

# 6G 全场景按需服务: 愿景、技术与展望

承楠<sup>1</sup>, 陈芳炯<sup>2</sup>, 陈文<sup>3</sup>, 程志密<sup>4</sup>, 杨清海<sup>1</sup>, 李长乐<sup>1\*</sup>, 沈学民<sup>5</sup>

1. 西安电子科技大学通信工程学院, 空天地一体化综合业务网全国重点实验室, 西安 710071

2. 华南理工大学电子与信息学院, 广州 510006

3. 上海交通大学电子工程系, 上海 200240

4. 大唐移动通信设备有限公司, 北京 100083

5. 鹏城实验室, 深圳 518000

\* 通信作者. E-mail: clli@mail.xidian.edu.cn

收稿日期: 2023-10-30; 修回日期: 2024-02-27; 接受日期: 2024-03-25; 网络出版日期: 2024-04-28

国家重点研发计划(批准号: 2020YFB1807700)资助项目

**摘要** 6G 移动通信系统面临着复杂动态的网络, 其多样化的场景和个性化的需求为网络服务的提供和价值的实现带来了极大挑战. 因此, 6G 网络应转向“以服务为中心”的新范式, 综合运用先进网络技术和多维网络资源在全域全场景实现按需服务. 本文对 6G 全场景按需服务研究展开综述, 首先对按需服务的必要性进行说明, 梳理了全场景按需服务的研究价值; 然后对其内涵、挑战和机遇进行了阐述; 接着对于 6G 全场景按需服务的架构和关键技术进行了分析; 最后展望了 6G 全场景按需服务的前景和潜在研究方向.

**关键词** 6G, 全场景按需服务, 知识, 意图驱动网络, 人工智能

## 1 引言

第六代移动通信系统 (the sixth generation mobile communication system, 6G) 是 5G 网络之后的下一代移动通信系统. 近年来, 随着移动网络需求的大幅提升, 低轨卫星、大规模天线以及人工智能等技术的快速进步, 6G 移动通信系统已经引起了世界各国的重视, 并已经开展了大量研究<sup>[1]</sup>. 6G 网络将是移动通信的范式变革, 相比现有万物互联的 5G 网络, 6G 网络将提供天地一体泛在、高效、安全、虚实结合的万物智联<sup>[2]</sup>. 尽管 5G 网络已经显著提高了移动用户的体验, 实现了许多新的应用, 如自动驾驶汽车、智慧城市和远程手术等, 但仍存在一些限制, 阻碍了进一步的技术发展<sup>[3, 4]</sup>. 6G 旨在通过提供先进能力来克服这些限制, 包括 T 比特 (Tbps) 的数据传输速率、超低延迟和高度安全可靠的连接等. 为了实现这些目标, 6G 网络必须满足一系列严格的要求. 首先是确保无缝连接和普及覆盖,

**引用格式:** 承楠, 陈芳炯, 陈文, 等. 6G 全场景按需服务: 愿景、技术与展望. 中国科学: 信息科学, 2024, 54: 1025–1054, doi: 10.1360/SSI-2023-0323  
Cheng N, Chen F J, Chen W, et al. 6G omni-scenario on-demand services provisioning: vision, technology and prospect (in Chinese). Sci Sin Inform, 2024, 54: 1025–1054, doi: 10.1360/SSI-2023-0323

这意味着要克服高频谱段带来的挑战, 并开发高效智能的网络管理系统. 其次, 6G 网络必须具备可持续性, 需要能源效率技术来最小化能耗, 以及可再生能源的集成. 第三, 6G 网络必须是适应性和灵活性的, 以适应各种用例和技术, 包括物联网 (Internet of Things, IoT)、人工智能 (artificial intelligence, AI) 和边缘计算等<sup>[2]</sup>.

6G 网络的典型特征之一是基于空天地一体化网络的全域融合及泛在服务<sup>[5]</sup>. 近年来, 低轨卫星星座、平流层通信、无人机通信等技术得到了迅速发展, 将通信网络由地面域扩展至空、天、海乃至地下与深空等空间, 从而实现了全域融合的泛在连接, 使得信息服务可以任何时间、任何地点无缝地提供给任何用户, 极大地扩展了人类生产生活的可能性<sup>[6]</sup>. 目前, 空天地一体化网络的研究重点主要在低轨卫星星座组网、立体多域网络融合、软件定义与网络功能虚拟化等方面<sup>[7~12]</sup>. 6G 网络的另一个典型特征是网络场景的复杂化、多样化和个性化<sup>[13~15]</sup>. 5G 网络定义了三大场景<sup>[16]</sup>, 即增强型移动宽带 (enhanced mobile broadband, eMBB)、极低时延高可靠通信 (ultra-reliable and low latency communications, uRLLC)、海量机器类通信 (massive machine type communications, mMTC), 实现了移动通信网络从高带宽数据服务向垂直行业服务的转变, 催生了自动驾驶、远程医疗、工业物联网、智慧城市等新型应用, 初步达成了移动通信系统服务社会的功能. 在 6G 网络中, 覆盖的广域化、能力的多维化以及业务的多元化使得网络场景进一步扩展, 为网络服务带来了新的挑战. 从网络性能需求分类角度考虑, Jiang 等<sup>[17]</sup> 提出了 6G 网络的新型场景类别, 包括泛在移动宽带 (ubiquitous MBB, uMBB)、极高可靠低时延宽带通信 (ultra-reliable low-latency broadband communication, ULBC), 以及海量极高可靠低时延通信 (massive ultra-reliable low-latency communication, mULLC), 以应对泛在智能、触觉互联网、无缝互联、元宇宙等新型服务. IMT-2030(6G) 推进组定义了 5 大类共 8 种 6G 网络潜在场景, 例如沉浸式云虚拟现实 (virtual reality, VR)<sup>[18, 19]</sup>、感官互联、智慧交互、普惠智能、全息通信等<sup>[20]</sup>. 可以预见, 随着网络能力的提升和用户需求的持续涌现, 6G 网络及其后继者将面临更加复杂、多样和难以预知的场景. 因此, 6G 网络的核心目标是在全域融合和全场景下提供各种泛在、智能、高效的服务. 聚焦全域全场景网络智能管控与按需服务关键问题, 从应用 – 按需服务 – 物理设施 3 个层次出发, 设计 6G 空天地一体化全域全场景网络架构, 如图 1 所示. 框架分为应用场景层、按需服务层以及基础设施层 3 部分. 首先, 在应用场景层中, 面对全场景沉浸式服务 (如元宇宙、虚拟现实等), 6G 转变为 XaaS 平台, 将基础设施即服务 (IaaS) 层的通信和计算基础设施服务、平台即服务 (PaaS) 层的基本组件服务以及人工智能应用服务即服务 (SaaS) 层的服务进行整合. 随后, 应用场景层将场景特征以及用户意图抽取并输入到按需服务层. 按需服务层由 3 个子层构成: 认知子层首先进行用户意图转译识别、网络能力识别以及全场景本体构建, 并进一步提取 6G 全场景关键信息, 从而实现网络全场景认知. 决策子层主要负责全场景动态资源感知、全域网络资源按需调配等. 智能子层由 AI 模型库、知识库以及意图库进行构成. 认知子层、决策子层与智能子层相辅相成, 认知与决策形成的智能知识存储到智能子层中, 而智能子层又进一步依靠库知识辅助认知与决策, 实现实时精准的认知与决策. 最后, 基础设施层为按需服务层提供网络能力, 即按需服务层将编排策略下发部署到基础设施层, 通过云原生技术在全域联通的基础上实现异构资源协同编排.

移动通信网络已经成为现代社会的基石, 是数字基础设施的重要组成部分. 理所当然地, 移动通信网络的设计和运营目标应是通过数据的高效可靠交互, 实现各种数字服务, 以支持个人、社会、国家的各种需求. 然而, 现有对移动通信网络系统的研究主要来自无线通信和无线网络领域, 其研究高度集中于对无线技术的原理分析、协议设计、算法优化, 以及通信硬件设计等, 而对于场景、需求、服务等要素在移动通信网络中的影响与作用, 长久以来疏于深入研究, 更缺乏对于具体场景, 应如何优化网络设计以实现场景价值的探讨. 对移动互联网的研究表明, 场景需求来源于场景, 满足需求产生价

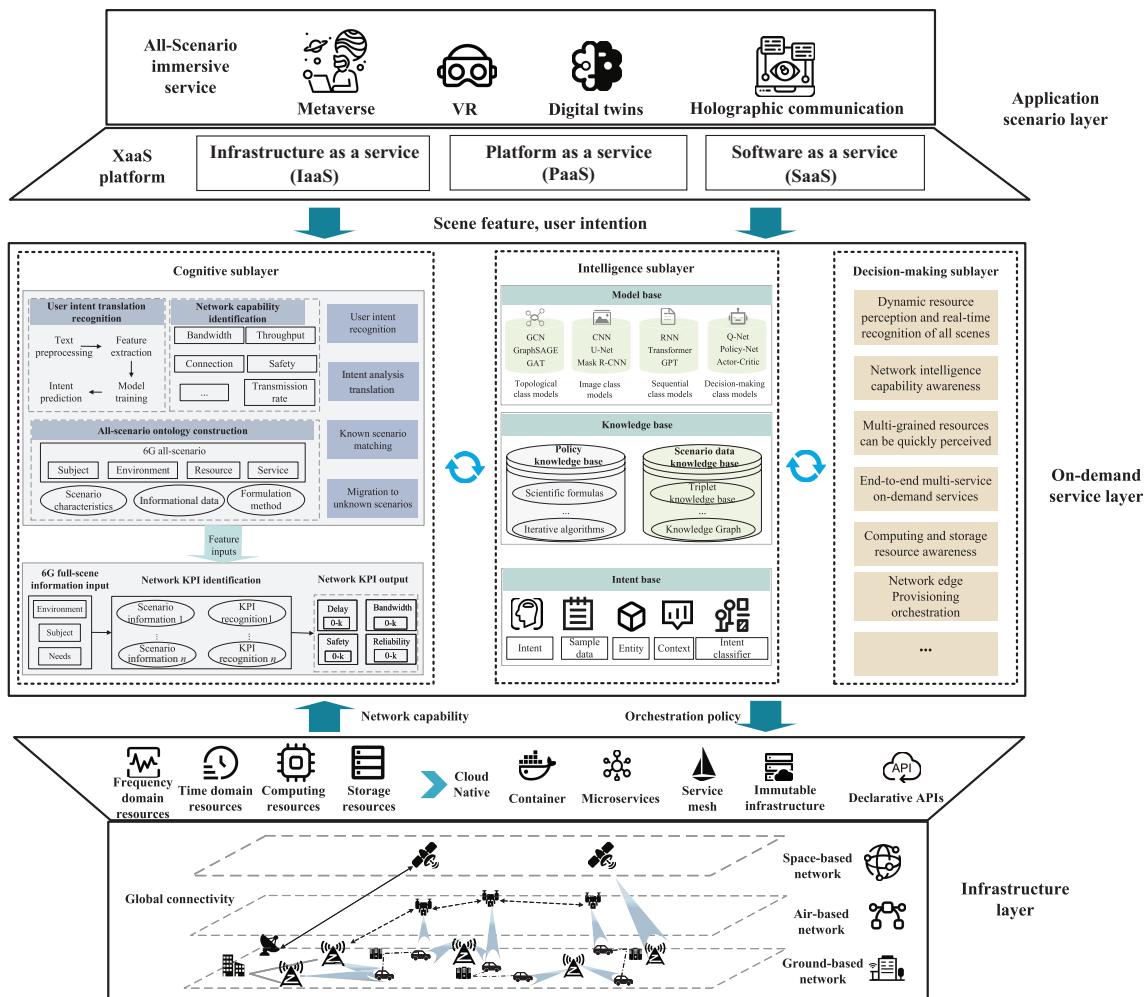


图1 (网络版彩图) 6G全域全场景网络架构

Figure 1 (Color online) 6G omni-domain omni-scenario architecture

值<sup>[21, 22]</sup>。场景化思维与方法为移动互联网带来了巨大的价值,有助于挖掘用户核心需求,产生沉浸式服务体验,提升个性化服务水平,通过满足用户需求实现用户价值,并促成企业实现商业价值。从具体场景出发,分析用户需求,并有针对性地设计相应服务与商业模式,已成为移动互联网行业的共识<sup>[23]</sup>。移动支付、购物直播、共享单车、外卖平台等成功的新型移动互联网服务在深刻改变社会与生活的同时,也证明了场景与需求在价值实现过程中的重要意义。

与移动互联网不同,移动通信网络长期处于“管道”定位,即负责将数据按照速率、时延、可靠性等要求在网络节点间进行传输,而不考虑数据的具体内容、含义等<sup>[24, 25]</sup>。具体的数据则依据服务提供商所提供的服务产生,在用户与分布在互联网上的大量服务器间传输。这种数据服务与数据管道的分工有其自身的优越性,即移动通信网络的设计者和运营者可以将资源集中在保障和优化网络自身的传输性能上,从而保证任何基于网络之上的服务都能得到一致的通信服务。然而,随着技术的不断发展,尤其在6G网络时代,这种方式的弊端也逐渐显现。首先,大量新型服务的出现,尤其是边缘服务与基于地理信息的服务,促使网络不但要作为数据传输的承载者,还要承担服务提供者的角色<sup>[26]</sup>。例如,在自动驾驶中,附近的自动驾驶车辆形成群组,协同执行自动驾驶服务,每个车辆平等地执行场景感

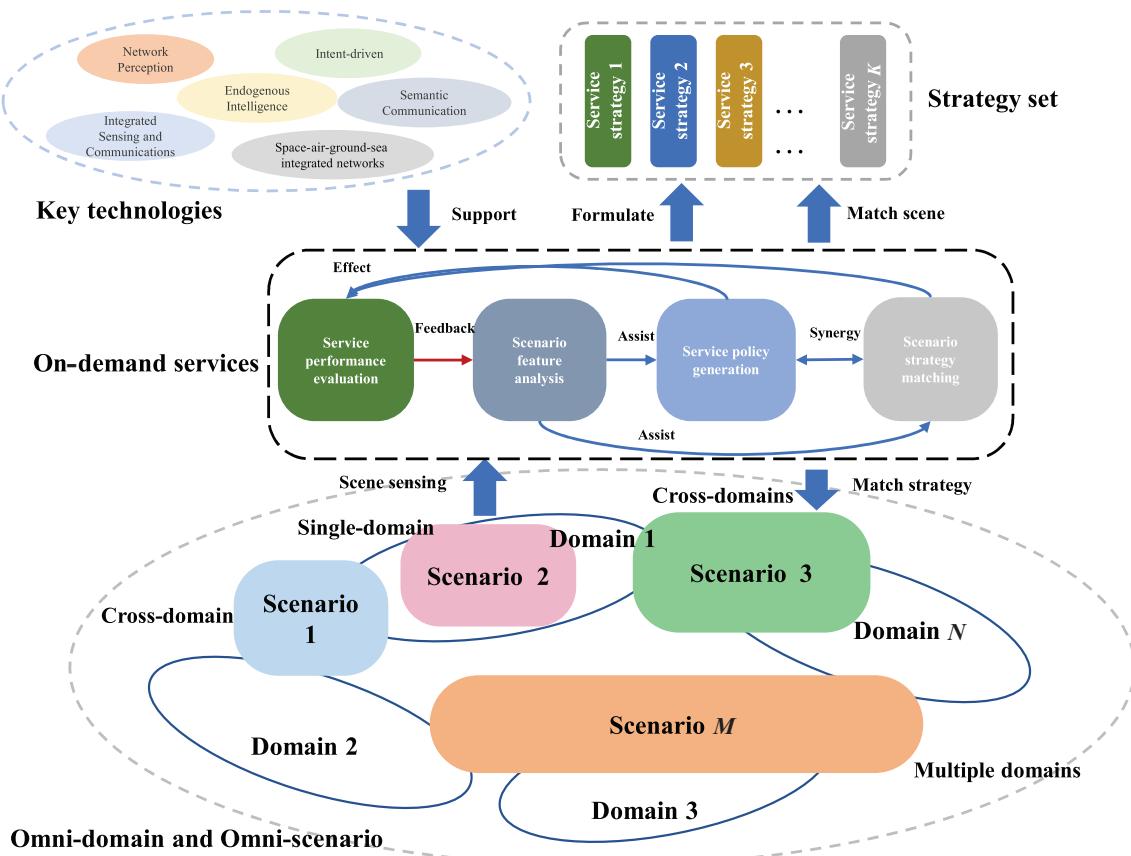
知、信息共享与行为决策等任务<sup>[27]</sup>. 在无人集群网络中, 也通常要求在没有互联网连接的场景下, 通过集群节点的互相协同执行复杂的任务<sup>[28]</sup>. 另外, 随着网络安全性越来越被重视, 用户对于数据隐私的要求越来越高, 与传统服务相比, 网络化服务不需要将包含其隐私的大量数据传送至服务器, 因而一定程度上保障了用户的隐私<sup>[29]</sup>. 其次, 这种服务与管道分离的方法, 难以进一步优化服务性能. 由于数据的解析和处理仅能在服务器上执行, 这会导致大量不必要的开销或时延. 实际上, 很多数据的处理是可以在网络边缘侧或者在数据传输“过程”中进行的. 在文献[30]中, 作者提出了一种邮箱理论, 即数据在传输过程中, 每个节点都可以对数据进行解读, 并根据节点自己的知识和情境信息(context information)对数据价值进行分析. 由此, 一些没有价值的信息, 即对于服务目标没有作用的价值, 会在传输的过程中被丢弃, 以节约网络资源. 再次, 相对于服务器, 网络更接近于用户, 因此对于用户场景和需求更加了解. 最后, 在这种网络作为数据管道的模式下, 移动通信网络运营商的价值难以实现. 运营商花费大量的开销进行网络的建设与维护, 仅获得了数据传输管道对应的收益. 因此, 6G 及未来的移动通信网络必然向移动互联网模式转变, 建立以服务为中心的模式, 更加深入地利用场景与需求信息, 实现全域全场景按需服务, 以实现价值. 当前, 在这一方面已有一些相关研究进展, 如基于服务的架构<sup>[31, 32]</sup> (service-based architecture, SBA)、算力网络与在网计算<sup>[33, 34]</sup>, 以及网络能力开放<sup>[35]</sup> 等. 然而, 对于无线网络场景的研究以及其在网络按需服务中的意义, 还没有得到广泛的关注. 为充分实现 6G 网络的价值, 应对基于网络场景的网络按需服务理论与关键技术进行深入的探讨和研究.

