

自智算力网络：架构、技术与展望

周晓茂¹, 贾庆民¹, 胡玉姣¹, 郭凯¹, 马千飘¹, 刘辉¹, 谢人超^{1,2}

(1. 网络通信与安全紫金山实验室, 江苏 南京 211111; 2. 北京邮电大学, 北京 100876)

摘要: 针对算力网络(CNC, computing and network convergence)中的新型业务场景和网络高度智能化的需求, 阐述了自智算力网络(Auto-CNC, autonomous CNC)的技术发展理念, 即通过将智能引入算网全生命流程的方式实现资源一体化、流程自动化、系统智能化, 简要分析了算力网络的研究现状和现存问题, 总结了自智算力网络应具备意图驱动的算网融合、算网闭环自治、网-算-智协同自适应演进的关键特征, 设计了自智算力网络的参考架构, 分析了其关键技术, 介绍了相关的初阶探索工作, 并对自智算力网络的未来发展进行了展望。

关键词: 自智算力网络; 意图驱动; 协同自适应演进; 自智运行

中图分类号: TN914

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2023.00350

Autonomous computing and network convergence: architecture, technologies, and prospects

ZHOU Xiaomao¹, JIA Qingmin¹, HU Yujiao¹, GUO Kai¹, MA Qianpiao¹, LIU Hui¹, XIE Renchao^{1,2}

1. Purple Mountain Laboratories, Nanjing 211111, China

2. Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: In view of the new service scenarios and the demand for high intelligence network in computing and network convergence (CNC), the concept of autonomous CNC (Auto-CNC) is elaborated, where intelligence was introduced into all the aspects of CNC, including resource integration, process automation, and system intelligence. The current research directions and remaining challenges of CNC were introduced, and three key features, i. e., intent-driven computing network, the autonomous system operation and the adaptive co-evolution of communication, computing intelligence, were summarized from the proposed Auto-CNC. Meanwhile, the reference architecture and key technologies of Auto-CNC were described, which were followed by several preliminary exploration cases. Finally, future research trends and technical advice were discussed and recommended.

Key words: autonomous CNC, intent-driven, adaptive co-evolution, autonomous operation

0 引言

随着 5G、云计算、大数据、人工智能(AI, artificial intelligence)等新一代信息技术的结合发展, 工业互联网、车联网、虚拟现实(VR, virtual reality)/增强现实(AR, augment reality)等新型业务正逐渐落地推广。为了满足复杂多元业务场景中

的不同服务质量需求, 新一代信息网络基础设施正向着通信、存储、计算等多要素深度融合的一体化服务方向发展。算力网络(以下简称算网)^[1]作为实现算网基础设施化的一个重要载体, 旨在将泛在的算力资源依托网络进行打通互联、协同调度, 并将不同的应用业务通过最优路径调度到最优的计算节点, 实现用户体验最优, 同时保证网络资源和

收稿日期: 2023-02-10; 修回日期: 2023-06-16

通信作者: 贾庆民, jiaqingmin@pmlabs.com.cn

基金项目: 国家自然科学基金资助项目(No.92267301); 紫金山英才-江宁百家湖计划项目(No.74072203-3)

Foundation Items: The National Natural Science Foundation of China(No.92267301), Purple Mountain Talents-Jiangning Baijia Lake Plan Program(No.74072203-3)

计算资源利用率最优。算力网络的核心思想是基于泛在分布的网络实现无处不在的算力资源,通过构建一张计算资源可感知、可分配、可调度的新型网络实现计算任务的统筹分配和灵活调度,算力资源云边端跨域分布和算网深度融合是其典型特征,为多元用户按需提供优质高效的算力资源服务是其最终目标。为实现上述愿景,学术界和产业界针对具体问题先后提出不同的解决思路,包括算力资源的统一抽象度量、算力优先网络、云边端分布式协同、算网资源的管理和编排等,上述技术思想都不断地推动着算力网络的创新发展和应用拓展。

尽管众多研究机构围绕算力网络在系统架构、流程机制、关键技术、行业应用等^[2-4]进行了大量的研究,并提出了人工智能赋能算力网络建设的重要研究方向,然而算力网络目前仍面临以下挑战。

1) 算网策略设计过度依赖人工经验,资源编排、算网配置、云边端协同等策略的设计通常依赖预设的规律准则,需要大量的经验和人工参与,而且无法应对动态复杂的应用场景需求。

2) 算网智能水平不高,通过 AI 加持形成的单模块智能缺少相互融合协作机制,无法达到更完整、更高级的智能水平,无法实现系统级的独立智能运行。

3) 智能固化问题,系统智能无法根据应用场景的动态变化,包括系统规模的扩缩、基础设施的增删、用户需求的变化、业务场景的更替等,进行自适应调整。

4) 注重业务功能方案设计,忽略系统实际运行的思考。系统的落地运行涉及规划、建设、运行、维护、优化等一系列过程,需要对感知、分析、决策、动作各环节进行综合考虑,而非局限于业务功能方案的设计。

针对上述问题,考虑新型业务场景对算网服务自动化与智能化要求,本文提出自智算力网络的建设思想。作为实现算网资源服务全流程自动化与智能化的关键解决方案,自智算力网络以网络为基础、算力为载体、智能为核心,通过多要素融合实现算网自动、自优、自愈、敏捷使能多元业务并提供极致用户体验,是利用 AI 技术使能算网深度融合与智能服务的新范式,也是智能算力网络构建的终极目标。自智算力网络以构建自智网络的思想使能算力网络的功能实现,在当前算力网络的体系框架下,通过将 AI 技术与算力网络的基础设施、

功能流程、服务应用等深度融合,把 AI 的解决目标和承载方式都设在算力网络内部,利用 AI 技术赋予算力网络基础设施智能化、业务流程一体化、服务能力自优化、算网运维自动化等能力,进而为多元应用提供泛在、高效、灵活、安全的服务化算力供给。在此基础上,自智算力网络还强调能力自主优化和智能自适演进,面向动态变化的应用场景和服务需求,能够通过自学习、自演进不断提升自身业务服务质量和智能化能力。

具体来说,自智算力网络围绕着算网融合一体化的建设目标,通过在系统全生命周期引入“智能”,利用前沿技术实现自动化的算力感知、算力评估、服务编排和调度、算力路由、算力交易等,让算网“自治”,提升业务服务质量和用户的服务体验。自智算力网络的本质是通过数据驱动进行自学习、自演进,对算力网络不同层面进行注智赋能,最大限度地消除算网功能实现对人力的依赖。从理论概念方面来说,自智算力网络是算力网络^[5-6]与自智网络^[7]的深度融合,基于自动化、智能化地实现系统功能的思想,最终实现算网自智,即系统全流程的自动化运行、智简的服务体验、多样化的业务承载、高效的资源利用率、自适应的优化调整等。从技术实现方面来说,算力网络中的基础资源感知和建模、资源的编排策略、算网运行的故障处理、算网系统的优化等关键动作,都可以通过智能算法实现自动化。从系统功能方面来说,算力网络的感知、分析、决策、调度、运维、安全等功能需要自主实现,而且需要不断提高智能化水平来满足日益复杂的功能需求,并进行“网-算-智”的协同迭代,逐渐实现自动化、自治化、自智化。从用户体验方面来说,获取极简的使用体验,根据自身需求/意图获得可靠性最优、资源最优、安全最优的算力服务,是自智算力网络的核心目标。

