



# 基于大数据的无线通信技术

梁应敞<sup>1</sup>, 谭俊杰<sup>1,2</sup>, 龚晨<sup>3,4</sup>, 张四海<sup>3,4</sup>, 朱近康<sup>3,4</sup>, 徐正元<sup>3,4\*</sup>, 李云洲<sup>5\*</sup>,  
钟晓峰<sup>6</sup>, 周世东<sup>6</sup>, 张倩倩<sup>1,2</sup>, 曹阳<sup>1,2</sup>, 葛君刚<sup>1,2</sup>, 杨刚<sup>1,2\*</sup>

1. 电子科技大学智能通信与网络研究中心, 成都 611731

2. 电子科技大学通信抗干扰技术国家级重点实验室, 成都 611731

3. 中国科学技术大学信息科学技术学院, 合肥 230026

4. 中国科学院无线光电通信重点实验室, 合肥 230026

5. 清华大学北京信息科学与技术国家研究中心, 北京 100084

6. 清华大学电子工程系, 北京 100084

\* 通信作者. E-mail: xuzhy@ustc.edu.cn, liyunzhou@tsinghua.edu.cn, yanggang@uestc.edu.cn

收稿日期: 2021-05-14; 修回日期: 2021-07-05; 接受日期: 2021-07-27; 网络出版日期: 2021-11-09

国家自然科学基金重点项目 (批准号: 61631005, 61631018, 61631013) 资助

**摘要** 随着各类丰富的移动应用普及, 可以预见, 未来的无线通信系统将需要满足更为苛刻的服务要求, 为无线通信技术的发展带来挑战. 另一方面, 大数据在众多领域的应用取得了令人瞩目的成功. 大数据如何赋能无线通信成为学术界和工业界关注的焦点. 过去 5 年, 我们承担了国家自然科学基金“无线通信大数据基础理论与技术研究”重点项目群, 致力于研究基于大数据的无线通信技术, 探索使用大数据突破无线通信发展瓶颈的新理论和新方法. 我们将这类在无线通信中产生、能够服务于无线通信的大数据定义为无线大数据. 本文是对项目群的研究和实践结果的总结. 具体地, 我们首先分析了无线大数据的来源和特征, 然后构建了与之匹配的无线知识学习范式、智能无线网络框架以及跨地域、跨机构、多用途、一体化的无线大数据平台, 最后提出并验证了基于大数据的无线传播环境分析、无线传输以及无线接入技术, 从而为基于大数据的无线通信技术的进一步发展提供基础理论支撑和应用范式.

**关键词** 无线大数据, 无线通信系统, 智能网络, 资源优化, 接入控制

## 1 引言

过去 20 年, 大数据的应用在计算机视觉、自然语言处理等领域取得了令人瞩目的进展, 颠覆性地改变了人们的生活. 与此同时, 移动通信系统也从第三代 (3G) 演变到当前的第五代 (5G), 其速率、时

**引用格式:** 梁应敞, 谭俊杰, 龚晨, 等. 基于大数据的无线通信技术. 中国科学: 信息科学, 2021, 51: 1946–1964, doi: 10.1360/SSI-2021-0165

Liang Y-C, Tan J J, Gong C, et al. Big data-based wireless communication technologies (in Chinese). Sci Sin Inform, 2021, 51: 1946–1964, doi: 10.1360/SSI-2021-0165

延和网络容量等网络服务质量指标有了质的提升. 可以预见, 为了适应越来越丰富的移动应用和庞大的流量需求, 未来的移动通信系统将需要满足更为苛刻的要求, 这为系统设计带来挑战<sup>[1]</sup>. 应用大数据是否能够为无线通信带来增益, 以及如何设计基于大数据的无线通信技术得到了学术界和工业界的广泛关注.

事实上, 作为传输信息的管道, 无线通信系统需要持续地处理海量数据, 其丰富的数据资源为大数据技术提供了充分的应用条件. 这些在无线通信中产生并能够服务于无线通信的海量数据被称为无线大数据<sup>[2]</sup>. 回顾传统的无线通信技术, 它们只依赖于专家知识所构建的普适模型, 提供在广泛场景下的性能鲁棒性, 但无法利用特定场景的特征来获得进一步的性能增益<sup>[3,4]</sup>. 为了解决这一问题, 数据驱动的无线通信技术成为了必然的选择. 原理上, 通过直接挖掘无线大数据中蕴含的规律, 可以得到比普适模型更符合真实通信环境的模型, 从而帮助系统作出更优的决策或判决. 然而, 无线通信系统的固有特殊性使得无线大数据也具有分布式、高效、强异构的特征, 导致传统的大数据处理方法难以直接应用. 因此, 有必要开展无线大数据的基础理论研究, 通过研究无线大数据的特征, 提出适应无线通信需求的大数据分析和应用范式, 并揭示无线大数据的潜力.

与此同时, 人工智能在计算机视觉、自然语言处理等众多领域的成功应用, 使得智能无线通信技术成为了当前的一大研究热点. 事实上, 智能无线通信技术与本文所关注的基于大数据的无线通信技术既有紧密联系也有区别. 一方面, 人工智能与无线大数据是相辅相成、难以分割的, 这是因为人工智能是分析和应用无线大数据的有效手段, 而无线大数据则为人工智能的应用提供了数据来源和平台基础. 但另一方面, 二者的侧重点有所不同: 智能无线通信技术强调人工智能在无线通信中的应用方法<sup>[3]</sup>, 而基于大数据的无线通信技术则是从收集、管理和应用无线大数据的角度展开研究的.

本文对过去4年来我们在基于大数据的无线通信技术的研究进展进行了概述和总结. 通过研究无线大数据的基础理论, 包括无线大数据的特征、无线知识学习范式、智能无线网络框架, 提出了新型的无线大数据平台架构, 总结了无线大数据在无线传播环境分析、无线传输理论方法以及无线接入技术的实践经验.

## 2 无线大数据基础理论

### 2.1 来源及特征

在计算机学科中, 大数据是指一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合, 具有大量化 (volume)、多样化 (variety)、快速化 (velocity)、价值密度低 (value) 这“4V”特征. 而无线大数据则是大数据在无线通信领域中的具体实例, 包括物理层的频谱信息、网络层的流量数据, 甚至应用层的用户行为记录等. 相应地, 无线大数据的来源非常广泛, 从数据生产者的角度大致可以分为终端、无线通信运营商以及无线应用.

- 终端: 终端包括各种需要进行无线通信的设备, 如手机、平板、传感器等. 一方面, 取决于终端需要完成的具体任务, 它们会产生大量的应用层数据, 如位置信息、传感器的探测信息等. 另一方面, 终端在完成数据传输的过程中也会产生大量相关数据, 如信道状态信息 (channel state information, CSI)、各基站信号强度等.

- 无线通信运营商: 运营商不但负责收发终端发送或请求的数据, 也需要将数据从接入网汇聚到核心网并最终连接互联网. 在这一过程中, 通信设备会产生大量的数据, 除了终端回传的信息外, 也包括对终端收发数据的流量日志、计费信息、深度报文解析 (deep packet inspection, DPI) 等<sup>[5]</sup>. 此外, 各

通信设备也会不断汇报其运行状态、负载等网络运维数据。

- 无线应用: 无线应用是用户与终端直接交互的场所, 它们记录了用户个性化的数据, 例如用户的位置信息、社交网络、对某些服务的使用频率等。

以上来源于真实环境的数据通常会含有用户或者商业上的敏感信息, 而获取这些数据门槛较高。无线大数据的另一种来源是在实验室环境的测试或仿真。这类实验室数据虽然并非来源于真实无线网络环境, 但是也能反映无线网络中的某些实际规律。同时, 它们大幅降低了数据获取门槛, 是研究无线大数据过程中不可或缺的部分。无线大数据除了具有“4V”特征之外还具有无线环境带来的特殊特征。

- 分布式: 无线通信系统具有天然的分布式特性, 其中的各个独立的实体自行产生大量数据。同时, 无线通信链路的有限容量也不允许将各个分布实体上的所有数据进行完整采集和集中储存<sup>[6]</sup>。

- 高时效: 在无线通信中, 无线环境和用户需求等因素快速变化。为了保证数据的准确性和处理结果的价值, 采集和处理无线大数据都有较高的实时性要求。

- 强异构: 系统中不同实体分布式产生的数据可能差异较大, 例如, 运营商生成的用户账单和终端处采集到的信道状态记录就具有完全不同的格式和时间尺度。即便对于同一运营商的设备, 它们的软硬件差异也会导致数据的异构性。此外, 不同实体所产生的数据可靠性也不同<sup>[7]</sup>。一般来说, 运营商内部采集的数据可信度比较高, 而终端上传的数据则可信度较低。

## 2.2 无线知识学习范式

为了度量无线大数据的作用, 文献 [8] 提出使用“无线知识熵”这一指标。设有无线事件 A 和 B, 其相似度记为  $S_{B/A}$ , 则当 A 的知识量已知时, B 相对于 A 的无线知识熵定义为

$$H_{B/A} = \mathbb{E}_B \left[ \log_2 \frac{1}{S_{B/A}} \right]. \quad (1)$$

由式 (1) 可知, 若 B 和 A 完全相同, 则二者的知识熵为 0。换言之, 当无线大数据所含有的知识量足够大时, 系统中的所有随机无线事件将能够被完美预知。发挥无线大数据的最大价值就是通过提取无线大数据中的有效知识最小化无线通信中特定无线事件或给定任务的知识熵, 这一过程被称为无线知识学习。