全场景按需服务, 是考虑网络场景与服务的特性与需求, 通过高效调度多域网络中的多维资源, 提供各种网络服务以实现需求的过程. 实现全场景按需服务, 核心是对网络的范式进行变革, 建立以服务为中心的新范式. 这种新型网络范式有两个要义, 即场景化与服务化. 对于场景化, 需要对于场景的类型甚至具体要素进行精确识别, 以分析场景的特性、资源的可用性、各类约束以及场景化的需求, 并针对此场景制定特定的服务策略; 对于服务化, 需要摒弃传统网络与服务分离优化的思想, 将网络及其支撑的服务考虑为一个整体 (即网络化服务系统), 从而实现面向服务的网络优化, 而非单纯对网络的指标 (如传输速率、丢包率等) 进行优化. 通过全场景按需服务, 可以提升网络的总体服务能力, 更加高效地利用网络资源, 针对复杂、个性化的场景提供精准的服务, 从而实现网络的价值. 然而, 新的按需服务范式也同时带来了新的挑战, 并要求网络提供新型能力, 比如场景感知能力、多域网络融合能力、多维资源协同能力, 以及服务策略快速制定及泛化能力等. 为实现这些能力, 网络需要对于各种新技术进行更加高效的支持、融合和应用, 例如通感一体技术、空天地海融合网络技术、意图驱动网络技术、AI 与通信深度融合技术、语义通信技术、数字孪生技术等. 6G 网络按需服务及其关键技术如图 2 所示.

本文旨在对于 6G 网络全场景按需服务进行深入探讨, 详细介绍其背景、挑战、架构与关键技术等. 第 2 节对相关定义的内涵进行了介绍, 并探讨了按需服务的挑战与在 6G 时代的机遇. 第 3 节对现有移动通信网络服务架构进行了调研, 并提出了一种新的基于知识和意图双驱动的 6G 全场景按需服务架构. 第 4 节对全场景按需服务的关键技术进行了深入讨论. 第 5 节指出了 6G 按需服务前景和潜在研究方向. 第 6 节对全文进行了总结.

## 2 内涵、挑战与机遇

网络按需服务的核心是通过将网络的能力与“需”进行高效匹配, 通过满足“需”从而实现价值. 其最终愿景为, 服务随心所想, 网络随需而变, 资源随愿共享. 因此, 为实现按需服务, 首先应对“需”进行深入理解.“需”可以分为两个层面, 即需要(desire)和需求(demand). 从经济学角度出发, 需要一



## 6G Omni-domain and Omni-scenario network

图 2 (网络版彩图) 6G 网络按需服务及其关键技术

Figure 2 (Color online) 6G on-demand service provisioning and key technologies

般是主体对于欲望和要求的主观表达,即“想要什么”,该表达不受客观因素(如负担能力、商品价格)的限制,且具有一定的模糊性<sup>[36]</sup>;而需求指在有能力负担和有意愿购买的前提下,对于某种特定商品或服务的要求。因此,按需服务也可分为两个层次,即低层次按需服务和高层次按需服务。低层次按需服务即用户的需要通过固定类型的服务需求表达,网络通过模板化的资源分配满足服务需求,这也是现阶段网络按需服务的实现形式<sup>[37~40]</sup>。具体来说,网络需求主要由用户使用的各种应用(application, APP)提供,如微信消息、语音通话、抖音直播等,用户在这些固定类型的服务中选择使用。网络主要通过一些模板化的资源配置实现服务关键性能指标(key performance indicator, KPI)或预定义的服务等级协议(service level agreement, SLA)。这种方式实现较为简单,但是难以满足动态变化和长尾分布的个性化需要;同时,网络运营商只起到管道作用,无法发挥其接近用户与场景的优势,因此低层次按需服务无法完全实现通信网络通过满足需求产生价值的愿景。相比而言,6G 应追求高层次按需服务,即深入分析场景化和个性化的需求,并同时通过灵活的能力调配和组合提供定制化的服务,从而尽可能满足用户需要,实现网络价值。

按需服务是一个非常广泛的命题,而实现按需服务的一个关键要素是场景。一般来讲,场景是用户与所在网络、所在环境的集合,是需求的实际来源<sup>[14]</sup>。因此,须对某一类、某一个特定的场景进行

分析, 确定其组成、需求、能力、约束等要素, 并在此之上实现场景化按需服务。移动通信网络在发展过程中曾对场景进行多次探讨与定义, 比如 3GPP 对于 5G 三大场景的定义<sup>[16]</sup>, 以及近来国际电信联盟 (International Telecommunication Union, ITU) 对于 6G 六大典型场景的定义<sup>1)</sup>等。本文认为, 这些对于网络场景的定义方式虽然较好地归纳了典型的场景类别, 并引导了相关赋能技术的发展方向, 但是其定义仍然较为粗糙, 仅表达了头部需求, 而无法表达大量存在的细粒度个性化需求, 这些需求甚至可能是未知或未被挖掘的, 难以满足 6G 按需服务的要求。因此, 为实现全场景按需服务, 满足行业化、个性化需求, 本文提出了一种基于本体论的场景表征方法, 以全面、详细地表征任一特定的场景。对于场景及其功能, 国内外不同领域的学者进行了探索, 并做出了不同的定义。例如, Scoble 等<sup>[41]</sup> 将场景定义为由空间环境、行为情境和心理氛围 3 个要素构成。Kenny 和 Marshall<sup>[42]</sup> 提出场景为用户所处的特定情境以及该情境下所产生的需求。Silver 等<sup>[43]</sup> 提出场景的要素为特定空间和时间下的地点、时间、时刻。这些相关研究表明, 场景的构成要素主要包含人、情感、空间环境、时间环境、事件、需求等要素。然而, 在移动通信网络领域, 对于网络场景进行系统化的表征尚未有相关研究。本体论<sup>[44]</sup> (Ontology) 起源于哲学领域, 研究世界本质的存在以及对于任何领域内真实存在的客观描述。因此, 利用本体论, 可以回答以下问题: 对于通信网络场景, 本质上有什么实体、过程、属性, 以及它们之间的依赖关系如何? 而这一问题, 正是对场景进行规范化描述和表征所需要回答的。

本文提出了一种 6G 网络场景本体架构, 如图 3 所示。网络场景的本体包括主体、环境和需求三大要素。其中, 主体是场景中需求的主要来源, 也是网络服务的对象。常见的场景主体包括人、车辆、无人机、公司、行业等。在网络场景中, 主体不仅请求服务内容或提供需求意图, 其客观属性和情感等主观属性还影响着服务提供商对于场景的分析, 如场景价值、服务优先级、服务模式与提供方式等。环境包括物理环境以及网络环境, 物理环境即主体所在的客观物理环境, 包括天气、位置、地形、场所、道路与建筑等, 而网络环境指能够为主体提供服务的网络环境, 主要包含网络状态、节点、覆盖情况、资源情况等。最后, 需要指主体请求网络提供的内容或业务, 或者以意图等形式表达的自身的主观需要。在引入场景本体后, 全场景按需服务即对于任意一个场景, 网络首先对场景本体各要素进行认知, 并进一步分析场景需求和网络能力。最后, 调配各类网络资源, 匹配场景需求和网络能力, 实现价值。

6G 网络需要在各方面性能指标上超越 5G 网络, 网络的复杂性和动态性也急剧增加。同时, 场景化会带来“千人千面”的需求。因此, 6G 网络在需求层面具有以下特点:

- 极致性。与 5G 相比, 6G 的各项 KPI 均需实现量级级别的提升。根据文献 [1], 6G 的峰值速率达到 Tbps, 网络延迟小于 1 ms 甚至实现“零时延”, 连接数量达到  $107/\text{km}^2$ , 能量效率实现 100 倍提升, 频谱效率实现 3 倍提升。由于现有物理层通信技术已经逼近信道容量, 如何实现这些要求极高的性能指标是 6G 技术发展需要面临的重要挑战。

- 维度扩展。6G 网络不但需要支持日新月异的新型服务, 如元宇宙、自动驾驶等, 还需要支持社会、行业, 以及个人的各类应用, 其多样化需求使得 6G 网络需要扩展 KPI 的类别, 定义新型 KPI 指标。已有相关研究对于 6G 新 KPI 进行了探讨<sup>[5, 45, 46]</sup>, 如覆盖度、定位精度、开销效率、智能等级、安全等级等。然而, 对于智能等级、安全等级等新指标, 当前暂没有公认的定义以及与具体服务性能的对应关系, 需要进一步研究。

- 动态性。在 6G 网络中, 以空天地海一体化网络为代表的高动态网络极大地改变了网络需求的特性。网络必须满足不同动态性情况下的需求, 如静止、步行、汽车、高铁、无人机、飞机等。不同的动态性情况下, 即使同样的业务也可能会产生不同的需求。例如直播业务, 在静止或慢速移动的情况下, 通常会对分辨率具有较高要求, 而快速移动时, 往往只对直播流畅性有要求。

1) <https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2030/Pages/default.aspx>.

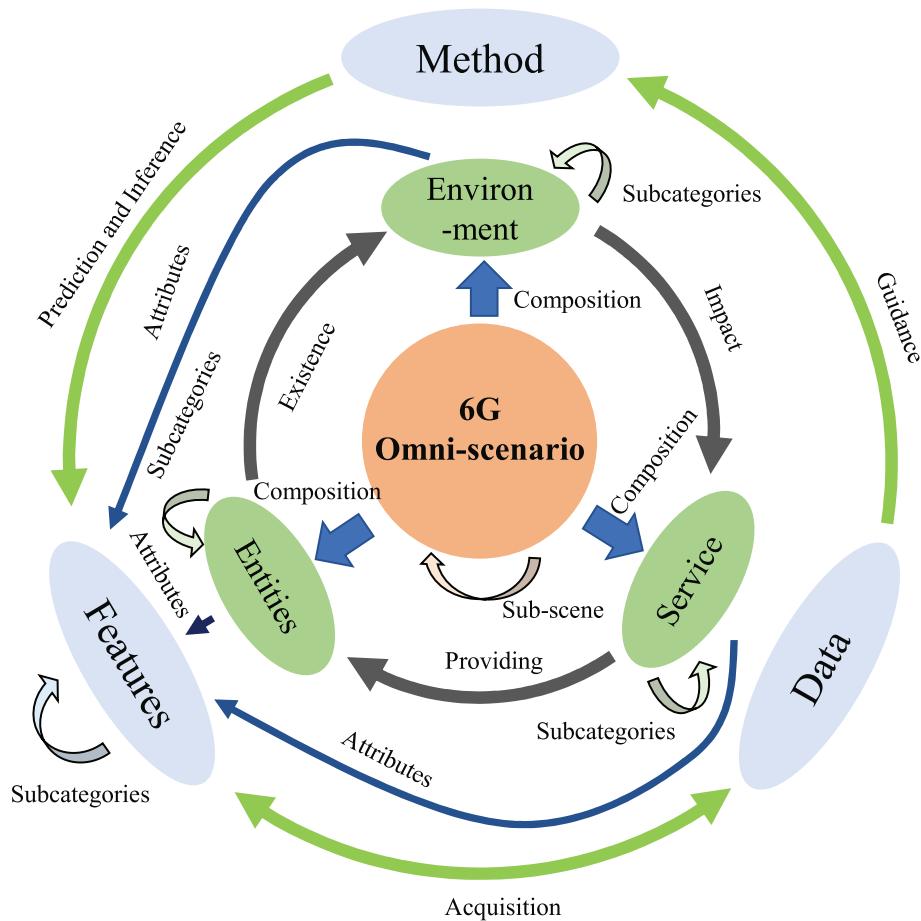


图 3 (网络版彩图) 6G 网络场景本体架构图  
**Figure 3** (Color online) 6G network scenario ontology architecture

• 个性化. 5G 网络开启了移动通信网络服务范围扩展的序幕, 将其服务对象由移动宽带通信业务扩展为社会各行各业. 6G 将进一步将服务对象扩展至每一个主体, 为每一个服务主体提供个性化, 定制化的服务. 然而, 由于服务主体千差万别, 其个性化需求必然是难以捕捉和处理的, 这对于 6G 网络的网络感知, 需求提取和智能处理能力提出了新的挑战.

• 场景化. 场景是 6G 网络进行按需服务的基本单元, 如前文所述, 场景的主体、环境和需求三大要素决定了其需求. 然而, 对于按需服务, 如何定义场景的粒度是十分重要的课题. 如使用较粗的粒度, 如 ITU 定义的 6G 典型场景, 其需求提取和资源调配的复杂度较低, 然而性能也会出现一定的下降. 如使用较细的粒度, 则需要对场景进行细致的感知, 以及对用户需求进行精确的分析, 并可能需要设计专有的服务模式或资源分配机制, 这会大大提升按需服务能力, 但同时会产生大量的开销.

• 模糊性. 用户的需求往往是模糊的, 而现阶段网络运营商或服务提供商仅仅在已有的服务集合中选择一种与用户需求最接近的提供给用户. 这种模式早已被证实是无法真正满足用户需要的. 如福特公司在对消费者进行调查时, 用户往往给出的是“想要一匹跑得快的马”, 而真正的需求则应该是“一种更快的交通工具或方案”. 福特公司认清了真正的需求, 催生了汽车的诞生. 因此, 只有对用户需求进行深入的分析和研究, 方能从模糊的表征中认清真正的需求.

