

Analysis of Computer Information Processing Technology in the Context of Big Data Era

Fuyou Cheng

Chinalco Intelligent Numerical (Hangzhou) Engineering Design and Research Institute Co., Ltd., Hangzhou, Zhejiang, 311101, China

Abstract

Against the backdrop of rapid growth in global data volume, computer technology must address the storage, processing, and transmission needs of massive amounts of data, driving innovation in fields such as data mining, machine learning, and artificial intelligence. This paper discusses in detail the characteristics of big data and its wide applications in fields such as finance, healthcare, and transportation, while also exploring the application and development of technologies such as cloud storage, distributed storage, and parallel computing. At the same time, data security and privacy protection have become important issues in big data applications. This paper analyzes existing technological means and future improvement directions, and finally summarizes the broad prospects and challenges faced by the development of big data technology, pointing out the importance of technological innovation and talent cultivation.

Keywords

big data; computer information processing technology; production efficiency

大数据时代背景下计算机信息处理技术的分析

程福友

中铝智能数维（杭州）工程设计研究院有限公司，中国·浙江 杭州 311101

摘要

在全球数据量迅猛增长的背景下，计算机技术必须应对海量数据的存储、处理和传输需求，推动了数据挖掘、机器学习和人工智能等领域的创新。论文详细讨论了大数据的特征及其在金融、医疗、交通等领域的广泛应用，同时探讨了云存储、分布式存储和并行计算等技术的应用与发展。与此同时，数据安全和隐私保护问题成为大数据应用的重要课题，论文分析了现有的技术手段和未来的改进方向，最后总结了大数据技术发展的广阔前景和面临的挑战，指出技术创新和人才培养的重要性。

关键词

大数据；计算机信息处理技术；生产效率

1 引言

在大数据时代，全球数据量迅猛增长，对计算机信息处理技术提出了更高要求。据国际数据公司（IDC）预测，全球数据总量将在 2025 年达到 175 泽字节。大数据的应用涵盖金融、医疗、交通、电子商务等领域，推动了数据挖掘、机器学习、人工智能等技术的发展^[1]。然而，海量数据带来的存储和处理挑战促使云存储、分布式存储和并行计算技术的不断创新。同时，数据安全和隐私问题也亟须解决，确保大数据的可靠应用成为各领域的重要课题。

2 大数据特征及影响

大数据的特征主要体现于三个方面：数据量、存储空间、

处理速度。数据量之大可谓浩如烟海，现代社会的每一次点击、每一条信息、每一个交易都在不断地生成庞大的数据，这些数据如同大江大河，源源不断地汇入信息海洋。存储空间的需求亦如日益增长的粮仓，不断地扩展以容纳这些日益增多的信息，传统的存储方式已经无法满足需求，云存储、分布式存储等新技术应运而生，为大数据的存储提供了可靠保障^[2]。处理速度的提升如同骏马飞驰，在数据处理的竞赛中，速度决定了成败。高效的数据处理技术，如并行计算、分布式计算等，使得庞大的数据在短时间内得到处理和分

析，迅速转化为有价值的信息。在计算机处理信息时，必须辨别和区分文字、图片、视频、网页格式等数据种类，这样才能提高处理的便利性，也使得信息处理更加准确和高效。举例而言，面对海量的文字信息，文本挖掘技术如同一把利剑，能够迅速从中提取关

【作者简介】程福友（1978-），男，中国浙江杭州人，本科，高级工程师，从事工业控制、人工智能研究。

键信息；而面对图像数据，图像识别技术则如同一双慧眼，能够准确识别和分类。应用这些技术，使得计算机处理信息的能力极大增强，为各行各业提供了强有力的支持。

大数据正在彻底重塑企业的运营模式，特别是在传统制造业中。过去，制造业的生产流程遵循“6+1”模式，涵盖了产品设计、原料运输、生产加工、订单处理、市场营销和零售六大关键环节。然而，随着大数据技术的融入，这一传统模式正经历着根本性的变化。大数据技术的介入，就像一位智慧的策略师，它可以重新配置和优化生产链的各个环节，催生了更加精简高效的“4+1”产业链模型。在这个新模型中，订单处理被推到了核心位置，利用先进的数据分析工具，可以实现生产流程的极大简化和加速，大幅提升了生产效率，并朝向零库存的目标迈进。这种模式不仅使生产过程更加高效，还极大地增强了企业的市场竞争力，由此带来的是企业流程的全面优化和竞争策略的根本刷新。

