

Research on the Application of Big Data Analysis Technology in the Network Field

Lichun Zhao

Guangxi Lvcity Water Group Co., Ltd., Nanning, Guangxi, 530000, China

Abstract

With the rapid development of information technology, the scale of data in network systems has shown explosive growth, and data types have become increasingly diversified. Big data analytics has gradually become a crucial means to support network operation and management. Focusing on the demands of massive data processing and value extraction in the network domain, this study systematically reviews the application pathways of big data analytics in network environments from the perspectives of data acquisition, storage and computing, data mining and analysis, and application optimization. Based on the analysis of network traffic characteristics, user behavior patterns, and resource scheduling mechanisms, the key roles of big data analytics in network performance optimization and security protection are further explored. Meanwhile, in combination with the requirements of data security and privacy protection, the supporting functions of related technologies in ensuring stable network operation are examined. By constructing a multidimensional application framework, this study provides technical references for improving the operational efficiency and security level of network systems.

Keywords

big data analytics; network systems; data mining; network optimization; network security

大数据分析技术在网络领域中的运用研究

赵力春

广西绿城水务集团股份有限公司, 中国·广西南宁 530000

摘要

随着信息技术的快速发展,网络系统中数据规模呈现爆发式增长,数据类型愈加多样化,大数据分析技术逐渐成为支撑网络运行与管理的重要手段。围绕网络领域中海量数据处理与价值挖掘需求,从数据采集、存储计算、挖掘分析及应用优化等环节出发,对大数据分析技术在网络环境中的应用路径进行系统梳理。在分析网络流量特征、用户行为模式及资源调度机制的基础上,进一步探讨其在网络性能优化与安全防护中的关键作用。同时,结合数据安全与隐私保护需求,分析相关技术在保障网络稳定运行中的支撑功能。通过构建多维度应用框架,为提升网络系统运行效率与安全水平提供技术参考。

关键词

大数据分析技术; 网络系统; 数据挖掘; 网络优化; 网络安全

1 引言

数字化进程的不断推进使网络基础设施逐步向高性能与高复杂度方向演进,网络空间中产生的数据不仅规模庞大,而且具有高频率、强动态与多源异构等显著特征。传统数据处理方式在面对海量网络数据时已难以满足实时分析与精细化管理需求,数据价值难以充分释放。大数据分析技术依托分布式计算、智能算法及多源数据融合等手段,为网络数据处理与应用提供了新的技术路径。网络运行状态、用户访问行为以及安全风险特征均可通过数据分析实现精细刻画,使网络管理由经验驱动逐步转向数据驱动。在复杂

网络环境中,如何通过高效的数据分析技术实现资源优化配置、提升服务质量并强化安全防护能力,已成为网络技术发展中的关键问题。

2 大数据分析技术的网络应用基础体系构建

2.1 大数据分析技术的核心内涵与关键特征

大数据分析技术是在海量数据环境下,通过高效计算与智能算法实现数据价值提取的综合技术体系,其核心在于对规模庞大、结构复杂的数据进行快速处理与深度分析。网络环境中数据呈现出体量巨大、类型多样及变化频繁等特征,数据规模可由TB级迅速扩展至PB级,传统集中式处理方式难以支撑实时分析需求。该技术具备高并发处理能力与横向扩展能力,通过分布式架构实现计算资源的动态调度,使数据处理效率显著提升。数据价值密度较低的特征要

【作者简介】赵力春(1974-),男,中国河北定兴县人,本科,工程师,从事水务计算机技术应用与安全生产等研究。

求分析过程具备强筛选与提取能力,通过模型训练与算法优化实现有效信息识别。在网络领域应用过程中,数据实时性要求不断提高,毫秒级响应逐渐成为常态,使大数据分析技术不仅承担数据处理功能,还承担决策支撑与运行优化的重要角色。

2.2 网络领域数据资源类型与结构特征分析

网络领域数据资源来源广泛,涵盖网络流量数据、用户行为数据、设备运行数据及安全日志数据等多种类型,不同数据在结构形态上存在显著差异。网络流量数据通常以数据包形式存在,具有强时序特征与高频采集特性,每秒数据产生量可达数百万条记录。用户行为数据多以访问记录、点击路径及停留时长等形式呈现,具有半结构化特征,便于进行行为建模分析。设备运行数据来源于路由器、交换机及服务器等网络设备,数据结构相对规范,但更新频率高,对存储与处理能力提出较高要求。安全日志数据包包含入侵检测、访问控制及异常记录等信息,数据量虽相对较小,但对准确性与完整性要求较高。多源异构数据并存的环境,使数据整合与统一建模成为关键问题,数据结构由结构化向半结构化与非结构化演进趋势明显^[1]。