6G 网络是对移动网络范式的一次重大革新, 从技术层面看, 6G 网络在空间上融合了空天地海多域网络<sup>[47~49]</sup>, 在频域上利用了全频谱<sup>[50,51]</sup>, 在资源上综合利用了通信、感知、计算等多维资源<sup>[52,53]</sup>, 其核心技术特征似乎难以归纳。然而, 如果从更深层面上看, 6G 的最终目的是为所有网络主体提供按需服务, 实现网络价值。因此, 6G 的范式转变应是从以网络为中心向以服务和价值为中心转变, 其最核心的特征是按需服务。按需服务, 即在 6G 移动通信系统中, 网络可以根据用户的需求实时提供不同类型和质量的服务。具体而言, 网络能够根据用户的请求动态调整服务参数, 以满足用户多样化的通信需求, 如低延迟、高带宽、大规模连接等。本文将按需服务视为一种网络范式, 其中服务的提供是基于实时和动态的用户需求, 并以定量的方式界定按需能力。定量需求包括具体的技术参数, 如延迟、带宽等。对于典型性 6G 场景, 可以直接映射相应的定量需求, 从而提供按需的资源和服务; 对于未知场景及模糊性需求, 本文利用意图驱动技术, 将其转译为网络可以识别的定量需求, 进而实现精准的资源调配和服务供应。然而, 6G 网络中, 场景和服务的多样性、需求的个性化和模糊性、网络的动态性和复杂性等特性给 6G 网络按需服务带了极大的挑战。本文对 6G 网络按需服务的挑战进行了分析, 并表述如下。

**挑战 1: 如何实现网络场景精准认知。** 全场景按需服务的前提是对网络场景进行精准地认知、分析其需求和能力。然而, 对网络场景进行全方位的认知, 受限于网络的感知能力和处理能力。传统无线网络的感知能力有限, 一般仅对通信的信道条件、用户位置、速度等简单信息进行感知, 难以完全体现网络场景信息。因此, 若对场景进行全方位感知, 需要网络感知技术的发展和感知能力的增强, 综合利用视频、图像、雷达、激光雷达、定位等感知技术提取多维场景信息。另一方面, 对场景的认知(如提取场景特性、需求、类别等)涉及大量场景数据的处理和学习。由于场景认知是按需服务的前置任务, 场景认知必须在尽量短的时间内完成, 以增强服务效率, 尤其对于时延敏感型业务, 敏捷快速的场景认知更为关键。这对于场景认知高效算法设计、网络数据处理、算力调配等方面均带来较大挑战。

**挑战 2: 如何针对极端多样化场景实现按需服务。** 6G 网络的重要特点之一是场景的多样化, “千人千面”的场景使得传统的服务模式与按需服务方法无法适用于每个场景。因此, 为实现按需服务, 需要对每个特定的场景设计有针对性的按需服务策略与方法, 甚至设计部署对应的应用或服务模式, 这对于人力和资源有限的网络服务提供商或运营商是极大的挑战。另一方面, 在实际部署中, 如此极大量独立的策略或算法也给网络运维带来了非常大的压力。因此, 设计可在不同场景之间实现泛化的算法, 或对不同场景进行合并等也是非常重要的研究课题。

**挑战 3: 海量场景共存时如何实现按需服务。** 随着 6G 全域网络的发展, 网络按需服务的对象逐渐由单一或较少场景向海量场景共存的情况转变, 再叠加上场景的多样性, 这一问题被称为“全域全场景按需服务”问题。例如, 6G 网络中可能同时存在地面大量用户需求高清视频业务场景、海洋用户需求卫星图像回传场景、空中用户需要接入移动互联网场景、多服务共享场景等, 这些场景特性不同, 需求迥异, 却需要网络利用有限的资源对所有场景同时进行按需服务。这对于网络的响应速度、场景优先级划分、资源调配效率与灵活性等能力都提出了很高的要求。

这些挑战都是困难且艰巨的, 在传统网络中几乎无法解决。幸运的是, 6G 网络同时带来了许多新的技术的发展与成熟, 为解决这些挑战带来了一丝曙光。例如, 随着空天地海一体化技术的发展, 卫星网络、空中网络、海洋网络等为 6G 带来更多资源和更大的灵活性, 因而可以容纳更多服务<sup>[54]</sup>。此外, 6G 网络可以综合利用包括通信、存储、计算、感知、拓扑、控制等多维资源, 这使得网络可以更加高效和灵活地分配资源, 以更好地满足多样化场景的需求。同时, 通过场景特征认知, 服务等级划分和多维网络资源的协同管控, 可优化部署系统对多服务共享的调度策略。更为重要的是, 6G 网络是智慧内生的, 天然可以利用各种智能方法, 以解决环境复杂、网络动态、需求多样等传统按需服务方法难以解

决的问题. 因此, 如何利用这些新的机遇, 更好地提供 6G 网络全域全场景按需服务, 是业界需要共同探讨的问题.

### 3 6G 全场景按需服务架构

近年来, 国内外研究机构不断提出各种 6G 网络架构, 相关研究不断演进. 这些架构旨在满足未来通信需求的不断增长, 每个都呈现出独特的特点和设计思想. 首先, 北美 NextG 联盟的 6G 网络架构<sup>[55]</sup>着重于可信、安全、弹性, 并强调数字世界体验、成本效率、分布式云网系统和 AI 内生无线网络的集成. 这一架构注重网络稳定性、用户体验和成本效益, 通过分布式云网和 AI 技术实现更广泛的覆盖和更智能的网络, 同时注重可持续性, 为未来通信技术奠定坚实基础. 而欧盟 Hexa-X 项目的 6G 网络架构<sup>[56]</sup>以“蓝色智能网络”用于人工智能和可编程网络, “绿色灵活网络”用于无线电资源管理和支持部署, 以及“橙色高效网络”用于简化无线接入网和核心网架构为特点. 这种多层次的设计将促进智能、高效和灵活的网络发展, 以满足未来多样化的应用需求. 其次, 日本 NICT 研究所的 6G 网络架构<sup>[57]</sup>则强调了扩展传统陆基移动通信系统, 光通信网络的高容量和低延迟支持, 以及引入数字孪生以处理多领域信息<sup>[58,59]</sup>. 数字孪生意图构造精确反映现实对象变化的虚拟模型, 数字孪生技术和网络切片范式之间相辅相成, 可以大幅度地提高 6G 网络的灵活性、可扩展性、适应性和智能性. 这一架构的特点有望实现更广泛的覆盖、更智能的网络和更多领域的应用. 此外, 中国移动提出的“三体四层五面” 6G 网络架构<sup>[60]</sup>为通信提供了多维度的设计思想. 它注重空间、逻辑和功能视图, 拓宽了通信空间, 提高了跨层协同, 加强了功能支持, 以满足未来通信需求. 中国电信则提出“三层三面”的 6G 网络架构<sup>[61]</sup>, 着重于资源泛在化、功能融合化和服务智慧化. 这一架构特点意味着 6G 网络将资源更广泛地分布, 功能更紧密地整合, 以提供更智能和个性化的通信服务. 总的来说, 这些不同国家和组织提出的 6G 网络架构都致力于推动通信技术的发展, 以满足未来社会日益复杂和多样化的通信需求. 它们的特点包括可信、安全、智能、资源分布、跨层协同、功能整合、成本效率和可持续性, 为 6G 技术的实现提供了多样的方法和视角.

为面向“全域全场景按需服务”需求, 本文在 6G 网络架构的基础上, 定义了按需服务架构, 并命名为知识与意图双驱动的 6G 按需服务架构, 旨在实现全域全场景网络向按需服务引擎提供场景数据用于全场景感知, 提供服务能力用于网络能力认知; 按同时需服务引擎向全域全场景网络下发服务策略, 并收集服务质量反馈. 这一架构以全场景感知、网络能力认知和智能服务编排为核心, 支持多领域智能服务的提供. 其特色是“知识与意图双驱动”, 基于全场景网络知识体系和网络意图管控体系, 建立知识库和意图库, 用于支撑按需服务引擎的自主化智能化运行.

如图 4 所示, 按需服务引擎包含 4 个要素, 分别是知识与意图、全场景认知、网络能力认知和智能服务编排. 其中, 知识与意图双驱动是本架构的核心理念, 它建立在全场景网络知识体系和网络意图管控体系的基础上. 这一理念构建了知识库和意图库, 它们是架构的关键支撑. 知识库储存了全场景网络知识体系, 如全场景知识图谱、关键特征推理方法、场景画像、知识驱动的资源调度以及知识迁移的服务编排, 这些知识对于全场景感知和智能编排至关重要. 意图库则涵盖意图驱动的网络管控体系, 包括意图感知、意图转译、意图传输和意图保障. 这两者共同支撑了按需服务引擎的自主化智能化运行. 而全场景感知是架构的另一个核心组成部分. 它的任务是感知各种场景的特征, 包含城市、农村、高空、深海、太空等各个领域. 这种全场景感知能力是 6G 网络的一个重要特性, 未来网络需要适应不同场景的需求. 全场景感知依赖于知识库, 通过对各种场景的知识学习和理解, 对全域全场景网络进行主体、需求以及环境的表征并将其提取为知识, 同时还对用户意图进行识别以及转译分析,

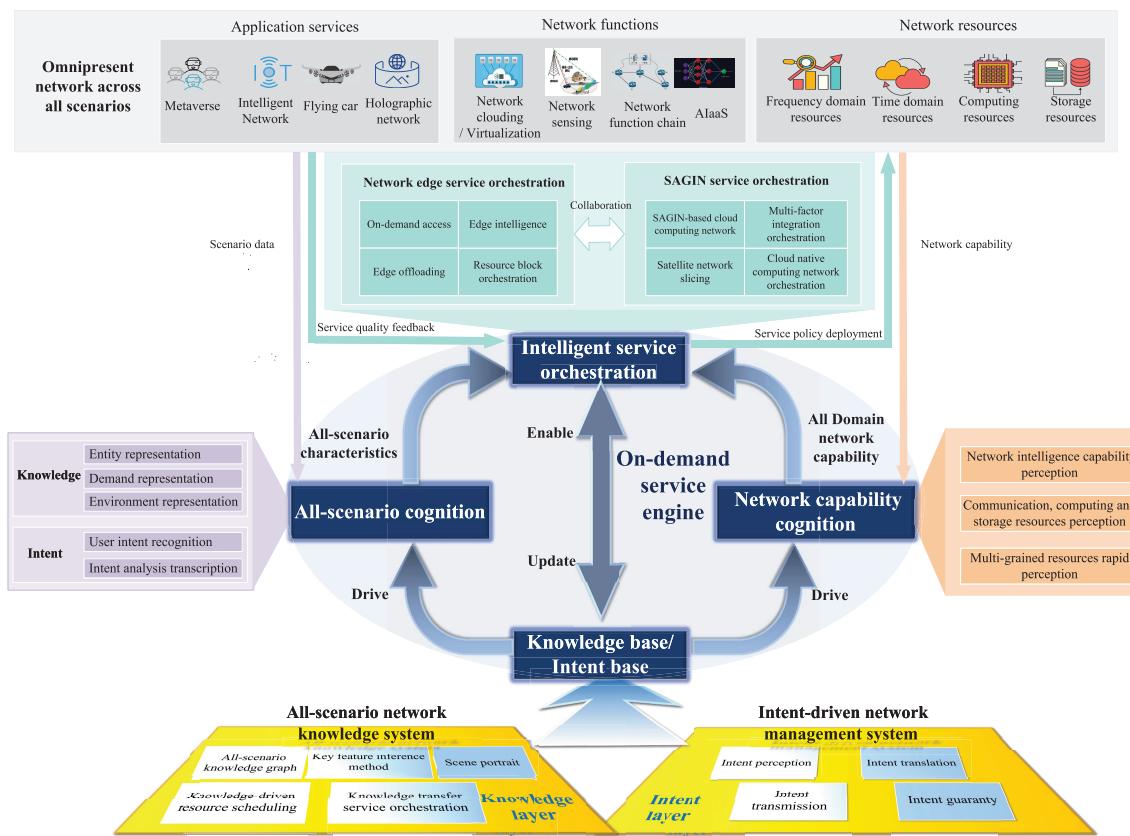


图 4 (网络版彩图) 按需服务引擎架构图

Figure 4 (Color online) On-demand service engine architecture.

实现全场景特性的提取, 为智能服务提供了重要信息.

网络能力认知则旨在了解全域网络的能力, 包括网络智能能力、通算存资源感知, 以及多粒度资源快速感知. 网络能力认知是一个动态的过程, 需要不断监测网络性能, 并根据不同场景的需求进行调整. 它依赖于意图库, 以了解网络的需求和意图, 从而进行智能编排, 以满足不同场景下的服务要求. 最后, 智能服务编排是根据全场景认知和网络能力认知, 以知识库和意图库为支撑, 实现对不同场景下的智能编排的过程. 这确保网络能够实现网络边缘服务编排, 包括按需接入、边缘智能、边缘卸载, 以及资源块编排; 同时还能协同实现空天地海服务编排, 包括空天地一体星云算网、星上切片、多要素融合编排, 以及算网原生编排.

在整个架构中, 知识库/意图库分别驱动全场景认知和网络能力认知; 根据全场景认知输入的全场景特征和网络服务能力输入的全域网络能力, 智能按需服务编排在知识与意图的使能下, 实现对网络边缘和空天地海全域网络的智能编排. 编排结果亦可以更新完善知识库和意图库. 整体而言, 按需服务引擎的 4 个要素形成双闭环, 循环反馈支撑网络的按需服务.

本文提出的知识与意图双驱动的 6G 全场景按需服务架构可适用于广泛的应用领域. 它可用于智能城市的建设, 实现城市各个领域的智能化管理, 包括交通、环境、能源等. 它还可以用于智能农业, 提高农业生产效率, 实现农业的智能化管理; 应用于智能医疗, 提供智能化的医疗服务, 包括远程医疗、智能诊断等. 最后, 它可用于智能交通, 实现交通的智能化管理, 提高交通的安全性和效率.

## 4 6G 全场景按需服务关键技术

本节对 6G 全场景按需服务的关键技术进行了阐述与讨论, 主要包括场景认知技术、意图驱动网络、知识驱动技术、云原生、语义与语用通信、智能网络优化及多维资源融合等.

### 4.1 场景认知技术

网络场景认知主要包括对场景本身的认知和场景相关的资源/网络能力认知两方面. 其中, 场景本身的认知主要是对场景的类型进行识别, 或者对场景本体进行认知, 并进一步提取场景需求、环境、约束等属性; 场景相关的资源/网络能力认知主要是对于场景对应的网络所具有的服务能力进行认知. 而按需服务则是将场景需求与能力进行合理匹配, 实现需求以及对应的价值. 传统的移动网络场景认知主要体现在对信道条件的感知上, 如利用信道估计或信道测量的方法对上下行无线信道的增益进行感知, 以进行资源分配或多输入多输出 (multiple-input multiple-output, MIMO) 预编码等<sup>[62]</sup>. 然而, 利用对环境的感知, 将提升信道感知的准确性或降低其复杂度. 例如, Sun 等<sup>[63]</sup> 提出了一种基于环境特征的快速路损预测方法, 仅通过距离、障碍等数个低维度环境信息与一个决策树模型, 即可实现传统需要大量数据拟合的路损模型, 这对于 6G 网络新型场景数量众多的状况尤为重要. 除了信道信息, 对于网络场景更全方面的认知方法目前的研究还比较初级. 在文献 [13] 中, 将知识图谱技术应用于网络场景的认知, 通过对网络场景要素的分解形成了对 6G 网络场景的规范化描述, 定义了 6G 网络场景本体, 并根据 3GPP 标准文档数据提出了一种基于本体的 6G 网络场景知识图谱构建方法. 同时, 提出了一种基于知识图谱嵌入的网络场景认知方法, 通过对场景知识图谱节点和关系的高效嵌入学习, 实现了对场景特征节点的预测推理, 并取得了较高的准确度, 最终完成了对网络场景的初步认知. 在文献 [64] 中, 针对 6G 网络中业务多样性及需求个性化的挑战, 提出了基于关键性能指标的多维度场景聚类分析方案. 鉴于 6G 场景数据的庞大和缺失问题, 该方案利用不完全多视图聚类技术, 并运用肘部法和轮廓系数法进行参数调优. 实验证明, 该方案在不完整的场景数据集中保持了稳定的聚类效果, 取得了显著的轮廓系数值. 同时, 通过对比不同缺失数据比例的聚类实验, 表明了其对多维度数据缺失的有效处理能力. 最终, 结合原始数据和聚类标签, 成功提炼出了 11 类场景的关键性能指标特征, 为未来 6G 网络中新兴场景及业务的发展提供了实质性的方法基础和理论指导.