图 1 展示了一种基于特征变量的无线知识学习范式, 其中包括用户群聚 (user grouping) 和任务计算 (task calculation) 两个部分。在用户群聚中, 海量用户的服务需求通常会在多个维度上存在相关性, 因而可以根据群聚变量将所有用户的服务需求聚类到少数几个用户群的数据模型中, 将通信方式从针对单个用户转变为面向群体用户。为了适应无线大数据的分布式特征, 这一过程只需要选择合适的特征变量作为群聚变量, 以用户自身数据来计算用户需求相似度, 能够降低数据收集和处理的开销。任务计算负责基于任务变量为每个分群实现任务的具体执行, 目标是 minimized 资源开销。任务变量可以是信息内容、信道衰落、不同蜂窝接入容量等特征变量。

## 2.3 智能无线网络框架

实现无线网络的智能化不仅需要从无线大数据中学习和提取知识, 还需要对网络资源进行高效利用, 这就要求网络有足够大的资源调度自由度。特别是随着蜂窝基站的密集部署以及不同制式网络的演进, 无线网络逐渐变成复杂的异构网络, 使得每个终端都同时被不同制式、不同基站的信号所覆盖。这些异构的网络之间是相对独立的, 它们不但无法弹性地共享资源也无法交互信息。为了打破不同网络之间的固有屏障, 文献 [7] 提出了一个协作框架, 利用无线大数据、人工智能、软件定义无线电

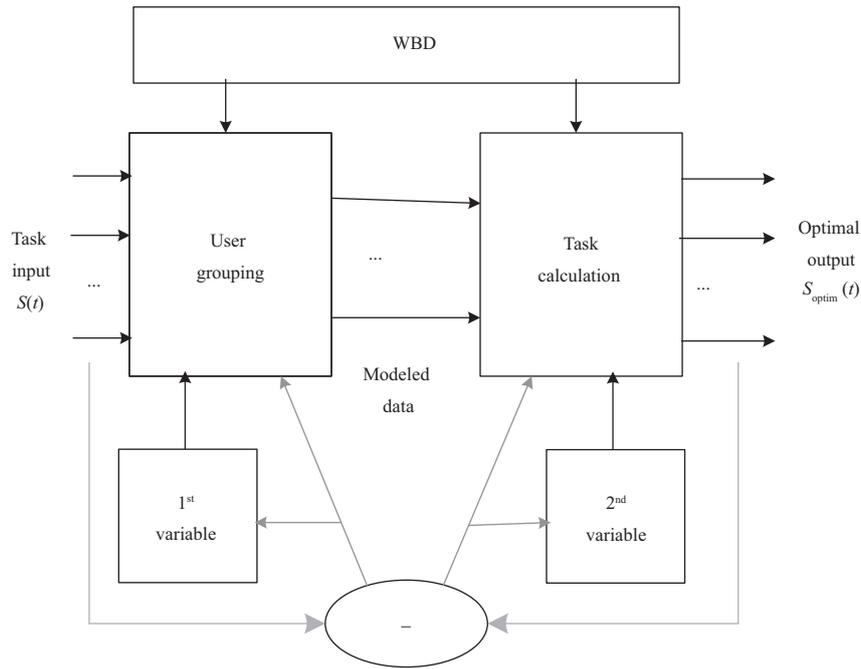


图 1 无线知识学习的基本范式 [8]

Figure 1 A basic paradigm of wireless knowledge learning [8]

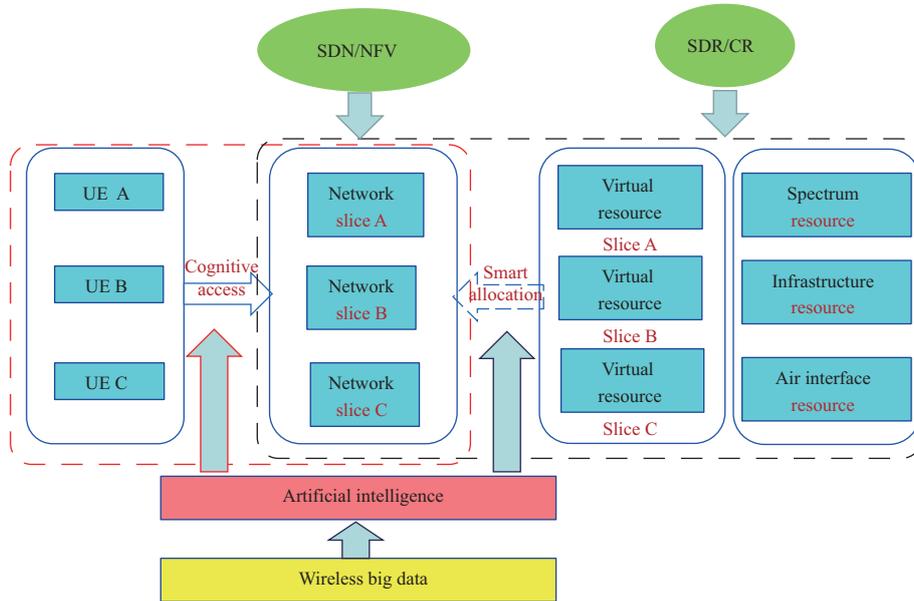


图 2 (网络版彩图) 智能无线网络框架 [7]

Figure 2 (Color online) A framework for smart wireless networks [7]

(software defined radio, SDR) 和软件定义网络/网络功能虚拟化 (software defined networking/network function virtualization, SDN/NFV) 将异构无线网络转变为协同的智能无线网络, 如图 2 所示。

在这一架构中, SDN 将原有异构的无线网络的控制面和数据面进行了分离, 将所有网络单元的控

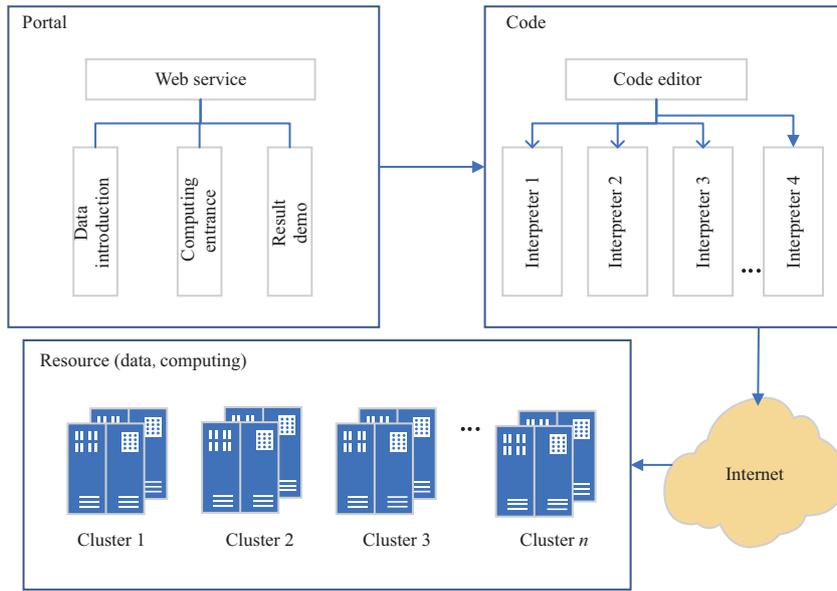


图 3 (网络版彩图) 无线大数据平台架构

Figure 3 (Color online) The architecture of a wireless big data platform

制面进行了整合, 打通了分离的设备之间的信息交互通道, 为采集无线大数据和下发决策指令提供了基础. 在 SDN 之上, NFV 将网络单元的计算、储存等物理资源进行了整合以及按需调度. 类似的, SDR 对异构网络所拥有的频谱、功率等空口资源进行了整合. 基于 NFV 和 SDR, 网络和空口资源的调度都有了极大的自由度, 能够根据用户的需求动态地为各个网络或者通信任务划分资源切片. 为了让所有用户能够在任何时刻都得到其预期的服务质量, 智能无线网络需要持续地对用户、网络和资源切片作出最优映射决策, 这就需要利用人工智能 (artificial intelligence, AI) 方法对无线大数据进行分析, 并对通信需求、流量等动态因素进行精确预测, 从而实现智能无线网络中的用户认知接入 (cognitive access) 和资源智能分配 (smart allocation) 两个必要功能. 此外, 这一框架创新性地提出了处理无线大数据的在线 - 离线双层引擎. 具体地, 部署在远端云服务器的离线引擎用于分析和提取复杂和海量数据中的知识, 而边缘侧的在线引擎负责整合离线引擎的处理结果并根据环境特征进行少量更新. 利用在线引擎, 边缘侧的各实体能够对无线环境作出快速响应, 从而满足了无线大数据的高时效特征.

### 3 无线大数据平台

无线大数据平台负责数据的采集、清洗、存储、共享以及分析, 是支撑挖掘和利用无线大数据的关键要素. 由于无线大数据具有海量、分布式以及强异构等特征, 无线大数据平台的建设需要考虑其特征作出相应的设计. 特别地, 无线大数据的强异构特征决定了其来源广泛而且拥有权高度多样化, 因而设计无线大数据平台时需要考虑计算和数据资源的解耦, 并严格控制数据的分级访问权限.

