教育大数据隐私保护机制与技术研究

乐洁玉[1]，罗超洋[2]，丁静姝[3]，李卿[2]
1. 华中师范大学国家数字化学习工程技术研究中心，湖北 武汉 430079；
2. 华中师范大学教育大数据应用技术国家工程实验室，湖北 武汉 430079；
3. 华中师范大学法学院，湖北 武汉 430079

摘要
随着大数据技术在教育领域应用的不断深入，教育数据采集、分析及共享等环节可能带来的个人隐私安全等潜在风险不容忽视。如何保障教育数据安全，对教育数据进行合理、合规的分析和利用是当前亟须解决的问题。基于教育大数据的特征，重点阐明了教育大数据的隐私内涵，围绕教育数据生命周期内各利益相关方的隐私保护需求，提出了教育大数据隐私保护框架，并针对性地梳理了目前可用于教育领域的隐私保护技术体系，以期为教育大数据各个应用环节提供支撑，推动教育大数据的规范有序发展。

关键词
教育大数据；隐私保护；利益相关方；数据生命周期；隐私保护技术

中图分类号：G40-057 文献标识码：A doi: 10.11959/j.issn.2096-0271.2020054

Research on privacy protection mechanism and technology of educational big data

YUE Jieyu[1], LUO Chaoyang[2], DING Jingshu[3], LI Qing[2]
1. National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China
2. National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan 430079, China
3. Law School of Central China Normal University, Wuhan 430079, China

Abstract
With the deepening application of big data technology in the field of education, the potential risks such as personal privacy security that may be brought by the collection, analysis and sharing of educational data can not be ignored. How to ensure the safety of educational data, reasonable and compliance analysis and utilization of educational data are the urgent problems to be solved. Based on the characteristics of education big data, the privacy connotation of education big data was focused on. The privacy protection framework of education big data was put forward around the privacy protection needs of stakeholders in the life cycle of education data. And the current privacy protection technology system that can be used in education field was combed, so as to provide support for various application links of education big data, promote the standardized and orderly development of education big data.

Key words
educational big data, privacy protection, stakeholder, data life cycle, privacy protection technology

1 引言

作为教育领域的基础性战略资源，教育大数据为教育管理者制定教育决策提供了科学依据，为教育创新和变革提供了巨大推动力。与此同时，在教育大数据的共享和挖掘过程中，数据的敏感性不可避免地给教育数据的应用发展带来了诸多障碍。教育部印发的《教育信息化2.0行动计划》指出，要深入应用教育大数据助力教育教学，同时需要重点保障师生数据安全，加强隐私保护。

国外学者较早关注教育数据的道德隐私问题，Slade S等人[1]关注数据主体的知情同意权，提出6项原则指导教育数据的采集和使用过程；Daniel B K[2]认为需建立使用教育数据的全球伦理和道德义务，学习分析研究中必须获得学生的“知情同意”，并考虑数据所有权和访问权。在教育数据治理上，李青等人[3]认为应从组织架构、业务领域、技术和平台3个方向推进教育大数据的治理框架；彭雪涛[4]也针对美国圣母大学、麻省理工学院和纽约大学数据治理的实例，提出应正确识别数据的利益相关方，从顶层设计规划，全面落实各方的权责机制，确保信息安全技术的支撑，推进教育数据的有效治理。而在技术层面上，学者们更加关注隐私保护机制改进，Gursoy M E等人[5]提出学习分析过程中的隐私保护机制，将匿名和差异隐私两种大数据隐私保护技术运用到教育领域，解决教育数据发布和挖掘中的隐私泄露问题；Askinadze A等人[6]则针对教育领域内数据挖掘算法的透明度进行了优化，让学生可自由选择数据存储及与第三方共享时的信息内容，从而尊重学生的数据隐私。可见，国内外学者主要从技术支撑、组织管理、伦理法律三大部分探讨教育大数据的安全与隐私保护问题，他们普遍认为结合管理和技术手段对教育大数据进行隐私保护十分必要。因此，亟须研究教育大数据隐私保护机制，以防止学生隐私信息泄露和学习分析技术滥用等事件的发生，规范教育大数据的应用过程和边界。

本文重点围绕教育大数据的隐私保护内涵、框架以及技术展开研究，旨在为教育大数据的有效应用提供隐私保护机制支撑和技术支持。

2 教育大数据的特征与隐私内涵

教育大数据涉及庞大规模的受教育者与教育者群体，对于这些人群，尤其是对于大量的未成年学生而言，隐私保护至关重要。明确教育大数据不同于一般大数据的独特性，厘清其隐私内涵，是推进教育大数据隐私保护的基础。