在未来 6G 全场景背景下, 对网络资源各个维度的精细化感知是针对场景化个性化需求, 实现按需服务的技术关键. 从粒度、时空频域、通信感知计算等多维度对网络资源感知, 进而完成资源的精细调度, 提高网络端到端服务质量, 以满足未来丰富的业务应用和极致的性能需求. 基于通信网络架构, 6G 网络感知可以从以下 5 个层面进行: 首先是物理资源感知, 具体来说是能动态感知网络链路带宽资源的状态, 如可用带宽、时域态势等, 针对小区密度大、网络状态信息 (network status information, NSI) 获取困难等问题, 通过多层次、细粒度感知技术获取物理层资源使用状态<sup>[65,66]</sup>, 并在此基础上为授权的业务分配网络资源. 第二是 MAC 层资源感知, 主要是感知资源的饱和度, 对时变网络负载进行快速估计. 6G 应用场景将会在 5G 的基础上继续向高级版本演进, 有更高的服务质量 (quality of service, QoS) 要求, 为避免时变网络负载带来的资源利用效率低下或资源冲突问题, 需要感知资源饱和度, 对网络负载进行估计<sup>[67~69]</sup>. 第三是网络层资源感知, 主要是感知网络的连接性, 发现网络拓扑结构和进行拓扑预测<sup>[70,71]</sup>. 未来高速移动用户会导致端到端链路连接不稳定, 此外复杂网络对安全性的要求与日俱增, 网络拓扑识别和预测是应对挑战的关键技术, 一方面网络拓扑识别是网络系统安全态势感知的重要方法, 能准确快速地发现网络的安全状态需求, 另一方面可通过感知网络连接性, 预测网络拓扑结构来感知通信链路变化, 对网络资源进行按需分配<sup>[72~75]</sup>. 第四是传输及应用层的网络

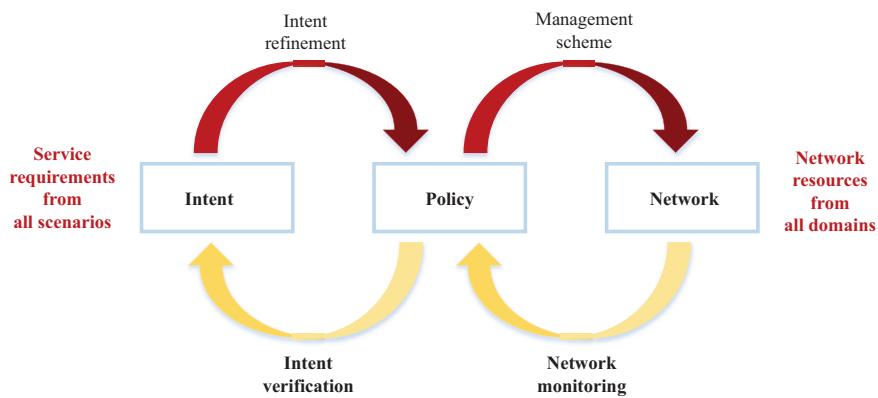


图 5 (网络版彩图) 意图驱动网络架构图  
Figure 5 (Color online) Intent driven network architecture

感知, 需要感知资源对于业务的可持续性, 并根据感知结果考虑有效、智能的任务调度算法 [76, 77], 以满足大量的移动设备和应用对实时通信和密集型计算的要求 [78~80]. 第五是感知资源的可关联性, 挖掘不同维度资源的隐含关系. 资源关联检测是指发现潜在的资源之间的关联性, 这种关联对于网络性能优化、异常检测、网络安全、服务质量改进等方面具有重要价值. 第七是信息感知的集约化, 6G 对场景感知的集约化程度也提出了更高的要求, 因此探索如何使用通信信号回波, 完成对环境信息的感知, 并最终实现通信 – 感知一体化是未来 6G 感知技术的一个重要研究方向 [81~83]. 第八是感知的智能化, 当前数字孪生技术的发展提供了对真实物理世界的数字模拟与真实再现, 因此可以通过数字孪生系统, 实时了解当前的环境状态并在数字孪生系统中预测分析环境的未来状态变化, 从而实现感知的智能化 [84~87].

## 4.2 意图驱动网络

如图 5 所示, 意图驱动网络依托意图自主转义和意图闭环验证两种关键技术, 实现意图领会、策略定制和网络管控三级功能. 意图驱动网络集全场景服务意图深度挖掘能力、全域网络状态全局感知能力、网络配置管理实时优化能力、网络资源精准调度能力为一身的, 可编程可定制的自动化网络 [88~93].

### 4.2.1 意图自主转译

作为意图驱动网络的基础, 意图自主转译可以结合当前网络状态将服务意图自主转化为计算机可识别的相应网络配置策略. 当前意图转译技术大多针对自然语言或者网络特定语言等进行设计, 比如 Merlin, PGA 和 JANUS 等, 缺乏场景和服务的通用性; 还有针对普通用户、网络管理员或者服务提供商等不同服务意图来源的意图转译技术设计, 在转译精度和泛化性等方面都不尽人意. 针对普通用户和专业用户以声明性语言描述意图, 屏蔽底层网络细节与操作实现细节的构想, 研究基于人工智能的意图驱动网络管理意图的自主解译技术, 提升意图解译的自主性、通用性与精确性. 具体地, 设计意图转译过程中的通用意图解译模型, 并进一步构建基于状态机的意图自主转译方法 [94].

- **模型设计及接口.** 针对意图在网络中形式难确定、意图原语难统一、意图实体难表征等问题, 研究基于扩展巴克斯范式的意图原语和意图多视图解析流程, 允许用户以图形、语音、文本等不同形式输入声明性意图, 普通用户可以使用自然语言的形式表达意图, 专业用户可直接描述与网络功能相关的参数要求 [88~91].

• **意图自主转译方法.** 针对北向服务接口难统一、端到端网络管理难、网络应用开发复杂等挑战, 面向多样化用户意图输入、多场景意图解译需求, 提出基于意图要素识别技术, 确定有限状态机、本体论的意图解译方法, 实现多类来源的意图转换为网络策略 [92~94].

#### 4.2.2 意图闭环验证技术

目前意图闭环验证技术主要基于形式化方法, 通过设置规则和约束等方式建模协助意图所有者进行意图验证, 导致意图驱动的网络智能化管理的效率较低. 也有部分研究基于数字孪生或者人工智能算法等实现更高效、更可靠的智能化意图验证方案; 然而, 尚未实现智慧网络管理意图的全生命周期的闭环验证. 针对多类来源的服务意图在转译过程中可能存在解译出错, 以及意图转译结果难以适配网络状态持续下发的问题, 研究面向意图驱动网络管理意图正确性与可行性验证技术, 实现意图闭环验证 [95].

• **意图正确性验证技术.** 在管理意图到网络策略的映射过程中, 基于信念 (belief) – 愿望 (desire) – 意图 (intent) 模型进行策略推理生成相关网络策略, 进一步构建意图解译正确性验证模型, 利用意图精准率、意图完备率、意图综合指数验证智慧网络管理意图解译的正确性 [95].

• **意图可行性验证技术.** 针对多用户意图并发易出错, 网络状态难以支持意图执行等技术挑战, 设计意图知识图谱构建与知识推理方法, 推理多个用户意图之间的潜在的逻辑一致性问题, 进而利用数学规则进行可行性验证, 保障智慧网络管理意图基于网络状态的可行性 [95].

### 4.3 知识驱动的网络优化

6G 网络的特点包括多元化应用的个性化需求、异构的空天地一体化广域覆盖、受限且多维立体化的网络资源以及高动态变化的网络环境. 这些特点使得 6G 网络优化问题变得更加复杂. 如何按需优化配置网络资源, 满足用户的个性化需求, 是一个亟待解决的问题.

现有的网络优化方法主要分为两类, 即基于模型的理论分析方法和数据驱动的深度学习方法. 基于模型的理论分析方法依赖于已经验证的理论模型, 如通信理论、信号处理和优化理论等, 以优化网络的技术和算法. 然而, 当这种方法被用于解决 6G 网络的优化问题时, 它面临着一些挑战. 首先, 这种方法常常对无线网络进行简化的假设, 如高斯噪声和线性系统, 这可能导致系统模型的不准确性和次优解. 其次, 在处理大规模网络优化问题时, 该方法的计算复杂度过高, 导致推理时间过长, 无法满足 6G 网络中用户的严格低时延要求.

数据驱动的深度学习方法通过利用神经网络从数据中学习网络的优化策略. 文献 [58] 详细地总结了 6G 网络中现有的数据驱动的深度学习方法, 由于神经网络强大的近似能力, 它可以逼近任何连续函数, 只要有足够的数据, 数据驱动的深度学习方法可以处理理论分析方法不能解决的复杂问题. 数据驱动的深度学习方法的另一个重要优势是其快速的在线推理, 有助于对动态网络的快速响应, 从而实现按需服务. 然而, 学习具有大量参数的神经网络需要大量高质量的训练数据, 这些数据的收集是耗时的, 甚至在无线网络中是不可能的, 这可能限制了其在满足按需服务需求的能力. 此外, 神经网络被视为“黑匣子”, 因为它们的内部工作原理和机制不容易理解. 这种可解释性的缺乏会导致在实际应用时产生风险, 降低用户的信任度.

基于此, 本文采用知识驱动的方法解决 6G 网络中的优化问题, 以按需满足 6G 网络中多样化服务的个性化需求, 如超高可靠低时延通信、无缝连接等. 知识驱动的神经网络方法利用几十年来在无线通信中发展起来的专业知识指导神经网络的构建和训练. 如图 6 所示, 一个典型的知识驱动的深度学习方法由 4 个主要部分组成, 即网络训练数据、先验网络知识、神经网络模型和学习算法. 其中先

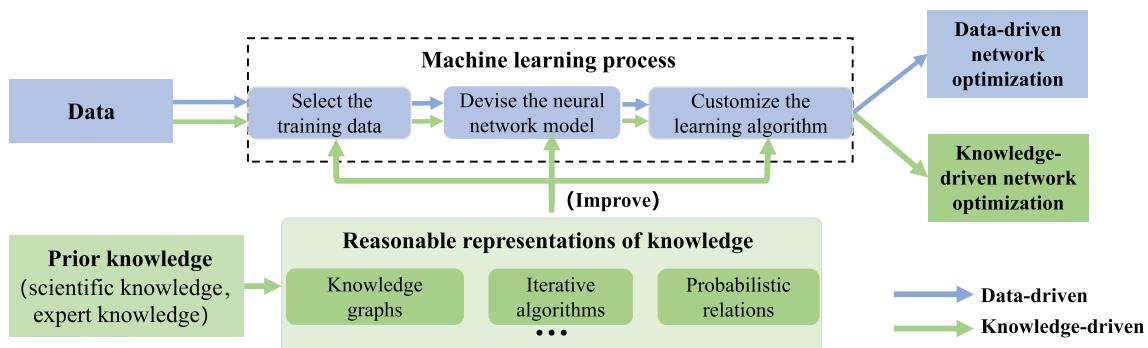


图 6 (网络版彩图) 知识驱动的网络优化结构

Figure 6 (Color online) Knowledge driven network optimisation architecture

验网络知识, 来源于物理空间中的通信网络, 是专家在探索和理解通信网络工作原理时通过逻辑分析或数据处理发现的. 除了几十年来积累的通信相关知识外, 对于 6G 的按需服务优化, 从个性化的用户需求、网络特征和历史决策中抽象出可以指导解决网络优化问题的知识也至关重要. 这些知识可以分为科学知识和专家知识. 科学知识旨在建立能够从理论上解释无线网络运行机制的普遍规律. 它具有系统性和客观性, 为通讯界所广泛接受, 比如香农 (Shannon) 定理、注水算法, 这对于满足按需服务的需求具有重要的指导意义. 专家知识是指一组专家通过广泛地实践、学习和了解无线网络而获得的知识, 通常具体到个人的经验, 这对于理解和满足用户的个性化按需服务需求具有重要的作用.

当构建一个知识驱动的网络优化方法时, 可以利用先验网络知识进行训练数据筛选、神经网络结构设计, 定制学习算法. 具体而言, 可以根据 6G 网络优化问题的具体特点, 选择合适的神经网络去解决, 对于 6G 网络中时间相关性较大的长期优化问题, 可以用长短期记忆 (long short term memory, LSTM) 网络去解决, 以更好地满足按需服务的需求, 比如最小化信息年龄<sup>[96]</sup> 和网络流量预测问题<sup>[97]</sup>. 知识驱动的深度学习方法兼有基于模型的理论分析方法和基于数据的深度学习方法的优点, 正成为网络优化领域的研究热点. 首先, 先验网络知识的融合显著增强了神经网络的可解释性, 有助于更好地理解和分析网络的行为. 这种可解释性也适用于轻量级的知识驱动神经网络, 减少了可学习参数, 减少了训练数据样本, 并减轻了离线计算负担. 其次, 与依赖于精确建模的基于模型的理论方法相比, 知识驱动的深度学习方法中使用的领域知识仅提供了任务的基本和一般特征, 而深度学习强大的通用近似能力弥补了模型不准确性造成的性能下降. 再次, 知识驱动的深度学习方法保持了神经网络的快速在线推理能力, 能够满足动态网络中实时服务的严格低延迟要求, 从而实现按需服务<sup>[98]</sup>. 基于以上, 知识驱动的深度学习具有在 6G 网络中高效智能地解决复杂大规模网络优化问题的潜力, 并将成为一种新的范式, 特别是在满足按需服务的需求方面具有巨大的潜力.

知识驱动的深度学习涵盖了知识指导、知识融合和知识嵌入 3 种集成方法. 知识指导注重运用领域内的先验知识来指导神经网络的架构选择与损失函数设计. 在网络优化任务中, 这种先验知识有助于挑选出与任务性质相匹配的模型, 例如 LSTM 和图神经网络 (graph neural network, GNN) 分别针对时间依赖性和空间依赖性问题. 此外, 知识指导损失函数通过将领域知识以数学公式的形式整合进损失函数, 使神经网络能够同时从数据和知识中学习, 例如在无线网络资源管理中应用原始 – 对偶学习策略. 知识融合方法结合了理论推导和数据驱动的神经网络, 以解决无线网络中的复杂挑战, 它将问题分解为多个互相关联的子问题, 并根据每个子问题的特性, 由知识模块或数据模块来解决. 在串行模式下, 知识与数据模块可以轮流处理不同的子问题, 或者先由理论模型提出初步解决方案, 再由数

据模块进行精细化处理。在并行模式下,知识和数据模块同时独立工作,提升了整个系统的鲁棒性和可靠性。知识嵌入方法则直接将领域知识内置于神经网络的结构之中,这不仅增强了模型的可解释性,也直接作用于网络输出。通过子结构设计,例如在输出层嵌入领域知识,或者通过整体结构设计,如利用深度展开技术将迭代算法转换为神经网络的各层,从而提升算法性能和网络的解释力<sup>[99]</sup>。知识指导着眼于利用先验知识改进学习过程,知识融合致力于融合理论与数据来解决复杂问题,而知识嵌入则专注于将领域知识直接融合入网络架构中,以增强模型性能和解释性<sup>[100]</sup>。

#### 4.4 云原生技术

云原生技术也是 6G 全场景按需服务中的一大关键技术。传统资源调度和服务分发方法往往难以应对动态变化的工作负载和复杂的信息技术 (information technology, IT) 环境,云计算技术应运而生。云原生技术通过其灵活性和可扩展性,为无线网络功能提供了按需轻量化部署的新范式。在基于云原生的服务化网络架构中,无线网络功能如核心网模块及边缘计算等,可以被封装为独立的微服务,运行在容器化环境中。这种微服务化的方法允许无线网络功能根据实际需求动态扩展或缩减,实现资源的最优分配。此外,云原生技术支持的自动化部署和管理能力,进一步增强了网络服务的灵活性和效率,使得无线网络可以更加敏捷地适应不断变化的用户需求和流量模式。云原生技术对网络架构的改动如图 1 所示,通过将网络基础设施层中各个节点的通信计算存储等资源进行容器化、云化、服务化,将网络资源与物理设备、操作系统解耦,实现更灵活、按需、敏捷的网络功能部署和资源统一编排。