图 3 展示了一种典型的无线大数据平台架构图, 主要包括门户、代码和资源三大子系统. 其中, 门户子系统主要提供公开数据集的展示 (如数据集简介和切片下载)、计算平台入口以及分析结果展示; 代码子系统提供交互式编程和交互式数据分析, 用户可采用各类编程语言在线编写和共享代码, 并调用资源子系统的数据和计算资源集群进行远程的数据访问和计算; 资源子系统是一个分布式的集群系统, 支持数据分布式存储和跨地域异构计算能力, 从而能够在保持数据独立性的前提下为多方提供

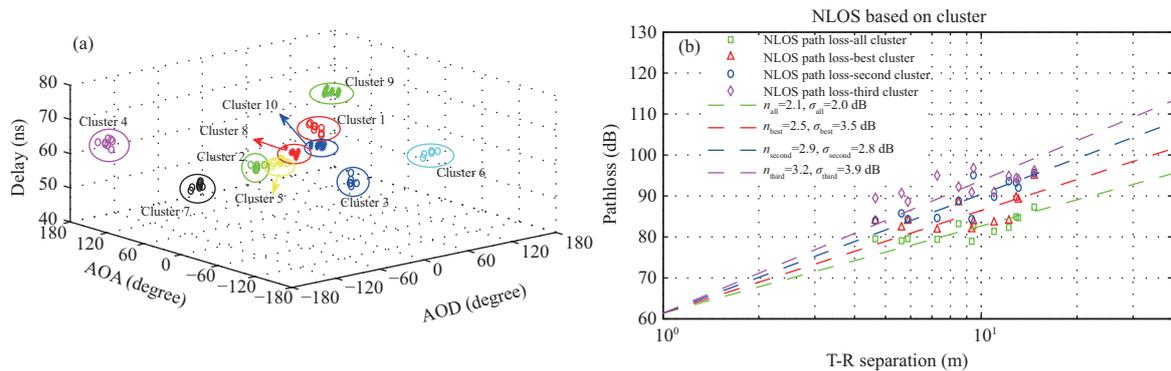


图 4 (网络版彩图) 办公室 NLOS 环境下, (a) 多径成分的聚类结果和 (b) 28 GHz 频点的路径损耗模型<sup>[10]</sup>

Figure 4 (Color online) In the office NLOS environment, (a) clustering the multipath components and (b) 28 GHz path loss model based on cluster<sup>[10]</sup>

处理数据的能力.

在开展研究过程中, 我们基于图 3 构建了统一的无线大数据平台, 各平台参与方可以便捷、高效和安全地共享自身的优势数据和计算资源. 此外, 响应学术界对推进数据与代码共享的提倡, 这一平台将持续进行资源扩充和开发, 目标成为国内重要的无线大数据资源共享平台<sup>1)</sup>, 进一步促进国内无线大数据研究. 在无线大数据平台的支撑下, 各方得以在无线传播环境分析、无线传输理论方法以及无线接入技术这 3 个重要课题上开展大量研究.

## 4 无线传播环境分析

### 4.1 信道建模与获取

进入 5G 及后 5G 时代, 无线通信面临高确定性、低延时、高谱效等新需求的挑战, 高效利用和改造无线信道成为无线通信进一步变革的突破点之一. 研究 CSI 与无线传播环境之间的影响机理, 通过 CSI 感知无线传播环境, 利用无线传播环境信息改进 CSI 的可预测性, 是高效利用和改造无线信道的主要思路. 采用机器学习的方法从无线大数据中挖掘具体无线传播环境与 CSI 之间的相关性能够帮助构建新型信道模型, 最终获取复杂无线环境下的 CSI. 为了探索具体环境下的 CSI 与无线传播环境的相互关系, 首先开展了无人机场景的信道测量与建模, 以及毫米波信道的参数提取工作.

文献 [9] 考虑了无人机 (unmanned aerial vehicles, UAV) 通信中的信道建模问题. 当 UAV 信道存在高层建筑等遮挡物体时, 其大尺度路径损耗主要由非视距 (non line-of-sight, NLOS) 因素决定. 因此, UAV 信道的大尺度路径损耗模型由信道是否为视距 (line-of-sight, LOS) 以及相应的出现概率决定. 这一概率可以通过马尔科夫 (Markov) 链模型进行建模, 并用实测数据学习得到其具体值. 文献 [10] 针对毫米波信道也提出了基于聚类的大尺度路径损耗模型. 首先, 由旋转定向天线构成的虚拟天线阵列测量出全向信道冲激响应, 然后通过高精度多径提取算法分离出每一条径的时延、波达角 (angle of arrival, AoA) 和发射角 (angle of departure, AoD). 鉴于时延域和角度域相近的多径会对信号产生相似的信道响应, 聚类算法可以将多径聚类为簇, 便于之后基于簇的信道建模和分析. 图 4(a) 展示了聚类算法将多径划分成簇的结果, 而图 4(b) 则展示了在 28 GHz 频点的典型办公室 NLOS 场景下接收能量最强的 3 个簇的路径损耗模型.

1) 平台链接. <https://wbd.ustc.edu.cn/gxpt/list.htm>.

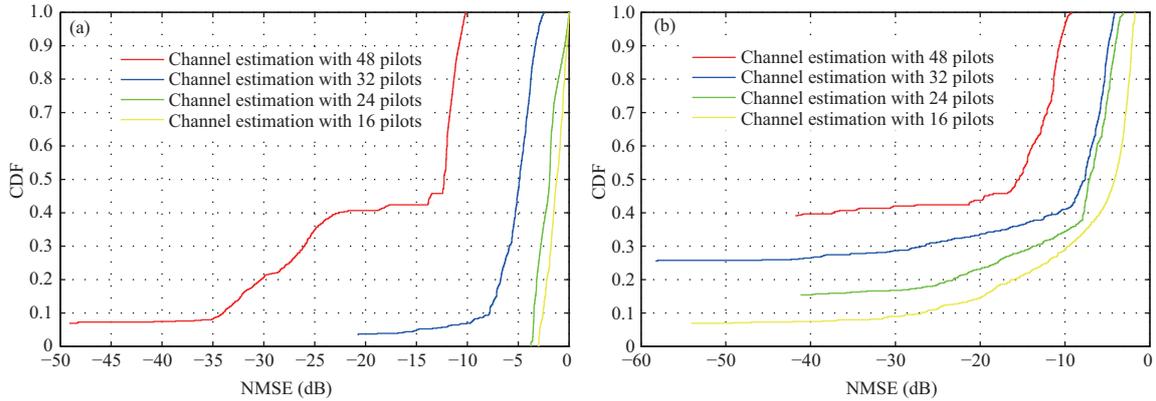


图 5 (网络版彩图) 归一化均方误差的累计分布函数. (a) 室外测量; (b) 室内测量<sup>[11]</sup>

Figure 5 (Color online) Cumulative distribution function (CDF) of normalized mean square error (NMSE). (a) In the outdoor measurement; (b) in the indoor measurement<sup>[11]</sup>

除大尺度路径损耗模型外, 文献 [11] 进一步地提出了层析信道模型, 其基本思想是将信道分成静态 (路径损耗、阴影衰落等)、动态 (相对运动导致) 和扰动 (随机物体产生的散射、反射等) 3 种分量并分别处理. 例如, 信道的静态分量可以直接对历史 CSI 求均值得到, 而 CSI 中的动态分量可以从接收机位置传感器回传的相对运动状态推算得到. 最终, 将静态、动态和扰动分量合并即可获得当前 CSI. 图 5(a) 和 (b) 分别展示了在室内外场景下估计 CSI 的归一化均方误差和导频数量的关系. 此外, 使用层析模型还能够根据一个频段的 CSI 推测另一频段的 CSI, 这是解决频分双工系统中上下行链路的互易性问题的可行手段<sup>[12]</sup>.

#### 4.2 物理环境推理

基于物理环境与无线信道之间存在的紧密关系, 除了可以利用物理环境特征帮助建模和获取信道外, 信道信息同样也可以用于推理和分析物理环境的特征. 在这一方面, 文献 [13] 提出了基于接收信号强度 (received signal strength, RSS) 的三维障碍物地图构建方法, 而文献 [14] 则研究基于信道特征的信号源定位问题. 在文献 [13] 中, 用于构建三维障碍物地图的数据来源于多个地面用户所接收到的蜂窝基站信号的 RSS 信息. 取决于障碍物是否存在, 不同用户接收到的信号可能来源于 NLOS 径或 LOS 径. 一种无监督的分类方法被提出用于分辨用户链路的类别. 对于 NLOS 径, 其障碍物高度可以通过参数估计得到. 最后, K 最近邻 (K-nearest neighbor, KNN) 算法被用于预测和补全地图上用户稀疏性造成的缺失值. 图 6(a) 和 (b) 分别展示了原始和重构得到的三维障碍物地图. 在文献 [14] 中, 多输入多输出 (multi-input multi-output, MIMO) 信道的角度域和时延域信息作为指纹特征被用于定位信号源. 具体地, 一个多层的卷积神经网络 (convolutional neural network, CNN) 以分类和回归两种模式学习从指纹特征到信号源坐标的映射关系. 从图 7(a) 的定位效果可以看出, 所提方案相较于传统 Two-stage<sup>[15]</sup> 方法能够有效提高定位精度, 并且图 7(b) 也说明了该方案能够明显降低计算开销.

