

云计算模型的物联网边缘数据挖掘面临的挑战与解决方案

赵玉林

赣州职业技术学院, 信息工程学院, 江西 赣州 341000

摘要: 随着物联网技术的飞速发展, 产生了海量的边缘数据。如何高效地挖掘这些数据的价值成为了关键问题。本文深入研究基于云计算模型的物联网边缘数据挖掘技术, 探讨了云存储技术、OpenStack 私有云搭建与管理、容器云技术、公有云技术以及 Linux 网络操作系统等在边缘数据挖掘中的应用, 分析了其优势与面临的挑战, 并提出了相应的解决方案, 旨在提高物联网边缘数据挖掘的效率和质量, 推动物联网应用的进一步发展。

关键词: 云计算模型; 物联网; 边缘; 数据挖掘技术; 研究

Challenges and Solutions of Edge Data Mining of Internet of Things based on Cloud Computing Model

Zhao, Yulin

Ganzhou polytechnic, School of Information Engineering, Ganzhou, Jiangxi, 341000, China

Abstract: With the rapid development of Internet of Things technology, a huge amount of marginal data has been produced. How to efficiently mine the value of these data has become a key issue. This paper deeply studies the edge data mining technology of Internet of Things based on cloud computing model, discusses the application of cloud storage technology, OpenStack private cloud construction and management, container cloud technology, public cloud technology and Linux network operating system in edge data mining, analyzes its advantages and challenges, and puts forward corresponding solutions, aiming at improving the efficiency and quality of edge data mining of Internet of Things and promoting the further development of Internet of Things applications.

Keywords: Cloud computing model; Internet of things; Edge; Data mining technology; Study

DOI: 10.62639/sspis03.20250203

引言

物联网 (IoT) 的广泛应用使得各种设备能够实时产生大量的数据, 这些数据大多在网络边缘产生和处理。边缘数据具有实时性、低延迟性和海量性等特点^[1]。云计算模型为物联网边缘数据挖掘提供了强大的计算和存储能力, 通过将云计算技术与物联网边缘计算相结合, 可以更好地实现对边缘数据的价值挖掘, 从而为各行业提供更精准的决策支持和智能化服务。

一、基于云计算模型的物联网边缘数据挖掘技术的应用

(一) 云存储技术在物联网边缘数据挖掘中的应用

云存储通过分布式文件系统和对象存储技术, 将数据分布在多个节点上, 从而提供高可靠性和可扩展性的服务。对于物联网边缘产生的海量数据, 云存储可以作为长期的存储解决方案, 满足数据归档和分析需求^[2]。借助云存储的弹性扩展能力, 能够有效应对数据量快速增长带来的压力, 同时通过开放的 API 接口提升数据访问效率, 支持高效的数据挖掘操作。

(二) OpenStack 私有云搭建与管理在边缘数据挖掘中的应用

OpenStack 是一个开源的云计算平台, 提供了计算、存储和网络服务的统一管理能力。在物联网边缘场景中, 私有云的部署为数据隐私和安全性提供了保障, 同时允许企业灵活调配资源以满足数据挖掘任务的动态需求。利用 OpenStack, 企业可以方便地对数据存储和计算资源进行监控与管理, 从而优化运行效率, 并减少资源浪费, 确保系统的长期稳定性^[3]。

(三) 容器云技术在物联网边缘数据挖掘中的应用

Docker 和 Kubernetes 等容器云技术, 为数据挖掘应用的部署提供了轻量级的解决方案。容器的可移植性和资源高效利用能力, 使得它在资源受限的边缘设备中也能够表现出色。借助 Kubernetes 等容器编排工具, 边缘计算系统可以实现容器化任务的集中管理和调度, 支持任务的动态扩展与升级。例如, 在智能物流中, 基于容器的边缘计算设备能够实时分析车辆数据, 优化运输路径, 从而提高整体效率。