3 大数据分析技术在网络数据处理中的关键应用

3.1 海量网络数据采集与预处理技术

网络数据采集过程依赖多点部署与高频抓取机制,通过流量镜像与日志采集方式实现数据获取,单个核心节点每秒可采集数据包数量达到 5×10^5 至 1×10^6 条,日均数据量超过2TB。在数据传输过程中采用高效压缩与分批传输技术,使带宽占用率控制在70%以内。预处理阶段通过数据清洗与格式转换对原始数据进行规范化处理,异常数据剔除比例通常控制在3%至8%之间。时间同步机制用于统一多源数据时间戳误差,将偏差控制在 $\pm 5\text{ms}$ 范围内,保证数据分析的时序准确性。数据去重与聚合处理使存储数据量降低约30%,同时提升后续分析效率。通过构建标准化数据格式,实现多源数据的统一输入,为后续建模与分析提供稳定基础。

3.2 分布式存储与计算框架应用机制

分布式存储系统通过多节点部署实现数据的冗余存储与负载分担,典型集群由50至200个节点组成,单节点存储容量约为20TB,总体存储能力可扩展至数PB级。数据副本机制将重要数据复制3份以上,使系统容错率提升至99.9%以上。计算框架采用并行处理模式,将任务划分为多个子任务在不同节点同时执行,使数据处理效率提升5至10倍。在典型网络分析任务中,处理1TB数据所需时间可由传统模式的6小时缩短至40分钟以内。资源调度系统根据任务负载动态分配计算资源,使CPU利用率维持在65%至85%之间,避免资源浪费。通过计算与存储协同优化,

实现数据访问延迟控制在100ms以内,保障网络分析的实时性与稳定性^[2]。

3.3 网络数据挖掘与特征提取方法

网络数据挖掘过程以算法模型为核心,通过对海量数据进行模式识别与规律提取,实现对网络运行状态的深入分析。常用方法包括聚类分析、关联规则挖掘及分类预测等,其中聚类算法可将用户行为数据划分为5至10类典型模式,使分析准确率提升至85%以上。特征提取过程中,通过降维技术将原始数据维度从上千维压缩至100维以内,同时保留90%以上的信息量,有效降低计算复杂度。流量特征分析中,关键指标包括包长度、传输速率及连接时长等,通过对这些指标进行统计分析,可识别异常流量占比约为2%至5%。在安全检测场景中,特征向量匹配精度可达到92%以上,使潜在威胁能够被及时识别。通过多算法融合与模型优化,网络数据分析能力得到持续提升,为网络运行优化与安全防护提供可靠支撑。

4 大数据分析技术在网络运行优化中的应用机制

4.1 网络流量分析与性能优化技术

网络流量分析依托高频数据采集与深度包检测技术,对网络中数据传输过程进行精细化刻画,通过对数据包长度、传输速率及连接持续时间等关键指标进行统计建模,实现流量分布特征识别。核心链路中流量峰值可达到10Gbps以上,采用流量采样率为1:100的方式可在保证分析精度95%以上的前提下降低计算压力。通过引入时序分析模型,对流量波动进行趋势预测,使网络拥塞提前识别率提升至88%以上。性能优化环节通过动态带宽分配与拥塞控制算法,将链路利用率由60%提升至85%,同时将平均时延由120ms降低至45ms。结合流量分类技术,将视频流、数据流及控制流进行区分处理,使关键业务数据丢包率控制在0.5%以内。数据驱动的优化机制使网络运行状态能够实时调整,从而实现高效稳定运行^[3]。

4.2 用户行为分析与服务质量提升路径

用户行为分析通过日志数据与访问记录构建行为特征模型,对用户访问路径、停留时长及操作频率进行量化分析,单日用户行为数据规模可达 5×10^7 条记录。通过构建用户画像模型,将用户划分为高活跃、中活跃及低活跃等不同类别,使行为识别准确率达到90%以上。结合聚类分析与序列模式挖掘技术,可识别典型访问路径长度为5至8步的行为模式,从而优化页面结构与服务响应逻辑。服务质量评估以响应时间、成功率及用户满意度为核心指标,通过实时监测将页面平均响应时间由2.5s优化至0.8s,服务请求成功率提升至99.2%。通过个性化推荐算法,将内容匹配度提升至85%以上,使用户停留时长增加约30%。行为数据与服务质量指标的深度融合,使网络服务由被动响应转向主动优

化,提升整体用户体验水平^[4]。

4.3 网络资源调度与负载均衡优化方法

网络资源调度依托分布式计算与动态调度算法,对计算资源、存储资源及带宽资源进行协同管理,通过实时监测节点负载状态,实现资源分配优化。典型数据中心包含 100 至 300 个计算节点,单节点 CPU 利用率波动范围在 30% 至 95% 之间,通过引入负载预测模型,将资源利用率稳定在 70% 至 85% 区间。负载均衡技术采用加权轮询与最小连接数算法,将请求分配至不同服务器节点,使单节点负载差异控制在 ±10% 以内。通过动态迁移机制,将高负载节点任务转移至空闲节点,使任务响应时间缩短约 40%。在高并发场景下,每秒请求量可达到 2×10^5 次,通过调度优化可将系统吞吐能力提升至原有水平的 1.8 倍。资源调度与负载均衡的协同应用,使网络系统在复杂环境下仍保持稳定高效运行。图 1 为 deepseek 系统网络资源调度的流程部署。

