

边缘计算与大数据融合的实时处理机制及优化研究

陈健军

深圳市非同凡响电子科技有限公司 广东深圳 518100

摘要: 随着物联网、5G等技术的快速发展,大数据呈现爆发式增长态势,传统云计算架构在实时数据处理场景中面临延迟过高、带宽消耗大等瓶颈。边缘计算凭借其靠近数据源头的分布式架构优势,为大数据实时处理提供了新的解决方案。本文深入研究边缘计算与大数据融合的实时处理机制,首先分析两者融合的技术互补性与核心挑战;其次设计“数据分层处理-任务动态调度-云端协同优化”三级实时处理机制,重点解决数据传输效率与计算资源分配问题;最后从资源调度、数据预处理、安全防护三个维度提出优化策略,并通过实验验证机制的有效性。实验结果表明,所提机制在处理延迟、吞吐量及资源利用率方面均优于传统云计算处理模式,为工业互联网、智能交通等实时性要求较高的场景提供技术支撑。

关键词: 边缘计算; 大数据; 实时处理机制

一、引言

(一) 研究背景

据IDC预测,到2025年全球数据圈规模将达到175ZB,其中80%以上的数据来自边缘设备。这些多源异构数据具有“4V”特征,尤其是在工业控制、智能驾驶等场景中,数据生成的高时效性要求处理延迟控制在毫秒级。传统云计算采用“集中式处理”模式,数据需传输至远端云中心进行处理,导致平均延迟通常超过100ms,无法满足实时性需求^[1]。同时,海量边缘数据传输至云端还会造成核心网络带宽拥堵,据统计,工业物联网场景中无效数据传输占比高达60%以上,严重降低了数据处理效率。

边缘计算作为一种新型计算范式,将计算、存储资源部署于网络边缘节点,实现数据“就近处理”,可使处理延迟降低至10ms以内。边缘计算与大数据的融合,能够充分发挥边缘计算的低延迟优势与大数据的价值挖掘能力,成为突破实时处理瓶颈的关键路径^[2]。然而,两

者融合过程中面临诸多挑战:边缘节点资源受限导致计算能力不足、多源数据异构性增加处理复杂度、边缘与云端协同机制不完善等。因此,研究两者融合的实时处理机制及优化策略具有重要的理论意义与实际应用价值。

(二) 研究现状

国内外学者已针对边缘计算与大数据融合展开研究。国内学者在《中国宽带》中提出移动边缘网络资源管理框架,通过基站侧调度缓解带宽压力,但未优化数据预处理导致存储浪费;部分研究验证了轻量级算法的边缘适配性,但未解决传输与调度协同问题^[3]。现有研究在实时处理机制系统性、动态调度及端云协同方面仍存不足,本文据此构建优化方案。

现有研究虽验证了融合的可行性,但在实时处理机制的系统性设计、动态资源调度及端云协同优化方面仍存在不足。本文针对上述问题,构建高效的实时处理机制并提出优化策略,提升融合系统的整体性能。

(三) 研究内容与技术路线

本文研究内容包括三部分:一是边缘计算与大数据融合的技术互补性分析,明确融合过程中的核心瓶颈;二是设计三级实时处理机制,构建数据处理、任务调度与云端协同的全流程架构;三是提出资源、数据、安全三维优化策略,并通过实验验证效果。

技术路线遵循“问题分析-机制设计-优化提升-实验验证”的逻辑展开:首先通过文献研究与场景调研明确核心问题;其次基于边缘计算分层架构设计处理机制;

作者简介: 陈健军(1984.12),男,汉族,户籍:湖北省利川市,深圳市非同凡响电子科技有限公司技术部数据分析师,深耕数据分析领域多年。主要研究方向为大数据处理、边缘数据实时分析及多源异构数据整合,聚焦边缘计算与大数据融合场景下的数据分析技术落地应用,为企业数据处理及业务决策提供技术支持。现任职于广东深圳。

然后结合智能算法与加密技术提出优化策略；最后搭建仿真实验平台，对比验证所提方案的优越性。

二、边缘计算与大数据融合的技术基础及核心挑战

（一）核心技术基础

边缘计算采用“云-边-端”三级架构，边缘节点作为核心枢纽，具备数据采集、本地计算、短期存储及与云端交互的能力，其核心技术包括边缘节点虚拟化、分布式任务调度等。大数据技术以Hadoop、Spark等框架为支撑，实现数据的分布式存储与并行计算，核心技术涵盖数据清洗、特征提取、流计算等。两者融合的技术互补性主要体现在：边缘计算为大数据提供“就近处理”的硬件基础，解决实时性问题；大数据技术为边缘计算提供智能分析能力，提升边缘节点数据处理精度。