### 5 无线传输理论方法

#### 5.1 信号检测

信号检测需要从经过信道畸变的接收信号中恢复出发送信号, 是一个典型的分类问题. 经典的信

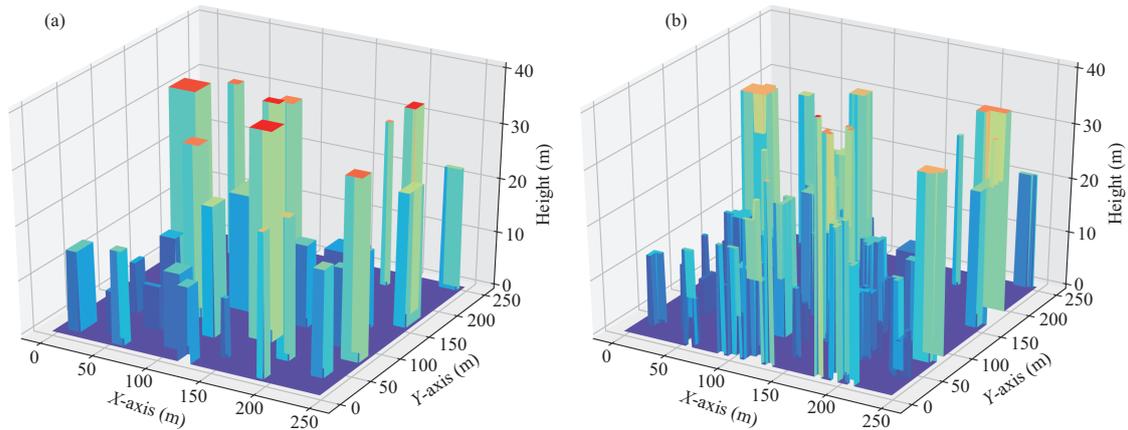


图 6 (网络版彩图) (a) 仿真的障碍物地图; (b) 重构的障碍物地图 [13]

Figure 6 (Color online) (a) Simulated obstacle map; (b) reconstructed obstacle map [13]

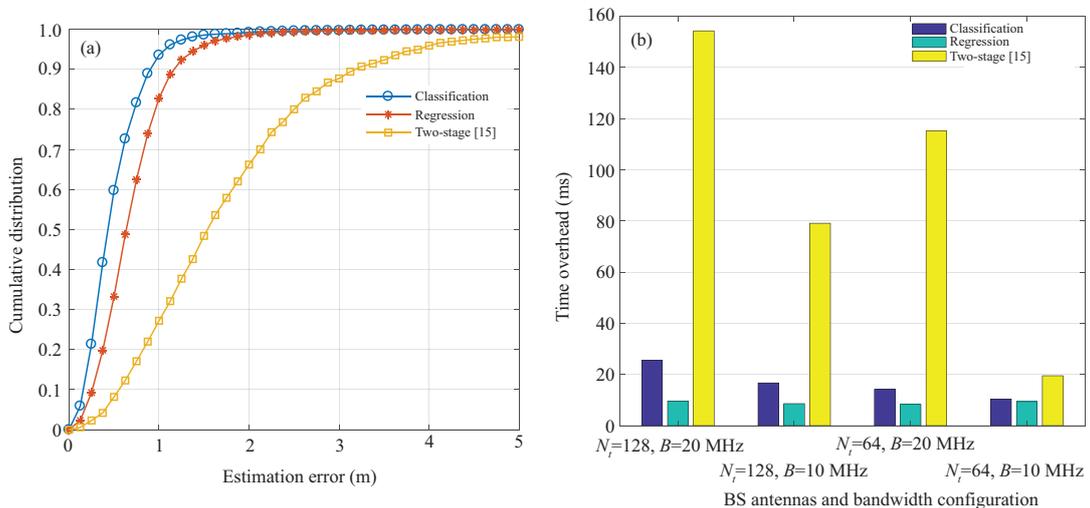


图 7 (网络版彩图) (a) 不同方法导致的定位误差的累积分布函数; (b) 不同定位方法所需的在线定位时间 [14]

Figure 7 (Color online) (a) The CDF of the location errors with different localization methods; (b) the online localization time with different localization methods [14]

号检测算法是在一般性的信道和信源假设下推导出判别公式, 无法利用实际信道或信源的特征来提高检测性能. 因此, 直接从数据中学习最优的信号检测器无疑是一种更好的替代选择. 在可见光频段中, 由于光波的方向性较强, 终端接收到的信号对于物体遮挡和终端移动非常敏感, 能否充分利用信道特征成为在光通信中提高信号检测准确率的关键因素. 文献 [16] 以二阶环境物体反射的极弱接收信号场景为例, 通过采集 0/1 符号下的接收信号波形并输入支持向量机 (support vector machine, SVM) 直接训练学习得到非线性接收信号波形与发送符号之间的关系. 结果表明, 相比于高斯逼近 (Gaussian approximation) 的信号检测, 基于 SVM 的信号检测器具有更低的错误检测概率. 进一步地, 文献 [17] 将双向长短期记忆 (bi-directional long-short term memory, BiLSTM) 神经网络用于提高非线性系统 0/1 符号检测性能. 图 8 展示了接收终端位于桌子下方时借助环境二阶反射时使用不同信号检测方法得到的误码率 (bit error rate, BER). 可以看出, 在没有硬件预均衡的条件下, 基于 BiLSTM 的方法得

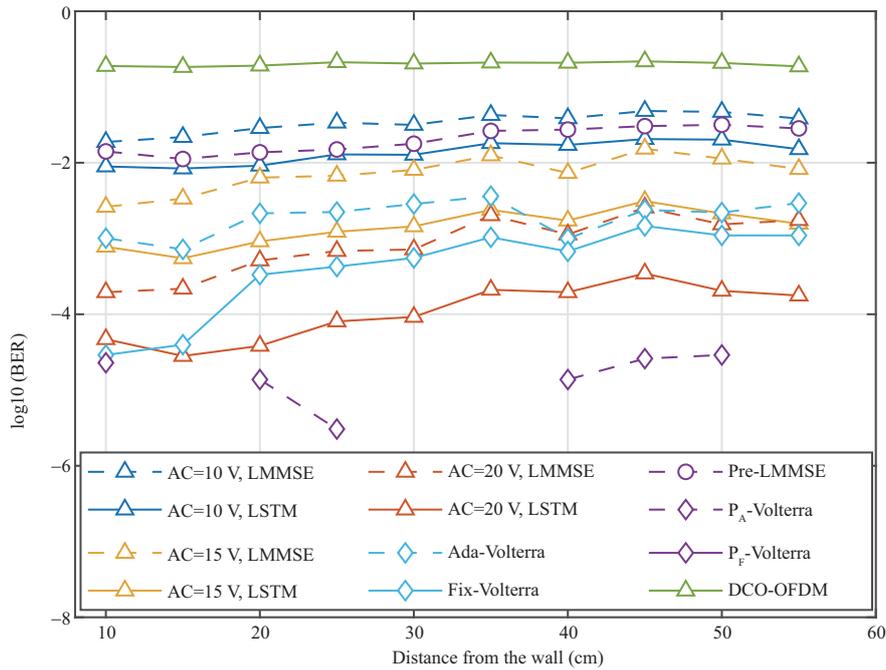


图 8 (网络版彩图) 二阶反射通信中接收机在不同位置得到的 BER [17]

Figure 8 (Color online) The BERs under different receiver locations for the second-order reflection communication [17]

到的 BER 最低; 在有硬件预均衡的条件下, 传统用于对抗系统非线性的 Volterra 级数均衡方法得到的 BER 最低。

此外, 文献 [18] 以环境散射通信 (ambient backscatter communication, AmBC) 为例也说明了数据驱动的信号检测方案的优势. 在 AmBC 系统中, 反射设备通过将信号调制到环境中的射频信号上传输信息. 由于接收机难以获得环境信号的导频信息进行信道估计, 在接收端恢复反射设备的发送信号是一项具有挑战性的任务. 传统的非相干检测方案, 如能量检测 (energy detection, ED), 会因为反射链路远弱于直射链路而产生严重的性能衰减 [19]. 为了解决这一问题, 文献 [20] 提出了共生无线电的概念, 通过将射频源系统与反射传输系统建立合作, 实现高可靠反向散射通信. 文献 [18] 提出了一种依赖发送数据本身的半盲信号检测方法. 具体地, 射频源将自身调制方式告知反射设备及其接收机, 反射设备在传输数据之前会发送两个接收端已知的标签数据, 然后接收机将接收到的信号与标签数据共同进行聚类分析, 最终根据标签匹配各类别并恢复出反射设备的发送信号. 这一方法称为 CL-LUS (constellation learning with both labeled and unlabeled signal) 算法. 图 9 对比了 CL-LUS 和其他几种算法在不同直射链路和反射链路相对信噪比 (signal-to-noise ratio, SNR) 下的 BER 性能, 包括 ED、已知信道下的最优检测 (optimal detection, OD)、传统的 Standard GMM-based EM 聚类算法 [15] 以及 CL-LS (constellation learning with labeled signal) 算法. 其中, CL-LS 算法和 CL-LUS 算法类似, 但仅利用了标记信号来聚类 and 反推出信号检测模型中的未知参数并据此恢复未标记信号. 可以看到, CL-LUS 算法的性能远优于 ED 和传统聚类算法, 并且也好于 CL-LS 算法. 这说明了 CL-LUS 算法能够利用待检测信号中存在的隐含相关性提高检测性能。

## 5.2 自适应调制编码

自适应调制编码 (adaptive modulation and coding, AMC) 是一类根据无线信道状态来动态调整调

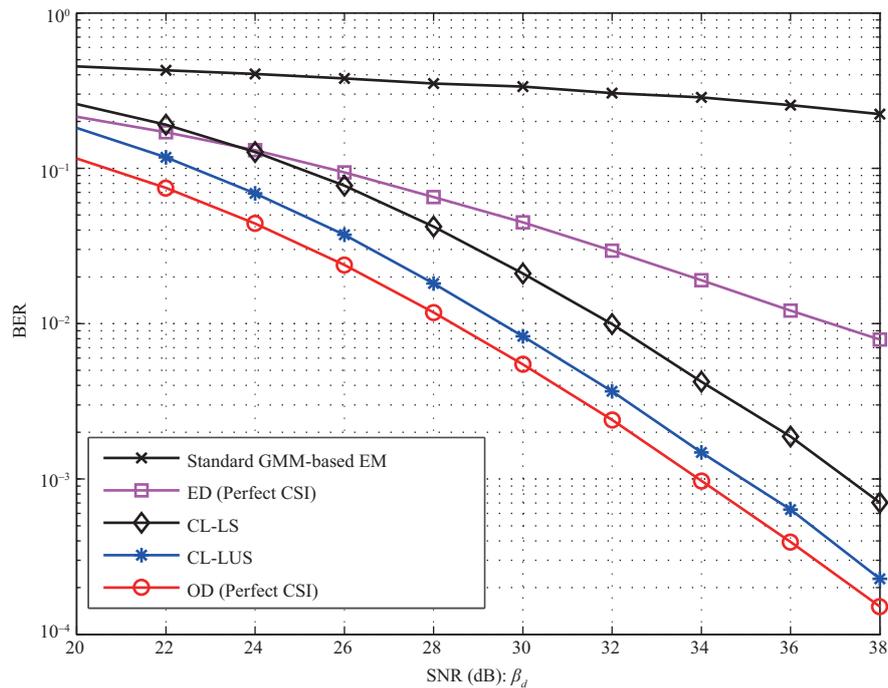


图 9 (网络版彩图) BER 随直射链路 SNR  $\beta_d$  的变化 (假设环境射频信号源采用 QPSK 调制方式)<sup>[18]</sup>

Figure 9 (Color online) BER versus direct link SNR  $\beta_d$  under the assumption of QPSK modulated ambient RF source<sup>[18]</sup>

制和编码方式的传输方法, 以使得通信链路可以持续获得可靠的最大通信速率. 显然, AMC 的效果取决于对无线环境的感知和追踪, 例如来自其他设备的同频干扰. 那么利用无线环境的历史数据是否可以帮助 AMC 提升性能呢? 文献 [21] 回答了这一问题.