2.1 教育大数据的特征

作为大数据的一个子集，教育大数据广义上泛指一切与教育活动相关的行为数据，而狭义上指学习者的行为数据[7]。根据教育数据的来源，一般可将教育大数据分为教学数据、数字资源、管理数据、生活数据、其他领域数据5个类型。教学数据来源于不同的教学活动，如教研活动、户外活动、课程教学、户外教学等；数字资源包括多媒体素材、在线课程、学科工具等；管理数据涉及学生、家长、学校、其他机构等不同数据主体的数据；生活数据涵盖图书借阅、健康运动、社交、娱乐等数据；而其他领域数据渗透到医疗、经济、就业、市政等生活的各个方面。可以看出，教育大数据来源多样，其采集和存储阶段汇聚了各种不

同类型和信息源的数据，包括非结构化、半结构化和结构化数据。总体而言，教育大数据具有层级多、维度高、跨度长等特性。

（1）层级多

教育大数据范围宽广，可分为教育管理五层级（即国家、区域、学校、班级、个体）、学习结果六层次（即识记、理解、应用、分析、综合和评价）、学习资源多粒度（如选项、题目、试卷、知识单元、课程等）、数据敏感度分级（即高、中、低），数据层级从上至下、从高到低逐步汇聚。

（2）维度高

教育涉及教学、管理、生活、服务等方面，包含学校、家庭和社会多个场景，以培养全面发展的人为核心。因此，教育大数据是数据类型多样、教育场景复杂、核心素养繁多的高维度数据集合。

（3）跨度长

教育大数据跨越学前教育、初等教育、高等教育、成人教育、终身学习5个阶段，是面向所有人、学习全过程的数据。

2.2 教育大数据的隐私内涵

（1）教育大数据中的隐私和安全问题

个人数据隐私与个人数据保护密切相关，无论在何时何地采集、存储或使用数据，都可能出现隐私问题。大数据环境下，隐私的存在空间从现实扩展到数据，但内容上仍是个人的、私人的、不愿被公开知晓的活动、信息及空间。

教育大数据的隐私保护问题可被认为是保护学习者可识别行为、内容等个人敏感信息安全。强调合理适当地使用和管理学习者的数据，在未经数据主体同意时，数据拥有者不得将学生信息出售或共享给第三方，并且需保证以正确的方式采集、共享和使用学生个人信息，使学习者的隐私权免受其他方侵害[8]。

（2）教育大数据下的学习者隐私特征

Rao P R M等人[9]认为大数据分析行为存在监视（通过技术手段持续收集用户行为）、披露（不可信的第三方识别用户敏感信息）、歧视（对私人信息产生偏见）和滥用（信息推送）隐私的威胁。而学习者个人的敏感信息涉及学习记录、考试测评等与教学活动直接相关的信息，也包含健康状况、家庭信息等学生管理数据，还包括餐饮消费、上网情况等学生在校园生活中产生的其他敏感数据。教育大数据的分析和挖掘也存在隐私泄露和滥用的风险。

一方面，教育大数据处理技术的应用可为学习者提供个性化服务，但在分析和挖掘海量、零碎教育大数据的过程中，学生个人隐私存在泄露风险。尤其是传感器等智能设备采集到的学习者人脸、体征等可识别学习者个人行为的敏感信息，具有独特性和不变性，一旦出现数据泄露和滥用的行为，将可能影响学习者的人身安全和权益。值得注意的是，当学习者的零碎数据被非法窃取，并进行二次重组关联应用时，会产生具有新价值的学习者数据链，让学习者无时无刻不被“监视”，从而出现学习者隐私披露风险。

另一方面，学习分析技术可增强对学生学习方式和学习目标的理解，表征学生当前的课堂表现，预测学生未来完成课程的成功率，可用的学生数据越多，数据可视化结果越好，学习反馈越及时。但是，机器学习的训练数据在分布上存在一定偏差，若仅仅使用学习者的历史数据，忽视学习者动态的成长过程，很可能为学习者提供固化标签，产生数据偏见，有碍于发掘学习者的发展潜力以及创造力。

尽管教育领域在数据采集、传输、存储和应用阶段有规范的处理措施，但教育行业仍是非常容易受到公开披露的行业之一。然而，实施教育数据的隐私保护措施

仍然是一个非常庞大繁杂的过程，随着大数据技术的不断提升，隐私泄露风险也在不断增加，亟须采取可靠的安全防范措施和隐私保护技术。因此，必须建立一套完整的隐私保护方案，从源头上遏制学生数据隐私泄露的问题，形成隐私保护管理机制，满足对学习者敏感信息使用的合规性要求。

