

The Key Role and Application Research of Big Data in Artificial Intelligence

Yingping Fu

Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi, 710061, China

Abstract

With the rapid development of the digital economy, the deep integration of big data and artificial intelligence has become the core driving force for promoting social change and industrial upgrading. Starting from the collaborative relationship between the two, this article systematically explains the key role of big data as a fundamental element in the development of artificial intelligence. It deeply analyzes the specific applications of big data in AI core technologies such as machine learning, deep learning, and reinforcement learning. Typical cases such as intelligent recommendation, financial risk control, and medical imaging diagnosis are combined to verify their practical value. At the same time, corresponding solutions have been proposed to address prominent issues such as data quality and privacy security, and the future development trend is finally discussed. Research shows that the collaborative evolution of big data and artificial intelligence is reshaping the intelligence process of various industries, and achieving a balance between efficient utilization of data elements and security protection will be a key direction for future development.

Keywords

big data; artificial intelligence; Collaborative development; Machine learning; Data Governance

大数据在人工智能中的关键作用与应用研究

付影平

西安邮电大学, 中国·陕西 西安 710061

摘要

当下数字经济飞速发展, 大数据与人工智能的深度融合已成为推动社会变革和产业升级的核心动力。本文从两者的协同关系出发, 关键作用作为人工智能发展基础要素的大数据被系统阐释, 深入分析了大数据在机器学习、深度学习、强化学习等AI核心技术中的具体应用, 智能推荐、金融风控、医疗影像诊断等典型案例被结合以验证其实际价值。同时, 针对数据质量、隐私安全等突出问题, 相应的解决策略被提出, 最后未来发展趋势被展望。研究表明, 大数据与人工智能的协同演进正在重塑各行业的智能化进程, 而实现数据要素的高效利用与安全保护的平衡, 将是未来发展的关键方向。

关键词

大数据; 人工智能; 协同发展; 机器学习; 数据治理

1 引言

信息技术的迅猛发展使得数据正以前所未有的速度爆炸式增长。据 IDC 相关预测, 到 2025 年全球数据总量将攀升至 175ZB, 这一海量数据资源为人工智能的蓬勃发展提供了坚实的“养料”基础。回顾人工智能的发展历程, 从击败世界围棋冠军的 AlphaGo 到能生成流畅文本的 GPT-3, 这些突破性成果的背后, 高质量、大规模的大数据支撑无一不被依赖。

不仅为人工智能模型的训练提供了丰富的素材, 大数据更通过构建有效的反馈闭环, 持续推动算法性能的优化, 两者相互促进、协同发展的良性循环得以形成。在这一背景

下, 深入探究大数据在人工智能各技术领域中的关键作用, 分析两者融合产生的实际价值, 探讨发展过程中面临的技术挑战及解决方案, 对于推动大数据与人工智能的协同创新具有重要的理论意义和实践价值。^[1] 本文正是基于此, 对上述问题展开系统研究, 以为相关领域的发展提供有益的参考和指导。

2 大数据与人工智能的协同关系解析

2.1 大数据: 人工智能发展的核心基石

三大核心要素即数据、算法和算力离不开人工智能的发展, 其中数据作为训练智能模型的“原材料”, 其重要性不言而喻。4V 特征即数据量 (Volume)、数据时效性 (Velocity)、数据多样性 (Variety) 和数据价值密度 (Value) 被大数据所具备, 直接对人工智能模型的性能产生深远影响。具体而言, 模型的泛化能力由数据量的大小决定, 数据

【作者简介】付影平 (1973-), 男, 中国陕西西安人, 本科, 工程师, 从事大数据研究。

量越充足，模型在面对新的、未知的数据时，表现往往越稳定可靠；实时决策的效果与数据的时效性相关，在金融交易、自动驾驶等对实时性要求极高的领域，及时获取和处理最新数据至关重要；模型的鲁棒性能够被数据的多样性提升，让模型在面对不同类型、不同来源的数据时，依然能够保持较好的性能；而训练效率与数据的价值密度紧密相关，高价值密度的数据能够让模型在更短的时间内学到有效的知识，提高训练效率。

2.2 大数据技术：人工智能工程化的强大引擎

人工智能能够从理论走向实践，实现工程化应用，大数据技术栈为其实际落地提供了坚实的基础设施支撑。海量数据的存储难题被分布式存储技术如 HDFS 有效解决，让大规模数据的保存成为可能；模型的训练过程被并行计算技术如 Spark 极大地加速，模型开发和迭代的周期被缩短；实时的人工智能推理被流处理技术如 Flink 支持，诸多场景下对实时决策的需求被满足；而多源数据的统一管理被数据湖架构实现，数据孤岛被打破，更全面、更丰富的数据支持为人工智能模型所提供。