文献 [21] 考虑了一个认知异构网络, 其中存在主用户和次用户两类用户. 理想情况下, 次用户接入频谱时需要保护主用户, 即只有在感知主用户没有在传输时才会发送信息. 然而, 实际上频谱感知存在误差, 次用户可能会错误地在主用户传输时接入同一频谱进行传输, 对主用户产生干扰. 由于次用户的传输时间晚于主用户的传输时间, 所以主用户不能及时测量到干扰信息, 并调整传输策略. 为此, 深度强化学习 (deep reinforcement learning, DRL) 被提出用于分析次用户对主用户的干扰规律并预测干扰, 从而自适应地调整调制编码方案, 提高主用户的传输速率. 从图 10(a) 的仿真结果可以看出, 在干扰缓慢变化的情况下, DRL 可以使主用户的传输速率接近完美知悉干扰情况的最优性能, 且比其他算法高 50% ~ 100%. 当干扰快速变化时, DRL 的性能略差于最优性能但相比其他算法仍有相当大的性能优势. 这证明了数据驱动的 AMC 可以更好地学习无线环境的动态性并通过预判环境变化调整传输策略.

### 5.3 波束赋形

波束赋形技术利用了多天线发送信号叠加特性, 通过调整不同发射天线的相位和幅度使得特定方向的接收信号得到相长或者相消. 在蜂窝网络如今普遍采用相邻蜂窝全频谱复用的情况下, 干扰环境下的协同波束赋形是一种极具前景的方案, 它通过联合设计多个基站的波束赋形矢量抑制蜂窝间干扰并提高系统整体性能<sup>[22, 23]</sup>.

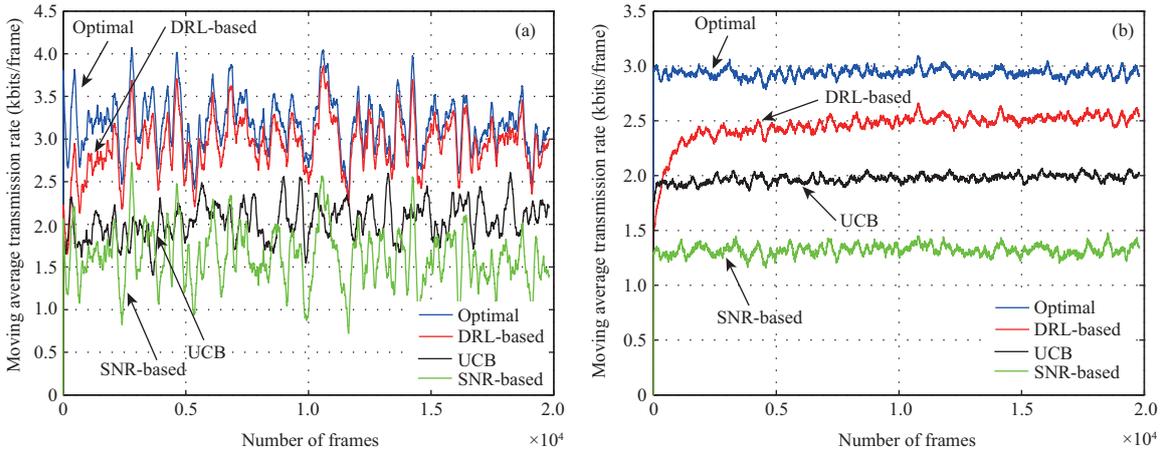


图 10 (网络版彩图) 传输速率比较. (a) 拟静态干扰变化场景; (b) 动态干扰变化场景 (图中每个值都是此前 200 个帧的瞬时值的滑动平均, 并且每条曲线是 20 次随机试验的平均值) [21]

Figure 10 (Color online) Transmission rate comparison in (a) a quasi-static interference scenario and (b) a dynamic interference scenario (each value is a moving average of the previous 200 frames and each curve is the average of 20 trials) [21]

显然, 让整个蜂窝网络中的所有基站进行协同波束赋形最直接和传统的方法就是在搜集全局 CSI 后执行集中式优化算法求解得到每个基站的波束赋形矢量. 然而, 这类集中式优化算法往往存在计算复杂度极高的问题, 并且在实际系统中也难以及时获得全局 CSI. 为了避免以上问题, 文献 [24] 提出了数据驱动的分布式波束赋形方案, 利用 DRL 让每个基站能够不断地观测和分析自身无线环境数据并据此自适应地调整波束赋形矢量, 有效地降低了基站决策所需的信道信息数据量以及计算复杂度. 图 11 展示了所提方案的仿真结果. 其中, 从图 11(a) 可以看出, 基于 DRL 的分布式波束赋形方案在学习过程中不断改善决策策略, 性能优于贪婪策略和随机方案, 并取得了接近传统集中式分式优化 (fractional programming, FP) 算法的性能, 但仍存在一定差距; 而图 11(b) 则说明了, 通过提高决策的精细程度 (图中括号内两个数字分别代表在功率和波束方向的决策精度, 数字越大表示相应的决策精度越高), 所提方案能够逼近传统集中式优化算法的性能.

## 6 无线接入技术

### 6.1 多址接入

作为未来无线通信的重要构成部分, 机器类通信 (machine type communications, MTC) 具有数据包短和海量接入的特点, 这与传统的语音和数据业务有极大不同. 因此, 如果在 MTC 中依旧采用诸如时分或频分等传统的正交多址接入方式会使得稀缺的频谱资源面临严峻挑战. 幸运的是, 具有相似特征的 MTC 设备所发送的巨量小数据通常具有比较高的内容相关性. 为此, 文献 [25] 从无线大数据的角度出发, 基于数据包的相关性提出了集群多址接入. 具体地, 所有 MTC 设备会被聚类算法分成众多具有相似特征的集群. 对于归属同一集群的 MTC 设备, 它们的历史发送数据被用于分析其数据包发送模式, 并据此预测出不同设备的数据包发送顺序. 接着, 通过延迟数据包传输调整数据包发送顺序, 使得集群内的 MTC 设备在同一时刻发送相同的数据包, 从而不但节省了无线资源也提高了频谱效率. 图 12 展示了所提算法的示意图, 其中图 12(a) 展示了 3 个终端用户在不同时刻的数据包相关性 (颜

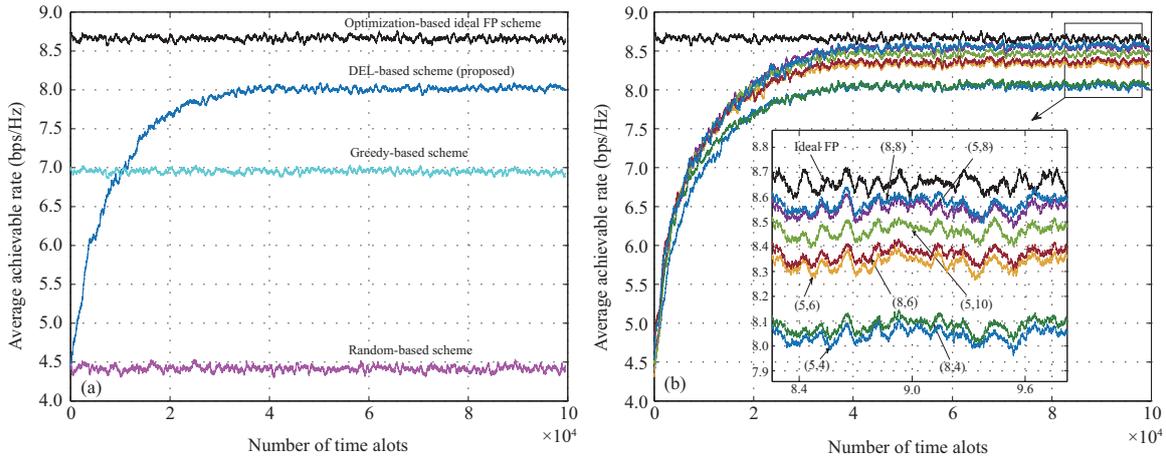


图 11 (网络版彩图) (a) 不同方法得到的平均可达速率; (b) 所提 DRL 方案以不同决策精度得到的平均可达速率 (图中每个值都是此前 500 个时刻的瞬时值的滑动平均) [24]

Figure 11 (Color online) (a) Average achievable rate of different approaches; (b) average achievable rate of the proposed DRL-based scheme with different sets of available actions (each value is a moving average over the previous 500 time slots) [24]