3 教育大数据的隐私保护机制

数据隐私性、真实性、完整性和访问控制是解决大数据安全保护的首要问题[10]。DONG X H等人[11]强调应基于现有的大数据技术，围绕整个数据生命周期考虑解决数据共享和隐私保护之间的问题，否则会危害大数据的应用环境。Salleh K A等人[12]认为在数据的传输、创建和处理过程中数据需被聚合或者匿名，以保证大数据技术应用环境中的隐私安全，而现有数据存储缺乏安全保障能力，必须引起重视。Xu L等人[13]提出了大数据安全模型，综合考虑了数据挖掘过程中不同角色类型（即数据提供者、数据采集者、数据处理者和数据决策者）的隐私保护需求；而Khaloufi H等人[14]提出大数据安全生命周期模型，包括大数据采集、存储、分析处理、知识创造4个阶段，旨在识别大数据各个生命阶段的隐私安全威胁和攻击，保证大数据的生命安全。本文结合教育大数据各利益相关方的隐私保护需求，识别数据的采集、存储、处理和可视化阶段的隐私风险，并提出教育大数据的隐私保护框架，以解决学习者的数据安全和隐私保护问题。教育大数据的隐私保护框架如图1所示。

3.1 面向教育大数据应用过程的利益相关方

教育数据的质量是学习分析与数据挖掘发挥最大价值的基本前提。了解面向教育大数据应用过程中利益相关方的隐私保护诉求，是保证数据质量完整性和价值性的基本保障。一般认为，学生、教育工作者、研究人员、教育机构和政府机构是面向学习分析过程的利益相关方[15]。实际中，利益相关方可能因目标需求不同而

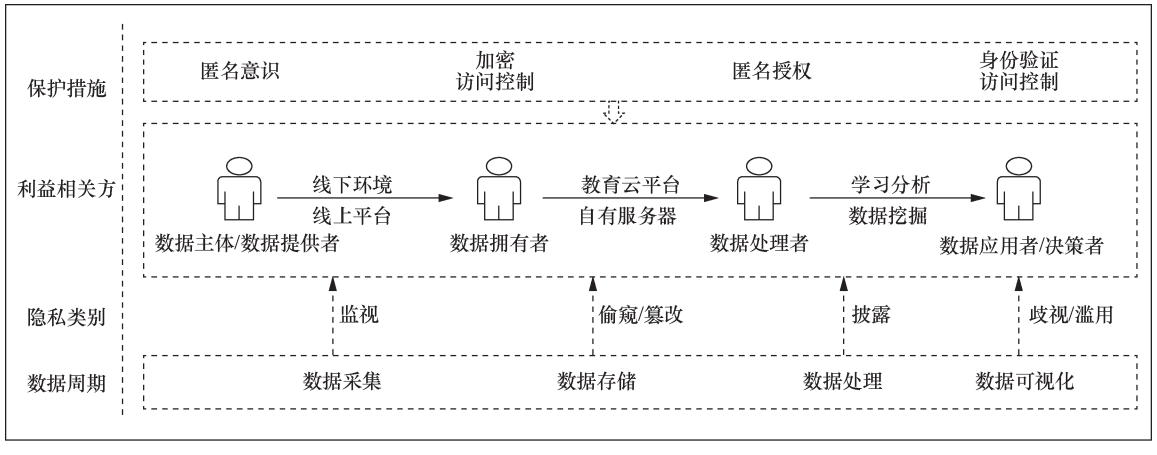

图1 教育大数据的隐私保护框架

出现矛盾、冲突。在教育大数据的应用过程中，围绕数据循环周期，主要参与的利益相关方可被分为数据主体/数据提供者、数据拥有者、数据处理者和数据应用者/决策者。

数据主体/数据提供者指被采集和分析数据的个体（存在潜在的敏感信息）。在教育领域中，学习者是主要的数据主体，而家长、教师和学校的敏感信息也应受到关注和保护，他们的数据一旦被滥用，则难以保证其敏感数据的隐私性，因此，数据主体/数据提供者主要关注所提供数据的敏感程度。

数据拥有者是与数据采集和存储相关的利益主体，包括政府、学校和相关的教育机构，其有责任确保学生数据的隐私性。如果直接公布采集到的原始数据或在数据挖掘之前不采取足够的隐私预防措施，学习者的敏感信息可能会被披露，因此，有必要对采集到的原始数据信息进行修改转换等隐私保护操作，防止其被恶意推断和修改。

