跨域数据授权运营研究及应用

张纪林[1]，顾小卫[2]，张亦钊[3]，郑小林[3]，陈超超[3]
1. 浙江省大数据发展管理局，浙江 杭州 310007；
2. 中共浙江省委网络安全和信息化委员会办公室，浙江 杭州 310007；
3. 浙江大学计算机科学与技术学院，浙江 杭州 310027

摘要
随着大数据和云计算的发展，数据管理正在打破“数据孤岛”，从面向单域的孤立服务发展到跨域的数据共享与协同服务。基于公共数据授权运营框架，给出了跨域数据授权运营全链路结构，并探讨了跨域数据加工过程中数据隐私和效率的挑战。针对这些挑战，提出了集中式和隐私计算两种数据加工模式，能够在保护数据隐私的同时提高数据加工效率。最后，给出了一个实际场景下跨域数据授权运营平台的应用案例。

关键词
跨域数据管理；数据授权运营平台；集中式计算；隐私计算

中图分类号：TP399 文献标志码：A doi: 10.11959/j.issn.2096-0271.2023050

Research and application of cross-domain data authorization and operation

ZHANG Jilin[1], GU Xiaowei[2], ZHANG Yizhao[3], ZHENG Xiaolin[3], CHEN Chaochao[3]
1. Zhejiang Big Data Development Administration, Hangzhou 310007, China
2. Office of the Zhejiang Cyberspace Affairs Commission, Hangzhou 310007, China
3. College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

Abstract
With the development of big data and cloud computing, data management is breaking down "data silos" and developing from isolated services for single domain to cross-domain data sharing and collaborative services. Based on the public data authorization and operation framework, this paper presented the full-link structure of cross-domain data authorization and operation, and discussed the challenges of data privacy and efficiency in the cross-domain data processing. In response to these challenges, two data processing models, centralized and privacy computing, were proposed, which could improve data processing efficiency while protecting data privacy. Finally, an application case of a cross-domain data authorization operation platform in a practical scenario was given.

Key words
cross-domain data management, data authorization & operation framework, centralized computing, privacy computing

0 引言

随着计算机技术和信息化的发展，大数据驱动的发展模式几乎遍布各行各业，由此也产生了数据管理以及数据安全等相关问题。海量数据由于产生和存储的场景不同，可以划分为公共（政务）数据和社会数据，两者主要区别在于数据是否归于政府或公共服务机构所持有。其中，公共数据包括政府部门持有的各类数据，例如人口、经济统计数据，医疗卫生数据，交通大数据等。公共数据的特点是数据量大、包含丰富的信息，可以被充分利用进行建模分析。但同时公共数据涵盖了多方面信息，其中也包括大量敏感信息，例如公民的收入状况、医疗记录、家庭住址以及交通出行记录等，需要在数据加工和利用过程中充分考虑隐私保护。而社会数据涵盖了电商数据、金融数据、社交平台数据等，主要由相关企业收集和存储。这些数据同样蕴含大量信息，部分还需要与公共数据联合建模分析。

两种数据持有者不同，数据自然划归于不同域。但是，从大数据的加工和进一步挖掘数据价值的角度来看，数据被要求能够在不同机构之间流通共享。由此需要引入跨域数据管理的模式以实现跨域数据联合加工和利用，同时又在一定程度上保护数据隐私。

2020年4月9日，中共中央、国务院发布了《关于构建更加完善的要素市场化配置体制机制的意见》，该意见将数据要素与土地要素、劳动力要素、资本要素等传统生产要素一并列为完善要素市场化配置的关键因素[1]。该意见指出，要在国家数据分类分级保护制度下，推进数据分类分级确权授权使用和市场化流通交易。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》也明确提出“开展政府（公共）数据授权运营试点”，鼓励第三方深化对公共数据的挖掘利用。政府（公共）数据授权运营是指授权特定的市场主体，在保障国家秘密、国家安全、社会公共利益、商业秘密、个人隐私和数据安全的前提下，开发利用政府部门掌握的与民生紧密相关、社会需求迫切、商业增值潜力显著的数据。与公有云运营相比，公共数据授权运营虽然也强调运营，但两者面向不同的具体需求，公有云通常指第三方提供商为用户提供的能够使用的云，这种云有许多实例，可在整个开放的公有网络中提供服务[2]（主要是计算服务）。而政府（公共）数据授权运营旨在提供合规的政府（公共）数据服务。在授权运营平台建立的基础上，可以进一步实现社会方授权运营单位和授权运营平台联合数据加工的过程。为贯彻这一要求，浙江省推出了建设公共数据平台授权运营域的方案，通过一体化、智能化公共数据平台，实现安全有序的公共数据授权运营工作，推动数据要素安全合规地流通。