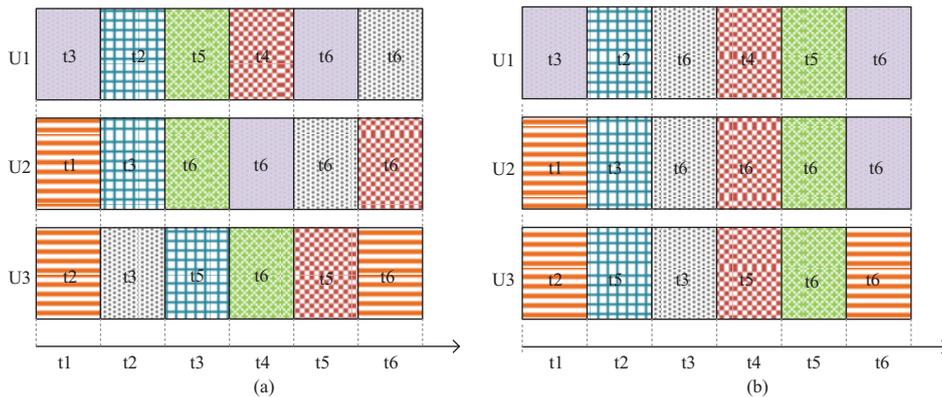


图 12 (网络版彩图) 集群多址接入方案. (a) 调整前的数据包传输顺序; (b) 调整后的数据包传输顺序 [25]

Figure 12 (Color online) Demonstration of the proposed scheme. (a) Before adjusting; (b) after adjusting [25]

色相同表示相关), 图 12(b) 则给出了调整传输顺序后的结果, 可以从中看出这一多址接入方案能够有效地节约时间开销.

### 6.2 动态频谱接入

频谱是承载无线信号最宝贵的资源, 而频谱固定分配是目前无线通信系统使用频谱最为广泛的方式. 然而, 由于用户流量存在空时差异特性, 固定分配的频谱无法根据系统实际负载进行伸缩, 降低了频谱利用率和网络整体接入容量. 针对频谱固化产生的问题, 能够支持动态接入频谱的认知无线电技术近年来得到了广泛关注 [26].

在接入频谱之前, 无线通信系统首先需要对频谱进行精确感知, 即识别频谱是否正在被其他系统占用, 以免对其他正在传输的系统造成干扰. 传统的频谱感知方法通过单一、固定的特征进行检测, 其性能受到 SNR 墙等因素限制. 此外, 这些传统方法是基于模型推导的, 而在实际环境中很难对信

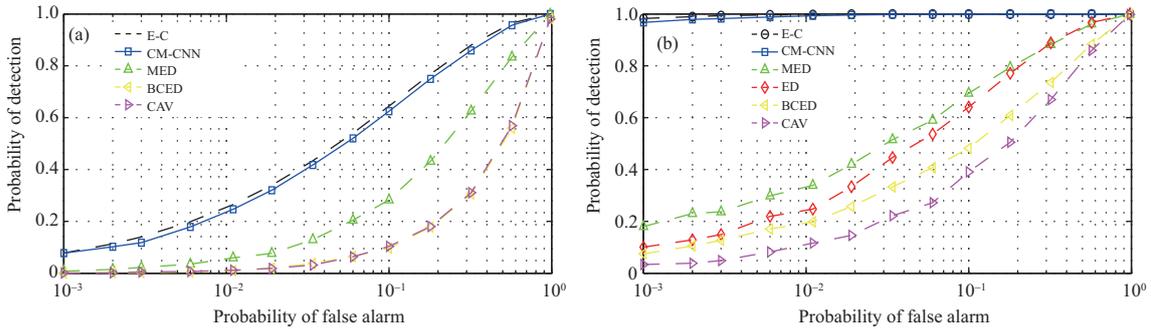


图 13 (网络版彩图) 在高斯噪声、28 根天线、100 个观察样本以及 SNR 为 -15 dB 下, 不同算法得到的 ROC 曲线. (a) 发送信号独立同分布; (b) 发送信号存在相关性 [27]

Figure 13 (Color online) ROC curves of different algorithms under Gaussian noise, 28 antennas, 100 observations, and -15 dB SNR: (a) i.i.d. model; (b) correlation model [27]

号进行精确的统计建模. 为了解决以上问题, 文献 [27] 提出了数据驱动的智能频谱感知方案, 借助深度学习 (deep learning, DL) 技术从原始频谱数据发掘深层特征. 具体来说, 卷积神经网络 (CNN) 被用于处理接收信号的协方差矩阵并得到检测统计量. 在离线训练中, 从训练数据可以训练得到检测统计量, 并依据所需的误警或检测概率搜索得到相应的检测阈值. 在训练完成后, 得到的神经网络可以从未知的接收信号中提取相应的检测统计量并根据阈值实现在线检测. 从图 13 的接收者操作特征 (receiver operating characteristic, ROC) 曲线可以看出, 无论发送信号服从独立同分布 (图 13(a)) 还是存在相关性 (图 13(b)), 这一智能频谱感知方案性能远远优于传统算法, 如最大特征值检测 (maximum-eigenvalue detection, MED)、盲合并能力检测 (blindly combined energy detection, BCED)、协方差绝对值 (covariance absolute value, CAV) 算法, 并且逼近已知统计协方差矩阵下理论最优的估计相关器 (estimator-correlator, E-C) 算法.

虽然频谱感知能够检测当前是否有信号正在传输, 但是它无法对频谱所承载的流量进行统计分析, 更难以相应地对频谱接入策略进行调整. 这在使用非授权频谱等具有公平接入要求的频谱时是极其重要的 [28]. 文献 [29] 以 LTE 蜂窝系统动态接入非授权频谱为例研究了这一问题. 由于 WiFi 系统是目前使用非授权频谱的主要系统, LTE 系统接入非授权频谱需要感知 WiFi 系统的流量并调节接入时间长度, 避免过多挤压 WiFi 传输时间导致 WiFi 系统服务质量下降. 具体来说, LTE 系统采用轮询的方式接入频谱, 即周期性地感知、接入和释放频谱. 为了在保护 WiFi 系统服务质量的同时最大化 LTE 传输速率, DRL 被提出用于从频谱观测数据 (如空闲、繁忙时间长度等) 推测真实 WiFi 系统流量并预测 WiFi 流量变化趋势, 从而能够对 LTE 系统的频谱接入时间进行自适应优化.

### 6.3 接入资源配置

无线用户的服务体验取决于系统能否及时地为用户提供所需资源. 然而, 无线通信系统的接入资源包含功率、时间、频谱, 甚至缓存等多域资源, 使得接入资源的配置问题是一个高维度的复杂决策问题. 此外, 无线通信普遍存在动态时变的无线环境和用户行为, 意味着捕捉系统动态需要由大量高实时性的信令交互作为支撑, 这也给高效地接入资源配置带来困难. 文献 [30~32] 尝试利用无线大数据中蕴含的系统动态变化规律来降低信令开销, 从而高效地求解接入资源配置问题.

文献 [30] 提出了一个 DL 和 DRL 协作决策的蜂窝接入网资源切片调度机制. 具体地, DL 分析和预测大时间尺度上的用户流量需求并提前配置资源到相应的切片. 由于业务需求在小时间尺度上的突

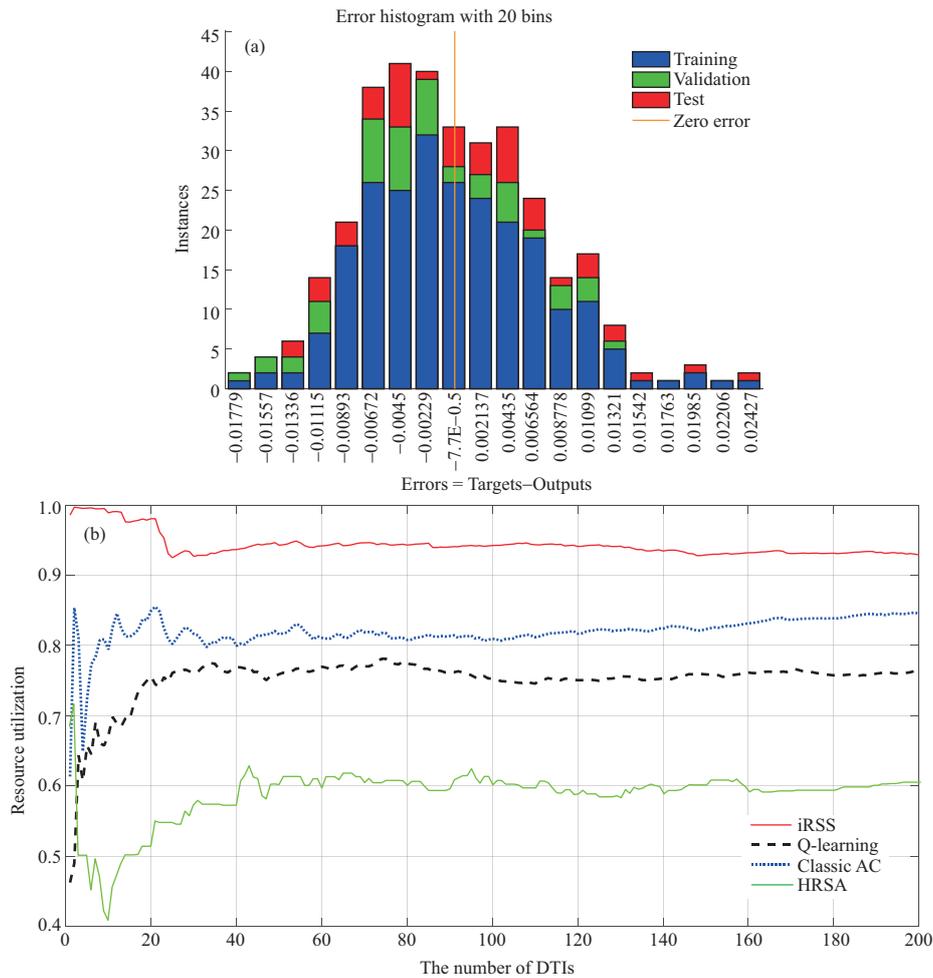


图 14 (网络版彩图) (a) DL 对用户流量预测的误差; (b) 资源利用率对比<sup>[30]</sup>

**Figure 14** (Color online) (a) User traffic prediction error by DL; (b) comparison of resource utilization ratios<sup>[30]</sup>. AC: actor-critic; HRSA: heuristic resource scheduling algorithm; DTI: decision time intervals

发性明显, 作者进一步提出采用具有更高动态性的 DRL 进行小时间尺度的在线资源调度. 该在线资源调度方法具有较低的算法复杂度以及较快的收敛速度, 但是可能在初期迭代阶段会获得一些较差的策略, 并且出现不稳定的性能抖动. 因此, 可以通过调整 DL 和 DRL 所作决策的相对重要性提高决策精度和网络性能. 图 14(a) 展示了 DL 关于用户流量需求预测的误差直方图, 可以看出预测结果是直接且有效的. 从图 14(b) 可以看出, 与经典的 Q 学习等算法相比, 所提的智能资源调度策略 (intelligent resource scheduling strategy, iRSS) 可以持续稳定地对资源进行高效利用.