数据处理者是有权访问学习者数据的系统设计人员、分析人员，也包括教育数据的管理人员。数据处理的目的是向数据应用人员提供有用的信息，需要采用强大的隐私保护挖掘和隐私保护学习分析算法来提取学习者的敏感信息，防止学习者敏感信息未经批准而被使用产生的披露行为，同时保留原始数据的客观性。

数据应用者/决策者包括教育管理人员、教师等所有有权使用数据的人员。

在教育领域中，学习者是整个教育系统中个人数据产生的主体和源头，如果个人缺乏对数据进行直接控制的意识，会导致数据过度滥用的潜在危险。教师、相关教育机构等其他面向教育大数据应用过程的利益相关方兼具数据拥有者、数据处理者、数据应用者/决策者多个角色。因此，对于隐私保护措施的制定来说，应考虑各个角色的隐私保护诉求，同时权衡各方的利益冲突，以实现教育大数据效用最大化。

3.2 基于利益相关方的教育大数据生命周期隐私保护框架

（1）教育数据采集阶段

数据采集是控制师生敏感信息泄露的源头。课堂、校园等线下学习环境仍然是师生数据采集的主要场景，从结构化学习环境（如智能导师系统）到越来越开放式的在线学习平台（如慕课网站），再到泛在学习空间，教育数据的采集内容和采集方式更加多样、实时和全面，教育数据多源异构，非结构化的教育数据越来越占据主导地位。从隐私安全的角度来看，可靠的数据源是数据采集的关键。因此，数据采集必须加强数据主体的隐私匿名意识，数据主体在合法受用教育信息化便利的同时，也要防止他人非法访问和窃取自己的敏感信息，确保数据隐私安全。

（2）教育数据存储阶段

数据存储阶段的授权访问应在不识别个人身份的敏感信息的前提下进行，并保证数据不被泄露和篡改。数据拥有者采集到数据后，需保证数据的完整性和客观性，利用相关隐私保护技术，对敏感数据进行脱敏、清洗、转换等预处理。除了高校和教育机构自有的服务器外，第三方教育云平台也是数据存储的另一选择方式。这一阶段中，未经授权的数据访问行为和基于数据挖掘的攻击行为是常见的挑战，需采取数据加密、访问控制等必要的隐私保护手段，并且数据拥有者应承担隐私信息泄露的主要责任，确保数据不被攻击篡改。

（3）教育数据处理阶段

数据处理阶段是教育大数据应用的中心环节，目标是及时识别并剔除异常数据。

在此阶段，数据挖掘技术和学习分析技术不仅能对学习者的数据进行分析和处理，而且经过分类、预测、聚合关联规则等操作，还能预测学习趋势，生成学习行为模型，有效检测到异常数据，并及时剔除。要防止数据处理过程中个人信息被识别和恶意提取敏感信息的行为，必须保证只有获得授权的数据处理者才可以从数据库中提取信息，将数据泄露与篡改的风险降到最低。另外，*k*-匿名（*k*-anonymity）、*l*-多样化（*l*-diversity）、*t*-贴近性（*t*-closeness）等匿名技术可隐藏识别数据主体的敏感信息，增强教育数据的隐私性。

（4）教育数据可视化阶段

数据可视化阶段的目的是更好地应用数据分析的结果，为数据决策者的行为提供科学依据，以便对学习者的行为活动进行有效干预和规划。如教育机构根据分析结果进行教学评价和决策，构建学生感兴趣的学习环境；教师可根据学习者数据增强教学实践，实时调整教学内容。但数据分析的结果（如学习者的教学评价、社交轨迹）可被认为是敏感信息，在教育数据的实际使用过程中，不透明的数据会导致数据滥用和歧视现象，影响学生身心发展，敏感数据并不会对外公布。差分隐私、安全检索及访问控制技术可保障学习资源的开放和共享。

4 教育大数据隐私保护技术

现有的隐私保护技术以数据的匿名化为主，加密、差分隐私、安全检索等是常用的关键技术[16]，数据生命周期的不同阶段涵盖许多隐私保护技术，每一种方法都各有优缺点，见**表1**。随着教育数据的应用场景和结构类型越来越复杂，隐私保护技术的开发成为新的研究热点。

4.1 数据存储

数据存储安全技术主要有数据加密和安全多方计算等，其中，数据加密包括静态数据加密和动态数据加密两种。

静态数据加密技术有对称加密、非对称加密（公钥加密）和混合加密3类[17]。对称加密算法适用于数据量小的数据加密，其安全性与密钥长度、算法轮次有关，算法效率高但安全性较低，且不具有可认证性和不可抵赖性，现用的算法主要有高级加密标准（advanced encryption standard，AES）、数据加密标准（data encryption standard，DES）等。公钥加密能够适应交互式环境，其安全性与其所基于的数学难题有关，主要算法包括RSA（基于大整数因子分解问题）、ECC（基于椭圆曲线离散对数问题）[18]。混合加密是对称加密和公钥加密两种方法的结合，先快速对数据进行对称加密，再进行公钥加密。

