

就地计算、云边协同：传感云与边缘计算的一体化框架(特邀)

王田, 李雨婷, 王文华

(北京师范大学人工智能与未来网络研究院, 广东 珠海 519087)

摘要: 面向大规模感知与智能应用场景, 集中式计算在时延、带宽、能耗与隐私保护的多重约束下逐渐呈现边际效益递减, 计算范式因此由单一的“万物上云”模式, 逐步转向“就地计算与云边协同”的新形态。在此背景下, 本文首先梳理集中化计算路径在不同发展阶段所具备的优势及其适用边界, 进而界定边缘计算在端-云之间所扮演的关键角色。在此基础上, 进一步概述“传感云-边缘-端”协同计算框架, 重点分析其中的核心机制, 包括数据“必要即上行”的传输原则、面向服务级别协议(SLA)感知的任务分配与双层调度策略, 以及边侧即时闭环执行与云侧全局策略治理之间的分工与协同关系。随着计算与智能能力向边缘侧持续下沉, 本文进一步讨论边缘智能的发展方向, 涵盖模型轻量化与本地学习机制、联邦学习与知识蒸馏的协同范式, 以及面向边缘环境的智能运维(AIOps for Edge)与多级降级机制所支撑的自治能力。同时, 强调构建以端到端闭环效率、系统韧性与可追责性为导向的综合评价体系的重要性。最后, 结合教育等典型应用场景以及产业实践, 论证就地计算与云边协同在保障确定性时延、提升系统整体韧性以及实现跨域一致性方面的现实有效性, 并据此指出计算范式由边缘计算向云边智能协同演进的必然趋势与发展方向。

关键词: 云边协同; 传感云; 边缘智能; 就地计算; 大小模型协同

中图分类号: TP18

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0253308

Near-Source Computing and Cloud-Edge Collaboration: An Integrated Architecture for Sensor Cloud and Edge Computing (Invited)

WANG Tian, LI Yuting, WANG Wenhua

(Institute of Artificial Intelligence and Future Networks, Beijing Normal University, Zhuhai 519087, Guangdong, China)

【Abstract】 Centralized computing exhibits diminishing returns under latency, bandwidth, energy, and privacy constraints in large-scale sensing and intelligent applications. Consequently, the architectural focus shifts from an “everything to the cloud” to near-source computing combined with cloud-edge collaboration. This paper reviews the stage-specific advantages and limitations of centralization. It characterizes edge computing as a near-data layer situated between endpoints and the cloud that uses local processing and closed-loop control to satisfy deterministic latency and resilience. From this perspective, the paper outlines a sensor-cloud-edge-device collaborative framework. This framework adopts upload-when-necessary data paths, Service Level Agreement (SLA)-aware task placement with two-tier scheduling, and a division of labor in which the edge closes loops instantly while the cloud performs policy and model governance. The paper then discusses the trajectory toward edge intelligence, including lightweight and on-device learning; federated learning and knowledge distillation; AIOps for Edge with multilevel degradation; and an evaluation regime oriented to end-to-end closed-loop efficiency, resilience, and auditability. Evidence from educational scenarios and current industrial practices demonstrate the practical effectiveness of near-source computing and cloud-edge collaboration in ensuring deterministic latency, enhancing overall system resilience, and achieving cross-domain consistency, and accordingly identify the inevitable evolution of the computing paradigm from edge computing toward cloud-edge intelligent collaboration.

【Key words】 cloud-edge collaboration; sensor cloud; edge intelligence; near-source computing; small-large model collaboration

0 引言

在信息技术的演化史中, 架构之争从非单一技术选项之优劣, 而是集中与分布、即时响应与整体最优

之间的长期均衡问题^[1]。云计算与边缘计算正构成这一张力的最新表征: 早期“万物上云”的范式, 依托集中化运维与算力规模效应释放了显著的性价比红利; 伴随物联网与智能终端的指数级增长, 时延-带宽-

基金项目: 国家自然科学基金联合重点基金(U25A20436); 广西重点研发计划(FN2504240036, 2025FN96441087); 国家自然科学基金(62372047); 广东省科技计划(2025B0101120006); 广东省自然科学基金(2024A1515011323)。