1 相关技术研究现状

1.1 算力网络

算力网络的概念自提出以来就得到了广泛的认可,产业界和学术界高度重视算力网络体系架构的创新研究,纷纷加大投入力度对相关技术进行探索。在标准制定方面,中国移动、中国联通、中国电信以及华为等单位分别在国际电信联盟(ITU, International Telecommunications Union)、因特网工程任务组(IETF, Internet Engineering Task Force)、

欧洲电信标准组织（ETSI, European Telecommunications Standards Institute）及世界宽带论坛等围绕架构、服务、安全、评估等多个维度开展了多系列行业标准的立项^[8-10]，具体包括架构需求、路由协议、管理编排、算力建模与度量、可信认证、评测指标与方法等。在生态建设方面，在相关政策的大力扶持下，电信运营商依托自身的基础设施资源和产业链资源等方面的优势，结合实际业务需求，协同多方力量推进算力网络生态建设，包括发布算力网络领域相关白皮书、成立算力网络研究组、开展商业模式试运营等。在落地应用方面，算力网络全面助力各行业数智化转型，在VR互动、新媒体直播、智慧医疗、车联网、跨域算力共享等场景中发挥着重要作用^[5-6]，例如，为工业物联网提供基础设施支撑，满足设备互联和智能计算对算网资源的需求。

在学术创新方面，各研究单位分别围绕算力网络的系统架构、关键技术、机理流程、应用模式等方面展开了丰富且深入的研究^[11]，通过结合AI、区块链、大数据等新技术突破算力网络的资源感知与纳管、服务编排与调度、网络控制策略、业务保证与服务开放、安全保障等核心技术。文献[12]提出了基于SRv6的算力网络体系，将算力网络从逻辑上分为算网基础设施层、编排管理层和运营服务层，在此基础上对算力网络的网络控制技术、异构计算资源纳管等机制机理进行了分析探讨。文献[13]和文献[14]分别探讨了算力网络的应用案例和一体化编排与路由。文献[15]提出了基于深度强化学习（DRL, deep reinforcement learning）的一体化任务调度策略，将计算任务分配的问题分为节点选择和节点内部任务分配两个阶段，分别以整合算力资源来提高总体资源利用率和优化节点内部的服务器与计算任务的匹配度来优化调度策略，能够有效降低总成本和提高资源使用效率。此外，文献[16]提出了算力资源、网络资源、智能模型及算法协同共生的融

算力网络建设理念，主要从内生智能和业务智能两个维度进行阐述。文献[1]深度分析了算力网络与现有的软件定义网络（SDN, software defined network）、信息中心网络（ICN, information-centric networking）、区块链、确定性网络等技术的结合思想与技术路线，总结出面向软件定义、命名、可信、确定性的算力网络未来发展方向。算力网络研究现状见表1，相关的研究成果都对算力网络的发展具有重要的参考意义。

1.2 自智网络

近年来，基于AI+Network的融合模式使网络智能已经成为构建新一代网络的研究共识。作为网络智能化的终极目标，自智网络的建设愿景由电信管理论坛（TMF, TeleManagement Forum）率先提出，旨在通过将AI技术深度嵌入网络的硬件、软件、系统、流程等层面，助力网络流程自动化和智能化。随着ETSI、ITU、TMF、第三代合作伙伴计划（3GPP, 3rd Generation Partnership Project）、全球移动通信系统协会（GSMA, Global System for Mobile communications Association）和中国通信标准化协会（CCSA, China Communications Standards Association）等标准化组织在自智网络的愿景目标、参考架构、功能和管理要求、分级测评等方面工作的展开^[17-19]，目前，自智网络在产业愿景、目标架构和分级标准等方面已达成了广泛的产业共识，相关技术已被网络运营商视为实现业务提质、增效、降本，使能业务敏捷创新的重要手段，自智网络研究现状见表2。同时，随着AI算法的突破创新，各类AI算法在自智网络的感知、分析、决策、执行等环节中发挥着关键作用，例如，基于时空长短期记忆（LSTM, long short-term memory）的网络态势感知^[20]、基于变分自编码器（VAE, variational auto encoder）的网络异常检测^[21]、基于知识图谱的网络故障传播推理^[22]、基于DRL的资源按需分配等^[23]。

表1

算力网络研究现状

研究方向	维度	内容
行业标准	功能架构、网内需求、传输协议、隐私安全、应用案例	算力网络框架与架构标准、算力网络总体技术要求、算力网络的信令需求、泛在算力调度管理技术和算力度量、算力网络控制技术需求、算力网络标识解析技术要求、算力网络交易平台技术研究
学术研究	系统架构、流程机制、关键技术	基于多维资源融合的虚拟化架构、基于云网边协同的算网融合策略、基于SRv6的算网一体编排与路由、算力感知路由分配策略、代价感知的任务调度算法
应用场景	生活场景、行业场景、社会场景	VR互动、云游戏、新媒体直播、智慧医疗、车联网、工业物联网、智能科学模拟、平台型算力共享

表 2

自智网络研究现状

研究方向	维度	内容
行业标准	体系架构、关键技术、智能等级、评估方法	自智网络体系架构、自智网络意图管理技术研究、通信网络智能化分级及评估方法、人工智能在电信网络演进中的应用研究、传输网智能化运维研究
学术研究	智能感知、智能分析、自主决策、自动执行	基于时空 LSTM 的网络态势感知、基于 VAE 的网络异常检测、基于知识图谱的网络故障传播推理、基于 DRL 的资源按需分配、基于生成对抗网络 (GAN, generative adversarial network) 的任务策略生成
应用场景	智能网元类、运维智能类	端到端的全流程自动化、意图驱动的按需服务、以用户为中心的闭环运维

在构建自智网络的实践过程中,意图网络^[24](IBN, intent-based network)被认为是实现网络智能化的关键技术,其是由 Gartner^[25]提出的一种能够把意图转换为针对基础设施配置的网络技术,主要流程包括意图表达/收集、转译和验证、自动化部署/实施、网络状态感知、保障和自动调优。意图网络的引入为自智网络确定了清晰的发展目标,即通过对基础设施进行自动配置来实现网络自规划、自适应、自优化、自管理,最终达成“网随意动”。类似地,自智算力网络在面向多元业务的算网自动化全场景按需服务时,也需要构建一种能有效挖掘用户或业务意图并实现算网资源自动化按需分配的方法,对规划设计、自动化配置、仿真验证、监控分析、故障修复和策略优化等进行有效组织,最终实现以意图为中心的全生命周期闭环。

2 自智算力网络愿景与特征要求

2.1 自智算力网络愿景

自智算力网络旨在通过自动化和智能化的手段,实现基础设施智能化、业务流程一体化、服务场景定制化、算网系统自动化等,为多元应用提供泛在、高效、实时、灵活、安全的服务化算力供给,并使能网络自身的自治运行与持续演进,最终基于算网自智实现网络无所不在、算力无所不达、智能无所不及。