动态数据加密主要采用同态加密（homomorphic encryption，HE），关键技术是全同态加密（fully-homomorphic encryption，FHE）。同态加密技术能够在加密的环境下处理数据，但其计算复杂度较高，导致效率较低。目前的全同态加密技术主要基于R-LWE问题进行研究[19]。

安全多方计算是指多名参与者共同安全计算某个约定函数，每名参与者除了自己的输入和输出及可推断的信息，无法得到任何额外的信息。常用的安全多方计算协议有4类：基于健忘传输（oblivious transfer，OT）的安全多方计算协议、使用可验证秘密分享（verifiable secret sharing，VSS）的安全多方计算协议、基于同态加密的安全多方计算协议、基于Mix-Match的安全多方计算协议。但这些协议还需更细致的研究和应用实现[20]。

表1 基于数据生命周期的隐私保护技术对比

生命周期	相关技术		特点	教育应用场景
数据存储	静态数据加密	对称加密	适用于数据量较小的数据的加密，效率高，但安全性低	教育信息业务系统数据管理、教育信息系统文件安全保障方案、交互式环境下的共享安全
		公钥加密	能够适应交互式环境，效率较低，但安全性较高	
		混合加密	高效且可保证较高的安全性	
	动态数据加密	类同态加密 全同态加密	在不解密的情况下，能对数据进行处理，但计算复杂度高，致使效率较低	
	安全多方计算	基于OT的安全多方计算协议 基于VSS的安全多方计算协议 基于同态加密的安全多方计算协议 基于Mix-Match的安全多方计算协议	理论上安全，但缺乏细致研究和实现	
数据处理	数据匿名化	*k*-匿名 *l*-多样化 *t*-贴近性	安全性较高，但会出现一定程度的信息损失	教育数据脱敏
		m-invariance，HD-composition算法	填补了经典方案仅适用于静态关系数据的局限，保证数据在非静态发布时的隐私安全	
数据应用	差分隐私技术		不局限于对抗性背景知识，使数据的公开量有限，但关键参数ε难以控制	教育数据发布、学习数据分析隐私保护
	区块链技术		能够去中心化，支持匿名交易，但安全性受算力影响	在线学习系统的数据安全与隐私保护
	安全检索技术	对称密文检索	安全性强，运算效率较高	教育信息系统文件安全保障
		非对称密文检索	功能性强，但效率较低	
	授权和访问控制技术	基于属性的授权与访问控制	可以实现细粒度的授权和访问控制，但是定义授权规则的操作困难、烦琐	在线学习和在线考试系统的安全保障
		基于角色的授权与访问控制	集成效率高，但安全管理员需具备多领域的专业知识	
		密码访问控制	效率高，开销小，但很难解决不同用户的密钥管理和分发问题	
		自主访问控制	权限管理复杂度爆炸式增长，使得执行困难	

上述技术的应用场景有教育信息业务系统数据管理、教育信息系统文件安全保障，以及交互式环境下的共享安全[21]。

4.2 数据处理

数据匿名化是数据处理的关键安全技术，主要用于数据脱敏。

经典的数据匿名化技术[22]有：*k*-匿名、*l*-多样化、*t*-贴近性。*k*-匿名模型在发布关系型数据时，要求每一个泛化后等价类最少包含*k*条相互难分辨的数据，它未对等价类中的敏感属性进行约束，可被两种手段攻击（同质攻击和背景知识攻击）；*l*-多样化在对关系型数据进行匿名处理时，会确保每个等价类至少包含*l*个不同的敏

感数据值，这可以防止同质攻击，但忽视了敏感属性的全局分布，可能遭受类群攻击；t-贴近性模型要求所有等价类中敏感数据值的分布与该属性的全局分布保持一致，t-贴近性通过敏感属性计算得出，该方法可以保证数据的公开，但是不能保证每次数据的合理分布，算法时间复杂度高，不适用于高实时性场景，且对数据价值有一定的破坏。

此外，m-invariance[23]和HD-composition[24]算法弥补了上述方法仅适用于静态数据的不足，其他数据匿名化技术还有随机化技术[25]、p-敏感匿名等。

