape and Urban Planning, 2009, 92(3):187–198.
- [65] Lu Dengsheng and Weng Qihao. Urban classification using full spectral information of landsat etm+ imagery in marion county, indiana [J]. Photogrammetric Engineering and Remote Sensing, 2005, 71(11):1275–1284.
- [66] Herold Martin, Liu XiaoHang, and Clarke Keith C. Spatial metrics and image texture for mapping urban land use [J]. Photogrammetric Engineering and Remote Sensing, 2003, 69(9):991–1002.
- [67] Voorde Tim Van de, Jacquet Wolfgang, and Canters Frank. Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data [J]. Landscape and Urban Planning, 2011, 102(3):143–155.
- [68] Zhang Wangsheng, Li Shijian, and Pan Gang. Mining the semantics of origin-destination flows using taxi traces [C]. // Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM, 2012:943–949.
- [69] Pan Gang, Qi Guande, Wu Zhaohui, Zhang Daqing, Li Shijian, et al. Land-use classification using taxi GPS traces [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1):113–123.
- [70] Soto Vctor and Fr òs-martnez Enrique. Robust land use characterization of urban landscapes using cell phone data [C]. // Proceedings of the 2011

- Workshop on Pervasive Urban Applications. Citeseer, 2011:1–8.
- [71] Pacifici Fabio, Chini Marco, and Emery William J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification [J]. *Remote Sensing of Environment*, 2009, 113(6):1276–1292.
- [72] Luck Matthew and Wu Jianguo. A gradient analysis of urban landscape pattern: a case study from the phoenix metropolitan region, arizona, usa [J]. *Landscape ecology*, 2002, 17(4):327–339.
- [73] 冷彪 和 赵文远. 基于客流数据的区域出行特征聚类 [J]. to appeared in *计算机研究与发展*, 2014.
- [74] Bram Jason and McKay Alisdair. The evolution of commuting patterns in the new york city metro area [J]. *Current Issues in Economics and Finance*, 2005, 11(10):1–7.
- [75] Zhang Fuzheng, Wilkie David, Zheng Yu, and Xie Xing. Sensing the pulse of urban refueling behavior [C]. // *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2013:13–22.
- [76] Momtazpour Marjan, Butler Patrick, Hossain M Shahriar, Bozchalui Mohammad C, Ramakrishnan Naren, and Sharma Ratnesh. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles [C]. // *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. ACM, 2012:126–133.
- [77] Tulusan Johannes, Staake Thorsten, and Fleisch Elgar. Providing eco-driving feedback to corporate car drivers: what impact does a smartphone application have on their fuel efficiency? [C]. // *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012:212–215.
- [78] 国务院关于印发大气污染防治行动计划的通知[EB/OL]. 中华人民共和国国务院, 2013[2013-10-11]. http://www.gov.cn/jzwgk/2013-09/12/content_2486773.htm
- [79] Zheng Yu, Liu Furui, and Hsieh Hsun-Ping. U-air: When urban air quality inference meets big data [C]. // *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013:1436–1444.
- [80] O’sullivan Arthur. *Urban Economics*. McGraw-Hill/Irwin, 2007.
- [81] Karamshuk Dmytro, Noulas Anastasios, Scellato Salvatore, Nicosia Vincenzo, and Mascolo Cecilia. Geo-spotting: Mining online location-based services for optimal retail store placement [C]. // *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013:793–801.
- [82] Porta Sergio, Latora Vito, Wang Fahui, Rueda Salvador, Strano Emanuele, Scellato Salvatore, Cardillo Alessio, Belli Eugenio, Cárdenas Francisco, Cormenzana Berta, et al. Street centrality and the location of economic activities in barcelona [J]. *Urban Studies*, 2012, 49(7):1471–1488.
- [83] Porta Sergio, Latora Vito, Wang Fahui, Strano Emanuele, Cardillo Alessio, Scellato Salvatore, Iacoviello Valentino, and Messori Roberto. Street centrality and densities of retail and services in bologna, italy [J]. *Environment and Planning B: Planning and Design*, 2009, 36(3):450–465.
- [84] Davics JJ, Beresford Alastair R, and Hopper Andy. Scalable, distributed, real-time map generation [J]. *IEEE Pervasive Computing*, 2006, 5(4):47–54.
- [85] Cao Lili and Krumm John. From GPS traces to a routable road map [C]. // *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2009:3–12.
- [86] Fathi Alireza and Krumm John. Detecting road intersections from GPS traces [J]. *Lecture Notes in Computer Science*, 2010, 6292:56–69.
