中国科学 : 信息科学 2026 年 第 56 卷 第 10 期: 2822–2824
SCIENTIA SINICA Informationis
刊讯

《中国科学》杂志社
SCIENCE CHINA PRESS
CrossMark click for updates

6G 网络大模型技术与智能体通信专题简介

崔琪楣[1*], 王可之[2], 潘存华[3], 刘云浩[4], 潘泉[5], 江沸菠[6], Marco DI RENZO[7], Mérouane DEBBAH[8]

1. 北京邮电大学, 北京 100876, 中国
2. Brunel University London, Uxbridge UB8 3PH, UK
3. 东南大学, 南京 210096, 中国
4. 清华大学, 北京 100084, 中国
5. 西北工业大学, 西安 710072, 中国
6. 湖南师范大学, 长沙 410081, 中国
7. CentraleSupélec, Gif-sur-Yvette 91192, France
8. Khalifa University of Science and Technology, Abu Dhabi 127788, UAE

* 通信作者. E-mail: cuiqimei@bupt.edu.cn

通信与人工智能 (artificial intelligence, AI) 的深度融合是第六代移动通信 (sixth generation mobile communication, 6G) 的重要发展方向. 随着大模型和智能体技术的兴起, 6G 网络正向支持智能体自主交互与协同的新范式演进. 传统判别式模型在复杂推理、跨场景泛化及不确定环境下的稳健决策方面仍面临挑战, 而主要依赖离线预训练的大模型也难以独立满足实时环境适应和持续多轮交互的需求. 基于大模型的智能体通过引入环境感知、记忆、规划与工具调用, 将模型的知识与推理能力融入任务执行闭环, 为面向目标的信息压缩、传输和协同决策提供了新的途径, 有望支撑自动驾驶、家庭服务机器人等具身智能应用.

智能体通信既要求网络为多智能体协作提供高效可靠的信息交互能力, 也推动智能体参与网络架构设计、资源管理和自主优化. 大模型与智能体的协同演进, 为 6G 网络适应动态、异构通信环境提供了新的思路, 同时也带来了交互开销、持续学习、安全隐私及互操作等问题. 为展示这一领域的研究进展, *SCIENCE CHINA Information Sciences* 组织了 “6G 网络大模型技术与智能体通信专题”(Special Topic: 6G Networks: Large AI Model Technologies and Agentic AI Communication), 共汇集 12 篇文章, 涵盖语义智能体网络架构、多模态通信、通信与计算资源优化、低空与卫星网络、模型协同微调以及标准化等方向. 我们将这些研究成果呈现给读者, 以期促进 6G 智能体通信的理论探索与技术发展.

关于智能体通信网络的体系架构与关键技术, 综述 “Towards semantic-based agent communication networks: vision, technologies, and challenges” 系统探讨了语义通信对智能体网络的支撑作用. 文章提出由意图提取与理解、语义编码与处理、分布式自治与协作构成的三层架构, 涵盖具身、通信、网络和应用四类智能体, 并将感知、记忆、推理和行动组织为完整的认知循环. 在此基础上, 梳理了语义增强智能体通信的研究进展, 讨论了理论建模、知识库管理、安全隐私和标准化等方面的挑战与研究方向.

针对低轨卫星网络通信安全、星上功耗与终端能量供给之间的耦合约束, 论文 “LEO satellites meet movable antennas: an SEEI entropy driven agentic AI cross-domain optimization framework” 从信息论角度构

引用格式: 崔琪楣, 王可之, 潘存华, 等. 6G 网络大模型技术与智能体通信专题简介. 中国科学: 信息科学, 2026, 56: 2822–2824, doi: 10.1360/SSI-2026-0511
Cui Q M, Wang K Z, Pan C H, et al. Special topic: 6G networks: large AI model technologies and agentic AI communication. Sci Sin Inform, 2026, 56: 2822–2824, doi: 10.1360/SSI-2026-0511

© 2026《中国科学》杂志社

建安全能效熵度量, 将可移动天线、混合预编码与无线携能通信相结合. 所提智能体利用语义路由和检索增强生成, 根据自然语言描述配置系统模型, 并通过深度强化学习联合优化发射功率、预编码及天线位置. 仿真结果表明, 该方案在满足保密速率和能量采集约束的同时, 能够提高系统能量效率.

面向意图驱动的 6G 网络, 论文 "Agent-defined architecture: make 6G a conversational and thinking network" 提出了智能体定义架构, 使用户和应用能够通过自然语言等方式表达业务需求. 该架构由意图处理层、网络智能体层和网络工具层组成, 引入系统智能体完成意图解析、任务分解及领域智能体协调, 并将网络能力封装为可组合调用的工具. 文章阐述了架构的设计原则与关键挑战, 并通过扩展现实业务用例展示工作流程, 为通信、感知和计算资源的协同编排提供了新的架构思路.

针对视觉语言模型中视觉 Token 粒度受限、序列冗长和跨模态对齐不足的问题, 论文 "TokenCom: vision-language model for multimodal and multitask token communications" 提出了面向 Token 通信的 TaiChi 框架. 该框架采用双视觉 Tokenizer 协同提取高、低分辨率图像中的局部细节和全局特征, 通过双边注意力网络融合多尺度信息, 并利用基于 Kolmogorov-Arnold 网络的模态投影器实现视觉与文本语义空间的非线性对齐. 进一步结合视觉语言模型与信道联合编码, 构建多模态、多任务 Token 通信系统, 实验验证了所提方案的有效性.