教育数据内含有大量的敏感数据和隐私数据，数据匿名化技术能很好地解决教育数据脱敏问题。

4.3 数据应用

教育数据的应用层面广泛，主要的安全保障技术有差分隐私技术、区块链（blockchain）技术、安全检索技术、授权与访问控制技术等。

差分隐私技术是通过随机化处理，根据用户自行指定的参数ε在数据中加入噪声，从而决定隐私保护程度及数据失真损失程度的技术。差分隐私技术改善了数据匿名的不足，不局限于对抗性的背景知识，可保证大部分数据不会被攻击者看到，而且公开的信息在理论上是有限的，故而差分隐私技术比数据匿名化技术更能防止数据隐私的泄露。但是在该技术的实现过程中，控制隐私保护与数据失真程度的关键参数ε难以人为控制[25]。

区块链是一种将区块以链的形式聚集在一起的数据结构，具有去中心化、按时序记录数据、集体维护、可编程和安全可信等优势[26]。它能够防止网络窃听，同时能够实现匿名交易，而且基于去中心化的特点，其对网络攻击有较好的应对。但是区块链也面临许多安全威胁，如其节点容易遭受攻击，同时由于其具有关联性，在算力足够大的情况下，其安全性难以保障[27]。尽管有所不足，但是区块链在教育中的应用范围仍很广泛，如在线学习系统中，利用区块链可以对学习记录进行分布式存储，提供具有可信性高、计算成本低的学习证书系统，或者进行去中心化知识库的搭建等[28]。

近年来，安全检索技术聚焦于探索密文检索技术，以实现在密文数据上的直接检索操作。密文检索技术可被分成对称密文检索和非对称密文检索。对称密文检索技术中只有数据拥有者拥有密钥，并提交敏感数据，故而数据拥有者就是数据检索者，这使得该技术更适用于单用户的情形，具有安全性高、加密、搜索运算效率高的特点。具体的实例有基于全文扫描的方法、基于文档-关键词索引的方法、基于关键词-文档索引的方法等[28]。而非对称密文检索主要采取非对称密文关键词检索（public key encryption with keyword search，PEKS）方案，任何可以获得数据检索者公钥的用户都可以提交敏感数据，但是只有拥有数据检索者私钥的用户才可以生成陷门，因此更适用于多用户的情形，算法功能强，但与哈希函数和分组密码运算相比，效率较低。经典实例有BDOP-PEKS方案[29]、KR-PEKS方案[30]、DS-PEKS方案等。在教育应用方面，该技术主要用于教育信息系统文件安全保障。

授权与访问控制技术各有优劣。基于属性的授权与访问控制能够实现细粒度的授权与访问控制，但以非常细的粒度为每个用户定义授权规则是困难和烦琐的，且难以同时保证系统的访问效率和可用性。基于角色的授权与访问控制具有较高的集成效率，但是安全管理员一般不具有足够丰富的多领域知识来精确定义和授权管理角

色。密码访问控制可分为基于密钥管理的访问控制和基于属性加密的访问控制，而该技术目前的主要问题是不同用户的密钥分发与管理问题。自主访问控制在大数据背景下也面临权限管理复杂的挑战，相关访问控制模型的选择与构建亦应联系实际场景，而在教育领域的应用则主要是保障在线学习和在线考试系统的数据安全隐私。

5 结束语

教育大数据的创新应用推动着教学模式、教学评价和教学管理等的全方位变革。然而，在教育领域的开发利用过程中，教育大数据的隐私保护策略尚处于探索阶段。当前，教育大数据的发展应用仍面临隐私保护机制不完善、数据开放共享机制未形成、大数据安全技术和平台发展支撑技术待突破等挑战。

围绕隐私保护机制问题，当前仅仅基于整个数据生命周期的隐私安全引入，或根据不同利益角色的保护诉求展开。相较于传统的教育数据，教育大数据覆盖的时间跨度更广、汇聚的结构类型更杂、涉及的教育主体更多，原有的单一保护机制已无法满足教育大数据的动态性需求。因此，本文基于教育大数据的生命循环，平衡利益相关方的价值冲突，增加了教育大数据应用的合规性、透明性和可靠性。

针对隐私保护技术问题，由于教育数据的应用场景更纷繁复杂，教育信息系统对数据的隐私性要求更严格，大数据隐私保护技术虽已有一定的发展，但解决教育数据安全与隐私问题的研究较为零散，针对性不强。本文根据不同教育应用场景的安全需求，使用相应的隐私保护技术，降低了教育数据应用过程中的风险，从而保障了教育数据质量的完整性、安全性和私密性。