然而,算力网络的复杂性以及现有技术的局限性决定了实现完全算网自智不是一蹴而就的,只能是一个长期目标,需要循序渐进、逐步实现。从简单的重复执行替代到复杂的算网融合策略生成、从局部的单域自治到全局的协同自智、从部分需要人工辅助到完全智能自智,算网在相关技术的使能下能够进行全生命周期的自动、自主、自智运行,并能够精准感知用户意图,按需高质地提供算网资源服务。

2.2 自智算力网络特征要求

自智算力网络关键特征如图1所示,包括意图驱动的算网融合、算网闭环自治和网-算-智协同自适应演进。

2.2.1 意图驱动的算网融合

算力网络本质上是一种将多级算力资源与网络资源进行一体化融合的新型信息基础设施,按需为用户提供算力资源服务是其基本功能。为实现面向多元用户的算网自动化全场景按需服务,基于意图的算力网络需要根据用户的业务需求生成最优的算力服务提供策略,并通过对基础设施进行自动配置来进行功能实现。因此,需要实现意图驱动的算网融合。

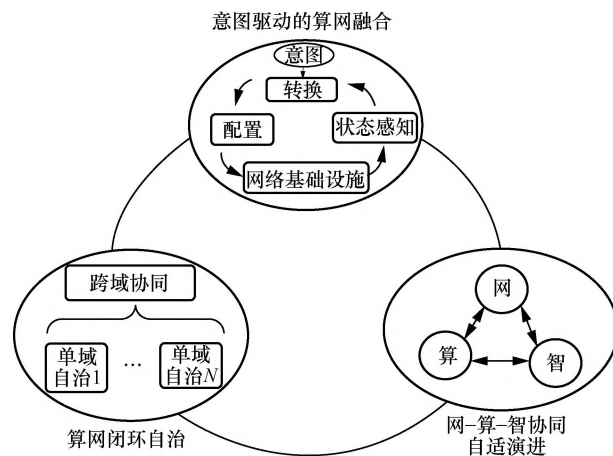


图1 自智算力网络关键特征

为实现上述目标,一方面,需要基于算力网络现有的研究,包括资源感知、算力节点协同机制、任务调度机制等,实现算网的深度融合;另一方面,结合意图网络相关技术,将用户的意图转译成网络可理解、可执行的指令,并根据当前算网状态进行策略验证、执行和结果反馈,确保用户意图正确实现。其中,如何进行用户意图解析是关键难点。目前,通过关键字原句和表达语句描述用户需求的方法^[26-27]存在缺少语义信息、抽象级别低等问题,无

法对复杂的用户需求进行充分表征。而自然语言处理技术^[28]能够准确地识别、解析和预测用户意图，是构建用户和算网之间沟通桥梁的有效手段。

此外，如何自动生成满足意图需求的算网融合策略也是实现算网自智的关键。传统的策略生成方法^[29]依赖大量的人工经验设计，而且无法适应动态变化的用户意图和网络状态。为解决此问题，基于意图-网络双向认知的策略生成机制^[30]通过对用户意图和网络状态信息进行特征挖掘和抽象聚类，提炼出满足各类意图共性和规律的算网融合策略，并将两者的匹配策略以知识的形式进行保存复用，从而为策略生成提供先验知识。

2.2.2 算网闭环自治

自智算力网络是一个高度智能的自动化网络，能够自动化、智能化地完成算网的规划设计、建设实施、维护优化、优化调整、运营管理等生命周期中各个阶段的任务，实现流程自动化、服务自优化和能力自主化，最终实现闭环自治。流程自动化是指算网能够在不依赖人工参与的情况下，利用 AI 技术将单个流程的重复性算网操作转换成由系统自动执行，并将多个环节打通串联，使能整个业务工作流的自动化运行。服务自优化是指算网能够对业务资源的历史数据进行智能分析和预测，实现算网资源和参数配置的自优化，并能够根据用户反馈提升服务质量，形成资源、业务、服务的多重闭环优化。能力自主化是指算网具备独立管理和控制算网的能力，基于信息化、自动化和智能化手段，通过深度感知、智能分析、自主决策和自动执行来对算网基础设施、业务流程、运营运维、服务质量等进行全方位自动化管控，保障系统和业务持续稳定的运行。

算网闭环自治目标的实现依赖 AI 技术的引入。一方面，AI 技术作为一种使能手段，能够显著提高算网感知分析和决策闭环能力。例如，通过联合规则引擎、智能化引擎等技术对算网数据进行挖掘分析，构建面向多元业务的算网故障发现、诊断、自愈、预测的闭环体系，并通过对控制器进行联动控制，从而实现对故障的自动修复。另一方面，在利用 AI 技术对单点或单域注智赋能的基础上，根据业务流程逻辑将各单域拉通互联，解决孤立系统和业务断点问题，将单域的能力赋予业务应用。算力网络在接收到上层指令与任务需求时，能够自动解析业务意图并生成最优策略来对基础设施进行自动化配置，实现全流程的自动化和业务闭环处理。

此外，基于 AI 技术将实时数据、运维经验、专家知识、用户反馈等进行数字化和知识提取，构建面向自智算力网络的机器可理解的知识图谱^[31]，进而赋予算网内生智能能力。上述形成的领域知识能够在算网进行推理分析、意图驱动、闭环决策、自动化运维等时提供技术支撑。

2.2.3 网-算-智协同自适演进

为实现全流程自动化和智能化的目标，自智算力网络需要进行多维度、全方位的智能协同融合与自适演进，形成以数据协同为基础、流程协同为路径、应用协同为目标的端到端的全域智能化闭环。基于此思想，文献[32]提出基于单域自治与跨域协同的解决方案，即首先实现单域全面自治，在此基础上利用不同单域感知的差异性、计算资源的互补性、数据的共享性、域间的交互性，通过跨域协同赋予算网更高层次的智能水平。

当前，大部分网络通过联合知识引导和数据驱动的方式进行智能学习^[33]，然而由于缺少泛化能力，基于特定数据和先验知识学习到的“智能”无法适应复杂多元的应用场景。同时，随着系统规模的扩缩、基础设施的添加删除、用户意图的变化、业务场景的更替等，智能也应该随着算网的动态变化进行自适应调整，形成可演化式智能。目前，基于迁移学习^[25]、终身学习^[26]等的可演进式智能技术正在被广泛研究，具有自适应、自学习、自演进能力的新型网络范式也逐渐成为研究热点。

3 自智算力网络参考架构与关键技术

自智算力网络通过整合自智网络和算力网络的功能诉求，以实现算网自智运行为终极目标，以单域自治+跨域协同为构建方针，以联合知识引导与数据驱动进行智能学习为技术手段，以算力网络中的具体功能为融合点，在实现用户体验最优同时，保证资源利用率最优化。同时，通过算网自治运行，能够在无人工参与的情况下实现流程自动化、参数自配置、故障自修复、系统自优化，提高算网智能化水平。