作者简介: 王田, 男, 教授、博士, 主研方向为边缘智能、物联网、移动计算; 李雨婷, 博士研究生; 王文华, 讲师、博士。

收稿日期: 2025-11-14 **修回日期:** 2025-12-30 **E-mail:** tianwang@bnu.edu.cn

能耗-隐私^[2]等约束叠加,使单一路径边际收益递减,由此,体系演化进入以就地计算承载近源闭环、以云边协同^[3]实现跨域一致与整体最优的分布式新阶段。

立足于物联网、边缘计算与人工智能(AI)的长期一线教研实践,本团队不仅见证了从“感知上云^[4]”到“智能在边^[5]”这一范式的技术转变,更以系统性研究与工程化落地推动其演进,并在此基础上出版了专著《传感云与边缘计算》^[6],对相关理论与工程实践进行了系统梳理与提炼。一方面,本团队在多场景中完成了可复用的方法论沉淀与工程实现;另一方面,在模型与治理层面推动了从单点最优到系统最优的认识升级。据此,本文围绕“为何需要协同、协同如何实现、如何落地验证”三类问题展开论证,形成从原理到方法再到实践的完整链条。

本文主要贡献在于:

1)从“传感云-边缘-端”一体化视角,对集中式计算的阶段性优势与边界进行系统梳理,提出以“就地计算+云边协同”作为面向大规模感知与智能应用的整体架构思路。

2)结合资源调度、边缘智能和大小模型协同等研究进展,归纳边缘智能的关键发展路径,形成从架构到运维的一套系统性问题图谱。

3)基于教育等典型应用场景和产业政策实践,凝练出以端到端闭环效率、韧性与可追责为导向的评价视角,为后续算法和系统研究提供需求牵引。

1 云起有序:集中式的黄金十年

回溯十余年前,企业数字化转型的核心路径是“上云”^[7]。彼时,受限于高额资本性支出/运营支出、复杂的硬件生命周期管理以及 7×24 h 运维压力,云计算以弹性扩容、按量计费、多租户与托管运维的能力体系^[6],带来了前所未有的效率革命。企业无需额外承担峰谷不均带来的资源闲置或短缺,只需通过应用程序接口(API)调用即可获得近似“无限”的计算与存储资源,从而实现以集中化为特征的规模化供给与效率提升。

在此阶段,云逐渐成为企业的信息中枢,“数据上云”几乎成为默认范式。生产线传感器、城市的摄像头与车载定位设备所产生的数据持续汇聚至云侧^[8],通过集中分析与统一决策实现标准化、可复制的业务能力。然而,集中化的另一面是距离。当所有智能都被推向远方,云的优势开始转化为代价——带宽、延迟与安全的压力悄然累积。随着设备数量的指数级增长,集中式路径带来的瓶颈逐渐显现。过去十年是云计算快速发展的阶段,也是下

一场计算范式变革的前夜。

2 云之所限:网络、实时、合规与“调度难题”

在实际应用中:家用智能音箱在网络波动时可能因无法访问云端服务而出现短暂失效;若自动驾驶车辆完全依赖远端数据中心决策,则难以规避链路时延带来的安全风险;工厂流水线中的机械臂也可能因为云端响应延迟而产生节奏抖动,影响生产稳定性。这些现象并非个别案例,而是集中式计算将智能和响应重心过度推向远端的结果。

问题主要有三点,首先是网络与成本:当数以亿计的终端同时上传视频流与传感数据时,带宽与出口费用被拉到极限。随之而来的是实时性的坍塌:工业机械动作、无人机避障、安防判定均难以承受“回云再判”的毫秒迟滞^[9]。更深层的裂缝在于合规与调度:一旦数据跨域流动,隐私监管的“红线”随之而至。在性能-成本-安全的联立约束下,集中式调度的可行域不断收缩,其适用性与可扩展性呈现结构性受限。

3 边到其位:就地计算、就地控制

当集中式计算的边界显现时,边缘计算应势而生。在架构层面,边缘计算是一种将数据处理前移至数据源近侧的计算范式^[10],作为端侧设备与云计算之间的中间层,依托近源处理与局部闭环满足对时延、带宽与韧性的确定性要求^[11]。

若将云比作“中枢”,边缘则相当于遍布全身的“神经末梢”,赋予各节点就地感知与快速响应的能力。以智能制造为例,机械臂、传感器与工业相机需在毫秒级时隙内完成检测与决策,以避免长链路回传引发的尾部时延。“识别-判断-控制”进而在边侧就地闭环^[12],仅保留必要的摘要/事件用于云侧的知识汇聚与策略治理以及后续的治理与复盘。