（二）融合过程中的核心挑战

通过场景调研与技术分析，明确两者融合的三大核心挑战：

数据传输效率低下：边缘设备生成的多源数据格式异构，未经过滤直接传输导致带宽占用率高。如智能交通场景中，单个摄像头每小时产生约10GB视频数据，直接传输会导致网络拥堵。

边缘资源调度失衡：边缘节点计算、存储资源有限，且数据生成具有突发性，传统静态调度策略易导致部分节点过载、部分节点闲置。工业场景中，设备故障突发时会导致局部边缘节点负载激增，处理延迟增加50%以上。

安全与隐私风险加剧：边缘节点分布广泛且环境开放，数据在采集、传输过程中易遭受攻击。大数据的高敏感性使隐私泄露风险倍增，如医疗场景中的患者生理数据在边缘处理时面临窃取风险。

三、边缘计算与大数据融合的实时处理机制设计

针对上述挑战，本文设计“数据分层处理-任务动态调度-云端协同优化”三级实时处理机制，核心流程包括数据接入、边缘处理、云端协同三个阶段。

（一）数据分层处理模块

采用“边缘预处理-本地计算-云端存储”分层模式提升效率。通过MQTT等轻量级协议接入多源数据，预处理环节对结构化数据均值滤波去噪，对非结构化数据提取关键特征，数据压缩率超80%。本地完成实时流计算后，仅上传处理结果与核心数据至云端，大幅降低带宽占用。

（二）任务动态调度模块

设计基于改进遗传算法的任务调度策略，实现边缘节点资源的动态分配。首先建立任务优先级评估模型，

根据任务截止时间、数据规模等指标将任务分为高（如工业控制指令）、中（如环境监测数据）、低（如日志数据）三个优先级。然后构建资源负载评估矩阵，实时采集边缘节点的CPU利用率、内存占用率、带宽使用率等指标。最后通过改进遗传算法求解最优调度方案：采用适应度函数（综合任务完成率与资源利用率）替代传统算法的单一指标，通过交叉变异操作优化任务分配结果，使边缘节点负载均衡度提升30%以上。

（三）云端协同优化模块

构建“边缘自主决策+云端全局优化”的协同机制。边缘节点具备本地决策能力，对于高优先级任务可直接处理并反馈结果；对于复杂任务（如大规模数据建模），采用“边缘预处理+云端训练”的模式：边缘节点完成数据清洗与特征提取后，将标准化数据上传至云端，云端基于Hadoop集群进行分布式建模，模型训练完成后下发至边缘节点用于本地推理。同时，云端实时监测边缘节点运行状态，当边缘节点负载超过阈值时，自动调度空闲节点分担任务，提升系统稳定性。

四、边缘计算与大数据融合的优化策略

基于上述处理机制，从资源调度、数据处理、安全防护三个维度提出优化策略，进一步提升融合系统的性能与安全性。

（一）资源调度优化：基于负载预测的动态分配算法

针对边缘节点负载失衡问题，提出基于LSTM（长短期记忆网络）的负载预测与动态调度优化策略。首先通过LSTM模型对边缘节点未来10分钟的负载情况（CPU、内存利用率）进行预测，预测精度可达92%以上；然后基于预测结果采用“过载预警-任务迁移-资源扩容”三级响应机制：当预测负载超过80%阈值时，提前将低优先级任务迁移至空闲节点；当多个节点同时过载时，触发云端临时资源扩容，确保高优先级任务的实时处理。该策略相比传统静态调度，任务完成率提升25%，平均处理延迟降低18%。

（二）数据处理优化：轻量级预处理与缓存策略

在数据预处理环节，提出“异构数据归一化+自适应过滤”优化方案。设计通用数据解析接口，支持不同格式数据的归一化处理；基于信息增益算法自适应筛选关键数据，对于低价值数据（如无变化的环境监测数据）直接过滤，进一步提升数据压缩率。在缓存策略方面，采用“热点数据本地缓存+冷数据云端迁移”的分层缓存机制：基于LRU（最近最少使用）算法识别边缘