目前,云原生技术在学术界和 IT 行业受到了广泛关注<sup>[6]</sup>。在学术界,研究者们主要研究云原生技术下网络架构的设计与计算资源的卸载。文献 [10] 中提出了一种基于天空地融合网络的管理和配置架构,该架构利用双向卸载技术,扩展了地面无线网络的同时并加强全网的容量和鲁棒性。文献 [101] 考虑了卫星的星上计算能力和卫星的动态特性,设计了轻量级卫星移动核心网,将核心网的移动性管理功能加载到卫星上,以提高组网的灵活性,并且设计了用户访问控制的交互协议,卫星基站的移动性控制以及地面网与卫星网的协同控制,以支持移动性管理。文献 [102] 主要针对于云原生下的算力网络,提出了一种天空地融合的“星算网络”。文献 [103] 将天空地融合网络与地面物联网相融合,提出了一种面向时延的物联网任务调度方案,该方案在满足无人机能量容量约束的同时减少了任务的卸载和计算延迟。文献 [104] 为天空地融合网络提出了一个面向任务的智能组网架构,应用边缘云计算技术和网络域划分技术,实现智能组网,降低延迟。文献 [105] 基于天地融合网络,提出了一个平面化的智能云原生组网架构。部分研究者针对于移动边缘节点算力匮乏的问题,研究云原生网络下的计算卸载决策。文献 [106] 以智慧城市为研究场景,提出了一种针对无人机的车辆辅助的计算卸载策略,该策略解决了智慧城市中无人机作为灵活传感器而面临的处理器性能不足的问题。文献 [107] 针对支持移动边缘计算的车联网提出了一种无线资源和移动边缘计算资源的联合分配算法,仿真结果表明该算法可以显著减少总时延并保证网络稳定性。文献 [19] 尝试在边缘设备上设计面向 VR 等设备的新型内容缓存和交付方法,显著提高 VR 视频流性能。另外,云原生技术已经在 IT 行业中实现了落地,如 Kubernetes 等容器编排工具成为了云原生应用的实体标准,该技术通过提供强大的容器编排功能,自动化了容器的部署和管理。云原生技术下游涵盖了云服务提供商、企业级应用开发、物联网应用、金融服务和电商平台等多个领域,常见的应用有微信和国外的流媒体平台 Netflix 等。通过云原生技术、云服务提供商为客户提供了高度可扩展的基础设施,企业可以更灵活地响应市场变化,而在物联网环境下,处理海量设备和数据也相较以往独立式处理变得更加高效。

#### 4.5 语义通信与语用通信

在不断发展的通信网络中, 理解和满足用户服务需求的能力至关重要。为实现高效的按需服务, 我们需要从通信中提取语义和语用信息, 这将成为未来按需服务网络的核心要素<sup>[108]</sup>。语义通信和语用通信是这一领域的重要概念, 语义通信强调了减少数据传输中的误解和歧义。它致力于提高通信信道的容量和可靠性, 确保信息在传输过程中不会失真或被错误理解<sup>[109]</sup>。通过语义通信, 我们可以更好地理解和解释传输的信息, 从而更好地满足用户的需求。语用通信则侧重于理解数据的作用和影响。它不仅分析数据的表面含义, 还深入挖掘数据中蕴含的深层信息和价值, 以更好地支持应用程序的运行和用户体验。语用通信将数据置于广泛的背景和上下文中, 以便更全面地理解其重要性。这两种通信方式的结合有助于网络更好地理解用户需求、数据传输的目的和影响, 从而实现更智能、更高效的服务。

然而, 尽管语义和语用通信在 6G 网络中带来了前所未有的机遇, 但也伴随着一系列挑战。这些挑战涵盖了多模态数据处理、复杂网络场景服务策略定制、隐私和安全等多个方面。首先, 6G 网络面临大规模的多模态数据传输, 包括文本、语音、图像和视频等<sup>[110]</sup>。如何高效处理、整合和提取这些多模态数据中的语义和语用信息仍然是一个重大挑战。其次, 6G 网络的复杂性表现在各种不同的网络场景和应用中<sup>[111]</sup>。这些多样的场景提供了大量的语义和语用信息, 但也带来了应用这些信息的困难。如何整合来自不同场景的语义/语用信息, 以更好地满足用户的需求并提供个性化的服务策略, 是一个需要深入研究的问题。此外, 随着语义和语用通信的应用, 隐私和安全问题也变得尤为重要<sup>[112]</sup>。在信息传输中, 如何避免语义和语用信息的泄密, 以及如何保护通信过程免受恶意篡改和攻击的影响, 是关乎用户权益和网络可信度的问题。

目前, 一些研究者已经应对上述挑战进行了初步的尝试。在语义通信方面, Luo 等<sup>[113]</sup> 基于深度学习 (deep learning, DL) 提出了一个端到端的语义通信基本架构, 通过构建语义知识库在源端提取用语义信息并且辅助目的端理解语义信息; 基于类似的想法, 研究者们将语义通信应用于文本<sup>[114~116]</sup>、语音<sup>[117~119]</sup>、图像<sup>[120~126]</sup>、视频<sup>[127~129]</sup> 等多种模态的数据传输之中, 有效地提高了网络的传输效率。此外, 面向任务/价值的通信可以视为对语用通信的一种初步探索, 其基本框架如图 7 所示。Zhang 等<sup>[130]</sup> 提出了一种面向任务的通信系统架构和一种自适应联合信源信道编码技术, 使用人工智能技术提炼信息中隐含的语用特征, 有效地降低了传输负载。在此基础上, Zhang 等<sup>[131,132]</sup> 还将类似的思路扩展到多设备协作传输和视频流传输领域。此外, 还有一些工作关注面向价值的通信系统。面向价值的通信系统不关注物理层编码, 着重对数据进行价值上的直接筛选。不同于面向任务通信提取可能影响语用信息的数据特征, 面向价值的通信试图将数据的语用信息直接映射到一个价值指标之上。Basagni 等<sup>[133]</sup> 和 Giordani 等<sup>[134]</sup> 基于专家知识和数学模型抽象了主观的信息价值评估机制, 在传感器网络和车联网中有效地降低了网络冗余负载。Chen 等<sup>[30]</sup> 提出了一种邮箱理论, 在保证隐私和安全的基础上分布式地提取了 6G 网络中数据的信息价值。在 6G 网络中, 应用语义和语用通信将不仅仅是一种趋势, 更是通信技术的未来方向。

#### 4.6 网络智能优化技术

面对虚实融合、沉浸式、情境化、个性化、泛在化的业务需求, 以及空天地海一体化通信的网络需求<sup>[135]</sup>, 6G 网络应该为了满足新业务的需求而不断改变, 这对 6G 网络的智能化提出了非常高的要求。当前以规则算法为核心的网络运行范式难以动态适应用户需求和网络环境的不断变化。而且网络运营经验无法有效积累, 制约了网络能力的持续提升<sup>[136]</sup>。在网络中加入“智能基因”, 形成网络的智

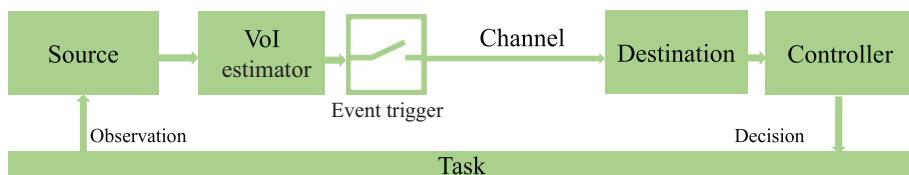


图 7 (网络版彩图) 面向价值的通信系统

Figure 7 (Color online) Value oriented communication system

能和自我进化能力,是解决上述问题的重要途径<sup>[137]</sup>. 内生智能就是这个“智能基因”. 传统的 AI 系统通常需要在训练期间接收大量的标记数据和指导来执行特定任务. 而内生智能的概念建立在自主学习和自适应性的思想之上, 这使得 AI 系统能够在没有大量外部干预的情况下不断提高其性能<sup>[138]</sup>. 它借鉴了生物智能系统中的一些原则, 尤其是类似于生物大脑如何通过不断的感知、学习和适应来实现智能的概念.

实现内生智能, 可以从按需、敏捷、分布式和自动化 4 个方面着手. 按需本质上是围绕用户和网络需求, 适时分配和提供各种资源和服务, 为用户提供更加沉浸且更富个性化的智能信息服务<sup>[139]</sup>. 一些研究者已经做了相关的探索. Yang 等<sup>[140]</sup> 提出了一种具有集成网络资源和普遍人工智能能力的网络人工智能架构, 用于支持具有保证的体验质量 (quality of experience, QoE) 的定制服务. Tian 等<sup>[141]</sup> 设计了一个集成主观用户需求和客观环境需求的 6G 通感算资源按需调配架构. Cao 等<sup>[142]</sup> 提出了一种高效的 6G 网络资源分配架构, 通过识别切片资源类型并设计相应的优化算法进行切片资源调配及部署, 从而实现用户请求的定制化服务.

敏捷指 6G 网络的灵活可控制、融合可演进以及弹性可定制的特征, 便于网络和业务的快速部署和保障, 具有更高的适应性和灵活弹性. Zhao 等<sup>[143]</sup> 提出一种面向 6G 多载波多模态融合组网的元小区模型, 能支持高中低任意无线频段资源的智能编排组合, 从而灵活按需地满足任意组网场景需求. Shang 等<sup>[144]</sup> 通过遍历计算图提取卷积神经网络 (convolutional neural network, CNN) 的关键算子完成模型的自适应划分, 增强调度策略灵活性.

6G 要赋予网络设备和节点信息感知能力和计算处理能力, 实现网络内生智能; 同时 6G 网络将逐步从基站的超密集部署向边缘计算设备的海量分布式部署演进<sup>[145]</sup>, 以上两点都需要更先进的分布式 AI 技术支持. Shen 等<sup>[146]</sup> 提出了 RingSFL, 这是一种新颖的分布式学习方案, 它将联邦学习与模型拆分机制集成在一起, 以适应客户端异构性, 同时保持数据隐私. Yang 等<sup>[147]</sup> 提出了一种使用 GNN 架构在设备到设备 (device to device, D2D) 网络中进行资源分配的新型知识驱动方法, 为具有严格延迟要求的分布式动态网络开发高效且可扩展的无线资源管理提供了解决方案.

AutoML 技术为实现内生智能网络自动化提供了关键支持. 利用 AutoML, 我们可以自动选择最佳的模型结构和超参数, 从而实现自优化. 更进一步, AutoML 可以自我迭代, 不断学习并改进, 满足自学习的需求<sup>[148]</sup>. Perrone 等<sup>[149]</sup> 提出了基于集成学习技术扩展贝叶斯优化方法, 它可利用贝叶斯优化过程中产生的模型通过自动集成学习生成新模型. Lévesque 等<sup>[150]</sup> 也提出了基于集成学习技术对模型池的贪心选择来自动构建集成模型, 并不断向其中添加新的基础学习器, 同时构建性能模型用以评估其性能. 充分应用内生智能网络优化技术, 可以满足未来 6G 业务的高质量要求和用户的高体验质量要求, 实现网络的自动化运行和控制.

#### 4.7 多维资源融合

面向 6G 时代, 通信系统、感知系统和计算系统的深度融合已成为技术发展的重要趋势<sup>[151]</sup>. 6G

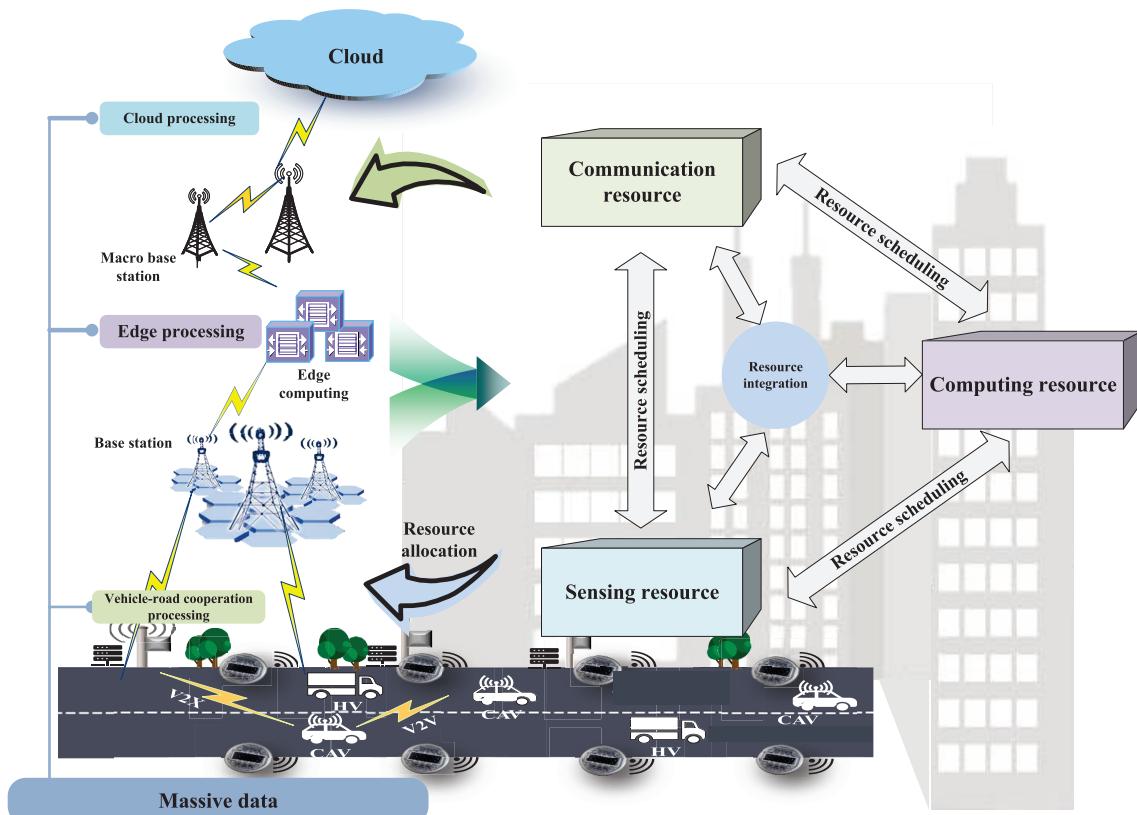


图 8 (网络版彩图) 通感算一体化网络架构图

Figure 8 (Color online) Integrated communication and computing network architecture

多维资源融合技术指的是通信系统、感知系统和计算系统的有机融合, 以实现感知、通信和计算能力的互补与协同, 从而构建具有强大智能感知与处理能力的网络系统, 如图 8 所示<sup>[152]</sup>。通感算融合网络可为各类新兴业务提供智能支持, 是实现 6G 全场景按需服务的关键技术之一。

然而, 实现通感算融合面临多方面挑战。首先, 缺乏通信、感知及计算资源之间相互平衡的信息理论支撑, 需要进一步研究如何以合适的指标来表征融合效果, 并平衡通信容量、感知性能和计算能力之间的关系<sup>[151]</sup>。其次, 通感算融合网络的架构设计也面临挑战。最后, 传统资源调配方案通常基于普适场景, 且网络中通信、感知、计算资源存在复杂的制约与耦合关系, 设计通感算资源的按需调配方案存在挑战<sup>[49]</sup>。

为推进 6G 通感算融合技术进展, 以实现未来 6G 网络资源按需调配研究的发展和应用, 现有研究主要集中于 6G 通感算融合的基础理论研究及设计按需资源调配策略两方面。在基础理论研究方面, 文献 [153] 针对通信感知一体化 (integrated sensing and communication, ISAC) 指标, 提出了一个分析感知与通信绩效权衡的一般框架, 有助于设计使用的帕累托 (Pareto) 最优 ISAC 信号策略。同时, 网络架构是信息通信网络的骨骼与中枢, 文献 [154] 提出了一种基于智能机器 (intelligent machines, IMs) 的新型 6G 网络通感算一体化框架, 该框架具有高可靠性、低延迟通信、高精度感知和快速环境适应的特点。另一方面, 现有研究通过设计通感算一体化的网络资源调配策略, 推进 6G 网络资源的按需调配。如文献 [83] 提出了一种统一的通感算资源调配框架, 通过设计通感算资源按需调配策略, 可优化 6G 无线网络中有限的系统资源, 从而在相同的无线资源上实现多目标感知、多流通信和多维计算的

并行性能。文献 [155] 设计一个集成用户和环境需求的 6G 通感算资源按需调配策略, 为未来 6G 网络资源按需调配研究的发展和应用提供参考。文献 [156] 提出了一种在线控制框架, 用于对未来的天-空-地一体化车载网络资源进行动态调度, 从而快速响应车辆环境, 提升网络平均性能。文献 [157] 提出了一种多视角感知通信策略, 在 UAV 网络中最小化平均信息年龄。文献 [158] 提出了一种任务卸载和传感资源调度策略, 辅助物联网进行自适应的边缘感知。

针对通感算一体化网络架构部署时, 集中云端训练的模型适配边缘亚性能设备的问题, 可采用多层次的模型适配策略。首先, 在云端进行模型训练时, 可采用模型压缩和知识蒸馏技术<sup>[159]</sup>, 以降低模型的复杂度, 使其能够在边缘设备上运行, 同时尽可能保留模型的性能。其次, 为了进一步适应边缘设备的计算能力, 可采用动态网络结构调整机制<sup>[160]</sup>, 该机制能够根据边缘设备的实时计算能力和任务需求, 动态调整网络的深度和宽度。此外, 为了应对边缘设备性能的多样性, 可采用基于设备能力的模型微调策略<sup>[161]</sup>。在此策略中, 边缘设备首先进行资源能力评估, 然后根据评估结果选择合适的模型版本进行微调, 以适应具体的性能要求。这一策略能够使模型更好地适配各种性能的边缘设备, 同时保证了在边缘计算环境中的 AI 服务质量。