从宏观层面来看,边缘计算正日益成为智能社会的基础支撑。从国务院《“十四五”数字经济发展规划》中提出“加强面向特定场景的边缘计算能力、强化算力统筹和智能调度”^[13]到 2023 年国家发展改革委、国家数据局等部门联合印发《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》再到 2025 年完善东西部算力协同调度机制、构建多元算力一体化高效调度体系^[14],政策层面已经从“提出边缘能力建设”进一步走向“构建全国一体化算力网与协同调度体系”。在产业技术层面,边缘 AI 平台与硬件栈快速演进:例如,NVIDIA 发布基于 Blackwell 架构的 Jetson AGX Thor^[15],将百瓦级功耗下的千级 TFLOPS 量级推理能力带

到边侧,用于机器人与实时视觉等场景,并在开发套件与能效指标上针对“在边侧运行大模型/生成式模型”的需求进行优化;国内生态亦在加速完善,华为于 2025 年公布“昇腾”系列中长期路线图,规划自研高带宽内存(HBM)模块与“超节点”平台,以提升边缘与云侧算力的系统集成能力。

在应用层面,边缘计算已在手机、个人电脑、智能驾驶、智能穿戴等领域率先落地,同时中国市场规模亦在快速增长,2022 年已突破 430 亿元。例如,京东方科技集团在其 14 条半导体显示生产线准确率达到 96%,人工替代率为 70%,生产准备时间缩短至 1 h;百度 Apollo 平台在路侧设备部署边缘节点,支持多车道交叉口的实时数据处理与本地决策,实现交通流量分析与智能信号控制,通行效率显著提升。

总体而言,边缘计算正由“硬件+算法”走向“芯片/系统+平台+行业方案”的系统化演进路径。其核心价值在于:以就地计算、就地控制消解集中式路径在网络与实时性上的约束,并通过近源闭环提供确定性响应。

4 云边协同:感知服务化与跨域治理框架

当集中式计算走向边界时,云的角色也由“集中处理”转向跨域治理与协同编排。在万物互联场景中,数据不再是被动上行的原料,而是可被编排、调度与复用的要素。“传感云”的概念由此出现——它并非单纯的数据汇聚中心,而是将感知能力服务化、治理化的系统层。

在这一体系中,数据的上行更加克制。传感节点通过语义索引与时空订阅机制仅在满足触发条件(异常、告警、策略请求)时上报摘要/事件;时延敏感任务在本地闭环,算力密集子任务在资源与合规约束下分摊至域边或云端。这一模式的关键不在算力大小,而在分配策略:边缘节点以队列/优先级与服务级别协议(SLA)感知的策略进行动态调度^[16],综合平衡时延、带宽与能耗,实现“随需而动”的执行选择。

在此基础上,控制回路相应重构为“边侧即时闭环,云侧策略治理”。近年来,本团队的研究与实践也充分表明,高效体系并非“云”与“边”的对立,而在于两者的协同。在工程实践中,这种协同多以“双层调度架构”实现:边侧以事件驱动队列与优先级机制保证关键路径时延,云侧以策略学习与模型治理支撑跨域一致;在治理层面,对应“去中心化执行、中心化治理”的目标形态——边缘在动、云在管;边缘执行、云定规矩。

为了更直观地展示本文所讨论的“传感云-边缘-端”一体化协同思路,图 1 给出了整体系统架构示意图:底层为各类感知设备与终端,承担数据采集与人机交互;中间层为边缘节点,通过就地计算与局部闭环满足实时性与韧性要求,部署轻量化小模型并作为联邦学习/知识蒸馏协同的本地参与端;上层为传感云与云中心,负责跨域数据与模型治理以及联邦/蒸馏协同的全局聚合与策略学习,同时通过 SLA 感知的双层调度与 AIOps 机制,将“必要即上行”的数据、策略与模型更新在端、边、云之间闭环流转。