3.1 自智算力网络参考架构

本节在融合算力网络和自智网络的系统架构基础上，提出自智算力网络的参考架构，并对相关的关键技术进行总结，期望为相关研究提供参考与启发，促进自智算力网络的技术创新与应用落地。自智算力网络参考架构如图 2 所示，设计采用分层

策略, 根据各层的功能抽象程度和业务逻辑, 自下而上包括基础设施层、算网管理层和业务应用层, 同时, 内生智能模块跨越所有层级, 为各个流程环节进行赋能。

基础设施层是自智算力网络的基石, 由各类硬件设施、软件系统、数据等构成, 对全网的算力资源、网络资源、存储资源等进行统一感知管理, 并能够根据业务需求对各类计算、存储资源进行高质量的管理和调度。通过内生智能, 基础设施层一方面能够拓展对自身信息的感知深度与维度, 包括资源感知、性能感知与故障感知等, 为算网管理层进行分析决策提供可靠全面的输入; 另一方面, 通过在数据源头进行分析决策, 能够实现实时不间断的业务响应、设备能耗的智能调节、毫秒级的算网故障感知和故障修复等功能, 提升系统自响应、自修复、自优化能力。

算网管理层是算力网络的大脑, 负责系统功能的具体实现。从功能内容来说, 算网管理层通过南北向接口分别对基础设施层状态信息和业务意图进行输入, 在此基础上进行分析决策和算网控制, 包括状态感知、资源调度、算力管理、服务编排、故障分析与自修复等, 从而实现感知、分析、决策、执行的全流程闭环管理。从实现方式来说, 算网管理层采用单域自治与跨域协同的分层渐进策略来实现系统自智。单域自治强调针对系统子功能模块的自动化、智能化执行, 并能够根据资源状态、用

户意图等进行自适演进, 从而实现针对单一自治域的局部自智。在此基础上, 跨域协同以各个自治域为基本单位, 将各独立的单域拉通互联, 通过多域协同的方式解决使用单一模块无法解决的复杂的问题, 进而实现功能流程的自动化和业务高层智能化闭环处理。算网管理层将 AI 技术深度嵌入算网的各个层面, 全方位提升不同功能模块的智能化学习及场景适应能力, 支持个性化智能服务能力的持续演进, 保证算网对当前业务和未来新业务的服务质量。

业务应用层用于实现面向用户的服务能力开放, 承载着抽象的业务功能。从用户的角度来说, 业务应用层能够根据用户意图自动地将服务应用调度到合适的节点, 实现资源利用率最优并保证极致的用户体验。

内生智能模块通过数据管理、学习训练、智能分发、持续学习等形成完整的智能闭环, 能够为基础设施层、算网管理层及业务应用层提供全方位的智能服务, 进而为多样化业务需求和算网自智提供智能化所需的基础能力。作为智能能力管理与知识统一中心, 内生智能模块摒弃传统“外挂式”AI 的方式, 在算力网络各个层级都实现感知、分析、决策、执行等功能与 AI 的深度融合, 将 AI 的设计训练、推理验证、部署应用、迭代优化等全生命周期都设在算网内部, 使得 AI 诞生于算网并服务于算网。基于 AI 强大的学习、分析和决策能力, 内生智能模块通过对算网运行过程中产生的数据进行深度挖

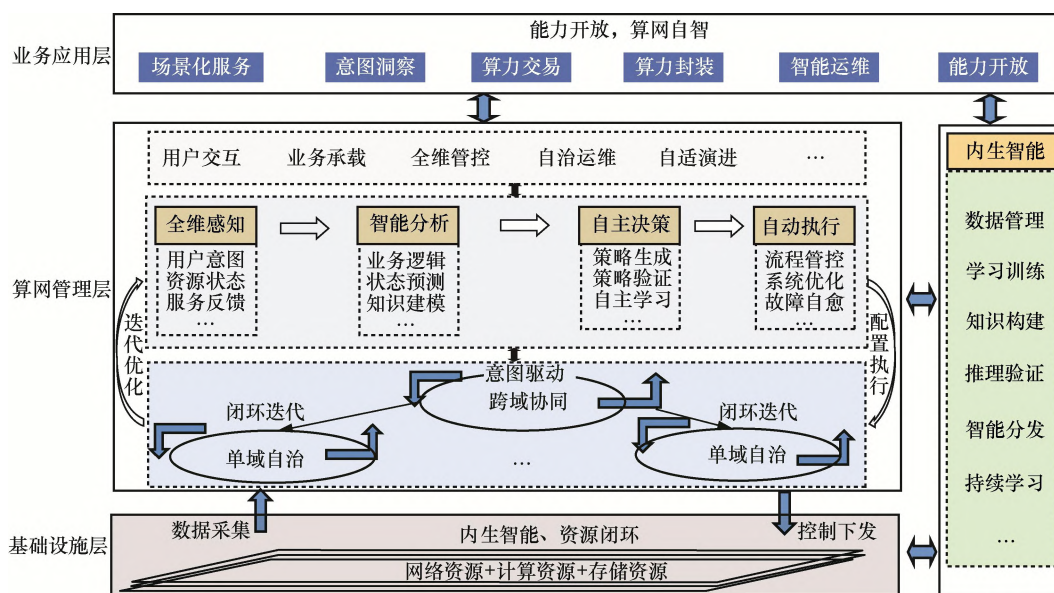


图 2 自智算力网络参考架构

掘, 并协同整合不同网络层之间的数据、资源、功能等方面的差异, 能够针对各种业务形成有效解决方案, 并综合考虑算网的运行效率、自动化水平、服务质量等。同时, 内生智能模块还支持 AI 的持续学习, 能够随着需求的改变自适应地进行知识演进, 并能够进行知识融合与推理, 从而产生新的知识, 这也是内生智能体需要具备的重要特征之一。

3.2 自智算力网络关键技术

自智算力网络的实现依托算力网络和自智网络两个领域的技术融合。自智网络为算力网络提供智能化解决方案, 算力网络为自智网络提供业务场景。一方面, 需要实现算力网络的基础功能, 如多节点异构资源整合、算力感知评估与度量、服务编排与调度、算力路由等^[27]; 另一方面, 需要参考自智网络的系统架构、标准规范、技术指标、评价体系等, 在数据、基础设施、算法、应用等方面进行积累迭代, 并联合知识引导与数据驱动使能智能化业务流程^[28]。在构建自智算力网络过程中, 需要将算力网络的具体业务功能与自智网络的智能化策略进行深度融合, 基于智能算法使能流程自动化、智能化, 并由单域自治演进到多域协同智能, 逐步推动算网自主运行、自适应演进, 最终实现算网自智。

基于上述思想, 以下从感知、分析、决策及执行 4 个方面对自智算力网络的关键技术进行阐述。

3.2.1 算网全维感知

算力网络利用网络手段将计算、存储等基础资源在云边端之间进行连接与协同, 从而提升业务服务质量和用户的服务体验。精准地对泛在异构、动态时变的计算资源的部署位置、实时状态、负载信息等进行感知, 以及对网络的传输时延、抖动、带宽资源利用率等信息进行实时动态获取, 是实现上述目标的前提。