综上, 通感算融合将成为促进 6G 全场景按需服务的关键技术方向, 其强大的感知、协作和处理能力将在工业互联网<sup>[162]</sup>、自动驾驶<sup>[163]</sup>、智慧医疗<sup>[164]</sup>等应用中发挥重要作用。因此, 为加速 5G 升级 6G, 行业与学术界需要加强合作, 以推进通感算融合对 6G 按需服务技术的研究与应用。

## 5 前景与展望

目前, 6G 网络的研究正处于快速发展中, 对于 6G 按需服务技术的相关研究和探讨也处于初始阶段。为更好地实现全场景按需服务这一 6G 网络的重要愿景, 需要研究者、设备商、运营商、服务商等各方面共同努力, 在需求、架构、发展范式等方面进行研究和探索。本节主要对一些该领域比较重要的潜在技术进行了探讨和展望。

### 5.1 面向网络决策的可解释性人工智能技术

6G 网络将在带宽、覆盖范围、可靠性、能效和延迟方面带来显著提升, 并引发大量实时自动化决策, 涵盖了从网络资源分配到自动驾驶汽车的碰撞避免等广泛应用。然而, 现有基于人工智能的模型缺乏可解释性, 设计者和用户难以理解和信任这些模型的决策行为<sup>[165]</sup>, 如何解释并确保自动化决策的可靠性成为 6G 网络的新需求。所幸近年来, 可解释性人工智能技术受到越来越多研究者的关注。可解释性人工智能 (explainable artificial intelligence, XAI) 是指一类人工智能系统或算法, 其决策过程和结果可以被理解并解释给人类用户或利益相关者。XAI 可以作为在 6G 网络中的指引者, 为设计者和用户解释和理解模型的决策行为。

针对 XAI 的应用方案, 主要有 5 种方法: 特征重要性分析、局部敏感性分析、可视化方法、解释性模型和对抗样本模型。特征重要性分析通过计算特征对模型预测结果的影响程度, 确定哪些特征对预测结果的影响最大。局部敏感性分析则在给定输入情况下, 通过改变输入中的某些特征来探索模型输出的变化, 揭示模型如何得出预测结果。可视化方法用于将模型的内部结构、参数等可视化呈现, 以更好地理解和解释模型的决策过程。解释性模型的方法则通过使用更易解释的模型 (如决策树、逻辑回归等) 替代复杂模型, 从而更好地理解和解释模型的决策依据。对抗样本分析的方法通过生成对抗样本来探究模型的漏洞和脆弱性, 并开发相应的防御措施<sup>[166]</sup>。

将 XAI 与 6G 网络有机结合可以为用户提供更高效、个性化、安全化的按需服务体验。XAI 可以

帮助解释和理解 6G 网络的运行情况和决策过程。6G 网络采用复杂的架构和智能算法, 以支持大规模设备连接和高速数据传输。用户可能面临更多网络参数和决策, 如优先级分配、无线电安全和频谱管理等。通过提供可解释性机制, XAI 能够增加网络决策的置信度和安全性, 提升用户对网络运行的信任。同时, XAI 还能根据用户需求、偏好和环境定制 6G 网络服务, 提高用户满意度和服务安全性。建立基于人工智能的解决方案的透明度和安全性是未来 6G 通信的关键问题, 必须为黑盒 AI 模型提供安全性、可解释性和鲁棒性的有效度量。同时, XAI 的研究不应局限于计算机视觉领域, 还应在物理层网络安全、自动驾驶、工业控制与优化等领域继续进行探索。

## 5.2 内生智能网络技术

作为第六代移动通信网络的关键技术之一, 内生智能网络技术近年来受到行业与学术界的广泛关注<sup>[167, 168]</sup>。内生智能网络指的是在 6G 网络架构设计之初, 将人工智能能力作为基础服务内嵌于网络体系结构, 以服务于网络的自主运行与优化。作为网络的决策者, 6G 内生智能的核心在于实现人工智能与通信网络的深度融合, 而非 5G 时代盛行的“打补丁, 外挂式”应用模式<sup>[169]</sup>。6G 内生智能理念的提出, 源于行业和学术界对人工智能在移动网络中的应用效果评估。研究表明, 5G 网络与人工智能的结合存在数据采集困难、模型实施验证周期长、AI 应用导致网络复杂度提高等问题, 导致应用效果不佳<sup>[170]</sup>。随着 6G 连接规模和应用场景的增多, 传统人工管理网络的模式难以为继。同时, 应用人工智能改进网络运维效率, 实现网络自主化运营的需求日益迫切。基于以上考量, 业界提出在网络架构设计中应考虑 AI 的内嵌, 形成内生智能网络体系, 以适应 6G 发展的需要。

当前, 内生智能网络的研究尚处起步阶段。国际电信联盟 (ITU) 已经提出支持内生智能网络的参考架构框架, 其中包括 AI 编排层面和网络协同层面<sup>[171]</sup>。欧洲电信标准化组织 (European Telecommunications Standards Institute, ETSI) 也在探索 AI 闭环式网络管理体系的设计方案<sup>[171]</sup>。中国通信标准化协会 (China Communications Standards Association, CCSA) 正在积极制定内生智能相关标准。以运营商中国移动为代表的部分企业, 也在探索内生智能技术与网络深度融合的新架构<sup>[172]</sup>。尽管研究初具规模, 内生智能网络仍面临多方面挑战: (1) 如何设计出支持 AI 全生命周期的网络内嵌架构; (2) 如何实现网络功能与 AI 功能间的逻辑解耦和物理融合; (3) 如何统一编排 AI 计算和网络通信资源; (4) 如何构建安全、可解释的 AI 决策模型。目前较为成熟的技术主要集中在强化学习驱动的网络自主优化方面, 还有待进一步拓展。后续研究可继续深化网络通信与 AI 计算的融合方案, 构建端到端的内生智能框架, 使网络向自配置、自优化、自愈合和自演进的高级智能体系靠拢。内生智能的框架设计和技术实现任重而道远, 需要产业界和学术界加强协同创新, 形成基础理论和关键技术共识。网络与人工智能领域的专家可围绕架构、安全、算法等维度展开交叉合作。运营商可从用户体验出发, 进行内生智能在特定网络场景和核心业务中的落地实践。一旦内生智能理念成熟, 技术可靠, 它将助力 6G 网络实现智能无所不在, 为用户提供“零等待、零接触、零故障”的服务体验, 也将推动网络经济实现新的发展。

## 5.3 空天地一体化下的通感融合技术

为满足第六代移动通信网络在覆盖范围和定位精度两个关键性能指标方面提出的全新的要求, 需要设计一种全新的架构, 在融合空天地异构网络基础上统筹通感资源。空天地一体化下的通感融合技术应运而生, 为满足高覆盖范围和定位精度提供了新的思路。当前研究主要集中在系统设计和关键技术创新两方面, 以实现这两项指标的要求<sup>[5]</sup>。

在覆盖范围方面, 6G 要求实现无所不在的全球覆盖, 覆盖陆地、海洋和天空的每一个角落<sup>[173]</sup>。这

需要打破地面基站覆盖的局限,构建立体化的通信系统。目前,卫星通信系统、高空平台系统和无人机系统被认为是扩展 6G 覆盖范围的三大重要手段<sup>[174]</sup>。卫星通信系统可部署在不同的轨道,包括低地球轨道 (low earth orbit, LEO)、中地球轨道 (medium earth orbit, MEO) 和地球同步轨道 (geostationary earth orbit, GEO), 从太空为地面用户提供服务。高空平台通常指飞行高度在 20~50 km 的各类平台,可搭载通信设备为地面用户提供覆盖。无人机系统也可在有限范围内为地面用户提供覆盖和中继,常见的有固定翼无人机、多旋翼无人机、飞艇无人机等。这些非地面系统可与地面蜂窝网络互联互通,共同扩展 6G 的覆盖范围。另外,智能反射表面也被认为是扩展 6G 覆盖范围的关键技术之一,其低成本和易部署的特点非常适合覆盖盲区<sup>[175]</sup>。当前研究主要从信号覆盖、容量覆盖和服务覆盖 3 个方面进行,以指导规划设计和评估系统覆盖性能。但是异构网络互操作性和频谱共享等问题仍需进一步研究。在定位精度方面,6G 要求实现厘米甚至毫米级的定位精度<sup>[176]</sup>,而 5G 网络的定位精度仅在米级。毫米波和太赫兹波等高频技术、大规模天线阵列以及智能表面等都是提升 6G 定位精度的重要手段<sup>[177]</sup>。另外,大规模 MIMO 和超密集网络也有利于通过提高视线信号所占比例来改善基于距离的定位性能<sup>[178]</sup>。通过多种定位技术和多源传感器信息的协同和融合,也可以进一步提高定位鲁棒性和精确度<sup>[179]</sup>。当前研究开始尝试联合通信与雷达,一体化的系统与波形设计以共同提升通信与定位性能。但设备成熟度和标准化等方面仍存在挑战<sup>[180]</sup>。6G 网络的高精度定位将推动诸如无人驾驶、元宇宙等新型应用的蓬勃发展,对经济增长和生活方式产生深远影响。当前研究已取得一定进展,但还需持续努力以实现 6G 的指标要求,并最终落地到商用产品中。这需要通信、电子、计算机等不同学科领域的密切合作和突破,以推进 6G 在定位等方面的技术发展。

#### 5.4 AIGC 和大语言模型辅助的无线网络技术

6G 网络被认为将引入新的通信网络范式,其中生成式人工智能内容 (artificial intelligence-generated content, AIGC) 和大型语言模型辅助的无线网络技术将在通信网络中发挥重要作用。区别于传统判别式 AI 技术,生成式 AI 技术在原理上不追求在超空间中寻找到一个超平面以区分数据类别或找到一种数据映射关系,而是致力于实现对特定概率分布或条件概率分布的拟合<sup>[181]</sup>。这其中的代表便是以拟合图片像素分布为基础实现图片或视频生成的 Diffusion 模型,或拟合语句中不同单词概率分布的大语言模型。因此利用 AIGC 对条件概率分布拟合的能力,可以实现语义通信、网络资源分配与网络管控技术。例如可以通过用户之间共享 AIGC 模型和从其视图图像中提取的语义信息来实现语义通信<sup>[182]</sup>。这使用户可以避免繁重和重复的计算任务,例如用户可以将生成的内容和从其视图图像中提取的语义信息传输给附近的用户,后者可以使用此信息获得其视图图像下的计算结果的空间匹配。除此以外,对于虚拟现实、混合现实和元宇宙等对资源调配能力要高的业务,传统的基于强化学习的资源管理策略难以实现高性能的资源调配,因此可以将 AIGC 与强化学习相结合。以 AIGC 为主干网络,实现对最优调配策略分布的拟合,并通过强化学习的方式,实现无标签的 AIGC 训练,从而在提高资源调配算法性能的同时,降低神经网络训练对于标签数据的依赖<sup>[183]</sup>。并且利用大语言对于自然语言的理解能力,可以实现基于自然语言的网络管控设置。尽管当前 AIGC 的发展为基于自然语言的按需服务提供了技术基础,然而当前大模型的训练方式往往依赖海量标注数据,这对于网络管控和资源管理而言是难以实现的,因此必须探索基于无标签的大模型训练方法,以实现其实际应用。除此以外, AIGC 本质上学习的是数据的概率或条件概率分布,因此一旦数据特征的概率分布发生改变,则 AIGC 的性能会明显下降,所以如何实现 AIGC 的数据分布自适应,也是一个有待探索的关键问题。

## 6 结论

本文对 6G 网络全场景按需服务领域的研究进行了全面的综述。为实现网络价值, 6G 正向“以服务为中心”的范式转变, 然而, 复杂多变的场景和个性化需求带来了极大挑战, 亟需对按需服务技术进行深入研究。本文对场景和按需服务的概念和内涵进行了阐述, 提出 6G 网络按需服务必须针对具体的场景, 进而实现全场景按需服务。本文进一步总结了场景精准认知、场景多样化按需服务及海量场景共存按需服务三大挑战, 并针对这些挑战, 总结分析了 6G 按需服务的架构及相关关键技术的发展现状。最后, 由于 6G 按需服务仍然是一个处于发展初期的领域, 本文对该领域的前景与潜在研究方向进行了展望, 探讨了可解释性、内生智能、AIGC 和大语言模型等研究方向对于该领域可能的重要影响。

## 参考文献

- 1 Zhang P, Zhang J H, Qi Q, et al. Ubiquitous-X: constructing the future 6G networks. *Sci Sin Inform*, 2020, 50: 913–930 [张平, 张建华, 戚琦, 等. Ubiquitous-X: 构建未来 6G 网络. 中国科学: 信息科学, 2020, 50: 913–930]
- 2 HuaWei. White paper on 6G: a new journey in wireless communications, 2022 [华为. 6G: 无线通信新征程白皮书, 2022] extension://cdonnmffkdaaoajfknoeeecmchibpmkmg/assets/pdf/web/viewer.html?file=https%3A%2F%2Fwww-file.huawei.com%2F-%2Fmedia%2Fcorp2020%2Fpdf%2Ftech-insights%2F1%2F6g-white-paper-cn.pdf%3Fla%3Dzh
- 3 Chowdhury M Z, Shahjalal M, Ahmed S, et al. 6G wireless communication systems: applications, requirements, technologies, challenges, and research directions. *IEEE Open J Commun Soc*, 2020, 1: 957–975
- 4 Nawaz S J, Sharma S K, Wyne S, et al. Quantum machine learning for 6G communication networks: state-of-the-art and vision for the future. *IEEE Access*, 2019, 7: 46317–46350
- 5 You X H, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. *Sci China Inf Sci*, 2021, 64: 110301
- 6 Shen X M, Cheng N, Zhou H B, et al. Space-air-ground integrated networks: review and prospect. *Chin J Int Things*, 2020, 4: 3–19 [沈学民, 承楠, 周海波, 等. 空天地一体化网络技术: 探索与展望. 物联网学报, 2020, 4: 3–19]
- 7 Chen C, Liao Z, Ju Y, et al. Hierarchical domain-based multicontroller deployment strategy in SDN-enabled space-air-ground integrated network. *IEEE Trans Aerosp Electron Syst*, 2022, 58: 4864–4879
- 8 Wu H Q, Chen J Y, Zhou C H, et al. Resource management in space-air-ground integrated vehicular networks: SDN control and AI algorithm design. *IEEE Wireless Commun*, 2020, 27: 52–60
- 9 Li J L, Shi W S, Wu H Q, et al. Cost-aware dynamic SFC mapping and scheduling in SDN/NFV-enabled space-air-ground-integrated networks for Internet of vehicles. *IEEE Int Things J*, 2022, 9: 5824–5838
- 10 Zhou S, Wang G C, Zhang S, et al. Bidirectional mission offloading for agile space-air-ground integrated networks. *IEEE Wireless Commun*, 2019, 26: 38–45
- 11 Wu H Q, Chen J Y, Zhou C H, et al. Learning-based joint resource slicing and scheduling in space-terrestrial integrated vehicular networks. *J Commun Inf Netw*, 2021, 6: 208–223
- 12 Han D, Ye Q, Peng H X, et al. Two-timescale learning-based task offloading for remote IoT in integrated satellite-terrestrial networks. *IEEE Int Things J*, 2023, 10: 10131–10145
- 13 Zhao Z Q, Cheng N, Chen J, et al. Research on 6G network scenario cognition based on knowledge graph. *Chinese J Intell Sci Technol*, 2023, 5: 494–504 [赵茁乔, 承楠, 陈勤, 等. 基于知识图谱的 6G 网络场景认知研究. 智能科学与技术学报, 2023, 5: 494–504]
- 14 Lu Y, Zheng X R. 6G: a survey on technologies, scenarios, challenges, and the related issues. *J Indust Inf Integration*, 2020, 19: 100158
- 15 Wang Z Q, Du Y, Wei K J, et al. Vision, application scenarios, and key technology trends for 6G mobile communications. *Sci China Inf Sci*, 2022, 65: 151301
- 16 IMT vision. Framework and overall objectives of the future development of IMT for 2020 and beyond. <http://www.itu.int/rec/R-REC-M.2083-0-201509-I>