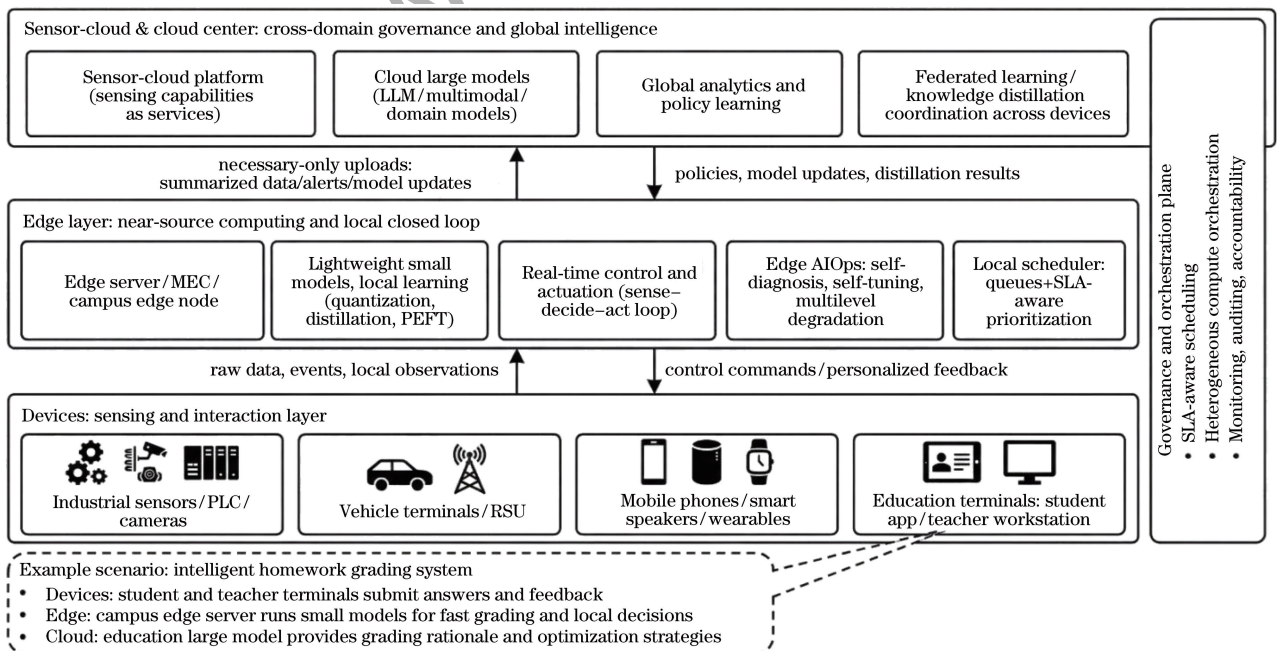


图 1 “传感云-边缘-端”一体化协同架构

Fig. 1 Integrated sensor-cloud-edge-device architecture for collaboration

上述协同理念正在产业侧加速落地。例如,腾讯云边协同智能制造平台在边缘层实现工厂车间的实时监控与数据采集,在云层进行 AI 训练与全局优化,通过“边缘执行、云端训练”的分工机制,显著降低了产品不良率并提升设备利用率。在政策层面,国家正自上而下推进“云-边-端”一体化布局,同时加快建设“边缘+云”一体化的国家计算平台,以提升区域算力服务能力与终端生态承载能力。《算力基础设施高质量发展行动计划》提出构建通用、智能与边缘算力协同的供给体系^[17],目标到 2025 年全国算力规模超过 300 EFLOPS、智能算力占比达 35%,并在 2027 年前实现百万级边缘节点规模化部署,政策焦点由“扩算力”转向“优结构、强协同”。

当技术演进至平台层时,竞争的焦点也在转移。云厂商依然掌握资源与算法的主导权,但新的增长点正在从中心向边缘蔓延。在智能制造、城市安防、车路协同等领域,设备商与行业方案商以“小模型+本地部署”的方式^[18],构建更具韧性、黏性的场景生态,形成近源闭环与云侧治理的良性循环。

5 大势所趋:从边缘计算到边缘智能

当基础设施逐步成熟,边缘计算的价值正由“算力供给”延伸至“模型与策略的持续学习”^[19]。系统不再局限于被动执行的终端,而是能够自主学习、持续优化与演化的智能体。这也是本团队近年来的重点方向。

这一演进的核心在于轻量化与本地学习。通过模型量化、蒸馏以及参数高效微调(PEFT)^[20]等方法,边缘侧的小模型在受限算力条件下实现个性化适配与渐进式更新。结合联邦学习^[21]与知识蒸馏^[22]的协同框架,设备无需上传原始数据,仅共享模型更新或中间表示,即可在隐私受控的前提下完成跨终端协同训练,从而使智能由“集中训练”走向“分布协同”,在端、边、云之间形成动态演化的学习链路。