此外, 文献 [31, 32] 基于 DRL 对接入资源进行配置. 其中, 文献 [31] 研究了蜂窝终端直通 (device-to-device, D2D) 网络中的频谱和功率配置问题. 鉴于频谱稀缺性, D2D 设备通常需要复用有限的蜂窝频谱来进行通信. 因此, D2D 设备需要对所使用信道以及发送功率进行精确的优化, 否则会产生相互间的强干扰. 这一联合信道选择及功率控制问题是一个复杂度极高的 NP-Hard 问题, 且需要在获得全局信道信息后才能求解. 为此, 作者提出一种分布式的资源配置框架, 让各个 D2D 用户自主地进行大数据分析以及资源配置决策. 在各个 D2D 用户上, DRL 通过分析本地和历史信息, 从而学习到自

身无线环境的变化规律并将其用于配置发送信道和功率. 与传统方法相比, 所提方法在运算复杂度和信令开销两方面都能取得良好的可扩展性. 文献 [32] 则研究了 MTC 中的非连续接收 (discontinuous reception, DRX) 自适应调整机制. DRX 是一种周期性断开终端与基站连接以达到节能目的的机制. 鉴于 MTC 中的业务类型多样, 相应的特征和需求差异也非常大, 此时采用传统 DRX 机制会造成接入延迟长且能耗高. 为此, 作者采用 DRL 对 DRX 周期进行自适应调整, 使终端的休眠时间适配于多样化的业务模式. 具体地, 时间被分割成多个时间窗, DRL 学习历史时间窗内数据流的统计规律, 然后在每个时间窗到来之初确定当前时间窗长度内采用的 DRX 周期, 并将这个 DRX 周期应用于当前时间窗内到达的所有数据流.

#### 6.4 接入或切换控制

接入和切换都是指用户被多个基站 (或接入点) 的信号覆盖时选择一个基站进行接入的问题, 其区别是接入发生在通信连接建立之初而切换发生在通信过程中. 我们将这类如何在多个基站中选择基站进行连接的问题统称为接入控制问题. 在传统的接入控制机制中, 用户首先选择最强接收信号的基站, 没有考虑未来的链路质量及流量情况. 当基站部署得更稠密以及用户数量激增时, 传统的方法容易导致用户扎堆, 造成拥塞和服务质量下降. 因此, 需要考虑如何挖掘并利用用户的移动性及网络流量动态性来建立更好的接入控制机制<sup>[33,34]</sup>. 无线大数据为实现这一目标提供了可行途径.

文献 [35] 以 UAV 网络为例, 提出了一种能够预测未来链路状态并考虑长期收益的接入控制机制. 在 UAV 网络中, UAV 基站的特定运动规律会导致网络环境动态变化. 为了在保证服务质量的同时减少不必要的切换, 用户需要对各个 UAV 基站未来的服务能力进行预测. 为此, 基于长短期记忆 (long short-term memory, LSTM) 神经网络的 DRL 算法被提出用于捕捉和预测用户接收信号强度和基站接入用户数变化. 除此之外, 文献 [36] 研究了毫米波超密集蜂窝网络中的用户接入控制问题. 在这种网络结构下, 用户的最强接收信号会频繁地在多个基站之间变动, 此时使用传统方法就会导致频繁切换和服务中断. 为此, 作者通过分析网络历史流量数据, 对网络中热点区域流量变化规律进行了合理建模, 然后在考虑用户服务质量的情况下使用置信区间上界 (upper confidence bound, UCB) 算法来进行基站选择, 整个接入控制方案又被称为 SMART 算法. 从图 15 中的仿真结果可以看出, 在不同的用户移动速度下, 所提出的 SMART 算法与基于传输速率 (rate-based handover, RBH) 以及基于 SINR (SINR-based handover, SBH) 的切换算法相比, 可以实现切换次数的有效降低, 并且在系统吞吐量上接近 SBH 算法.

为了进一步提高无线通信系统的接入性能, 用户接入控制也可以与接入资源配置一起进行联合优化. 文献 [37] 以三层异构蜂窝网络为例, 采用 DRL 来解决联合优化接入控制和资源配置时产生的决策空间维度爆炸的问题. 具体地, 网络中的所有用户将本地信息汇合并形成联合观测, 使得用户在进行接入决策时可以对网络整体的动态变化进行预测. 与遗传算法等传统方法相比, 这一方案在系统容量、系统效能以及计算时间等方面均能体现出显著的性能优势.

## 7 结束语

本文总结和概括了我们在过去 5 年对基于大数据的无线通信技术的研究进展. 首先, 我们将无线通信中产生并能够服务于无线通信的海量数据定义为无线大数据, 并对无线大数据的来源和特征进行了总结. 然后, 为了适应无线大数据所具有的分布式、高时效和强异构特征, 提出了相应的无线知识学习范式和智能无线网络框架, 并构建了跨地域、跨机构、多用途、一体化的无线大数据平台. 最后,

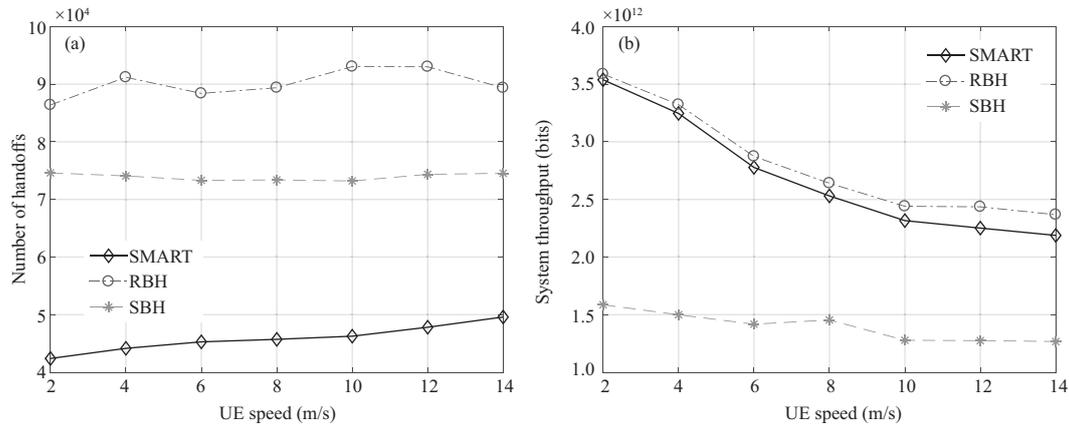


图 15 切换性能与用户移动速度的关系. (a) 切换次数; (b) 系统速率<sup>[36]</sup>

**Figure 15** Relationship between the handoff performance and user equipment (UE) speed. (a) The number of handoffs; (b) system throughput<sup>[36]</sup>

总结了无线大数据在无线传播环境分析、无线传输理论方法以及无线接入技术的实践经验,证明了大数据能够为无线通信带来显著增益,同时能够为未来基于大数据的无线通信技术的进一步演进提供参考.

## 参考文献

- Chen S Z, Liang Y C, Sun S H, et al. Vision, requirements, and technology trend of 6G: how to tackle the challenges of system coverage, capacity, user data-rate and movement speed. *IEEE Wireless Commun*, 2020, 27: 218–228
- Yi Z L, Cui C F, Han S F, et al. Wireless big data and intelligent 5G. White Paper of Future Mobile Communication Forum, 2017 [易芝玲, 崔春风, 韩双锋, 等. 无线大数据与智慧 5G. 未来移动通信论坛白皮书, 2017]
- Liang Y C, Tan J J, Niyato D. Overview on intelligent wireless communication technology. *J Commun*, 2020, 41: 1–17 [梁应敞, 谭俊杰, Niyato D. 智能无线通信技术研究概况. *通信学报*, 2020, 41: 1–17]
- Tan J J, Liang Y C. Deep reinforcement learning for intelligent communications. *J Univ Electron Sci Technol China*, 2020, 49: 169–181 [谭俊杰, 梁应敞. 面向智能通信的深度强化学习方法. *电子科技大学学报*, 2020, 49: 169–181]
- Wang Z, Wei G F, Zhan Y L, et al. Big data in telecommunication operators: data, platform and practices. *J Commun Inf Netw*, 2017, 2: 78–91
- Qian L J, Zhu J K, Zhang S H. Survey of wireless big data. *J Commun Inf Netw*, 2017, 2: 1–18
- Huang Y D, Tan J J, Liang Y C. Wireless big data: transforming heterogeneous networks to smart networks. *J Commun Inf Netw*, 2017, 2: 19–32
- Zhu J K, Gong C, Zhang S H, et al. Foundation study on wireless big data: concept, mining, learning and practices. *China Commun*, 2018, 15: 1–15
- Yang Z, Zhou L, Zhao G Y, et al. Blockage modeling for inter-layer UAVs communications in urban environments. In: *Proceedings of the 25th International Conference on Telecommunications (ICT)*, Saint-Malo, 2018. 307–311
- Zhou L, Xiao L M, Li J H, et al. Path loss model based on cluster at 28 GHz in the office and corridor environments. In: *Proceedings of IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Montreal, 2016. 1–5
- Li H H, Li Y Z, Zhou S D, et al. Wireless channel feature extraction via GMM and CNN in the tomographic channel model. *J Commun Inf Netw*, 2017, 2: 41–51
- Li H H, Li Y Z, Zhou S D, et al. Static CSI extraction and application in the tomographic channel model. *China Commun*, 2019, 16: 132–144
- Fan C C, Zhong X F, Wei J. BS-to-ground channel reconstruction with 3D obstacle map based on RSS measurements. *IEEE Access*, 2019, 7: 99633–99641
- Sun X Y, Wu C, Gao X Q, et al. Fingerprint-based localization for massive MIMO-OFDM system with deep convo-