当前, 算力网络主要围绕如何提高对计算资源和网络资源的感知精度、感知效率、感知范围及感知智能性等方面展开研究^[34-36]。针对单个节点存在感知范围有限、效率低等问题, 通过将多节点感知的算网信息进行聚合来扩展感知的维度和范围, 并利用智能算法构建全局统一的算网状态视图。另外, 基于资源状态预测机制^[37]也经常被用来提高网络的感知效率和精度。通过对计算与网络资源状态的历史和当前数据进行采集、处理和分析, 结合机器学习相关算法, 构建各节点算力状况以及网络状况的预测模型, 进而实现对计算与网络资源状态的

预测性感知。此外, 借助先进算法模型将多节点感知数据进行定制化的特征提取及融合处理, 将原始感知信息转换为可被终端或用户直接理解的意图及语义信息, 实现感知的认知能力增强。

同时, 业务意图^[38]也是算网资源感知的重要部分。算力网络需要对业务内容和意图进行全面感知, 综合考虑当前的网络和计算资源状态, 通过匹配算力和服务, 将不同的计算任务调度到合适的节点进行处理, 实现通信和算力在网络的全局最优。其中, 算网与业务的交互方式包括被动式和主动式, 前者由业务或用户直接向网络提供所需的资源信息, 后者则需要网络自身对业务或用户意图进行解析评估。一般来说, 业务意图以抽象的符号、文字、语音等方式存在, 算力网络需要借助自然语言处理技术将其转换成算网意图表达模型, 从而生成满足业务所需要的资源配置策略。

3.2.2 算网智能分析

算网智能分析需要在感知的基础上对用户业务、算网状态、功能流程等多个方面进行深度理解, 并在此基础上形成面向算网自智的知识空间。首先, 针对算力网络中泛在化的异构算力资源以及多样化的业务需求, 如何有效地对算力进行标识和度量、对任务内容进行分类解析、对用户的满意度进行测评量化, 进而对业务需求-资源状态-服务质量之间的相关性进行准确评估是算网融合的基础步骤。其次, 算力网络需要对自身的状态进行实时分析, 洞悉当前的业务能力、性能水平、安全状态等, 并进行风险预判和预测性资源配置。同时, 算网还需要对不同流程、环节之间的逻辑关系进行解析, 从而为多元业务构建最优的实现流程。在发生故障时, 算网需要对故障根因进行分析, 快速准确地定位故障节点, 并提供解决意见或方案。此外, 算网还需要对业务承载、系统运行、故障分析等功能背后的机理进行认知, 在解决问题的过程中将相关的规律、法则、策略等凝练为可复用的知识, 从而支撑算网进行推理分析、闭环决策、自治运行等。

3.2.3 算网自主决策

在算力网络中, 将复杂多元的计算任务分派并调度到最匹配的计算节点进行高效处理, 需要算网具备实时精确、灵活智能的决策能力, 能够根据任务需求和算网状态信息实现资源编排、路由选择、任务调度等功能。传统的策略生成通常基于人为设定的规则和经验数据, 如基于链路基础度量值的路

由选择^[39]、基于分时的计算节点选择^[40]、基于加权代价函数的任务调度^[41]等。然而,随着算网规模和业务类型的飞速扩增,上述的算网策略生成方式成本越来越高,而且无法保证最优的服务质量和用户体验。为解决此问题,基于知识的算网自主决策通过利用构建的领域知识库对任务需求进行分析,使算网能像人类一样智能、自动地做出决策。目前,基于 AI 的策略自动生成机制、基于数字孪生等技术的策略验证,以及基于用户反馈的策略自动优化等技术已经成为相关研究的热点。

3.2.4 算网自动执行

算力网络的基本目标是根据自身资源状态最优地进行用户意图执行和方案部署,而自智运行概念的引入又赋予其自动化、智能化的发展目标,其自动执行涵盖动作的自动完成、故障的自动修复、服务的自动优化、能力的自主学习等多个层面。考虑传统算力网络缺少系统功能运行层面的思考,通常无法有效应对算网规模动态化调整、意图适应性承载、故障自动化修复、智能自适应演进等需求。为解决上述问题,首先,需要对算网全元素进行多粒度、全维度的精准控制,满足复杂多元应用的不同服务需求。其次,需要对算网进行全局协同控制,在对单域内元素进行管控的基础上,在业务运行、系统优化、监控排障等具体场景中通过联合多域构筑端到端的智能化闭环,进一步支撑算网实现自动化运行、故障自修复等。另外,引入智能控制策略,利用机器学习与深度学习等方法对算网数据进行分析和学习,绕过复杂的数据特征提取和网络功能建模,基于生成的普适性、智能化的控制策略完成端到端的控制流程,最终实现算网运行的智能化。

此外,高度智能化的算力网络还需要具备自优化和自学习的能力。在业务的承载以及与用户的交互过程中,算网能够根据用户的反馈自动化地对资源配置、业务流程、服务质量等进行改进,从而为用户提供更好的体验。同时,算网还需要利用自主学习能力不断扩展和改善自身能力水平,在不受外界支配与干扰的情况下,自动地选择学习目标、制定学习计划、构建学习方法、评价学习结果等,从而获得应对新业务需求的能力。

4 自智算力网络研究挑战与发展方向

自智算力网络受算力网络和自智网络的双重

驱动,同时也面临着来自两个领域的技术挑战。一方面,尽管在系统架构、关键技术以及应用场景等方面取得了诸多进展,算力网络仍需要解决度量、感知、路由、交易、编排等多方面的技术难题,以及面临应用落地过程中的服务质量、系统效率、智能化等诸多挑战;另一方面,自智网络的研究还处于初级阶段,在系统架构、关键技术、标准规范以及应用场景等方面尚未形成统一、完整的解决方案。

同时,自智算力网络在融合算力网络与自智网络的过程中又会衍生出新的挑战,例如,智能化能力建设缺乏完备的基础设施支撑,核心算法缺少领域知识、数据以及场景验证,融合机制缺少统一标准和评测规范等。在应对上述挑战的过程中,自智算力网络需要将不同领域的问题进行统一融合,在 AI、算力网络、自智网络等领域现有研究成果的基础上,利用构建自智网络思想和手段解决算力网络的具体问题,把 AI 解决方案的目标与承载方式都建立在算力网络内部,在实现算网深度融合同时,使能算网自治运行。

针对构建自智算力网络过程中的研究挑战和实践难题,本节选取若干代表性的问题进行简要阐述分析,包括多维功能智能化与一体化融合、基于知识定义的算网自治、基于多节点协同的可演进式智能学习策略、基于深度学习的网络故障检测与根因分析,并介绍相关的技术探索方向,期望能够为今后相关的工作提供参考和启发。