在边缘生成式推理中, 设备性能未知、负载变化及设备动态上下线增加了资源管理的难度. 论文 "E^3-Agent: an executable and evolving agent for resource management of edge generative inference" 提出了可执行、可演化的资源管理智能体, 将毫秒级任务派发的快速路由器与事件驱动的大语言模型元控制器相结合. 元控制器通过工具接口调整风险门控、路由配置和性能校准, 系统则依据执行反馈持续更新设备性能认知. 基于设备与模型测量先验的仿真表明, 该方案能够降低动态场景下的平均时延, 并抑制服务卡顿.

关于多智能体协作中不同信息交互形式具有不同的推理与传输开销, 论文 "A token/KV-cache communication media selection and resource allocation strategy for multi-agent collaboration" 研究了 Token 传输与键值缓存传输之间的时延权衡, 揭示了两类方式的适用性取决于计算资源和信道条件. 以最小化协作端到端时延为目标, 文章联合优化通信媒介选择与无线资源分配, 并提出低复杂度求解算法. 数值结果表明, 通过在异构链路上自适应选择交互形式并分配带宽, 所提方案较单一交互方式能够有效降低协作时延.

在 6G 车载元宇宙中, 车辆位置与边缘智能体部署位置的关联可能泄露车辆轨迹, 论文 "Cross-reality location privacy protection in 6G-enabled vehicular metaverses: an LLM-enhanced hybrid generative diffusion model-based approach" 提出跨现实位置隐私保护框架, 利用跨现实位置熵度量隐私水平, 联合优化物理空间的位置扰动与虚拟空间的智能体迁移. 为求解混合动作决策问题, 设计了大语言模型增强的混合扩散近端策略优化算法, 结合奖励设计与双生成式扩散模型提升策略探索能力. 真实数据集上的实验验证了该方法在隐私保护与服务质量之间的平衡效果.

针对多任务、多模态语义通信中语义质量、服务约束与资源消耗的耦合问题, 论文 "Agentic AI-aided resource allocation for multi-task and multi-modal semantic communications" 提出智能体驱动的分层闭环资源优化框架. 框架由网络切片编排智能体确定切片结构与关键参数, 生成可解释的结构先验和偏好约束; 由网络切片资源优化智能体执行连续在线调度, 并通过反馈协调两层决策. 仿真结果表明, 该方案能够适应动态业务到达和时变信道, 在长期系统效用与约束满足程度方面取得较好的性能.

面对低空通感一体化网络的波束赋形复杂的非凸优化与多重约束, 论文 "From general-purpose LLM to specialized beamforming for low-altitude ISAC networks: a LoRA-based multi-expert framework" 提出基于低秩适配的多专家框架, 将通用大语言模型转化为面向通感任务的优化器. 该方法以自然语言描述场景配置和约束, 利用监督微调学习结构化解的生成, 再通过强化学习改善解的可行性与优化性能, 并以轻量化专家模块适配不同场景. 实验验证了该框架在波束赋形性能及跨场景适应方面的优势.

针对分离式信源信道编码在低信噪比条件下易因残余比特错误而出现性能骤降的问题, 论文 "In-context source and channel coding" 提出接收端上下文译码框架, 在保持发射端不变的条件下增强文本传输的鲁棒性. 该框架利用纠错码 Transformer 获取比特可靠性, 通过可靠性引导的比特翻转和多样性采样构造候选序列, 再结合大语言模型算术译码, 以可靠性与序列似然的融合准则选择重建结果. 文章给出了采样过程的稳定

性与收敛性分析, 并在高斯白噪声和瑞利衰落信道下验证了性能增益.

低空无线电地图重建中, 受限的上行传输速率可能造成观测延迟, 使测量数据与动态无线环境失配. 面对这一挑战, 论文 “Diffusion model-based radio map reconstruction for low-altitude wireless networks” 提出通信感知的无人机闭环采样与地图重建框架. 该方法在路径规划中考虑上传时延, 通过时空高斯过程将稀疏异步观测对齐至目标时刻, 再利用条件扩散模型生成无线电地图. 不同任务时限和带宽条件下的仿真表明, 该方案能够改善动态环境中的地图重建精度.

面向 6G 智能体网络中基础模型的协同适配, 论文 “Consensus-optimized decentralized fine-tuning of foundation models for 6G agentic AI networks” 研究了低秩适配与去中心化联邦学习结合时的共识误差放大问题. 文章提出 ConLoRA 方法, 在各智能体间共享并冻结一个非零初始化的低秩因子, 仅更新和同步另一因子, 从而消除双因子乘积引起的误差二次项, 并减少通信开销. 自然语言理解、文本生成及无线网络故障根因分析实验表明, 该方法能够在稀疏拓扑和非独立同分布数据条件下改善模型一致性与任务性能.

以上是本专题收录的全部文章, 衷心感谢所有撰稿作者, 并诚挚感谢所有匿名审稿人对稿件及时且细致的评审工作.