- [87] Watts Duncan J and Strogatz Steven H. Collective dynamics of ‘small-world’ networks [J]. *Nature*, 1998, 393(6684):440–442.
- [88] Barabási Albert-László and Albert Réka. Emergence of scaling in random networks [J]. *Science*, 1999, 286(5439):509–512.
- [90] Sen Parongama, Dasgupta Subinay, Chatterjee Arnab, Sreeram PA, Mukherjee G, and Manna SS. Small-world properties of the indian railway network [J]. *Physical Review E*, 2003, 67(3):036106.
- [91] Seaton Katherine A and Hackett Lisa M. Stations, trains and small-world networks [J]. *Physica A: Statistical Mechanics and its Applications*, 2004, 339(3):635–644.
- [92] Jung Woo-Sung, Wang Fengzhong, and Stanley Eugene. Gravity model in the Korean highway [J]. *Europhysics Letters*, 2008, 81(4):48005.
- [93] Lee Keumsook, Jung Woo-Sung, Park Jong Soo, and Choi MY. Statistical analysis of the metropolitan seoul subway system: Network structure and passenger flows [J]. *Physica A: Statistical Mechanics and its Applications*, 2008, 387(24):6231–6234.
- [94] Soh Harold, Lim Sonja, Zhang Tianyou, Fu Xiuju, Lee Gary Kee Khoon, Hung Terence Gih Guang, Di Pan, Prakasam Silvester, and Wong Limsoon. Weighted complex network analysis of travel routes on the singapore public transportation system [J]. *Physica A: Statistical Mechanics and its Applications*, 2010, 389(24):5852–5863.
- [95] Derrible Sybil and Kennedy Christopher. Evaluating, comparing, and improving metro networks [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2010, 2146(1):43–51.
- [96] 李英, 周伟, and 郭世进. 上海公共交通网络复杂性分析 [J]. *系统工程*, 2007, 25(1):38 – 41.
- [97] 赵金止, 狄增如, and 王大辉. 北京市公共汽车交通网络几何性质的实证研究 [J]. *复杂系统与复杂性科学*, 2005, 2(2):45 – 48.
- [98] Lu Huapu and Shi Ye. Complexity of public transport networks [J].

- Tsinghua Science & Technology, 2007, 12(2):204–213.
- [99] 赖丽萍. 城市轨道交通网络鲁棒性研究[D]. 北京:北京交通大学, 2012.
- [100] Zhang Jianhua, Xu Xiaoming, Hong Liu, Wang Shuliang, and Fei Qi. Networked analysis of the shanghai subway network, in china [J]. Physica A: Statistical Mechanics and its Applications, 2011, 390(23):4562–4570.
- [101] Cajueiro Daniel O. Optimal navigation for characterizing the role of the nodes in complex networks [J]. Physica A: Statistical Mechanics and its Applications, 2010, 389(9):1945–1954.
- [102] 周涛, 韩筱璞, 闫小勇, 杨紫陌, 赵志丹, and 汪秉宏. 人类行为时空特性的统计力学 [J]. 电子科技大学学报, 2013, 42(4):481-540.
- [103] Barabasi Albert-Laszlo. The origin of bursts and heavy tails in human dynamics [J]. Nature, 2005, 435(7039):207–211.
- [104] Brockmann Dirk, Hufnagel Lars, and Geisel Theo. The scaling laws of human travel [J]. Nature, 2006, 439(7075):462–465.
- [105] Hong Wei, Pan Xiao-Pu, Zhou Tao, and Wang Bing-Hong. Heavy-tailed statistics in short-message communication [J]. Chinese Physics Letters, 2009, 26(2):028902.
- [106] Zhao Zhi-Dan, Xia Hu, Shan Ming-Sheng, and Zhou Tao. Empirical analysis on the human dynamics of a large-scale short message communication system [J]. Chinese Physics Letters, 2011, 28(6):068901.
- [107] Wu Ye, Zhou Changsong, Xiao Jinghua, Kurths Jürgen, and Schellnhuber Hans J. Evidence for a bimodal distribution in human communication [J]. Proceedings of the National Academy of Sciences, 2010, 107(44):18803–18808.
- [108] Gonzalez Marta C, Hidalgo Cesar A, and Barabasi Albert-Laszlo. Understanding individual human mobility patterns [J]. Nature, 2008, 453(7196):779–782.
- [109] Song Chaoming, Qu Zehui, Blumm Nicholas, and Barabási Albert-László. Limits of predictability in human mobility [J]. Science, 2010, 327(5968):1018–1021.