4.1 多维功能智能化与一体化融合

目前,利用 AI 使能技术提高网络在感知、控制、通信、计算、服务等方面的能力已经成为相关领域的研究趋势。算力网络中海量的数据能够为 AI 提供数据支撑,而难以建模和无法精确求解的感知、预测、决策、优化等任务则为 AI 提供多样的应用场景,通过对大量数据进行自动特征挖掘和规则学习,AI 算法可以显著提高网络多维功能的运行效果和智能性。例如,基于图神经网络建立双向消息传递的网络状态认知模型^[20],对网络资源的时序特征和拓扑结构进行多维度感知。基于循环神经网络构建预测机制的通信优化策略^[16],对不同业务场景中的通信质量进行个性化优化。基于云边端协同计算与学习^[42],提升网络解决复杂任务的能力与效率。基于分层深度强化学习的跨域分层控制策略,对网络进行全维度、多粒度的控制。同时,将文档中的人工经验与知识数字化,通过数据驱动与知识

引导相结合的方式构建算网智能化解决方案，相关技术已经成为研究热点。

此外，随着复杂多元化业务对不同维度功能需求差异的增大，通过多维功能一体化将多维感知、协作通信、协同计算、实时控制、泛在智能等功能进行融合是实现算网深度融合的重要内容，也是构建自算力网络的基础。多维功能智能化与一体化融合架构如图3所示，多维功能智能化与一体化融合旨在将各功能模块的资源和功能进行一体化集成与融合，实现通信、感知、计算、智能各功能彼此之间的深度支撑、互助互惠，良性地循环、促进、提升系统综合性能，为多元化业务需求提供资源深度融合使用支撑，有效提升算网服务能力、安全性、可靠性、资源利用率等。

4.2 基于知识定义的算网自治

算网自治是构建自算力网络的重要内容和关键步骤，主要特征包括流程自动化、服务自优化和能力自主化。为实现此目标，向算力网络引入AI技术是主要手段，然而其关键不是实现单个环节或流程的自动化，而是在解决问题的过程中将网络相关的规律、机理、策略凝练为抽象的、可复用的知识，最终形成面向算网自治的完整的知识体系。基于形成的知识空间，自算力网络能够与用户进行智能交互、如同专家一样做出决策，并进行持续学习，真正意义上实现对算网的智能自主管控。

基于知识定义的算网自治^[43]旨在全域网络感

知的基础上，使用AI技术处理算网数据生成知识，并利用这些知识帮助算网进行闭环自治。作为数据的抽象升级，知识能够更好地揭示策略生成的可解释性和逻辑推理，并通过挖掘不同知识点之间的深层关系来推理探索新的知识。算网知识的生成依赖不同类型的数据，包括日志、性能指标、运维手册、业务需求、用户体验等，以及先验知识，如拓扑、专家规则、运维经验等。基于自动化或半自动化的知识抽取方法，算网知识可以以结构化数据、知识图谱、AI模型等方式进行表征，具体内容可归纳为历史记载、客观现状、主观体验和动作反馈4种类型。知识定义的算网决策流程示意图如图4所示，在算网运行过程中，知识的引入使得各种策略生成更加高效精准，并能够通过推理和再学习来适应动态的业务需求。作为长期愿景，算网可以学习如何自动化地管理和控制网络，最终实现自主学习、自治运行。目前，基于知识定义的方式实现算网自治/自智已逐渐得到业界的认可，相关的研究也在积极开展中^[43]。

4.3 基于多节点协同的可演进式智能学习策略

繁杂动态的运行环境、更高级别的智能化要求、日益复杂的个性化需求等对算网智能的准确度、效率、智能水平、安全性提出了更高的要求，如何构建具有自适应演进能力的新型智能学习系统已成为人工智能领域的演进热点，也是使能自算力网络进行网-算-智协同演进的关键技术。为实现此目标，可以利用不同节点功能的差异性、计算