- 17 Jiang W, Han B, Habibi M A, et al. The road towards 6G: a comprehensive survey. *IEEE Open J Commun Soc*, 2021, 2: 334–366
- 18 Shen X S, Gao J, Li M S, et al. Toward immersive communications in 6G. *Front Comput Sci*, 2023, 4: 1068478
- 19 Li M S, Gao J, Zhou C H, et al. User dynamics-aware edge caching and computing for mobile virtual reality. *IEEE J Sel Top Signal Process*, 2023, 17: 1131–1146
- 20 IMT-2030 (6G) Promotion Group. 6G typical scenarios and key capabilities [IMT-2030 (6G) 推进组. 6G典型场景和关键能力] extension:/cdonnmffkdaaoajfknoeeecmchibpmkmg/assets/pdf/web/viewer.html?file=http%3A%2F%2Fwww.caict.ac.cn%2Fkxyj%2Fqwf%2Fztbg%2F202207%2FP020220724437616473703.pdf
- 21 Raychaudhuri D, Gerla M. Emerging Wireless Technologies and the Future Mobile Internet. New York: Cambridge University Press, 2011
- 22 Fettweis G, Alamouti S. 5G: personal mobile internet beyond what cellular did to telephony. *IEEE Commun Mag*, 2014, 52: 140–145
- 23 Jia K, Kenney M. Mobile internet business models in China: vertical hierarchies, horizontal conglomerates, or business groups. In: Proceedings of Berkeley Roundtable on the International Economy Working Paper, 2016
- 24 Sharma P. Evolution of mobile wireless communication networks-1G to 5G as well as future prospective of next generation communication network. *Int J Comput Sci Mob Comput*, 2013, 2: 47–53
- 25 Liu G Y, Jiang D J. 5G: vision and requirements for mobile communication system towards year 2020. *Chinese J Eng*, 2016, 2016: 1–8
- 26 Kekki S, Featherstone W, Fang Y G, et al. MEC in 5G networks. *ETSI white paper*, 2018, 28: 1–28
- 27 Eckelmann S, Trautmann T, Ußler H, et al. V2V-communication, LiDAR system and positioning sensors for future fusion algorithms in connected vehicles. *Transp Res Procedia*, 2017, 27: 69–76
- 28 Liu X N, Li Z, Zhao N, et al. Transceiver design and multihop D2D for UAV IoT coverage in disasters. *IEEE Int Things J*, 2019, 6: 1803–1815
- 29 Ranaweera P, Jurcut A D, Liyanage M. Survey on multi-access edge computing security and privacy. *IEEE Commun Surv Tut*, 2021, 23: 1078–1124
- 30 Hao Y X, Miao Y M, Chen M, et al. 6G cognitive information theory: a mailbox perspective. *BDCC*, 2021, 5: 56
- 31 Zong J Y, Liu H T, Liu Y, et al. Service-based architecture evolution of radio access network towards 6G. In: Proceedings of the International Conference on Computer Engineering and Networks, 2022. 525–534
- 32 Li Y Z, Huang J, Sun Q B, et al. Cognitive service architecture for 6G core network. *IEEE Trans Ind Inf*, 2021, 17: 7193–7203
- 33 Duan X D, Yao H J, Fu Y X, et al. Computing power network technologies for computing and network integration evolution. *Telecomm Sci*, 2021, 37: 76–85 [段晓东, 姚惠娟, 付月霞, 等. 面向算网一体化演进的算力网络技术. 电信科学, 2021, 37: 76–85]
- 34 Tang X Y, Cao C, Wang Y X, et al. Computing power network: the architecture of convergence of computing and networking towards 6G requirement. *China Commun*, 2021, 18: 175–185
- 35 Lin L, Zhu B, Wang Q Q, et al. A novel 5G core network capability exposure method for telecom operator. In: Proceedings of the IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2020. 1450–1454
- 36 Grawe J E. Herzberg's theory of motivation and Maslow's hierarchy of needs. *Pract Assess Res Evaluation*, 1996, 5: 11–14
- 37 Slalmi A, Chaibi H, Chehri A, et al. Toward 6G: understanding network requirements and key performance indicators. *Trans Emerging Tel Tech*, 2021, 32: e4201
- 38 Liu G Y, Huang Y H, Li N, et al. Vision, requirements and network architecture of 6G mobile network beyond 2030. *China Commun*, 2020, 17: 92–104
- 39 Wang C X, You X H, Gao X Q, et al. On the road to 6G: visions, requirements, key technologies, and testbeds. *IEEE Commun Surv Tut*, 2023, 25: 905–974
- 40 Akhtar M W, Hassan S A, Ghaffar R, et al. The shift to 6G communications: vision and requirements. *Hum Cent Comput Inf Sci*, 2020, 10: 1–27
- 41 Scoble R, Israel S, Benioff M. Age of Context: Mobile, Sensors, Data and the Future of Privacy. 1st ed. Boston:

- Patrick Brewster Press, 2014
- 42 Kenny D, Marshall J F. Contextual marketing. *Harv Bus Rev*, 2000, 78: 119–125
- 43 Silver D, Clark T N, Navarro Yanez C J. Scenes: social context in an age of contingency. *Soc Forces*, 2010, 88: 2293–2324
- 44 Stalnaker R. *Propositions: Ontology and Logic*, volume 3. Oxford: Oxford University Press, 2022
- 45 Qiao J, He Y J, Shen X M. Improving video streaming quality in 5G enabled vehicular networks. *IEEE Wireless Commun*, 2018, 25: 133–139
- 46 Huang X Y, Li M S, Wu W, et al. Digital twin-assisted collaborative transcoding for better user satisfaction in live streaming. 2022. ArXiv:2211.06906
- 47 Guo H Z, Li J Y, Liu J J, et al. A survey on space-air-ground-sea integrated network security in 6G. *IEEE Commun Surv Tut*, 2022, 24: 53–87
- 48 Cui H X, Zhang J, Geng Y H, et al. Space-air-ground integrated network (SAGIN) for 6G: Requirements, architecture and challenges. *China Commun*, 2022, 19: 90–108
- 49 Cheng N, He J C, Yin Z S, et al. 6G service-oriented space-air-ground integrated network: a survey. *Chin J Aeronaut*, 2022, 35: 1–18
- 50 Alsharif Mohammed H, Albreem Mahmoud A M, Solyman Ahmad A A, et al. Toward 6G communication networks: terahertz frequency challenges and open research issues. *Comput Mater Continua*, 2021
- 51 Katz M, Pirinen P, Posti H. Towards 6G: getting ready for the next decade. In: *Proceedings of the International Symposium on Wireless Communication Systems (ISWCS)*, 2019. 714–718
- 52 Xu H, Klaine P V, Onireti O, et al. Blockchain-enabled resource management and sharing for 6G communications. *Digital Commun Netw*, 2020, 6: 261–269
- 53 Guan W Q, Zhang H J, Leung V C. Customized slicing for 6G: enforcing artificial intelligence on resource management. *IEEE Netw*, 2021, 35: 264–271
- 54 Cheng N, Quan W, Shi W S, et al. A comprehensive simulation platform for space-air-ground integrated network. *IEEE Wireless Commun*, 2020, 27: 178–185
- 55 Next G Alliance. Next G Alliance Report: Roadmap to 6G. Technical Report, 2022
- 56 EU. Hexa-X Vision on 6G and Research Challenges. Technical Report, 2020
- 57 NICT. Beyond 5G/6G (White paper). Technical Report, 2021
- 58 Shen X M, Gao J, Wu W, et al. Holistic network virtualization and pervasive network intelligence for 6G. *IEEE Commun Surv Tut*, 2022, 24: 1–30
- 59 Huang X Y, Wu W, Hu S S, et al. Digital twin based user-centric resource management for multicast short video streaming. 2023. ArXiv:2308.08995
- 60 China Mobile. China Mobile 6G Network Architecture Technical White paper. Technical Report, 2022 [中国移动. 中国移动6G网络架构技术白皮书, 2022]
- 61 China Telecom. China Telecom 6G Vision and Technology White Paper. Technical Report, 2022 [中国电信. 中国电信6G愿景与技术白皮书, 2022]
- 62 Gong Z J, Jiang F, Li C, et al. Simultaneous localization and communications with massive MIMO-OTFS. *IEEE J Sel Areas Commun*, 2023
- 63 Sun Y T, Zhang J H, Zhang Y X, et al. Environment features-based model for path loss prediction. *IEEE Wireless Commun Lett*, 2022, 11: 2010–2014
- 64 Zhang Q R, Cheng N, Chen W, et al. Incomplete multi-view clustering analysis of 6G business scenarios. *J Xidian Univ* [张茹倩, 承楠, 陈文, 等. 6G 业务场景的不完全多视图聚类分析. 西安电子科技大学学报]
- 65 Michelusi N, Nokleby M, Mitra U, et al. Multi-scale spectrum sensing in millimeter wave cognitive networks. In: *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, 2017. 1640–1644
- 66 Michelusi N, Nokleby M, Mitra U, et al. Multi-scale spectrum sensing in dense multi-cell cognitive networks. *IEEE Trans Commun*, 2019, 67: 2673–2688
- 67 Zhao Z X, Du Q H, Karagiannidis G K. Improved grant-free access for URLLC via multi-tier-driven computing: network-load learning, prediction, and resource allocation. *IEEE J Sel Areas Commun*, 2023, 41: 607–622
- 68 He H L, Ren PY, Du Q H, et al. Traffic-aware overload control scheme in 5G ultra-dense M2M networks. *Trans Emerging Tel Tech*, 2017, 28: e3146

- 69 Tello-Oquendo L, Pla V, Leyva-Mayorga I, et al. Efficient random access channel evaluation and load estimation in LTE-A with massive MTC. *IEEE Trans Veh Technol*, 2019, 68: 1998–2002
- 70 Zhou S P, Cui L G, Fang C, et al. Research on network topology discovery algorithm for internet of things based on multi-protocol. In: Proceedings of the International Conference on Modelling, Identification and Control (ICMIC), 2018. 1–6
- 71 Ni Y, Shi H Z, Zhu P D. Construct network topology based on OSPF database. *Comput Technol Dev*, 2009, 19: 1–4
- 72 Li J, Lu H C, Xue K, et al. Temporal netgrid model-based dynamic routing in large-scale small satellite networks. *IEEE Trans Veh Technol*, 2019, 68: 6009–6021
- 73 Li Z, Li X P. Destruction-resistant routing protocol based on topology prediction and backup path. In: Proceedings of the International Conference on Multimedia Communication Technologies (ICMCT), 2022. 36–40
- 74 Wang P, Zhang X S, Zhang S, et al. Time-expanded graph-based resource allocation over the satellite networks. *IEEE Wireless Commun Lett*, 2019, 8: 360–363
- 75 Chen Y, Cao H, Zhou Y Q, et al. A GCN-GRU based end-to-end LEO satellite network dynamic topology prediction method. In: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), 2023. 1–6
- 76 Tran-Dang H, Kim D S. Task priority-based resource allocation algorithm for task offloading in fog-enabled IoT systems. In: Proceedings of the International Conference on Information Networking (ICOIN), 2021. 674–679
- 77 Abdelrehim M M, Hossny E, Fatma Omara H. A survey of the most common task scheduling algorithms. In: Proceedings of the International Telecommunications Conference (ITC-Egypt), 2023. 126–133
- 78 Sun H. Resource deployment and task scheduling based on cloud computing. In: Proceedings of the 2nd International Conference on Computer Systems (ICCS), 2022. 25–28
- 79 Mishra P K, Chaturvedi A K. State-of-the-art and research challenges in task scheduling and resource allocation methods for cloud-fog environment. In: Proceedings of the International Conference on Intelligent Communication and Computational Techniques (ICCT), 2023. 1–5
- 80 Liao J X, Wu X W. Resource allocation and task scheduling scheme in priority-based hierarchical edge computing system. In: Proceedings of the International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), 2020. 46–49
- 81 Liu A, Huang Z, Li M, et al. A survey on fundamental limits of integrated sensing and communication. *IEEE Commun Surv Tut*, 2022, 24: 994–1034
- 82 Cui Y H, Liu F, Jing X J, et al. Integrating sensing and communications for ubiquitous IoT: applications, trends, and challenges. *IEEE Netw*, 2021, 35: 158–167
- 83 Qi Q, Chen X M, Khalili A, et al. Integrating sensing, computing, and communication in 6G wireless networks: design and optimization. *IEEE Trans Commun*, 2022, 70: 6212–6227
- 84 Wang X C, Ma L F, Li H C, et al. Digital twin-assisted efficient reinforcement learning for edge task scheduling. In: Proceedings of the 95th Vehicular Technology Conference, 2022. 1–5
- 85 Lu Y L, Huang X, Zhang K, et al. Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks. *IEEE Trans Ind Inf*, 2021, 17: 5098–5107
- 86 Wild T, Braun V, Viswanathan H. Joint design of communication and sensing for beyond 5G and 6G systems. *IEEE Access*, 2021, 9: 30845–30857
- 87 Elayan H, Aloqaily M, Guizani M. Digital twin for intelligent context-aware IoT healthcare systems. *IEEE Int Things J*, 2021, 8: 16749–16757
- 88 3GPP. Technical specification group services and system aspects; study on intent-driven management for network slicing. TR 28.836, 2023. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4036>
- 89 Mwanje S S, Banerjee A, Goerge J, et al. Intent-driven network and service management: definitions, modeling and implementation. *ITU J-FET*, 2022, 3: 555–569
- 90 Sebrechts M, Volckaert B, De Turck F, et al. Fog native architecture: intent-based workflows to take cloud native toward the edge. *IEEE Commun Mag*, 2022, 60: 44–50
- 91 Mehmood K, Kralevska K, Palma D. Contextual knowledge-based intent processing for management of next generation cellular networks. 2023. ArXiv:2302.08544
- 92 Zhang J W, Yang C G, Dong R, et al. Intent-driven closed-loop control and management framework for 6G open

- RAN. IEEE Int Things J, 2023. doi: 10.1109/jiot.2023.3312795
- 93 Yang C G, Mi X R, Ouyang Y, et al. SMART intent-driven network management. IEEE Commun Mag, 2023, 61: 106–112
- 94 Ouyang Y, Yang C G, Song Y B, et al. A brief survey and implementation on refinement for intent-driven networking. IEEE Netw, 2021, 35: 75–83
- 95 Song Y B, Yang C G, Zhang J M, et al. Full-life cycle intent-driven network verification: challenges and approaches. IEEE Netw, 2023, 37: 145–153
- 96 Leng S Y, Yener A. Learning to transmit fresh information in energy harvesting networks. IEEE Trans Green Commun Netw, 2022, 6: 2032–2042
- 97 Hua Y X, Zhao Z F, Li R P, et al. Deep learning with long short-term memory for time series prediction. IEEE Commun Mag, 2019, 57: 114–119
- 98 Shlezinger N, Whang J, Eldar Y C, et al. Model-based deep learning. Proc IEEE, 2023, 111: 465–499
- 99 Yang H, Cheng N, Sun R J, et al. Knowledge-driven resource allocation for wireless networks: a WMMSE unrolled graph neural network approach. IEEE Int Things J, 2024. 1–1
- 100 Sun R J, Cheng N, Li C L, et al. Knowledge-driven deep learning paradigms for wireless network optimization in 6G. IEEE Netw, 2024. doi: 10.1109/MNET.2024.3352257
- 101 Han Z Z, Xu C, Liu K, et al. A novel mobile core network architecture for satellite-terrestrial integrated network. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), 2021. 01–06
- 102 Zhai D S, Wang C, Zhang R N, et al. Energy-saving deployment optimization and resource management for UAV-assisted wireless sensor networks with NOMA. IEEE Trans Veh Technol, 2022, 71: 6609–6623
- 103 Zhou C H, Wu W, He H L, et al. Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN. IEEE Trans Wireless Commun, 2021, 20: 911–925
- 104 Liu J, Du X Q, Cui J H, et al. Task-oriented intelligent networking architecture for the space-air-ground-aqua integrated network. IEEE Int Things J, 2020, 7: 5345–5358
- 105 Xu Q C, Su Z, Lu R X, et al. Ubiquitous transmission service: hierarchical wireless data rate provisioning in space-air-ocean integrated networks. IEEE Trans Wireless Commun, 2022, 21: 7821–7836
- 106 Dai M H, Su Z, Xu Q C, et al. Vehicle assisted computing offloading for unmanned aerial vehicles in smart city. IEEE Trans Intell Transp Syst, 2021, 22: 1932–1944
- 107 Hou Y Z, Wang C R, Zhu M, et al. Joint allocation of wireless resource and computing capability in MEC-enabled vehicular network. China Commun, 2021, 18: 64–76
- 108 Wang X B, Mei J, Cui S G, et al. Realizing 6G: the operational goals, enabling technologies of future networks, and value-oriented intelligent multi-dimensional multiple access. IEEE Netw, 2023, 37: 10–17
- 109 Zhang P, Xu W J, Gao H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks. Engineering, 2022, 8: 60–73
- 110 Luo X W, Gao R B, Chen H H, et al. Multi-modal and multi-user semantic communications for channel-level information fusion. IEEE Wirel Commun, 2022, Early Access: 1–18
- 111 Wang C X, Lv Z, Gao X Q, et al. Pervasive wireless channel modeling theory and applications to 6G GBMSs for all frequency bands and all scenarios. IEEE Trans Veh Technol, 2022, 71: 9159–9173
- 112 Nguyen V L, Lin P C, Cheng B C, et al. Security and privacy for 6G: a survey on prospective technologies and challenges. IEEE Commun Surv Tut, 2021, 23: 2384–2428
- 113 Luo X W, Chen H H, Guo Q. Semantic communications: overview, open issues, and future research directions. IEEE Wireless Commun, 2022, 29: 210–219
- 114 Xie H Q, Qin Z J, Li G Y, et al. Deep learning enabled semantic communication systems. IEEE Trans Signal Process, 2021, 69: 2663–2675
- 115 Xie H Q, Qin Z J. A lite distributed semantic communication system for Internet of Things. IEEE J Sel Areas Commun, 2021, 39: 142–153
- 116 Wang Y N, Chen M Z, Luo T, et al. Performance optimization for semantic communications: an attention-based reinforcement learning approach. IEEE J Sel Areas Commun, 2022, 40: 2598–2613
- 117 Weng Z Z, Qin Z J. Semantic communication systems for speech transmission. IEEE J Sel Areas Commun, 2021, 39: 2434–2444

