 DOI: https://doi.org/10.12349/tie.v2i10.8454

Analysis of Innovation and Application Practice of Big Data Privacy Protection Technology

Yang Dai[1] Zhendong Niu[1] Yicheng Zhao[1] Tao Li[2]

1. Guizhou Digital Technology Co., Ltd., Guiyang, Guizhou, 550002, China
2. Guiyang Block Data City Construction Co., Ltd., Guiyang, Guizhou, 550000, China

Abstract

With the rapid development and widespread application of big data technology, data has become a key production factor driving social progress and economic growth. However, privacy leakage issues have become increasingly prominent during large-scale data collection, transmission, sharing, and deep mining, with their severity continuously escalating. This paper aims to systematically analyze data privacy protection technologies in the big data environment, focusing on their innovative achievements and practical applications. The article first elucidates the necessity and core challenges of big data privacy protection, then delves into technical innovations and practical implementations. It provides in-depth analysis of cutting-edge technologies such as differential privacy, federated learning, homomorphic encryption, and data desensitization, explaining their principles and characteristics. By examining typical industry scenarios including finance, healthcare, and government services, the paper evaluates their implementation effects and strategic value. Finally, it summarizes and prospects future development trends of big data privacy protection technologies, aiming to provide references for theoretical research and practical exploration in related fields.

Keywords

big data; privacy protection; differential privacy; federated learning; homomorphic encryption

大数据隐私保护技术的创新与应用实践分析

代杨[1] 牛振东[1] 赵一丞[1] 李涛[2]

1. 多彩贵州数字科技股份有限公司，中国·贵州 贵阳，550002
2. 贵阳块数据城市建设有限公司，中国·贵州 贵阳 550000

摘　要

伴随大数据技术的飞速发展与广泛应用，数据已成为推动社会进步与经济增长的关键生产要素。然而，在数据的大规模采集、传输、共享及深度挖掘过程中，隐私泄露问题日益凸显，其严峻性持续升级。本文旨在系统剖析大数据环境下的数据隐私保护技术，聚焦其创新成果与实践应用。文章首先阐明大数据隐私保护的必要性及核心挑战，进而从技术革新与应用实践两个维度展开深入探讨：重点解析差分隐私、联邦学习、同态加密及数据脱敏等前沿技术的原理与特性，并结合金融、医疗、政务等典型行业场景，分析其落地效果与战略价值。最后，文章对大数据隐私保护技术的未来发展趋势进行总结与展望，以期为相关领域的理论研究与实践探索提供参考。

关键词

大数据；隐私保护；差分隐私；联邦学习；同态加密

1 引言

大数据技术通过海量数据的采集、存储、挖掘与分析，为商业智能、社会治理、科学研究等领域带来了前所未有的发展机遇。然而，在数据价值释放的过程中，个人隐私数据、敏感数据被过度采集、滥用和泄露的风险日益凸显，成为制约大数据产业健康可持续发展的关键障碍。传统隐私保护手段（如匿名化处理、访问控制等）在面对大数据的关联分析能力与重识别攻击时，往往难以有效抵御隐私泄露威胁。在此背景下，如何在充分释放数据价值的同时，构建切实可行的个人隐私、敏感数据的安全防护体系，已成为学术界与业界共同关注的焦点。本文基于这一现实需求，旨在系统梳理当前主流的隐私保护技术创新方向，结合典型应用场景分析其实施路径与效能，探索数据价值挖掘与隐私安全防护之间的动态平衡机制。

【作者简介】代杨（1981–），男，土家族，中国贵州贵阳人，工程师，从事政务信息化、大数据、隐私计算、公数运营等研究。

2 现状描述

近年来，我国大数据隐私保护的法律法规体系不断完善，形成了以《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》为核

心的“三驾马车”，并出台了《网络数据安全管理条例》《个人信息出境标准合同办法》等配套法规，明确数据处理者的主体责任、数据分类分级保护等要求。

随着大数据的广泛应用，隐私保护技术不断创新，其中隐私计算（Privacy-Preserving Computation）成为解决数据孤岛与隐私保护矛盾的关键技术。

3 大数据隐私保护的核心技术创新

大数据环境下的隐私保护技术已从传统的数据加密与匿名化手段，逐步演进为一系列更契合复杂计算场景与多方共享需求的创新解决方案，隐私计算的核心逻辑是在不泄露原始数据的前提下，实现跨主体的数据协作与分析，通过差异化防护机制，为数据隐私安全构建了多层次、动态化的防御体系。

3.1 差分隐私技术

差分隐私作为一种基于严谨数学框架的隐私保护技术，其核心机制在于向查询结果注入符合特定概率分布的随机噪声。该技术确保对于任意可能的查询操作，在仅相差一条个体记录的两个相邻数据集上，其输出结果的概率分布高度近似，难以被有效区分。换言之，无论特定个体是否参与或退出数据集，其对最终发布的统计结果的影响均被控制在极低水平，从而使得攻击者即便掌握除目标个体外的所有辅助信息，也难以通过观察和分析查询结果来可靠地推断出该个体的敏感信息 [1]。