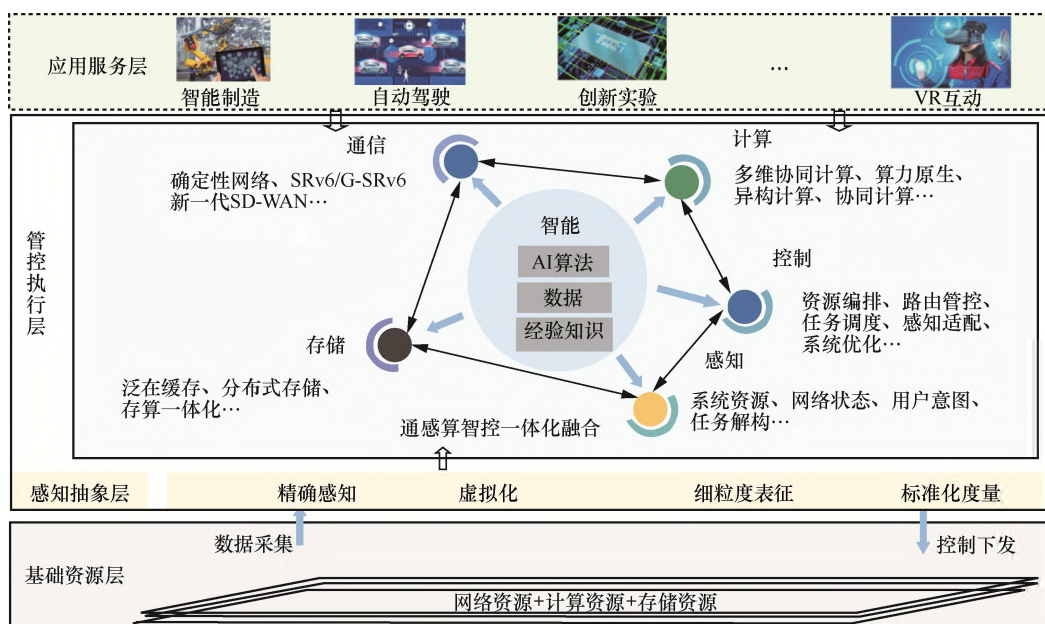


图3 多维功能智能化与一体化融合架构

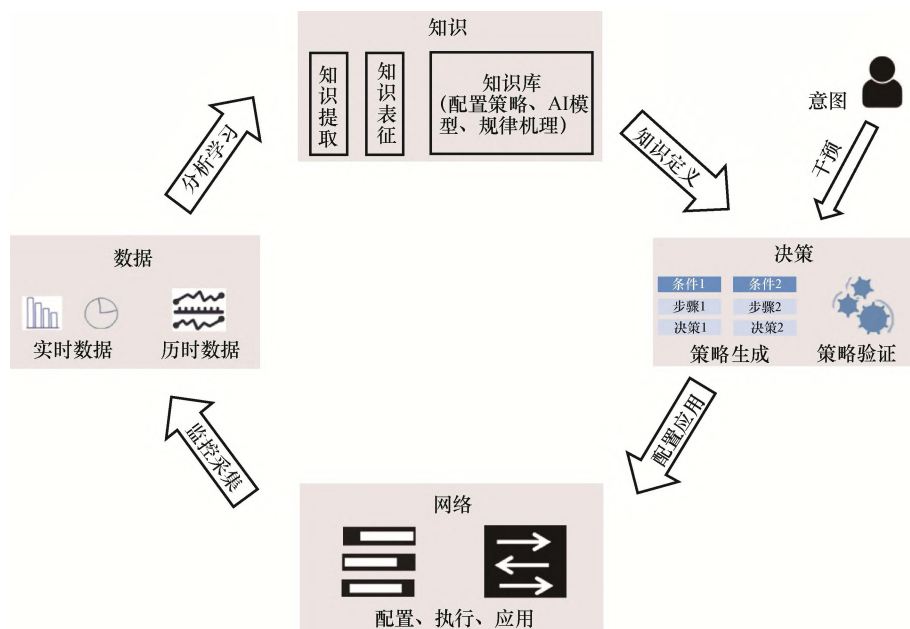


图 4 知识定义的算网决策流程示意图

资源的互补性、节点间的交互性，通过多节点协同自适应感知、智能在线适配学习、迁移学习与终身学习、群智能体分布式学习等赋予算网更高层次的智能水平。例如，针对单节点资源有限、任务动态变化、学习效率低等问题，基于多节点协同的自适应适配学习，能够联合自身数据和其他节点的先验知识进行高效学习，解决“智能固化”问题。此外，针对多节点协同学习存在学习效率低、数据异构、训练结果泛化能力差等问题，基于多容量权值共享的高效联邦学习策略^[44]，在提高协同学习效果的同时，可以最大化保留不同节点的个性化学习结果。相比于通过引入固化的 AI 模型为算网注智，可演进式智能学习策略能够赋予算网自适应演进和自主学习的能力，对进一步提升算网的智能水平具有重要意义。

4.4 基于深度学习的网络故障检测与根因分析

随着规模的扩大和功能复杂度的提升，自智算力网络需要以自动化技术提升故障发现、分析、处理、预测等流程的效率和智能性。针对现有技术中算网故障处理不力的问题，相关研究通过将 AI 技术应用用于故障发现、根因分析、故障自愈等领域，旨在实现自动化的智能运维，如基于时空图注意力网络的流量预测、面向多元时空数据的无监督混合异常检测、基于模型集成的网络故障发现及根因分析等^[45]。然而，目前的研究侧重于如何利用海量数据提升单一环节的算法检测或识别的准确率，在根因分析等复杂的推理任务中表现较差，距离算网自

治运维的要求还有很长距离。为此，还需要引入外部知识来弥补在业务流程、运维策略、故障因果等方面的不足，利用数字化手段将专家经验、领域知识与 AI 技术等相融合，构建面向自智算力网络自动化运维的专家系统。

5 结束语

算力网络作为面向算网融合演进的新型网络架构，是支撑泛在智能多元计算需求的关键。同样，自智网络是实现网络自动化、智能化，为用户提供极致的服务体验的必然要求。自智算力网络需要结合算力网络和自智网络两个研究领域现有的成果，并在此基础上进行融合创新。同时，实现算网自智是一个持续迭代、循环演进的系统工程，需要在以下方面持续推进。

1) 持续探索算力网络与自智网络的融合关键点与策略，将算力网络现存的问题挑战作为切入点，结合自智网络关键技术，构建自智算力网络解决方案。

2) 加强自智算力网络顶层设计，在系统架构、标准规范、技术指标、测试体系等方面形成统一规范和标准。

3) 突破关键技术，包括 AI、算网融合、智能运维等多方面，在数据、流程、应用等方面进行多维度深度的协同融合，构筑端到端的全域智能化闭环，逐渐实现网络自动化、自主化、自治化。

4) 应用落地、生态推广, 以用户商业应用和业务意图为驱动, 促进学术和产业的双向发展, 逐步进行功能落地。

参考文献:

- [1] 贾庆民, 丁瑞, 刘辉, 等. 算力网络研究进展综述[J]. 网络与信息安全学报, 2021, 7(5): 1-12.
JIA Q M, DING R, LIU H, et al. Survey on research progress for compute first networking[J]. Chinese Journal of Network and Information Security, 2021, 7(5): 1-12.
- [2] KROEL M, MASTORAKIS S, ORAN D, et al. Compute first networking: distributed computing meets ICN[C]//The 6th ACM Conference on Information-Centric Networking. New York: ACM Press, 2019: 67-77.
- [3] NOUR B, MASTORAKIS S, MTIBAA A. Compute-less networking: perspectives, challenges, and opportunities[J]. IEEE Network, 2020, 34(6): 259-265.
- [4] 曹畅, 唐雄燕. 算力网络关键技术及发展挑战分析[J]. 信息通信技术与政策, 2021, 47(3): 6-11.
CAO C, TANG X Y. Analysis on key technologies and development challenges of computing network[J]. Information and Communications Technology and Policy, 2021, 47(3): 6-11.
- [5] 中国联通网络技术研究院. 中国联通算力网络白皮书[R]. 2019.
China Unicom Network Technology Research Institute. White paper on China Unicom computing power network[R]. 2019.
- [6] 中国移动研究院. 算力感知网络技术白皮书[R]. 2019.
China Mobile Research Institute. White paper on computing-aware network technology[R]. 2019.
- [7] 中兴通信. 自智网络白皮书[R]. 2022.
ZTE. White paper on autonomous network[R]. 2022.
- [8] IETF. Computing in the network research group (coinrg) [EB]. 2019.
- [9] IETF. Routing area working group (rtgwg)[EB]. 2019.
- [10] IETF. Computing in network research group, requirement of computing in network draft-liu-coinrg-requirement-02[R]. 2020.
- [11] DI Z, CAO Y, QIU C, et al. New computing power network architecture and application case analysis[J]. Journal of Computer Applications, 2022, 42(6): 1656-1662.
- [12] HUANG G, SHI W, TAN B. Computing power network resources based on SRv6 and its service arrangement and scheduling[J]. ZTE Technology Journal, 2021, 27(3): 23-28.
- [13] TANG X Y, CAO C, WANG Y X, et al. Computing power network: the architecture of convergence of computing and networking towards 6G requirement[J]. China Communications, 2021, 18(2): 175-185.
- [14] ZHANG S, CAO C, TANG X. Computing power network technology architecture based on SRv6[J]. ZTE Technology Journal, 2022, 28(1): 11-15.
- [15] 梁冰, 纪雯. 基于次模优化的边云协同多用户计算任务迁移方法[J]. 通信学报, 2020, 41(10): 25-36.
LIANG B, JI W. Multiuser computation offloading for edge-cloud collaboration using submodular optimization[J]. Journal on Communications, 2020, 41(10): 25-36.
- [16] 胡玉姣, 贾庆民, 孙庆爽, 等. 融智算力网络及其功能架构[J]. 计算机科学, 2022, 49(9): 249-259.
HU Y J, JIA Q M, SUN Q S, et al. Functional architecture to intelligent computing power network[J]. Computer Science, 2022, 49(9): 249-259.
- [17] 通信世界. 中国移动牵头 3GPP 成立自治网络分级标准项目[EB]. 2020.
Communication World Network. China Mobile leads 3GPP to establish autonomous network classification standard project[EB]. 2020.
- [18] ITU-T. Focus group on machine learning for future networks including 5G[S]. 2020.
- [19] ITU-T. Framework for evaluating intelligence levels of future networks including IMT-2020:ITU-T Y.3173[S]. 2020.
- [20] XU X X, SUN J, WANG C Y, et al. A novel hybrid CNN-LSTM compensation model against DoS attacks in power system state estimation[J]. Neural Processing Letters, 2022, 54(3): 1597-1621.
- [21] CHEN W X, XU H W, LI Z Y, et al. Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE[C]//Proceedings of IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 1891-1899.
- [22] ZHANG R, MAO Y, ZHAO W. Knowledge graphs completion via probabilistic reasoning[J]. Information Sciences, 2020(521): 144-159.
- [23] GUO S Y, DAI Y, XU S Y, et al. Trusted cloud-edge network resource management: DRL-driven service function chain orchestration for IoT[J]. IEEE Internet of Things Journal, 2020, 7(7): 6010-6022.
- [24] WEI Y, PENG M, LIU Y. Intent-based networks for 6G: insights and challenges[J]. Digital Communications and Networks, 2020, 6(3): 270-280.
- [25] Gartner. Innovation insight: intent-based networking systems[EB]. 2017.
- [26] QIAO F F, ZHU X J. Domain intelligent Q&A user intention recognition based on keyword separation[C]//Proceedings of 2020 International Conference on Culture-oriented Science & Technology (ICCST). Piscataway: IEEE Press, 2020: 224-229.
- [27] PÉREZ-SOLER S, GUERRA E, DE LARA J. Model-driven chatbot development[M]. Conceptual Modeling. Cham: Springer International Publishing, 2020: 207-222.
- [28] WU S F. Internet public information an text data mining and intelligence influence analysis for user intent understanding[J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(1): 487-494.
- [29] WANG Y, FORBES R, CAVIGIOLI C, et al. Network management and orchestration using artificial intelligence: overview of ETSI ENI[J]. IEEE Communications Standards Magazine, 2018, 2(4): 58-65.
- [30] GU Z Y, SHE C Y, HARDJAWANA W, et al. Knowledge-assisted deep reinforcement learning in 5G scheduler design: from theoretical framework to implementation[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(7): 2014-2028.
- [31] CHEN X, JIA S, XIANG Y. A review: knowledge reasoning over knowledge graph[J]. Expert Systems With Applications, 2020(141): 112948.
- [32] 邓灵莉, 顾宁伦, 袁向阳, 等. 网络智能化标准、开源与产业研究[J]. 电信科学, 2021, 37(10): 12-21.
DENG L L, GU N L, YUAN X Y, et al. Network intelligence standards, open source and industry research[J]. Telecommunications Science, 2021, 37(10): 12-21.
- [33] SARKER I H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions[J]. SN Computer Science, 2021, 2(6): 420.
- [34] 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[J]. 电信科学, 2021, 37(10): 76-85.