总体来说，针对教育大数据的多源异构特征和数据应用服务的隐私伦理问题，为加强学习者的隐私安全，仍需从以下方面推进教育大数据的有效应用。首先，完善法律法规，从法律上界定、规范公开数据与私有数据的边界，落实教育数据使用主体的责任和权利归属，构建面向教育大数据研究应用的伦理准则，从而为各项数据业务提供依据，以推进所有利益相关方的道德自律；第二，加强技术攻关，通过对用户隐私信息的隐藏或混淆，构建有效的教育数据隐私保护技术体系和平台，以降低数据精确性和数据披露风险，在确保用户隐私信息不可还原和追踪的前提下，满足教育数据研究和应用的要求；第三，制定标准规范，围绕教育大数据采集、分析、应用过程，分层、分类进行规范的顶层设计，秉承相关性、唯一性、清晰性、有效性和易用性等原则[31]，规范教育数据的应用流程；最后，提升利益相关方、数据主体等人的数据素养，通过加大相关宣传力度，强化利益相关方的隐私保护意识与专业知识，提高各责任主体对数据安全的敏感性，增强其辨识能力，以保障数据主体的权利。

参考文献:

[1] SLADE S, PRINSLOO P. Learning analytics: ethical issues and dilemmas[J]. American Behavioral Scientist, 2013, 57(10): 1510–1529.

[2] DANIEL B K. Big data and data science: a critical review of issues for educational research[J]. British Journal of Educational Technology, 2019, 50(1): 101–113.

[3] 李青, 韩俊红. 数据治理: 提升教育数据质量的方法和途径[J]. 中国远程教育, 2018(8): 45–53.
LI Q, HAN J H. Data governance: methods and approaches to improve the quality

of education data[J]. China Distance Education, 2018(8): 45–53

[4] 彭雪涛. 美国高校数据治理及其借鉴[J]. 电化教育研究, 2017, 38(6): 76–81.
PENG X T. Data governance in American universities and its reference[J]. E–education Research, 2017, 38(6): 76–81.

[5] GURSOY M E, INAN A, NERGIZ M E, et al. Privacy–preserving learning analytics: challenges and techniques[J]. IEEE Transactions on Learning Technologies, 2016, 10(1): 68–81.

[6] ASKINADZE A, CONRAD S. Respecting data privacy in educational data mining: an approach to the transparent handling of student data and dealing with the resulting missing value problem[C]// 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises. Piscataway: IEEE Press, 2018: 160–164.

[7] 吴砥, 饶景阳, 王杨春晓. 教育大数据标准化的思考与建议[J]. 计算机教育, 2018(11): 12–15.
WU D, RAO J Y, WANG Y C X. Thoughts and suggestions on education big data standardization[J]. Computer Education, 2018(11): 12–15

[8] ABOUELMEHDI K, BENI–HESSANE A, KHALOUFI H. Big healthcare data: preserving security and privacy[J]. Journal of Big Data, 2018, 5(1): 1.

[9] RAO P R M, KRISHNA S M, KUMAR A P S. Privacy preservation techniques in big data analytics: a survey[J]. Journal of Big Data, 2018, 5(1): 33.

[10] ALMUTAIRI A M, ALBUKHARY R A T, KAR J. Security and privacy of big data in various applications[J]. International Journal of Big Data Security Intelligence, 2015, 2(1): 19–24.

[11] DONG X H, LI R X, HE H, et al. Secure sensitive data sharing on a big data platform[J]. Tsinghua Science and Technology, 2015, 20(1): 72–80.

[12] SALLEH K A, JANCZEWSKI L. Technological, organizational and environmental security and privacy issues of big data: a literature review[J]. Procedia Computer Science, 2016, 100: 19–28.

[13] XU L,JIANG C,WANG J, et al. Information security in big data: privacy and data mining[J]. IEEE Access, 2014, 2: 1149–1176.

[14] KHALOUFI H, ABOUELMEHDI K, BENI–HSSANE A, et al. Security model for big healthcare data lifecycle[J]. Procedia Computer Science, 2018, 141: 294–301.

[15] GRELLER W, DRACHSLER H. Translating learning into numbers: a generic framework for learning analytics[J]. Journal of Educational Technology & Society, 2012, 15(3): 42–57.

[16] 孟小峰, 张啸剑. 大数据隐私管理[J]. 计算机研究与发展, 2015, 52(2): 265–281.
MENG X F, ZHANG X J. Big data privacy management[J]. Computer Research and Development, 2015, 52(2): 265–281.

[17] 方滨兴, 贾焰, 李爱平, 等. 大数据隐私保护技术综述[J]. 大数据, 2016, 2(1): 1–18.
FANG B X, JIA Y, LI A P, et al. Summary of big data privacy protection technology[J]. Big Data Research, 2016, 2(1): 1–18.