建立大数据平台后，企业可以借助数据分析了解市场需求，制定更为精准的生产计划，避免资源浪费。订单处理、生产设计、原料运输、生产加工形成了新的闭环，使得企业运作更加灵活和高效^[1]。依托大数据方法，信息采集、数据分析、文字表述等步骤有条不紊地进行，以消费者为主导的制造生产计划得以实现，使得市场经济环境下的企业运作更加具有计划性和可控性。

3 计算机处理技术分析

3.1 获取—加工—传输

3.1.1 获取

在大数据时代，数据获取的复杂性和广泛性犹如采集珍珠，既要求广覆盖，又要求高精度。数据来源包括但不限于传感器网络、爬虫技术和物联网设备。传感器网络在各个环境中部署了大量传感器，实时收集物理世界中的温度、湿度、光照等数据。爬虫技术可模拟用户行为，从互联网上自动化抓取数据，尤其适用于结构化和半结构化数据的采集。物联网设备可连接智能家居、智能城市基础设施等各类智能设备，实现海量数据的实时收集。这些技术如同一双慧眼，从纷繁复杂的信息海洋中精准、快速地挑选出有价值的信息，为后续的数据处理奠定坚实的基础。

3.1.2 加工

数据加工是将原始数据转化为有用信息的关键环节，犹如炼金术。首先，需要对数据进行清洗，涵盖噪音数据和无效数据剔除两个环节。这一过程包括去重、补全缺失值、纠正错误数据等技术措施。使用正则表达式进行文本数据的格式校验，以算法填补缺失值，或利用数据去重算法识别和删除重复数据。其次，数据转换，将不同格式和类型的数据统一转换为标准化的格式，如将不同编码的文本数据统一转换为 UTF-8 编码，或者将各种日期格式统一转换为 ISO 8601 标准格式。数据转换也包括对数据实施规范化处理，

如利用归一化算法将数值数据缩放到同一范围内。最后，数据集成，将来自不同源头的数据融合在一起，形成完整的数据视图，使用 ETL（抽取、转换、加载）技术，在数据仓库或数据湖支持下将异构数据源整合在一起，利用数据融合算法处理数据冲突和不一致性。

3.1.3 传输

数据传输技术是保证数据在不同节点间高效、安全流动的关键。光纤通信利用光纤作为传输介质，具有高带宽、低延迟的优点，适用于大规模数据传输。5G 网络提供更高的传输速度和更低的延迟，适用于需要实时数据传输的应用场景。如 TCP/IP 协议等传输协议为数据传输提供了可靠性保障，基于握手机制、重传机制等技术支持数据包完整传输。数据压缩技术利用 Huffman 编码、LZ77 算法减少数据传输量，提高传输效率。数据传输安全性方面，AES（高级加密标准）、TLS（传输层安全协议）等加密技术为数据提供了机密性和完整性保障。虚拟专用网络（VPN）以加密隧道技术保护数据在公共网络中安全传输，防火墙技术负责过滤恶意数据包，保护数据传输链路的安全，这些技术如同信息高速公路上的护卫队，共同为大数据的分析和应用提供了可靠的支持。

3.2 存储—感知—安全

3.2.1 存储

在大数据时代，数据存储如同建造一座巨大的粮仓，需要同时满足容量大、效率高和可靠性的需求。传统的存储方式已难以应对海量数据的挑战，云存储和分布式存储技术应运而生。云存储犹如天空中的云朵，容量无限、灵活扩展，用户可以按需存取数据，如 Amazon S3、Google Cloud Storage 等平台均可提供高可用性和弹性，实现数据等随时随地访问。分布式存储如同一张密织的蜘蛛网，可将数据分布存储在多个节点上，实现数据的高可用性和容错性。Hadoop 的 HDFS（Hadoop Distributed File System）和 Apache Cassandra 等系统利用数据分片和复制使得技术即使在部分节点故障的情况下仍可安全可靠地存取。

3.2.2 感知

数据感知技术是大数据处理中的关键环节，犹如一双慧眼，使计算机能够感知和理解来自物理世界的各种数据。传感器技术可实时监测收集温度、湿度、光照、压力等物理参数。如智能城市中广泛应用的环境监测传感器，可用来实时收集城市各个角落的空气质量数据，为环境保护和治理提供科学依据。物联网（IoT）技术进一步拓展了数据感知的范围，将各种智能设备连接起来，形成一个庞大的数据感知网络。智能家居中的温控器、智能手环等设备也可实时收集用户的生活数据和健康数据，为个性化服务和健康管理提供了丰富的数据支持。应用机器学习和人工智能技术，也可使数据感知从简单的数据收集向高级的数据理解和预测转变。例如，计算机视觉技术可借助深度学习算法准确识别、分类