由于公共数据和社会数据分域存储，在建设公共数据授权运营平台的过程中，公共数据主管部门将数据授权给运营单位，授权运营单位在进行多方数据联合加工时，不可避免需要解决跨域带来的新问题。这些问题主要分为两个方面：在跨域数据加工过程中，如何保证隐私信息不泄露；在提供数据隐私保护的同时，如何保证跨域数据加工的效率。基于对以上论述的总结，本文的主要内容分为3点：

- 探讨跨域数据授权运营链路中数据加工过程面临的两方面挑战；
- 探讨集中式计算和隐私计算两种技术模式，以及各自如何解决上述挑战；
- 给出跨域数据加工模式的应用案例。

1 相关工作

1.1 跨域数据共享和应用的相关工作

跨域数据共享和应用的场景较为丰富。目前，针对这一领域有较多的现有工作。这些工作涵盖了多个应用场景，包括电子政务、推荐系统、医疗信息处理、城市治理以及作战联合分析等。邓磊等[3]分析了构成电子政务应用平台支撑之一的数据交换平台功能受到制约的关键因素及现有的解决办法，建立并形式化了域标识和域关系模型，并基于这一模型设计了可信数据交换模型框架。张彬等[4]从多个角度对跨域推荐中的知识融合研究进展进行了梳理和总结。卢红建等[5]探索了跨域医疗信息共享与业务协同标准建设，对未来其他地域间开展跨区域共享联动，以及形成完整的就诊数据链信息具有一定的参考价值。董晶等[6]针对当前城市数据治理的跨域安全、可信互通和共享问题，并基于复杂系统论提出了多视域城市跨域数据治理体系。王蒙蒙等[7]基于联邦学习技术，探索了一种面向联合作战场景下的跨域数据安全互联方法，打通各作战域之间的数据壁垒。

同时，也有一些相关研究重点关注跨域数据处理中的隐私保护问题，这些工作采用区块链等技术实现物联网场景下跨域数据处理的隐私保护。冯绮航[8]研究了物联网场景下数据跨域共享的安全模型，基于区块链技术设计跨域安全共享模块，并并设置监管中心、云存储器与联盟区块链3个实体实现数据的跨域共享。潘雪等[9]利用区块链技术构建了基于主从链的隐私数据跨域共享模型，解决了各个信任应用域之间的跨域安全问题，保证隐私数据的跨域共享。陈嘉熳[10]和郑佳伟[11]研究了基于区块链的跨域数据安全共享技术方案，并提出了数据共享平台的具体设计。