- [110] Jiang Bin, Yin Junjun, and Zhao Sijian. Characterizing the human mobility pattern in a large street network [J]. Physical Review E, 2009, 80(2):021136.
- [111] Liang Xiao, Zheng Xudong, Lv Weifeng, Zhu Tongyu, and Xu Ke. The scaling of human mobility by taxis is exponential [J]. Physica A: Statistical Mechanics and its Applications, 2012, 391(5):2135–2144.
- [112] Rambaldi Sandro, Bazzani Armando, Giorgini Bruno, and Giovannini Luca. Mobility in modern cities: looking for physical laws [C]. // Proceedings of the ECCS, volume 7, 2007:132–141.
- [113] Bazzani Armando, Giorgini Bruno, Rambaldi Sandro, Gallotti Riccardo, and Giovannini Luca. Statistical laws in urban mobility from microscopic gps data in the area of florence [J]. Journal of Statistical Mechanics: Theory and Experiment, 2010, 2010(5):P05001.
- [114] Riccardo Gallotti, Armando Bazzani, and Sandro Rambaldi. Towards a statistical physics of human mobility [J]. International Journal of Modern Physics C, 23(9):Article6.
- [115] Gallotti Riccardo, Bazzani Armando, Esposti Mirko Degli, and Rambaldi Sandro. Entropic measures of individual mobility patterns [J]. arXiv preprint arXiv:1305.1836, 2013.
- [116] Roth Camille, Kang Soong Moon, Batty Michael, and Barthélemy Marc. Structure of urban movements: polycentric activity and entangled hierarchical flows [J]. PLoS One, 2011, 6(1):e15923.
- [117] Yan Xiao-Yong, Han Xiao-Pu, Wang Bing-Hong, and Zhou Tao. Diversity of individual mobility patterns [J]. arXiv preprint arXiv:1211.2874, 2012.
- [118] Simini Filippo, González Marta C, Maritan Amos, and Barabási Albert-László. A universal model for mobility and migration patterns [J]. Nature, 2012, 484(7392):96–100.
- [119] Liang Xiao, Zhao Jichang, Dong Li, and Xu Ke. Unraveling the origin of exponential law in intra-urban human mobility [J]. arXiv preprint arXiv:1305.6364, 2013.
- [120] Yan Xiao-Yong, Zhao Chen, Fan Ying, Di Zengru, and Wang Wen-Xu. Universal predictability of mobility patterns in cities [J]. arXiv preprint arXiv:1307.7502, 2013.
- [121] 袁晓如. Thememap[OL]. 北京大学 PKUVIS, 2013[2013-10-11]. <http://vis.pku.edu.cn/weibova/thememap>.
- [122] 倪明选 和 罗吴蔓. 数据爆炸时代的技术变革 [J]. 中国计算机学会通讯 数据密集型计算专题, 2011, 7(7):12 – 20.
- [123] Xiong Zhang, Luo Wuman, Chen Lei, and Ni Lionel M. Data vitalization: a new paradigm for large-scale dataset analysis [C]. // Proceedings of the 16th IEEE International Conference on Parallel and Distributed Systems. IEEE, 2010:251–258.
- [124] Fan Wei, Chen Zhengyong, Xiong Zhang, and Chen Hui. The internet of data: a new idea to extend the iot in the digital world [J]. Frontiers of Computer Science, 2012, 6(6):660–667.
- [125] Yuan Yuanming, Qin Xue, Wu Chanle, and Tang Taozhou. Architecture and data vitalization of smart city [J]. Advanced Materials Research, 2012, 403:2564–2568.
- [126] Mei Lin, Cai Xuan, Zhang Hongzhou, Hu Chuanping, and Qi Li. Video structured description—vitalization techniques for the surveillance video data[G] // CCIS 331: Advances on Digital Television and Wireless Multimedia Communications. Berlin: Springer, 2012: 219–227.
- [127] Data revitalization [OL]. Bardess Group Ltd , 2013[2013-10-11]. <http://www.bardess.com/pages/solutions/data-management-data-revitalization.html>.
- [128] 《中国城市发展报告》编委会. 中国城市发展报告(2012) [R]. 中国城市出版社, 2013.
- [129] 中国特大城市列表 [OL]. Wikipedia, 2013[2013-10-11].

<http://zh.wikipedia.org/wiki/中国特大城市列表>



Wang Jingyuan, Assistant Professor at School of Computer Science and Engineering, Beihang University, China. His main research interests include the Smart City, multimedia signal processing and computer networks.