3 大数据在人工智能核心技术中的实践应用

3.1 机器学习：数据驱动下的模型进化

标注大数据在监督学习领域起着至关重要的作用。以计算机视觉领域为例，ImageNet 数据集的出现，为相关算法的研究和突破提供了强大的数据支撑，计算机视觉技术的飞速发展被推动；在推荐算法方面，千万级用户的行为数据被用于优化算法，推荐结果更加符合用户的喜好和需求。^[2] 潜在的模式和规律从海量数据中自主被发现是无监督学习的侧重点。通过聚类分析技术，客户能够被有效地分群，不同客户群体的特征和需求被企业更好地了解；而异常检测技术则能够从大量的交易数据中发现金融欺诈等异常行为，金融安全被保驾护航。

3.2 深度学习：大数据规模与模型性能的正相关

明显的正相关关系在深度学习中的神经网络性能与数据规模之间呈现。在自然语言处理领域，高达 45TB 的文本数据被 GPT-3 模型的训练所使用，这使得其在语言生成、理解等任务上表现出色；在计算机视觉领域，包含 1400 万张图像的 ImageNet 数据集被 ResNet 模型训练所使用，强大的图像识别能力从而具备。数据不足的问题为了应对，数据增强技术应运而生，有效数据量被其有效扩展。在图像处理中，样本的多样性通过对图像进行旋转、裁剪等操作可以增加；在文本处理中，语言模型的泛化性通过采用文本回译等方法能够提升，使其在不同的语言场景下都能有较好的表现。

4 大数据与人工智能融合的典型应用案例

4.1 智能推荐系统：精准对接用户需求

大数据的作用在电商平台的智能推荐系统中得到了充

分体现。每日新增 TB 级的用户行为数据被该系统作为基础，包括用户的浏览记录、购买历史、收藏行为等。基于协同过滤的混合推荐模式被其算法架构采用，用户的个性化特征和群体行为模式被结合。这些海量数据通过分析和挖掘，用户的喜好和需求被系统精准把握，合适的商品为用户所推荐。实际应用效果显示，采用该智能推荐系统后，电商平台的点击率提高了 35%，商品交易总额（GMV）增长了 22%，平台的运营效率和用户满意度被极大地提升。

4.2 金融风控模型：筑牢金融安全防线

大数据在金融领域应用的典型案例是银行的反欺诈系统。超过 10 亿条的交易记录以及详细的客户画像信息被该系统的数据来源广泛包括，客户的基本信息、信用记录、消费习惯等多个方面被客户画像涵盖。2000 多个维度的风险特征在特征工程阶段被系统从这些海量数据中提取出，通过对这些特征的分析建模，强大的金融风控模型被构建。^[3] 潜在的欺诈行为被该模型准确识别出，经实际验证，其欺诈识别准确率达到了 99.2%，银行的金融风险被有效降低，金融系统的安全稳定运行被保障。

4.3 医疗影像诊断：提升疾病诊断效率与准确性

医疗行业带来了显著变革通过 AI 辅助诊断系统在医疗影像诊断领域的应用。百万级标注的医学影像数据被该系统作为基础，包括 X 光片、CT 图像、核磁共振图像等。3D CNN 网络架构被其算法训练采用，医学影像中的深层次特征能够被深入挖掘。强大的疾病检测能力通过在大量标注数据上的训练被该系统具备。在临床验证中，其对肺癌的检测灵敏度达到了 91.3%，疾病被医生更快速、更准确地发现，诊断效率被提高，患者的早期治疗为宝贵时间所争取。

5 大数据与人工智能发展面临的技术挑战及应对策略

5.1 数据质量治理：提升数据有效性

制约模型性能的重要因素在大数据与人工智能的发展过程中成为数据质量问题。常见的问题包括样本偏差导致模型产生偏见，例如在招聘相关的 AI 模型中，如果训练数据存在性别偏见，可能会导致模型在招聘时对某一性别的候选人产生歧视；监督学习的效果被标注错误影响，模型的学习被错误的标注信息误导；模型的预测精度被数据噪声降低，模型的决策可靠性下降。

针对这些问题，有效的解决方案需要被采取。首先，完善的数据质量评估体系被建立，对数据的准确性、完整性、一致性等进行全面评估；其次，自动化清洗工具被开发，数据中的错误和噪声被快速识别和处理；此外，鲁棒性更强的算法被设计，模型对数据质量问题的容忍度被提高，数据质量问题对模型性能的影响被减少。