(四) 公有云技术在物联网边缘数据挖掘中的应用

公有云提供商通过大规模的数据中心和丰

(稿件编号: IS-25-3-1001)

作者简介: 赵玉林 (1994-), 男, 汉族, 籍贯: 江西省赣州市, 硕士研究生, 高职助教, 研究方向: 软件工程, 云计算技术应用。

富的计算资源,为物联网边缘数据挖掘提供了可靠的支持。企业可以按需租用公有云资源,快速部署数据挖掘任务,降低基础设施建设的成本压力^[4]。在智能城市场景中,公有云通过分析海量传感器数据,帮助挖掘空气质量趋势和污染源分布,为环境治理决策提供了可靠依据。同时,公有云的全球覆盖能力进一步拓展了数据挖掘的空间和深度。

(五) Linux 网络操作系统在物联网边缘数据挖掘中的作用

Linux 凭借其开源性、稳定性和强大的网络支持,已成为物联网边缘设备的首选操作系统。在边缘设备上, Linux 为数据采集和预处理提供了基础运行环境,支持数据清洗、特征提取等任务,并以其强大的网络协议栈功能支持设备之间和设备与云端的高效通信。这种优势使 Linux 成为物联网数据挖掘中不可或缺的一部分,为系统运行的可靠性和效率提供了基础保障。

二、基于云计算模型的物联网边缘数据挖掘面临的挑战与解决方案

(一) 网络延迟与带宽限制

网络延迟与带宽限制的问题,就像一条道路在高峰期遭遇严重拥堵,即便边缘计算已经在数据传输总量上做了优化,但边缘与云端之间的通信需求依旧频繁,特别是在实时性要求较高的场景中,延迟和带宽不足往往成为掣肘之处。为解决这一问题,边缘缓存无疑是一种颇具成效的策略。设想边缘设备像一个小型的本地仓库,它能够优先存储那些使用频率高或时间敏感的数据,这样一来,无需每次都从云端调用相关信息^[5]。例如,在智能工厂中,设备运行参数的历史记录和常用预测模型如果能够长期留存在边缘设备中,显然能够大幅减少频繁的数据请求操作。需要注意的是,缓存空间的规划不宜采取“一刀切”的方式,而是应当根据数据的重要性、调用频率以及缓存设备的硬件条件加以灵活调整,甚至可以引入分层缓存机制,将极其关键的数据存放在本地,而次级重要的数据则通过中间节点传递给云端进行进一步处理。这样的设计,不仅让数据交互更加高效,也在某种程度上缓解了带宽不足的压力。

与此同时,网络架构的优化也是解决延迟问题的必由之路。在当今的技术背景下,5G 网络无疑成为了应对这一难题的强有力武器。相比传统网络,5G 的超低延迟和高吞吐量特性为数据传输提供了近乎“高速公路”般的体验。设想一个智能交通系统中,车辆与路边单元、再到云端的多方通信,若仍依赖传统网络,很容易因延迟导致关键决策被推迟,甚至酿成安全隐患。此时,5G 网络的部署能够为这些场景提供一种全新的解决思路,不仅加快了数据传输速度,还能够在带宽压力较大的情况下,分配出更为合理的通信资源。然而,5G 的普及需

要巨大的基础设施投入,并非所有场景都能立刻受益。为此,可以考虑在关键节点优先部署 5G,例如城市的交通枢纽、工业园区的核心生产线等区域。如此一来,有限的资源得到了最有效的利用,既避免了过度投入,又能够迅速为网络优化带来明显成效。

除了硬件和架构上的调整,软件层面的调优同样不可忽视。当前,越来越多的系统开始采用边缘和云端协同的分布式处理模型,这种模型的核心在于将延迟敏感的计算任务分配给边缘节点,而将复杂的非实时任务转移至云端进行处理。这一设计理念的落脚点在于精细化的任务划分与调度算法。如果能够根据任务类型、数据流量以及网络状态实时调整计算负载的分布,就可以让有限的带宽资源得到最大化利用,同时让任务处理的时效性保持在合理范围内。尤其是在动态负载场景中,例如应急事件的处理,需要构建灵活的负载预测模型,根据历史数据与实时流量,提前分配好计算资源和带宽,为高峰时段预留出足够的处理能力。