节点的热点数据（如工业生产中的关键工艺参数），缓存至本地闪存；冷数据定期迁移至云端分布式存储，释放边缘节点存储资源。实验表明，该优化使边缘节点存储利用率提升40%，数据传输带宽占用降低50%。

（三）安全防护优化：全流程隐私保护机制

构建“采集-传输-处理”全流程安全防护体系。采集阶段用差分隐私技术扰动敏感数据；传输阶段通过TLS 1.3加密通道并验证节点身份；处理阶段采用联邦学习，边缘节点仅上传模型参数，避免原始数据泄露，显著提升数据安全性。

五、实验验证与结果分析

（一）实验环境搭建

采用NS-3搭建“云-边-端”实验平台模拟工业物联网场景。端设备：100个传感器（采样频率10Hz）、20个摄像头（720P）；边缘节点：10个服务器（16GB内存、512GB SSD）；云端：Hadoop集群（3主10从节点）。对比对象：传统云计算（CK1）、未优化融合模式（CK2）、本文优化模式（EXP）。

（二）实验指标与方案设计

实验选取三个核心指标：（1）处理延迟：从数据生成到结果输出的总时间；（2）系统吞吐量：单位时间内处理的数据量；（3）资源利用率：边缘节点的CPU平均利用率。实验设计三个场景：场景1（低负载）：50个传感器+10个摄像头运行；场景2（中负载）：80个传感器+15个摄像头运行；场景3（高负载）：100个传感器+20个摄像头运行。每个场景运行1小时，采集指标数据并取平均值。

（三）实验结果分析

处理延迟对比结果显示：在低、中、高负载场景下，EXP模式的平均延迟分别为8.2ms、12.5ms、18.3ms，相比CK1（45.6ms、68.2ms、92.5ms）分别降低82.0%、81.7%、80.2%，相比CK2（15.6ms、23.8ms、35.6ms）分别降低47.4%、47.5%、48.6%。这表明本文设计的分层处理机制与负载预测调度策略有效降低了处理延迟，即使在高负载场景下仍能满足实时性需求。

系统吞吐量对比结果显示：EXP模式在三个场景下的吞吐量分别为128Mbps、195Mbps、242Mbps，相比CK1（62Mbps、85Mbps、102Mbps）分别提升106.5%、129.4%、137.3%，相比CK2（95Mbps、142Mbps、

185Mbps）分别提升34.7%、37.3%、30.8%。原因在于数据预处理优化降低了传输数据量，提升了单位时间内的数据处理能力。

资源利用率对比结果显示：EXP模式的边缘节点CPU平均利用率为68.5%，相比CK1（32.2%）提升112.7%，相比CK2（45.8%）提升49.6%，且各节点负载标准差为5.2%，远低于CK2的18.6%，表明动态调度优化实现了资源的均衡利用。

六、结论与展望

（一）研究结论

本文深入研究边缘计算与大数据融合的实时处理问题，主要成果包括：（1）明确两者融合的技术互补性及“传输效率、资源调度、安全隐私”三大核心挑战；（2）设计三级实时处理机制，通过数据分层处理减少传输压力，基于改进遗传算法实现任务动态调度，构建云端协同优化架构；（3）提出三维优化策略，负载预测调度提升资源利用率，轻量级预处理提升传输效率，全流程防护保障数据安全。实验验证表明，所提方案在处理延迟、吞吐量及资源利用率方面均优于传统模式，为实时大数据处理提供有效解决方案。

（二）研究不足与展望

本文研究仍存在不足：一是边缘节点的能源消耗优化未涉及，在电池供电的边缘设备场景中适用性有限；二是负载预测模型未考虑极端天气、设备故障等突发因素，预测精度有待进一步提升。

未来研究方向：（1）结合强化学习算法设计能耗优化策略，在保证实时性的前提下降低边缘节点能耗；（2）引入注意力机制改进LSTM预测模型，提升突发场景下的负载预测精度；（3）开展实际工业场景的试点应用，验证方案的工程实用性。

参考文献

- [1] 王健, 刘芳, 张伟. 工业4.0场景下边缘计算实时处理优化方法研究[J]. 计算机学报, 2023, 46(5): 1023-1038.
- [2] 李娜, 王明, 赵刚. IoT边缘网络中轻量级算法优化与实时处理应用[J]. 电子学报, 2024, 52(2): 389-401.
- [3] 陈志, 林达, 黄勇. 大数据环境下边缘网络资源调度与隐私保护机制[J]. 通信学报, 2023, 44(8): 156-170.