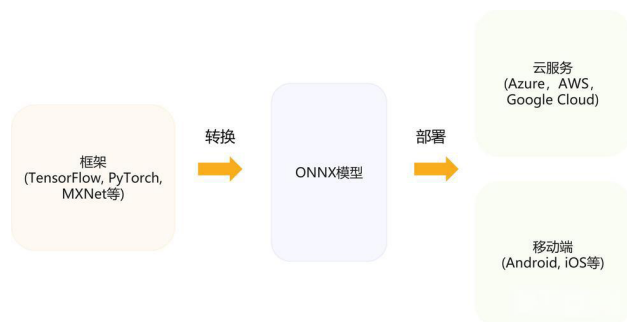


图 1 deepseek 系统网络资源调度的流程部署

5 大数据分析技术在网络安全保障中的应用路径

5.1 网络异常检测与入侵识别技术

网络异常检测通过对流量特征与行为模式进行持续监测,实现对异常活动的快速识别。基于机器学习算法构建分类模型,将正常流量与异常流量进行区分,训练样本规模可达到 10^6 条以上,使识别准确率达到 93% 以上。异常检测指标包括连接频率、数据包大小及访问时间间隔等,通过对这些特征进行统计分析,可识别异常行为占比约为 3% 至 6%。入侵识别系统采用规则匹配与行为分析相结合的方式,对已知攻击模式与未知攻击行为进行综合识别,使检测响应时间控制在 50ms 以内。通过引入深度学习模型,对复杂攻击行为进行特征提取,使误报率降低至 2% 以下。实时检测机制使潜在威胁能够在早期阶段被识别,从而提升网络整体安全防护能力。

5.2 安全威胁预测与风险评估方法

安全威胁预测依托历史攻击数据与实时监测数据构建预测模型,通过时间序列分析与概率模型对攻击趋势进行评估。历史数据规模可达到数 TB 级,通过数据清洗与特征提

取,将关键特征维度控制在 200 以内,使模型训练效率显著提升。风险评估以威胁发生概率与影响程度为核心指标,通过构建风险矩阵模型,将风险等级划分为低、中、高三级,使评估准确率达到 88% 以上。预测模型可对未来 24 小时内攻击发生概率进行估计,其误差范围控制在 ±7% 以内。通过对关键节点风险进行量化分析,将高风险节点识别率提升至 90% 以上,从而实现重点防护。数据驱动的预测与评估机制使安全管理由事后处理转向事前预防,提升整体防控能力^[5]。

5.3 数据安全与隐私保护技术应用

数据安全与隐私保护通过加密算法与访问控制机制实现对敏感信息的有效保护,在数据传输过程中采用对称加密与非对称加密相结合的方式,使数据泄露风险降低 80% 以上。访问控制系统通过身份认证与权限管理,对不同用户设置访问级别,使非法访问拦截率达到 98% 以上。数据脱敏技术通过对敏感字段进行替换与扰动处理,使数据可用性保持在 85% 以上的同时保障隐私安全。隐私保护模型通过差分隐私机制控制数据发布风险,使隐私泄露概率降低至 0.01 以下。存储安全方面采用多副本与分布式存储策略,使数据可靠性达到 99.99%。多层次安全技术的综合应用,使数据在采集、传输与存储全过程中均处于可控状态,从而为网络系统安全运行提供坚实保障。

6 结语

大数据分析技术在网络领域中的深入应用,正在重塑网络运行与管理模式,使数据由被动积累转变为主动驱动的重要资源。围绕数据采集、处理分析及应用优化等关键环节,技术体系不断完善,网络性能提升与安全保障能力显著增强。在复杂多变的网络环境中,通过数据挖掘与智能分析实现资源配置优化与风险预判,已成为提升系统稳定性的重要路径。面向未来,随着数据规模持续扩大与应用场景不断拓展,大数据分析技术将在网络领域发挥更加关键的支撑作用,推动网络系统向高效、安全与智能化方向持续发展。

参考文献

- [1] 李家潇.大数据分析技术在网络领域中的应用分析[J].科技资讯,2025,23(20):49-51.
- [2] 刘子华.网络安全分析中的大数据技术应用研究[A].2025数字时代的社会结构变迁与治理创新学术交流会论文集(上)[C].中国智慧工程研究会:2025:319-320.
- [3] 孙伟.网络安全分析中的大数据技术应用[J].石子科技,2024,(02):17-19.
- [4] 王芳.大数据支撑下计算机应用技术的分析与研究[J].现代计算机,2023,29(12):60-64.
- [5] 王宇韬,钱妍竹.Python大数据分析机器学习商业案例实战[M].机械工业出版社:202301.