(二) 数据异构性与互操作性

数据异构性与互操作性一直是物联网边缘数据挖掘中的难题,这就像在一个语言不通的世界中搭建高效沟通的桥梁。物联网设备种类繁多,各自的数据格式、通信协议、编码标准五花八门,而这些数据又必须被整合在一起,为更高层次的分析和决策服务^[6]。问题的核心在于,异构数据无法直接对接,系统间彼此“听不懂”,甚至误解信息,这不仅拖累了处理效率,还可能导致数据丢失或错误分析。解决这一问题,首先需要以统一的数据标准为切入点。行业标准的制定不宜单打独斗,而应借助多方协作的力量,促成标准的共识。例如,在智能交通领域,车辆数据的标准应涵盖速度、位置信息、交通信号的统一编码,既服务于本地设备,又具备跨地区、跨平台的可读性。与此同时,针对已有的多样化设备,标准化也不能一刀切,而应保持一定的兼容性,为此可以设计中间件来进行数据转换。中间件的作用类似于翻译官,将各种设备产生的异构数据“翻译”成通用格式,以便后续处理。以物联网数据平台为例,开发一个支持多协议的数据网关模块,通过灵活的插件机制适配不同的协议标准,这种模块化设计能极大提高系统的兼容性和可扩展性。

除了技术解决方案,底层架构的选择也至关重要。如果将整个系统比作一座大厦,数据互操作性的底层支持就是那坚实的地基。数据库技术的选择上,传统的关系型数据库往往难以处理物联网数据的高并发性和实时性,而非关系型数据库,如 MongoDB 或 Apache Cassandra,因其在处理非结构化数据时的灵活性,成为一个值得探讨的方向。同时,在数据互操作的过程中,资源消耗和性能瓶颈往往成为隐藏的问题,特别是在设备规模较大时,数据转换和适配工作可能带来巨大的计算开销。对此,分布式架构的引入便显得顺理成章。将

数据处理任务分散到多个边缘节点上运行,既减轻了单点压力,也提升了整体效率。设想在智慧农业的场景中,不同农田传感器的温湿度数据被分布式地预处理,再统一汇总到中央系统进行全局分析,这样既保留了实时性,又实现了灵活的数据协同。

考虑到实际系统的运行场景,数据流向的动态变化也需加以重视。一个合理的解决方案是引入语义层次的数据标注体系,为数据赋予语义标签。例如,在工业物联网中,可以为设备的温度数据打上“设备ID”、“时间戳”和“故障预测”等语义标签,使数据在不同系统间的流转和匹配变得更加智能。语义标签的引入不仅方便了机器的自动处理,也为人机协作提供了更直观的接口。

最后,任何技术方案的落地都离不开实践中的权衡与调整。在部署过程中,应建立一套评估体系,从兼容性、处理效率、数据完整性等多个维度,对方案的实际效果进行监测,并根据反馈持续优化。例如,可以定期举办跨行业的测试对接活动,模拟真实场景下的互操作性问题,找到系统薄弱环节并及时改进。这不仅能够帮助技术方案更贴近实际需求,也能推动全行业对数据互操作性的重视,从而为物联网的长远发展奠定基础。数据异构性与互操作性问题虽然复杂,但正是这些挑战促成了技术的革新与进步,而上述策略无疑为我们提供了更多解决这一问题的可能性。