差分隐私的重大创新在于首次为隐私保护提供可严格量化的理论支撑。其核心参数 -- 隐私预算（ε），直接定义了隐私保护的强度：当两个相邻数据集生成的输出分布越接近时，所施加的隐私保护程度越高，需添加的噪声量通常也越大。这种精准的数学关联使数据管理者能明确量化隐私风险与数据效用之间的权衡关系，为差分隐私在统计数据库查询、公共数据发布等场景中的应用提供理论依据。通过构建隐私减损与数据有效性的科学平衡，能在保障个体隐私安全的同时，确保发布数据在宏观层面仍具备较高的统计价值

3.2 联邦学习技术

联邦学习的核心在于突破传统机器学习 " 数据集中、模型训练 " 的固有范式，实现了 " 数据不动、模型动 " 的分布式协同学习架构。在该框架下，多个拥有本地数据的数据持有方（称为客户端）无需向服务器或第三方上传原始数据，仅需在本地利用私有数据独立训练模型，并将计算得到的模型更新（如梯度或权重变化量）进行加密后传输至协调方。

协调方（通常为中央服务器）负责收集来自各客户端的模型更新，并通过特定聚合算法（如联邦平均算法）将分散的更新整合和和优化为一个统一的全局模型。该全局模型经优化后重新分发至各参与方，开启新一轮本地训练，形成 " 本地训练 - 全局聚合 - 模型迭代 " 的闭环循环。这种工作模式从架构层面彻底阻断原始数据跨本地环境传输的路径，从根本上消除因数据集中收集、传输和存储引发的隐私泄漏风险。

3.3 同态加密技术

同态加密技术是密码学领域的里程碑式突破，其核心在于支持对加密状态下的数据进行特定代数运算 -- 无论是基础的加法、乘法运算，还是复杂的函数计算，经解密后的结果与直接在原始明文上执行相同运算的结果完全一致。这一特性使得数据处理方在完全无法接触明文的前提下，仍能完成有意义的计算任务，真正实现了 " 数据可用不可见 " 的安全范式，为破解数据隐私保护与价值挖掘之间的固有矛盾提供了革命性技术路径。

早期的全同态加密方案因极高的计算与通信开销而难以实用化，但随着理论算法的持续优化及专用硬件加速技术的突破，正逐步跨越实验室阶段走向实际应用。当前，其处理效率与资源消耗已取得显著进步，为规模化落地创造了现实条件。最具潜力的应用场景包括云计算环境下的安全外包计算（用户可将加密数据委托给云服务商处理而无需泄露任何敏感信息），以及多方参与的联合数据分析（实现隐私保护下的协同建模与知识发现）。这充分表明，同态加密正在为构建兼顾隐私安全与数据价值释放的新型计算生态奠定坚实基础 [2]。

4 大数据隐私保护技术的应用实践

4.1 金融风控领域

在金融领域，数据驱动已成为业务发展的核心引擎，尤其在风险管控与精准营销等关键环节中，跨机构数据共享与联合建模的需求日益迫切。但客户隐私保护法规与商业秘密要求，使得银行间及与外部公司间的原始数据直接交换面临重大合规挑战。因此，联邦学习技术通过 " 数据可用不可见 " 的加密计算机制，有效破解跨机构数据共享的合规困境。以商业银行反欺诈场景为例，多家机构可基于联邦框架构建联合风控模型：各参与方仅上传本地客户交易数据的梯度更新信息（如特征权重调整参数），通过加密通道进行聚合优化。可在确保用户账户信息与交易细节始终保留在本地的前提下，显著降低数据泄露风险，大幅提升金融网络对新型欺诈模式的识别与防御能力。如江苏银行与腾讯安全合作的实践表明，该技术使信用卡反欺诈模型的 K-S 值提升 25.1%，同时实现黑灰产特征变量在加密状态下的安全融合。

在金融数据开放场景中，差分隐私技术通过向统计结果中注入经过精密数学计算的随机噪声，实现统计信息披露与隐私保护的量化平衡。如中国人民银行向研究机构或公众发布宏观统计数据时，系统自动向统计结果添加符合 (ε, δ)-差分隐私约束的高斯噪声，确保任何个体客户的敏感交易行为或身份信息无法从汇总数据中被逆向推导。

4.2 医疗健康领域

医疗数据作为高度敏感的个人信息，其安全共享机制

对推动精准医学研究及创新药物研发具有关键意义。针对基因数据隐私保护需求，同态加密技术提供了一种创新解决方案：该技术允许研究人员在加密状态下直接对基因序列执行统计分析、对比运算、突变检测及相关性研究等操作，全程无需接触原始明文数据。这一机制在切实保障患者隐私权益的同时，为跨机构、跨地域的基因数据协同分析提供了技术支撑，既显著降低数据泄漏风险，又加速遗传疾病研究与个性化治疗方案开发的进程 [3]。