5.2 隐私安全保护：平衡数据利用与安全

隐私安全问题日益凸显随着大数据和人工智能的广泛

应用。面临的风险主要有敏感信息泄露,如用户的个人信息、健康数据等被非法获取和利用;模型逆向攻击,攻击者通过对模型的输出进行分析,训练数据中的敏感信息可能被反推出;成员推断攻击,攻击者能够判断某一个体是否属于模型的训练数据集中。这些隐私安全风险为了应对,一系列创新技术应运而生。联邦学习技术实现了数据的“可用不可见”,不同机构在不共享原始数据的情况下进行联合建模被允许,数据的隐私被保护;差分隐私技术通过在数据中加入适当的噪声,数据的统计安全被保障,个体的敏感信息无法被攻击者从统计结果中获取;同态加密技术则支持在密文上进行计算,数据在处理过程中的安全性被确保。

6 大数据与人工智能的未来发展趋势展望

6.1 技术创新方向:突破现有瓶颈

多个重要方向将在未来大数据与人工智能的技术创新方面呈现。多模态数据融合是其中之一,文本、图像、语音等多种来源的数据被其进行融合处理,信息能够被更全面地理解和分析,模型的性能和应用范围被提升。对大量数据的依赖瓶颈被小样本学习技术旨在突破,模型能够在数据量有限的情况下,通过少量样本快速学习并掌握相关知识和技能,这对于一些数据获取困难的领域具有重要意义。^[4]数据高效利用也是未来的重要发展方向,数据处理流程被优化、数据价值挖掘能力被提升,单位数据的价值产出被提高,数据资源的最大化利用被实现。

6.2 行业应用深化:赋能各领域升级

大数据与人工智能的融合将进一步深化在行业应用方面,各个领域的升级发展被注入新的动力。在工业互联网领域,设备运行数据通过实时监测和分析,设备的预测性维护被实现,智能制造的发展被推动,生产效率和产品质量被提高。在智慧城市建设中,城市大数据被充分利用,公共服务如交通管理、能源分配、环境保护等被优化,城市的运行效率和居民的生活质量被提升。在元宇宙领域,用户行为数据将被用于构建更加真实、丰富的虚拟世界,沉浸式的体验被用户提供,元宇宙相关产业的发展被推动。

6.3 生态体系完善:规范行业发展

大数据与人工智能的健康有序发展为了保障,相关的生态体系需要不断完善。数据要素市场的规范化是重要的一环,健全的数据交易规则和机制被建立,数据的流通和共享

被促进,数据的要素价值被充分发挥。AI数据标准体系被建立,数据的格式、质量要求等被统一,数据的兼容性和可用性被提高,AI技术的标准化发展被推动。此外,数据安全立法被健全,数据收集、使用、存储等环节的安全责任和规范被明确,大数据和人工智能的发展被坚实的法律保障所提供,数据安全风险被防范。

7 结论

以下主要结论通过对大数据与人工智能协同发展的系统研究,深入分析两者之间的紧密关系、大数据在AI核心技术中的应用、典型案例以及面临的挑战和未来趋势得出:

首先,人工智能模型性能的上限被大数据的质量与规模直接决定,高质量、大规模的数据是训练出高性能AI模型的基础。其次,必要的基础设施支持被大数据技术为人工智能的工程化落地所提供,AI技术能够在实际场景中得到广泛应用。再次,人工智能进步的关键是数据驱动与算法创新的协同,两者相互促进,共同推动AI技术的不断发展。最后,隐私计算等技术在平衡数据利用与安全保护方面发挥着重要作用,是未来大数据与人工智能健康发展的重要保障。

基于以上结论,为推动大数据与人工智能的协同发展,未来应加强高质量数据资源建设,更优质的“原材料”为AI模型所提供;数据要素市场化配置被推进,数据的流通和利用效率被提高;数据安全治理体系被完善,数据的安全使用被保障;跨领域数据融合应用被促进,AI技术的应用场景和价值被拓展。通过这些举措,大数据与人工智能的深度融合和创新发展有望实现,为社会经济的进步做出更大贡献。

参考文献

- [1] 贺俊宾,龙建伟.数智时代的企业商业秘密保护——基于大数据与人工智能的战略探索[J].中国价格监管与反垄断,2025,(07):44-47.
- [2] 杨业凯.“智能机器人及控制”课程教学改革[J].纺织服装教育,2025,40(01):46-50+69.DOI:10.13915/j.2095-3860.2024.0240.
- [3] 汪雨菲.基于大数据技术的机器人自动化控制系统设计与实现[J].科学技术创新,2025,(08):77-80.
- [4] 杨朗悦.产业智能化与企业出口国内附加值[D].浙江大学,2023. DOI:10.27461/d.cnki.gzjdx.2023.001259.