进一步地,边缘智能运维(AIOps for Edge)^[23]使边缘节点具备自诊断、自调优与自愈能力。面对网络波动、突发负载或能耗约束,系统可依据任务优先级与服务级别目标自动评估与重排,动态调整算力与带宽分配,按需迁移或降级,从而实现自治运维与弹性调度。此类智能边缘体系减少对中心化调度的刚性依赖,更倾向于以分布式智能体^[24]的方式进行自组织与协同。

相应地,评价体系也需重构^[25]。系统性能不应仅以准确率或吞吐量衡量,而应关注端到端闭环效

率与韧性:例如异常发现率与虚警率、处置与闭环完成时延、在不同工况下的闭环达成率、单瓦推理效率以及在资源受限情形下的服务稳定性与可恢复性等。这些指标反映的并非单一算法的优劣,而是系统在长期运行中的可持续智能化能力。

总体来看,边缘计算正迈向“边缘智能”的新阶段。问题的焦点不再是“功能部署在边还是在云”,而在于如何使端敢于决策、边善于协同、云长于升维。智能的重心前移,能力仍源于云边端三者的有效协同之中。

6 思想落地:从“就地计算”到“协同智能”的行动实践

理念的生命力在于可验证的实践。作为“云边协同”、“大小模型共生”的长期践行者,本团队持续承办中国计算机大会(CNCC)“大小模型协同智能计算”专题论坛,构建跨学科、跨行业的交流平台。

在 CNCC 2025 上,笔者所在团队牵头组织了相关主题论坛,由笔者担任论坛主席。论坛围绕“大小模型如何协同”、“边缘计算与智能设备的融合”、“降低智能计算的成本与时延”等议题展开研讨,多位受邀嘉宾从端-边-云协同、边缘智能、联邦学习与类脑计算等方向分享最新进展,形成了理论与工程的多维对话。

本团队也分享了基于教育场景的实践:以智能作业批改系统为代表,围绕“边侧小模型+云侧大模型”的协同机制开展了持续性试点。传统作业批改高度依赖教师逐题处理,重复性强、反馈滞后;在开放性/主观题场景中,又常因缺乏标准答案而使自动批改难以直接落地。为此,本团队将系统目标明确为:在保证判分一致性与可解释性的前提下,将流程重构为“机器自动批改+教师抽样复核”,以显著降低教师工作量并缩短处理时延。

具体而言,本团队采用端-云协同架构:云侧大模型(如 72B 级)负责跨模态语义理解与全局推理,并面向主观题引入“伪答案预生成”机制——从题干图文信息中生成参考答案与评分依据(rubric/评分点),再将其结构化为“批改提示词/批改依据”下发至端侧;边侧小模型(如 7B 级)聚焦特定题型或知识点的细粒度判别与执行式打分,仅依据云端下发的结构化依据完成快速批改,从而兼顾速度、成本与可部署性。

为保证输入稳定、提升批改的可落地性,边侧专用小模型首先完成试卷预处理闭环,包括试卷角度校正、基准试卷特征对齐、作答区域热力区识别、学

生作答 OCR,以及连贯题组的语义检测与重归类。该预处理链路可在端侧完成一份试卷的全流程处理,耗时低于 0.2 s,为后续批改提供稳定、可复用的结构化输入。

在对比评测中,若由云端大模型端到端独立完成批改,单题推理通常需 30~60 s;采用“云端生成批改依据+边侧快速执行”的协同方案后,单份作业平均处理时延可由 30~60 s 量级降至秒级(约 1~2 s),在保持判分一致性的前提下实现约 200 倍的效率提升。系统批改准确率和一致性可稳定达到 92%,针对数学等结构化题型的实测准确率可达 96%。配合教师侧流程改造,整体实现从“逐题全量批改”向“抽样复核”的转变,显著减少教师重复性劳动。总体而言,该实践在不牺牲一致性与可解释性的前提下有效降低了平均处理时延与人力负担,为教育场景验证“就地计算+云边协同”提供了可复用的范式基础。

这一系列实践使“就地计算、云边协同”的理念从技术架构走向产业落地。“就地计算”体现为端侧小模型的即时响应与自适应优化;“协同智能”依托云端大模型的知识整合与策略治理得以实现。两者形成互补共生的“智能循环”,使系统既能在本地高效执行,又能在全局持续演化。