在临床诊疗数据应用领域，联邦学习技术通过分布式模型训练模式实现多方数据价值的整合利用。各医疗机构可基于本地存储的患者电子病历，在不共享原数据的前提下，共同优化疾病诊断模型与预后预测系统。该框架仅需交换加密后的模型参数更新信息，而非直接传输诊疗记录，从而在数据物理隔离的状态下实现知识聚合。这一模式既严格遵循 HIPAA 等法规对患者健康信息的保护要求，又有效打破医疗机构间的数据壁垒，为构建更精准的医疗人工智能模型提供合规高效的实现路径。

4.3 智慧政务领域

在数据驱动决策的背景下，政府部门在利用人口、经济等大规模数据资源进行公共政策分析和服务提升时，面临着数据公开与隐私保护的双重挑战。差分隐私技术通过向统计查询结果或数据集中添加精确计算的随机噪声，有效防止从宏观数据中逆导识别特定个体或家庭信息的风险，成为人口普查报告、区域经济统计等官方数据发布的核心技术。该技术既保障数据的统计效用，为政策制定提供可靠依据，又从根本上切断从汇总信息反推个人身份的可能性。

为赋能公共数据运营，公数运营平台创新采用联邦学习与差分隐私的协同防护机制，在传统数据脱敏技术基础上构建起 " 数据可用不可见、价值可控可溯源 " 的多层次安全体系。该技术通过联邦学习的分布式架构确保原始数据不出域，同时利用差分隐私在模型参数传输阶段注入可控噪声，有效抵御模型逆向攻击与数据重识别风险。在保留数据核心分析价值的前提下，将个人隐私泄露风险降至最低，为公共管理决策、学术研究创新及社会应用开发提供了兼具安全性与可用性的数据支撑。

5 大数据隐私保护技术的发展趋势与展望

伴随技术进步和应用深化，大数据隐私保护技术正呈现多维度、融合化与标准化的发展趋势，其核心演进路径可归纳为以下三个层面。

5.1 技术融合与协同增强：构建体系化防护体系

单一隐私保护技术难以应对复杂多变的数据安全风险，技术融合成为突破方向。联邦学习虽能实现数据 " 可用不可见 "，但仍面临模型逆向攻击等安全漏洞；而差分隐私通过可控噪声注入可有效抵御此类攻击。二者协同应用可在参数传输阶段构建双重防护机制，显著提升系统安全性。此外，同态加密与安全多方计算的结合，能够在密文状态下实现多方安全计算，在保障数据安全的同时兼顾计算效率。这种技术协同不仅提升整体防护能力，更推动隐私计算从单点防御向体系化防护的跨越，为跨机构、跨地域的数据协作提供可靠技术支撑。

5.2 性能优化与实用化突破：破解 " 效率 – 安全 " 悖论

尽管全同态加密等前沿技术虽具备理论安全性，但计算与通信开销过大制约了应用成效。未来研究将聚焦三大方向：一是研发轻量级加密算法与近似计算策略，在满足基本隐私需求的同时提升处理速度；二是开发专用硬件加速芯片，通过异构计算优化资源分配；三是构建可调节的隐私 - 效率平衡机制，实现安全性与实用性的动态适配。结合边缘计算等新型架构，这些优化将推动隐私保护技术向物联网、实时推荐系统、医疗诊断等高时效性场景渗透，完成从 " 技术可行 " 到 " 应用好用 " 的跨越。

5.3 标准规范与生态构建：完善制度保障体系

隐私保护技术的健康发展需要技术、法律与生态的协同推进。全球范围内，GDPR、国内《个人信息保护法》等法规体系正倒逼企业采用合规技术方案。行业亟须建立统一的隐私度量标准、可审计的技术评估体系及跨系统互操作协议。随着国际监管政策逐步接轨和技术认证体系完善，正在形成 " 政府引导 - 企业实践 - 学术支撑 " 的协同生态，涵盖技术研发、合规运营、伦理审查等全链条，为隐私保护技术的大规模应用奠定制度基础与社会信任根基。

6 结语

综上所述，大数据时代的隐私保障是一项兼具复杂性与长期性的系统工程。本文系统剖析了以差分隐私、联邦学习和同态加密为代表的创新性隐私防护技术，这些技术从数据可用性、安全性及计算效率等维度，为平衡数据价值挖掘与隐私保护提供多元化解决方案。在金融风控、医疗数据共享、政务协同等领域的实践表明，相关技术已从理论探索迈入规模化应用阶段，并展现出显著的社会经济效益。

参考文献

[1] 张晗,谢鹏,武兰.人工智能技术应用中的数据隐私保护技术研究[J].互联网周刊,2025,(20):21-23.

[2] 辛翠平.大数据环境下的隐私保护与数据脱敏技术研究[J].网络安全和信息化,2025,(09):121-123.

[3] 邱景,赵笑尘,胡徐茜,张丽杰.大数据医疗时代的人工智能与患者个人隐私保护的技术方案研究[J].国外电子测量技术,2025,44(07):294-300.