(三) 算力资源的动态调度与优化

算力资源的动态调度与优化一直以来都是物联网边缘数据挖掘中绕不开的话题,它不仅牵动着技术实现的效率,更关乎整个系统的稳定性和资源的经济性。尤其在边缘计算与云计算结合的场景下,算力需求波动大、任务多样化、资源分布不均等问题无时无刻不在考验系统的弹性能力和调度策略。那么,究竟如何破解这一难题?首先,灵活的算力分配机制是核心所在。在边缘计算节点有限的硬件条件下,单靠固定分配资源显然捉襟见肘。更明智的做法是引入任务优先级模型,将任务按重要性和时效性分级处理。例如,对于突发性的实时任务,如交通信号优化或灾害预警,优先分配高性能计算资源;而对于容忍延迟的数据分析任务,如历史数据挖掘,则可以暂时搁置,等待资源空闲时再行处理。这种任务优先级策略不仅能够提升关键任务的完成率,还能避免低优先级任务占用宝贵资源。

另一方面,算力供需之间的不匹配也是一个不可忽视的现实。在这一点上,容器化技术的优势尤为突出。以Kubernetes为例,其容器编排能力就像是一位精通资源分配的调度员,能够在多个节点之间动态分发工作负载。然而,仅仅依赖这种工具本身并不够,调度规则的控制才是决定成败的关键。为了让资源调度更加高效,可以设计一种基于预测的算力需求模型。通过对历史数据和当前任务队列的分析,预判

未来可能的算力需求,并提前释放或预留资源。譬如,在智能工厂的生产线上,如果历史数据显示某一时段内设备故障率高,系统可以主动增加该时段的边缘节点计算能力,以应对突发性数据处理需求。这种未雨绸缪的做法不仅减少了资源浪费,还能避免因资源不足导致的任务延迟。

资源调度过程中常常陷入单点性能瓶颈的窘境,尤其在边缘设备能力有限的场景中,单个节点的负载过高会迅速拖垮整体性能。为了解决这一问题,边缘节点之间的协同显得至关重要。打个比方,如果我们将边缘节点比作棋盘上的棋子,那么这些节点之间的协作就像棋子的排兵布阵。建立节点之间的资源共享机制,将算力需求从过载节点转移到闲置节点,无疑是有效的解决思路。这样的机制可以通过设定一个负载阈值来触发转移操作。当某个节点的负载超出阈值时,系统自动将部分任务转移至其他节点,这既避免了过载,也使闲置资源物尽其用。

三、结语

综上所述,基于云计算模型的物联网边缘数据挖掘技术展现出显著的应用潜力与技术优势,为实时性要求高、数据量庞大的物联网场景提供了强有力的支持。然而,面对网络延迟、数据异构性以及资源动态调度等挑战,技术实施仍需在理论与实践不断磨合。未来的研究与实践应更加关注方案的经济性与可扩展性,同时从行业协作和标准制定入手,为技术普及与创新创造更有利的条件。这一探索的过程不仅是技术发展的见证,更是推动社会各领域智能化变革的重要驱动力。

参考文献:

- [1] 陈江. 基于物联网技术的传感器网络样本数据定向挖掘方法[J]. 自动化技术与应用, 2024, 43 (10): 104-107+143.
- [2] 张颖玮, 桑世庆, 陈祥. 基于大数据技术的藏红花栽培分析模型的构建[J]. 中国新技术新产品, 2021, (17): 16-18.
- [3] 冯傲岸. 基于物联网和大数据应用的高速公路机电系统数字监测与运维[J]. 信息系统工程, 2023, (09): 20-23.
- [4] 杨晓娟. 基于物联网和云计算技术的数据挖掘模型设计[J]. 无线互联科技, 2023, 20 (14): 113-115.
- [5] 晏永刚. 大数据及物联网技术在网络安全系统中的应用[J]. 信息与电脑(理论版), 2024, 36 (08): 232-234+238.
- [6] 周挺. 基于物联网技术的一卡通用户信息数据挖掘模型研究[J]. 自动化与仪器仪表, 2021, (03): 58-60+64.