7 结语:云边协同范式的收束与展望

技术演进并非线性前行,而是呈现出阶段性的重心转换:云计算阶段表明,集中化与标准化能够带来可观的规模红利;当前的边缘智能实践则显示,唯有就地执行与可验证治理,才能支撑关键场景的韧性。真正的未来,不在“云”与“边”的对立,而在两者的共生。以传感云为底座,统一数据与服务的秩序;以边缘智能为引擎,赋予系统自主判断与演化能力;再以云端治理为收束,确保整体安全与可追责。

未来,本团队也将以此为契机,继续深化云-边-端协同智能体系的研究,探索更高效、更安全、更具可解释性的智能计算范式,让理念真正转化为推动产业智能化进程的实践力量。

参考文献

- [1] 张晓丽, 杨家海, 孙晓晴, 等. 分布式云的研究进展综述[J]. 软件学报, 2018, 29(7): 2116-2132.
- [2] ZHANG X L, YANG J H, SUN X Q, et al. Survey of geo-distributed cloud research progress[J]. Journal of Software, 2018, 29(7): 2116-2132. (in Chinese)
- [3] NIU C Y, DING Y C, LU J H, et al. Collaborative learning of on-device small model and cloud-based large model: advances and future directions [EB/OL]. [2025-10-25]. <https://arxiv.org/abs/2504.15300>.
- [4] 曾建电, 王田, 贾维嘉, 等. 传感云研究综述[J]. 计算机研究与发展, 2017, 54(5): 925-939.
- [5] ZENG J D, WANG T, JIA W J, et al. A survey on sensor-cloud[J]. Journal of Computer Research and Development, 2017, 54(5): 925-939. (in Chinese)
- [6] 梁玉珠, 梅雅欣, 杨毅, 等. 一种基于边缘计算的传感云低耦合方法[J]. 计算机研究与发展, 2020(3): 639-648.
- [7] LIANG Y Z, MEI Y X, YANG Y, et al. A low-coupling method in sensor-cloud systems based on edge computing[J]. Journal of Computer Research and Development, 2020(3): 639-648. (in Chinese)
- [8] 王田, 王文华, 徐旸. 传感云与边缘计算[M]. 北京: 人民邮电出版社, 2025.
- [9] WANG T, WANG W H, XU Y. Sensor-cloud and edge computing[M]. Beijing: People's Posts and Telecommunications Press, 2025. (in Chinese)
- [10] WANG Z Y, GOUDARZI M, BUYYA R. TF-DDRL: a Transformer-enhanced distributed DRL technique for scheduling IoT applications in edge and cloud computing environments [J]. IEEE Transactions on Services Computing, 2025, 18(2): 1039-1053.
- [11] 王思轩. 数字化转型架构[M]. 北京: 电子工业出版社, 2021.
- [12] WANG S X. Digital transformation architecture: methodology and cloud-native practice [M]. Beijing: Publishing House of Electronics Industry, 2021. (in Chinese)
- [13] LIU Z Y, CHEN X, WU H, et al. Integrated sensing and edge AI: realizing intelligent perception in 6G [J]. IEEE Communications Surveys & Tutorials, 2025, 28: 2725-2770.
- [14] LUO Q Y, HU S H, LI C L, et al. Resource scheduling in edge computing: a survey [J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2131-2165.
- [15] HUANG F Y, WANG W H, LIU Q, et al. DRMQ: dynamic resource management for enhanced QoS in collaborative edge-edge industrial environments [J]. IEEE Transactions on Services Computing, 2025, 18(2): 743-757.
- [16] 唐朝刚, 李召, 肖硕, 等. 一种面向车载边缘计算基于服务缓存的任务协同卸载算法[J]. 计算机学报, 2025, 48(4): 864-876.
- [17] TANG C G, LI Z, XIAO S, et al. Service caching based collaborative task offloading algorithm in vehicular edge computing[J]. Chinese Journal of Computers, 2025, 48(4): 864-876. (in Chinese)
- [18] 国务院. 关于印发“十四五”数字经济发展规划的通知(国发〔2021〕29号)[J]. 中国军转民, 2022(1): 6-12.
- [19] State Council. Notice on issuing the “14th Five-Year” plan for digital economy development (state council document No. 29 [2021]) [J]. China Defence Industry Conversion, 2022(1): 6-12. (in Chinese)
- [20] 国家发展改革委, 国家数据局, 中央网信办, 工业和信息化部, 国家能源局. 关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见(发改数据〔2023〕1779号)[EB/OL]. [2025-10-25]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202312/t20231229_1363000.html.
- [21] National Development and Reform Commission, National Data Administration, Cyberspace Administration of China, Ministry of Industry and Information Technology, National Energy Administration. Implementation opinions on further deepening the “East Data, West Computing” project to accelerate the construction of a national integrated computing power network (development and reform data [2023] No. 1779) [EB/OL]. [2025-10-25]. <https://www.ndrc.gov.cn/>