- DUAN X D, YAO H J, FU Y X, et al. Computing force network technologies for computing and network integration evolution[J]. Telecommunications Science, 2021, 37(10): 76-85.
- [35] THAKKAR H K, SAHOO P K, VEERAVALLI B. RENDA: resource and network aware data placement algorithm for periodic workloads in cloud[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(12): 2906-2920.
- [36] HAGHNEGAHDAR L, JOSHI S S, DAHOTRE N B. From IoT-based cloud manufacturing approach to intelligent additive manufacturing: industrial internet of things—an overview[J]. The International Journal of Advanced Manufacturing Technology, 2022, 119(3): 1461-1478.
- [37] SUN Z J, YANG H, LI C, et al. Cloud-edge collaboration in industrial internet of things: a joint offloading scheme based on resource prediction[J]. IEEE Internet of Things Journal, 2022, 9(18): 17014-17025.
- [38] DE S, DEY S, BHATIA S, et al. An introduction to data mining in social networks[J]. Advanced Data Mining Tools and Methods for Social Computing, 2022: 1-25.
- [39] VIJAYABASKAR, KUMANAN T. Link quality and energy-aware metric-based routing strategy in WSNS[M]. Intelligent Computing in Engineering. Singapore: Springer Singapore, 2020: 533-539.
- [40] SAMAL T, KABAT M R. Energy-efficient time-sharing multichannel MAC protocol for wireless body area networks[J]. Arabian Journal for Science and Engineering, 2022, 47(2): 1791-1804.
- [41] LI C, TANG J, TANG H, LUO Y. Collaborative cache allocation and task scheduling for data-intensive applications in edge computing environment[J]. Future Generation Computer Systems, 2019(95): 249-264.
- [42] 刘泽宁, 李凯, 吴连涛, 等. 多层次算力网络中代价感知任务调度算法[J]. 计算机研究与发展, 2020, 57(9): 1810-1822.
- LIU Z N, LI K, WU L T, et al. CATS: cost aware task scheduling in multi-tier computing networks[J]. Journal of Computer Research and Development, 2020, 57(9): 1810-1822.
- [43] 王敬宇, 周铨, 张蕾, 等. 知识定义的意图网络自治[J]. 电信科学, 2021, 37(9): 1-13.
- WANG J Y, ZHOU C, ZHANG L, et al. Knowledge-defined intent-based network autonomy[J]. Telecommunications Science, 2021, 37(9): 1-13.
- [44] ZHOU X, JIA Q, XIE R. NestFL: efficient federated learning through progressive model pruning in heterogeneous edge computing[C]//The 28th Annual International Conference on Mobile Computing and Networking. [S.l.:s.n.], 2022: 817-819.
- [45] SHUAI L, WANG K, ZHANG L, et al. Global-local integration for GNN-based anomalous device state detection in industrial control systems[J]. Expert Systems With Applications, 2022(209): 118345.

[作者简介]



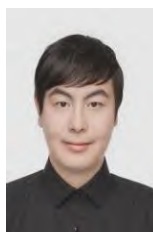
周晓茂 (1993-), 男, 博士, 网络通信与安全紫金山实验室研究员, 主要研究方向为边缘计算、人工智能、算力网络、工业互联网等。



贾庆民 (1990-), 男, 博士, 网络通信与安全紫金山实验室研究员, 主要研究方向为算力网络、工业互联网、确定性网络、人工智能等。



胡玉姣 (1993-), 女, 博士, 网络通信与安全紫金山实验室研究员, 主要研究方向为信息物理融合系统、算力网络、人工智能、云边协同等。



郭凯 (1991-), 男, 博士, 网络通信与安全紫金山实验室研究员, 主要研究方向为算力网络、工业互联网、确定性网络、人工智能等。



马千飘 (1993-), 男, 博士, 网络通信与安全紫金山实验室研究员, 主要研究方向为边缘计算、联邦学习、分布式机器学习等。



刘辉 (1983-), 男, 博士, 网络通信与安全紫金山实验室高级工程师, 主要研究方向为边缘智能体系架构、算力网络、云计算等。



谢人超 (1984-), 男, 博士, 北京邮电大学教授、博士生导师, 主要研究方向为未来网络体系架构设计、信息中心网络、移动网络内容分发等。