[18] 冯登国. 大数据安全与隐私保护[M]. 北京: 清华大学出版社, 2018: 103–112.
FENG D G. Big data security and privacy protection[M]. Beijing: Tsinghua University Press, 2018: 103–112.

[19] 黄霖, 黎源, 汪星辰, 等. 数据自治开放的加密技术挑战[J]. 大数据, 2018, 4(2): 50–62.
HUANG L, LI Y, WANG X C, et al. Encryption technology challenges of data autonomy and opening[J]. Big Data Research, 2018, 4(2): 50–62.

[20] 李强, 颜浩, 陈克非. 安全多方计算协议的研究与应用[J]. 计算机科学, 2003(8): 52–55.
LI Q, YAN H, CHEN K F. The research and application of secure multi–party computing protocol[J]. Computer Science, 2003(8): 52–55.

[21] 刘梦君, 姜雨薇, 曹树真, 等. 信息安全技术在教育数据安全与隐私中的应用分析[J]. 中国电化教育, 2019(6): 123–130.
LIU M J, JIANG Y W, CAO S Z, et al.

Application analysis of information security technology in education data security and privacy[J]. China Educational Technology, 2019(6): 123–130.

[22] MURTHY S, BAKAR A A, RAHIM F A, et al. A comparative study of data anonymization techniques[C]// 2019 IEEE 5th International Conference on Big Data Security on Cloud (Big Data Security), IEEE International Conference on High Performance and Smart Computing (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS). Piscataway: IEEE Press, 2019: 306–309.

[23] XIAO X, TAO Y. M-invariance: towards privacy preserving re-publication of dynamic datasets[A]. ACM, 2007: 689–700.

[24] BU Y, FU A, WONG R C W, et al. Privacy preserving serial data publishing by role composition[J]. Proceedings of the VLDB Endowment, 2008, 1(1): 845.

[25] JAIN P, GYANCHANDANI M, KHARE N. Differential privacy: its technological prescriptive using big data[J]. Journal of Big Data, 2018, 5(1): 15.

[26] 李青, 张鑫. 区块链: 以技术推动教育的开放和公信[J]. 远程教育杂志, 2017, 35(1): 36–44.
LI Q, ZHANG X. Blockchain: promote the openness and credibility of education with technology[J]. Journal of Distance Education, 2017, 35(1): 36–44.

[27] 祝烈煌, 高峰, 沈蒙, 等. 区块链隐私保护研究综述[J]. 计算机研究与发展, 2017, 54(10): 2170–2186.
ZHU L H, GAO F, SHEN M, et al. A review of blockchain privacy protection research[J]. Computer Research and Development, 2017, 54(10): 2170–2186.

[28] CHANG Y C, MITZENMACHER M. Privacy preserving keyword searches on remote encrypted data[C]// International Conference on Applied Cryptography and Network Security. Heidelberg: Springer, 2005: 442–455.

[29] BONEH D, FRANKLIN M. Identity-based encryption from the Weil pairing[J]. SIAM Journal on Computing, 2003, 32(3): 586–615.

[30] HENG S H, KUROSAWA K. K-resilient identity-based encryption in the standard model[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2006, 89(1): 39–46.

[31] 王正青, 但金凤. 大数据时代美国教育数据质量管理流程与保障[J]. 现代远程教育研究, 2019, 31(5): 96–103.
WANG Z Q, DAN J F. American educational data quality management process and assurance in the era of big data[J]. Modern Distance Education Research, 2019, 31(5): 96–103.

作者简介

乐洁玉(1996–), 女, 华中师范大学国家数字化学习工程技术研究中心硕士生, 主要研究方向为教育大数据、学习行为分析。

罗超洋(2000–), 男, 华中师范大学教育大数据应用技术国家工程实验室本科生, 主要研究方向为教育大数据。

丁静姝（1999– ），女，华中师范大学法学院本科生，主要研究方向为民商法学、经济学。

李卿（1982– ），女，博士，华中师范大学教育大数据应用技术国家工程实验室副教授、硕士生导师，主要研究方向为教育科学战略、教育大数据、感知计算。

收稿日期： 2020-09-04

通信作者： 李卿，viven_a@mail.ccnu.edu.cn

基金项目： 国家自然科学基金资助项目（No.61807012）；中央高校基本科研业务费专项资金项目（No.CCNU20QN027）；2019年大学生创新创业训练计划（国家级）（No.20190417016）

Foundation Items: The National Natural Science Foundation of China (No.61807012), The Project of Special Funds for Basic Scientific Research Operating Expenses of Central Universities (No.CCNU20QN027), Innovation and Entrepreneurship Training Program for College Students (No.20190417016)