- xxgk/zcfb/tz/202312/t20231229_1363000.html. (in Chinese)
- [15] SALI S, MERIBOUT A, MAJEED A, et al. Real-time object detection and associated hardware accelerators targeting autonomous vehicles: a review[EB/OL]. [2025-10-12]. <https://arxiv.org/abs/2509.04173>.
- [16] KAREEM AWAD W, ZAINOL ARIFFIN K A, AHMAD NAZRI M Z, et al. Resource allocation strategies and task scheduling algorithms for cloud computing: a systematic literature review[J]. *Journal of Intelligent Systems*, 2025, 34: 20240441.
- [17] 工业和信息化部, 中央网络安全和信息化委员会办公室, 教育部, 国家卫生健康委员会, 中国人民银行, 国务院国有资产监督管理委员会. 算力基础设施高质量发展行动计划(工信部联通信〔2023〕180号)[EB/OL]. [2025-10-25]. https://www.gov.cn/zhengce/zhengceku/202310/content_6907900.htm.
Ministry of Industry and Information Technology, Office of the Central Cyberspace Administration of China, Ministry of Education, National Health Commission, People's Bank of China, State-owned Assets Supervision and Administration Commission of the State Council. Action plan for the high-quality development of computing power infrastructure(joint MIIT communication document〔2023〕No. 180)[EB/OL]. [2025-10-25]. https://www.gov.cn/zhengce/zhengceku/202310/content_6907900.htm. (in Chinese)
- [18] 鲁飞鸿, 罗杨一飞, 高士洪, 等. 基于不确定性校准的云边协同推理框架[J]. *计算机学报*, 2025, 48(10): 2487-2507.
LU F H, LUO Y Y F, GAO S Q, et al. An uncertainty-calibrated cloud-edge collaborative inference framework[J]. *Chinese Journal of Computers*, 2025, 48(10): 2487-2507. (in Chinese)
- [19] WANG Z, WU F, YU F, et al. Federated continual learning for edge-AI: a comprehensive survey[EB/OL]. [2025-10-25]. <https://arxiv.org/abs/2411.13740>.
- [20] WANG L P, CHEN S, JIANG L N, et al. Parameter-efficient fine-tuning in large language models: a survey of methodologies[J]. *Artificial Intelligence Review*, 2025, 58(8): 227.
- [21] LI Y T, WANG W H, WANG T. pFSSL-D: generalization meets personalization in dual-phase federated semi-supervised learning[C]//*Proceedings of the IEEE 41st International Conference on Data Engineering (ICDE)*. Washington D.C., USA: IEEE Press, 2025: 2188-2200.
- [22] YANG C P, ZHU Y, LU W, et al. Survey on knowledge distillation for large language models: methods, evaluation, and application[J]. *ACM Transactions on Intelligent Systems and Technology*, 2025, 16(6): 1-27.
- [23] BECKER S, SCHMIDT F, GULENKO A, et al. Towards AIOps in edge computing environments[C]//*Proceedings of the IEEE International Conference on Big Data (Big Data)*. Washington D.C., USA: IEEE Press, 2020: 3470-3475.
- [24] AHMED I, SYED M A, MAARUF M, et al. Distributed computing in multi-agent systems: a survey of decentralized machine learning approaches[J]. *Computing*, 2024, 107(1): 2.
- [25] SUSNJAK T, MCINTOSH T R, BARCZAK A L C, et al. Over the edge of chaos? Excess complexity as a roadblock to artificial general intelligence[J]. *IEEE Transactions on Cybernetics*, 2026, 56(1): 29-41.

文字编辑 陆燕菲
栏目编辑 宋圆