图像中的物体和场景，为自动驾驶、安防监控等领域提供技术支持（图1）。自然语言处理（NLP）技术可基于语义分析和理解，使计算机能够准确感知和处理人类语言，为智能客服、语言翻译等应用提供了有效的解决方案。



图1 计算机视觉技术在各领域的应用

3.2.3 安全

数据安全是大数据处理的基石，犹如建造一座坚固的城堡，既要防止外敌入侵，又要防止内部泄密。随着数据量的急剧增加，数据泄露、黑客攻击等安全问题层出不穷，对数据安全提出了更高的要求。加密技术是数据安全的第一道防线，对数据进行加密处理后，即使数据在传输过程中被截获，攻击者也无法解读数据内容。访问控制技术以身份验证、权限管理方式确保只有授权用户才能访问敏感数据。例如，基于角色的访问控制（RBAC）模型可分配不同的角色和权限，严格控制用户对数据的访问范围。数据脱敏技术在处理敏感数据时，会对数据进行伪装转换，避免数据在共享分析中泄露隐私信息。隐私保护计算（PPC）技术以加密计算、差分隐私等方法保护数据隐私，使数据可安全地计算分析。

3.3 发展—问题

3.3.1 发展

在大数据时代，计算机信息处理技术的发展可谓日新月异，如同奔腾的江河，既有波澜壮阔的前景，也有曲折迂回的挑战。大数据技术的发展经历了从数据采集、存储、

处理到分析的全过程，每一步都离不开技术的创新和进步。首先，数据采集技术的进步使得数据源更加多样化，物联网（IoT）设备、传感器网络、社交媒体等多种渠道的涌现，极大地扩展了数据的覆盖面和获取能力。存储技术的突破为大数据提供了坚实的基础，Hadoop HDFS、Amazon S3 云存储、分布式存储技术的成熟应用也使得海量数据的存储变得更加高效和经济。

随着边缘计算和雾计算技术的兴起，大数据技术的发展更为迅速，边缘计算可将数据处理和计算任务分散到靠近数据源的设备上，减少数据传输延迟，提高了实时处理的能力。而雾计算可在云计算和边缘计算之间建立一个中间层，实现更灵活、更高效的数据处理架构。

3.3.2 问题

大数据技术的发展也面临着诸多问题和挑战，如同在广阔的海洋中航行，既有惊涛骇浪，也有暗礁险滩。首先，数据隐私和安全问题日益突出，随着数据量的急剧增加，数据泄露、黑客攻击等安全事件频发，对个人隐私和企业数据安全构成了严重威胁。尽管加密技术、访问控制和区块链技术为数据安全提供了一定的保障，但仍需不断改进和完善，更好地应对日益复杂的安全威胁。

数据处理和分析的效率问题也不容忽视，尽管分布式计算框架和先进的算法提高了数据处理的效率，但随着数据量的不断增长，计算资源的需求也在急剧增加。这不仅对计算设备的性能提出了更高要求，也增加了能源消耗和成本。如何在保证处理效率的同时降低资源消耗，成为大数据技术发展的重要课题。

4 结语

总的来说，大数据时代给计算机信息处理技术带来了前所未有的机遇和挑战。随着全球数据量的迅猛增长，技术创新不断推进，使得数据采集、存储、处理和分析变得更加高效和智能。然而，在享受大数据带来的便利和优势的同时，我们也必须面对诸如数据隐私、安全、质量和效率等方面的挑战。这就如同在大海中航行，既要借助技术的风帆，乘风破浪，又要警惕隐患，确保航程的安全和顺利。

参考文献

- [1] 李正君.计算机信息处理技术在大数据时代中的有效应用[J].科技创新与应用,2019(21):159-160.
- [2] 崔亮亮.大数据时代背景下计算机信息处理技术的分析[J].现代工业经济和信息化,2022,12(2):122-123+152.
- [3] 熊志强,张娴静.基于大数据背景下的计算机信息处理技术研究[J].北京印刷学院学报,2020,28(1):142-144.