- 118 Tong H N, Yang Z H, Wang S H, et al. Federated learning based audio semantic communication over wireless networks. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), 2021. 1–6
- 119 Han T X, Yang Q Q, Shi Z G, et al. Semantic-preserved communication system for highly efficient speech transmission. *IEEE J Sel Areas Commun*, 2023, 41: 245–259
- 120 Huang D L, Tao X M, Gao F F, et al. Deep learning-based image semantic coding for semantic communications. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), 2021. 1–6
- 121 Bourtsoulatze E, Burth Kurka D, Gündüz D. Deep joint source-channel coding for wireless image transmission. *IEEE Trans Cogn Commun Netw*, 2019, 5: 567–579
- 122 Yang M Y, Kim H S. Deep joint source-channel coding for wireless image transmission with adaptive rate control. In: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2022. 5193–5197
- 123 Dai J C, Wang S X, Tan K L, et al. Nonlinear transform source-channel coding for semantic communications. *IEEE J Sel Areas Commun*, 2022, 40: 2300–2316
- 124 Zhang H W, Shao S, Tao M X, et al. Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data. *IEEE J Sel Areas Commun*, 2023, 41: 170–185
- 125 Dong C, Liang H T, Xu X D, et al. Semantic communication system based on semantic slice models propagation. *IEEE J Sel Areas Commun*, 2023, 41: 202–213
- 126 Huang D L, Gao F F, Tao X M, et al. Toward semantic communications: deep learning-based image semantic coding. *IEEE J Sel Areas Commun*, 2023, 41: 55–71
- 127 Tung T Y, Gündüz D. DeepWiVe: deep-learning-aided wireless video transmission. *IEEE J Sel Areas Commun*, 2022, 40: 2570–2583
- 128 Jiang P W, Wen C K, Jin S, et al. Wireless semantic communications for video conferencing. *IEEE J Sel Areas Commun*, 2023, 41: 230–244
- 129 Wang S X, Dai J C, Liang Z J, et al. Wireless deep video semantic transmission. *IEEE J Sel Areas Commun*, 2023, 41: 214–229
- 130 Shao J W, Mao Y Y, Zhang J. Learning task-oriented communication for edge inference: an information bottleneck approach. *IEEE J Sel Areas Commun*, 2022, 40: 197–211
- 131 Shao J W, Mao Y Y, Zhang J. Task-oriented communication for multidevice cooperative edge inference. *IEEE Trans Wireless Commun*, 2023, 22: 73–87
- 132 Shao J W, Zhang X J, Zhang J. Task-oriented communication for edge video analytics. *IEEE Trans Wireless Commun*, 2023, Early Access: 1–1
- 133 Basagni S, Bölöni L, Gjanci P, et al. Maximizing the value of sensed information in underwater wireless sensor networks via an autonomous underwater vehicle. In: Proceedings of the IEEE Conference on Computer Communications, 2014. 988–996
- 134 Giordani M, Higuchi T, Zanella A, et al. A framework to assess value of information in future vehicular networks. In: Proceedings of the 1st ACM MobiHoc Workshop on Technologies, MOdels, and Protocols for Cooperative Connected Cars, 2019. 31–36
- 135 Rasheed A, San O, Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access*, 2020, 8: 21980–22012
- 136 Yu Q, Ren J, Zhou H B, et al. A cybertwin based network architecture for 6G. In: Proceedings of the 6G Wireless Summit (6G SUMMIT), 2020. 1–5
- 137 Zhou F Q, Li W J, Yang Y, et al. Intelligence-endogenous networks: innovative network paradigm for 6G. *IEEE Wireless Commun*, 2022, 29: 40–47
- 138 Shafin R, Liu L J, Chandrasekhar V, et al. Artificial intelligence-enabled cellular networks: a critical path to beyond-5G and 6G. *IEEE Wireless Commun*, 2020, 27: 212–217
- 139 Sun Z J, Liao Y W, Lu Z M, et al. Implicit semantic-aware communication for 6G ai-native networks. *Mob Commun*, 2023, 47: 7–13 [孙子剑, 廖逸玮, 鲁智敏, 等. 面向 6G 智能内生的隐性语义认知通信. 移动通信, 2023, 47: 7–13]
- 140 Yang Y, Ma M L, Wu H Q, et al. 6G network ai architecture for everyone-centric customized services. 2022. ArXiv:2205.09944
- 141 Tian M Q, Hui Y L, Cheng N, et al. On-demand allocation of 6G communication-sensing-computing resources: a

- scene-recognition-based framework. *Mob Commun*, 2023, 47: 14–18 [田梦秋, 惠一龙, 承楠, 等. 6G 通感算资源按需调配: 一种基于场景识别的架构. *移动通信*, 2023, 47: 14–18]
- 142 Cao H T, Du J B, Zhao H T, et al. Toward tailored resource allocation of slices in 6G networks with softwareization and virtualization. *IEEE Int Things J*, 2021, 9: 6623–6637
- 143 Zhao Z, Wang F, Zhang K J. Research on 6G multi-carrier and multi-modal integration networking and key technologies. *Mob Commun*, 2023, 47: 52–58 [赵喆, 王菲, 张康杰. 面向 6G 多载波多模态融合组网及关键技术探究. *移动通信*, 2023, 47: 52–58]
- 144 Shang S F, Jiang L, Li Y C, et al. Adaptive partitioning and scheduling method of convolutional neural network inference model on heterogeneous platforms. *J Comput Appl*, 2023, 43: 2828–2835 [尚绍法, 蒋林, 李远成, 等. 异构平台下卷积神经网络推理模型自适应划分和调度方法. *计算机应用*, 2023, 43: 2828–2835]
- 145 Geng G L, Gao B, Xiong K, et al. A survey of federated learning for 6G networks. *Chin J Int Things*, 2023, 7: 50–66 [耿光磊, 高博, 熊轲, 等. 联邦学习赋能 6G 网络综述. *物联网学报*, 2023, 7: 50–66]
- 146 Shen J L, Cheng N, Wang X C, et al. RingSFL: an adaptive split federated learning towards taming client heterogeneity. *IEEE Trans Mob Comput*, 2024, 23: 5462–5478
- 147 Yang H, Cheng N, Sun R J, et al. Knowledge-driven resource allocation for D2D networks: a wmmse unrolled graph neural network approach. 2023. ArXiv:2307.05882
- 148 Doke A, Gaikwad M. Survey on automated machine learning (AutoML) and meta learning. In: *Proceedings of the International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021. 1–5
- 149 Perrone V, Jenatton R, Seeger M W, et al. Scalable hyperparameter transfer learning. In: *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2018. 1–11
- 150 Lévesque J C, Gagné C, Sabourin R. Bayesian hyperparameter optimization for ensemble learning. 2016. ArXiv:1605.06394
- 151 Yang Y, Ma M L, Wu H Q, et al. 6G network AI architecture for everyone-centric customized services. 2022. ArXiv:2205.09944
- 152 China Institute of Communications. Frontiers of integrated networks for sensorimotor computing report. Technical Report, 2021 [中国通信学会. 通感算一体化网络前沿报告. 2021]
- 153 Xiong Y F, Liu F, Cui Y H, et al. On the fundamental tradeoff of integrated sensing and communications under gaussian channels. *IEEE Trans Inform Theor*, 2023, 69: 5723–5751
- 154 Fang X R, Feng W, Chen Y F, et al. Joint communication and sensing toward 6G: models and potential of using MIMO. *IEEE Int Things J*, 2023, 10: 4093–4116
- 155 Tian M Q, Hui Y L, Cheng N, et al. On-demand allocation of 6G communication-sensing-computing resources: a scene-recognition-based framework. *Mob Commun*, 2023, 47: 14–18
- 156 Lyu F, Yang P, Wu H Q, et al. Service-oriented dynamic resource slicing and optimization for space-air-ground integrated vehicular networks. *IEEE Trans Intell Transp Syst*, 2022, 23: 7469–7483
- 157 Jiang W W, Ai B, Shen C, et al. Age-of-information minimization for UAV-based multi-view sensing and communication. *IEEE Trans Veh Technol*, 2024, 73: 1100–1114
- 158 Lyu L, Zhao L H, Dai Y P, et al. Adaptive edge sensing for industrial IoT systems: estimation task offloading and sensor scheduling. *IEEE Int Things J*, 2023, 10: 391–402
- 159 Deng L, Li G, Han S, et al. Model compression and hardware acceleration for neural networks: a comprehensive survey. *Proc IEEE*, 2020, 108: 485–532
- 160 Ma L F, Cheng N, Wang X C, et al. On-demand resource management for 6G wireless networks using knowledge-assisted dynamic neural networks. In: *Proceedings of the IEEE International Conference on Communications (ICC)*, 2022. 1–6
- 161 Dai X L, Zhang P Z, Wu B C, et al. Chamnet: towards efficient network design through platform-aware model adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019
- 162 Padhi P K, Charrua-Santos F. 6G enabled industrial internet of everything: towards a theoretical framework. *Appl System Innovation*, 2021, 4: 11
- 163 He J H, Yang K, Chen H H. 6G cellular networks and connected autonomous vehicles. *IEEE Netw*, 2021, 35: 255–261
- 164 Batista E, Lopez-Aguilar P, Solanas A. Smart health in the 6G era: bringing security to future smart health services.

- IEEE Commun Mag, 2023. doi: 10.1109/MCOM.019.2300122
- 165 Kumar J S, Quoc-Viet P, Rukhsana R, et al. Explainable AI over the internet of things (IoT): overview, state-of-the-art and future directions. IEEE Open J Commun Soc, 2022, 3: 2106–2136
- 166 Guo W S. Explainable artificial intelligence (XAI) for 6G: improving trust between human and machine. IEEE Commun Mag, 2020, 58: 39–45
- 167 Shen X M, Gao J, Wu W, et al. AI-assisted network-slicing based next-generation wireless networks. IEEE Open J Veh Technol, 2020, 1: 45–66
- 168 Wu W, Zhou C H, Li M S, et al. AI-native network slicing for 6G networks. IEEE Wireless Commun, 2022, 29: 96–103
- 169 Cheng R T, Zhang H T, Wang L. Deployment strategy of 5G wireless network. Telecomm Sci, 2018, 1: 9–16
- 170 Sun Y H, Peng M G, Zhou Y C, et al. Application of machine learning in wireless networks: key techniques and open issues. IEEE Commun Surv Tut, 2019, 21: 3072–3108
- 171 Liu G Y, Deng J, Zheng Q B, et al. 6G native intelligence: technical challenges, architecture and key features. Mob Commun, 2021, 45: 68–78 [刘光毅, 邓娟, 郑青碧, 等. 6G 智慧内生: 技术挑战、架构和关键特征. 移动通信, 2021, 45: 68–78]
- 172 China Mobile Research Institute. 2030+ White Paper of Vision and Requirements, 2019 [中国移动通信有限公司研究院. 2030+ 愿景与需求白皮书. 2019] <http://cmri.chinamobile.com/wp-content/uploads/2020/11/2030%E6%84%BF%E6%99%AF%E4%B8%8E%E9%9C%80%E6%B1%82%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf>
- 173 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
- 174 Wang Y M, Feng W, Wang J, et al. Hybrid satellite-UAV-terrestrial networks for 6G ubiquitous coverage: a maritime communications perspective. IEEE J Sel Areas Commun, 2021, 39: 3475–3490
- 175 Long W X, Chen R, Moretti M, et al. A promising technology for 6G wireless networks: intelligent reflecting surface. J Commun Inf Netw, 2021, 6: 1–16
- 176 De Lima C, Belot D, Berkvens R, et al. Convergent communication, sensing and localization in 6G systems: an overview of technologies, opportunities and challenges. IEEE Access, 2021, 9: 26902–26925
- 177 Chen H, Sarieddeen H, Ballal T, et al. A tutorial on terahertz-band localization for 6G communication systems. IEEE Commun Surv Tut, 2022, 24: 1780–1815
- 178 Xiao Z Q, Zeng Y. An overview on integrated localization and communication towards 6G. Sci China Inf Sci, 2022, 65: 131301
- 179 Elzanaty A, Guerra A, Guidi F, et al. Toward 6G holographic localization: enabling technologies and perspectives. IEEE Int Things M, 2023, 6: 138–143
- 180 Wymeersch H, Shrestha D, de Lim C M, et al. Integration of communication and sensing in 6G: a joint industrial and academic perspective. In: Proceedings of the 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2021. 1–7
- 181 Ajay A, Du Y L, Gupta A, et al. Is conditional generative modeling all you need for decision-making? 2022. ArXiv:2211.15657
- 182 Du H Y, Wang J C, Niyato D, et al. AI-generated incentive mechanism and full-duplex semantic communications for information sharing. IEEE J Sel Areas Commun, 2023, 41: 2981–2997
- 183 Du H Y, Li Z H, Niyato D, et al. Generative AI-aided optimization for AI-generated content (AIGC) services in edge networks. 2023. ArXiv:2303.13052

## 6G omni-scenario on-demand services provisioning: vision, technology and prospect

Nan CHENG<sup>1</sup>, Fangjiong CHEN<sup>2</sup>, Wen CHEN<sup>3</sup>, Zhimi CHENG<sup>4</sup>, Qinghai YANG<sup>1</sup>,  
Changle LI<sup>1\*</sup> & Xuemin (Sherman) SHEN<sup>5</sup>

1. *School of Telecommunications Engineering, State Key Laboratory of Integrated Services Network, Xidian University, Xi'an 710071, China;*

2. *Electronic and Information Engineering, South China University of Technology, Guangzhou 510006, China;*

3. *Department of Electronics Engineering, Shanghai Jiaotong University, Shanghai 200240, China;*

4. *Datang Mobile Communication Equipment Co. Ltd., Beijing 100083, China;*

5. *Pengcheng Laboratory, Shenzhen 518000, China*

\* Corresponding author. E-mail: clli@mail.xidian.edu.cn

**Abstract** The 6G mobile communication system is facing a dynamic and complex network environment, with diverse scenarios and personalized demands that pose significant challenges to the provision of network services and value realization. Therefore, 6G networks should transition to a new paradigm that is “service-centric”, integrating advanced network technologies and multi-dimensional network resources to achieve on-demand services across all domains and scenarios. This article provides an overview of research on 6G on-demand services across all scenarios. Firstly, it highlights the necessity of on-demand services and outlines the research value of on-demand services in all scenarios. It proceeds to expound on the essence, challenges, and foundations of on-demand services in all scenarios. Subsequently, it analyzes the architecture and key technologies for 6G on-demand services in all scenarios. Finally, the paper offers a prospective outlook on the future and potential research directions for 6G on-demand services in all scenarios.

**Keywords** 6G, omni-scenario on-demand service provision, knowledge, intent-driven networks, AI