- lutional neural networks. *IEEE Trans Veh Technol*, 2019, 68: 10846–10857
- 15 Bishop C M. *Pattern Recognition and Machine Learning*. New York: Springer, 2016
  - 16 Zhu Y, Gong C, Luo J, et al. SVM-assisted realization and demonstration of indoor 4 Mb/s non-line-of-sight visible light communication with second-order reflection. *IEEE Photonics J*, 2019, 11: 1–17
  - 17 Zhu Y H, Gong C, Luo J H, et al. Indoor non-line of sight visible light communication with a Bi-LSTM neural network. In: *Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops)*, Dublin, 2020. 1–6
  - 18 Zhang Q Q, Guo H Y, Liang Y C, et al. Constellation learning-based signal detection for ambient backscatter communication systems. *IEEE J Sel Areas Commun*, 2019, 37: 452–463
  - 19 Liu V, Parks A, Talla V, et al. Ambient backscatter: wireless communication out of thin air. *SIGCOMM Comput Commun Rev*, 2013, 43: 39–50
  - 20 Liang Y C, Zhang Q Q, Larsson E G, et al. Symbiotic radio: cognitive backscattering communications for future wireless networks. *IEEE Trans Cogn Commun Netw*, 2020, 6: 1242–1255
  - 21 Zhang L, Tan J J, Liang Y C, et al. Deep reinforcement learning-based modulation and coding scheme selection in cognitive heterogeneous networks. *IEEE Trans Wireless Commun*, 2019, 18: 3281–3294
  - 22 Rashid-Farrokhi F, Liu K J R, Tassiulas L. Transmit beamforming and power control for cellular wireless systems. *IEEE J Sel Areas Commun*, 1998, 16: 1437–1450
  - 23 Zhang H Y, Dai H Y. Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks. *J Wirel Com Netw*, 2004, 2004: 202654
  - 24 Ge J G, Liang Y C, Joung J, et al. Deep reinforcement learning for distributed dynamic MISO downlink-beamforming coordination. *IEEE Trans Commun*, 2020, 68: 6070–6085
  - 25 Sun Y H, Zhao M, Zhang S H, et al. Aggregation transmission scheme for machine type communications. *Sci China Inf Sci*, 2017, 60: 100305
  - 26 Liang Y C. *Dynamic Spectrum Management: From Cognitive Radio to Blockchain and Artificial Intelligence*. Singapore: Springer, 2020
  - 27 Liu C, Wang J, Liu X M, et al. Deep CM-CNN for spectrum sensing in cognitive radio. *IEEE J Sel Areas Commun*, 2019, 37: 2306–2321
  - 28 Tan J J, Xiao S, Han S Y, et al. QoS-aware user association and resource allocation in LAA-LTE/WiFi coexistence systems. *IEEE Trans Wireless Commun*, 2019, 18: 2415–2430
  - 29 Tan J J, Zhang L, Liang Y C, et al. Intelligent sharing for LTE and WiFi systems in unlicensed bands: a deep reinforcement learning approach. *IEEE Trans Commun*, 2020, 68: 2793–2808
  - 30 Yan M, Feng G, Zhou J H, et al. Intelligent resource scheduling for 5G radio access network slicing. *IEEE Trans Veh Technol*, 2019, 68: 7691–7703
  - 31 Tan J J, Liang Y C, Zhang L, et al. Deep reinforcement learning for joint channel selection and power control in D2D networks. *IEEE Trans Wireless Commun*, 2021, 20: 1363–1378
  - 32 Zhou J H, Feng G, Yum T S P, et al. Online learning-based discontinuous reception (DRX) for machine-type communications. *IEEE Internet Things J*, 2019, 6: 5550–5561
  - 33 Zhang Q Q, Liang Y C, Poor H V. Intelligent user association for symbiotic radio networks using deep reinforcement learning. *IEEE Trans Wireless Commun*, 2020, 19: 4535–4548
  - 34 Yi W W, Zhang X, Wang W B, et al. Multi-agent deep reinforcement learning based adaptive user association in heterogeneous networks. In: *Proceedings of International Conference on Communications and Networking in China*, Chengdu, 2018. 57–67
  - 35 Cao Y, Lien S Y, Liang Y C. Deep reinforcement learning for multi-user access control in non-terrestrial networks. *IEEE Trans Commun*, 2021, 69: 1605–1619
  - 36 Sun Y, Feng G, Qin S, et al. The SMART handoff policy for millimeter wave heterogeneous cellular networks. *IEEE Trans Mobile Comput*, 2018, 17: 1456–1468
  - 37 Zhao N, Liang Y C, Niyato D, et al. Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks. *IEEE Trans Wireless Commun*, 2019, 18: 5141–5152

## Big data-based wireless communication technologies

Ying-Chang LIANG<sup>1</sup>, Junjie TAN<sup>1,2</sup>, Chen GONG<sup>3,4</sup>, Sihai ZHANG<sup>3,4</sup>, Jinkang ZHU<sup>3,4</sup>, Zhengyuan XU<sup>3,4\*</sup>, Yunzhou LI<sup>5\*</sup>, Xiaofeng ZHONG<sup>6</sup>, Shidong ZHOU<sup>6</sup>, Qianqian ZHANG<sup>1,2</sup>, Yang CAO<sup>1,2</sup>, Jungang GE<sup>1,2</sup> & Gang YANG<sup>1,2\*</sup>

1. *Center for Intelligent Networking and Communications, University of Electronic Science and Technology of China, Chengdu 611731, China;*

2. *National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China;*

3. *School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China;*

4. *Wireless-Optical Communications Key Laboratory, Chinese Academy of Sciences, Hefei 230026, China;*

5. *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China;*

6. *Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

\* Corresponding author. E-mail: xuzhy@ustc.edu.cn, liyunzhou@tsinghua.edu.cn, yanggang@uestc.edu.cn

**Abstract** With the popularization of an assortment of mobile applications, it can be predicted that future wireless communication systems have to satisfy increasingly demanding service requirements, rendering the development challenges to wireless communication technologies. On the other hand, applications of big data have gained dramatic successes in numerous areas. It becomes the focus in both academia and industry how big data can enable wireless communications. In the past five years, supported by the key program group of National Nature Science Foundation of China, we devoted to investigating the big data-based wireless communication technologies and exploring the new theories and new methods of applying big data to break the development bottlenecks of wireless communications. We define wireless big data (WBD) as the big data that are generated in wireless communications and meanwhile can assist wireless communications. This paper summarizes our research and practices. Specifically, we first analyze the sources and the characteristics of WBD, and then correspondingly propose the paradigm of wireless knowledge learning, the framework for smart wireless networks, and a cross-region, cross-institution, multi-role, and unified platform for WBD. Finally, the big data-based wireless environment analysis, wireless transmission and access technologies are developed and examined. This paper aims to provide the fundamental theories and application paradigms for the further development of big data-based wireless communication technologies.

**Keywords** wireless big data, wireless communication systems, intelligent networks, resource optimization, access control



intelligence, and machine learning techniques.

**Ying-Chang LIANG** is currently a professor with the University of Electronic Science and Technology of China. He was a professor with The University of Sydney, Australia, a principal scientist and technical advisor with the Institute for Infocomm Research, Singapore, and a visiting scholar with Stanford University, USA. His research interests include cognitive radio, symbiotic networks, dynamic spectrum access, the Internet-of-Things, artificial



and founding director of Wireless-Optical Communications Key Laboratory of Chinese Academy of Sciences. His research interests include optical wireless communications, mobile networking, artificial intelligence, wireless big data, sensing, ranging, and localization.

**Zhengyuan XU** received his B.S. and M.S. degrees from Tsinghua University, China, and his Ph.D. degree from Stevens Institute of Technology, USA. He was a tenured full professor at University of California at Riverside and later at Tsinghua University before he joined University of Science and Technology of China (USTC). He was founding director of the multi-campus Center for Ubiquitous Communication by Light (UC-Light), University of California,



**Yunzhou LI** received his Ph.D. degree from Tsinghua University, Beijing, China in 2004. He is currently a professor in Beijing National Research Center for Information Science and Technology of Tsinghua University. He mainly focuses on signal processing technologies and protocol design in wireless and mobile communications, including transmission, networking, positioning, and RF-sensing technologies.



**Gang YANG** received his Ph.D. degree from Nanyang Technological University, Singapore, in 2015. He is currently an associate professor with University of Electronic Science and Technology of China. His research interests include reconfigurable-intelligent-surface empowered wireless communications, and Internet-of-Things communications based on backscatter communications.