现有研究主要聚焦于上述场景，并未对数据授权运营过程中出现的跨域数据加工和隐私保护进行特别的论述，后者是本文讨论的具体问题。

1.2 数据授权运营相关工作

目前针对数据授权运营的相关研究主要包括授权运营的具体定义、平台建设、技术架构、制度规定和授权模式等方面。高丰[12]开创性地提出对数据授权运营的运营产出和运营行为两个概念予以精细化拆解的必要性，厘清数据授权运营的内涵。林镇阳等[13]基于数据应用的现实需求，探究了数据要素市场化的核心机制、运行机理和法律制度需求。胡业飞等[14]分析了基于传统API技术构建政府数据授权运营模式的局限性，进而提出利用联邦学习技术来建构政府数据授权运营新模式。文章从横向联邦学习、纵向联邦学习、联邦迁移学习3种技术路径出发，分别结合电力供给与配置、个人与企业信贷评估以及医疗服务3个实际情景，呈现出基于联邦学习的政府数据授权运营模式在不同情景下达成公共数据资源开发利用目标的过程。冯洋[15]指出公共数据授权运营的行政许可属性与制度建构方向，提出此项制度构建的4项建议，具体包括鼓励多种所有制市场主体参与授权运营、授权开展数据运营的市场主体数量应控制在合理范围、开展收取公共数据资源使用费试点、授权公共数据的类型范围应当适度限缩。童楠楠等[16]探索研究了一种适配公共数据资源化、资本化、资产化价值生成路径的数据财政、数据税收、数据金融的三层次利益分配框架，在制度层面上为数据要素市场的建设和数字

经济发展提供了一些参考，优化数据价值的分配。黄丽华等[17]提出了基于市场逻辑的数据要素流通价值链模型，在“创生赋权”的设计思路下为数据产权结构性分置提供了一种解释。

综上所述，目前对数据授权运营平台的研究并未结合跨域数据管理的场景，也没有对跨域加工问题和隐私保护问题进行深入的讨论。相较于已有的研究，本文在跨域场景下探讨了数据授权运营过程中可能出现的具体问题。

1.3 隐私计算与数据安全技术

现有的隐私计算技术主要包括安全多方计算、联邦学习以及差分隐私等随机化方法。联邦学习[18]是一种特殊的分布式机器学习技术，其核心思想是通过在多个拥有本地数据的数据源之间进行分布式模型训练，在不需要交换本地个体或样本数据的前提下，仅通过交换模型参数或中间结果的方式，构建基于所有客户端数据的全局模型。借助联邦学习可以满足“数据不出域”的要求，同时完成模型联合训练的任务，保护本地隐私数据。与联邦学习类似，安全多方计算（secure multi-party computation，MPC）也是常用的隐私计算技术。MPC是一个密码学概念，表示多个参与方共同计算一个函数，且除去函数计算结果外，不暴露关于各自输入的信息。一般的安全多方计算概念最初由姚期智在1986年提出，在此之前也已经有许多针对特定函数的多方计算研究。姚期智提出的混淆电路[19]，通过加密算法和不经意传输（oblivious transfer），理论上可以安全计算一切布尔电路。除此之外，同态加密[20]作为一种可以在密文上进行运算的技术，也被广泛应用在安全多方计算领域。差分隐私是一种对数据添加噪声从而实现隐私保护的技术。不同于安全多方计算，差分隐私允许暴露一定量的隐私数据。差分隐私的核心在于给数据添加了随机性，使不同的样本运算有可能得到同样的结果。Google的专家们[21]提出差分隐私机制可以被运用于深度学习的随机梯度下降（SGD）过程中，并且提出了基于动量叠加（moments accountant）的隐私预算上界算法。也有研究者[22]在安全多方计算的物联网场景下提出了各个数据方本地进行差分隐私的方法，使各个数据方直接保证了自身数据的隐私性。

数据脱敏是指在从原始环境向目标环境进行敏感数据交换的过程中，通过一定的方法消除原始环境数据中的敏感信息，并保留目标环境业务所需的数据特征或内容的数据处理过程[23]。在跨域数据进入授权运营平台进行数据加工前，需要采取数据脱敏方法对敏感信息进行预处理。目前，已有的数据处理方法包括面向数据库类型数据的k-匿名（k-anonymous）[24]、l-多样性（l-diversity）[25]、t-保密（t-closeness）[26]等方法，以及针对异构大数据改进的数据脱敏方法[27]等方法。

安全数据沙箱（Sandbox）被广泛应用于浏览器的安全中，防止病毒木马通过浏览器入侵本地。沙箱机制通过进程和内存等计算机资源隔离，控制沙箱内的进程对本地系统资源的调用。在数据跨域加工的过程中，可以在授权运营平台内部建立起数据安全沙箱，防止内部加工人员对隐私数据的非法访问和窃取，限制对数据的操作方式。目前已有工作[28]讨论了在大数据场景下安全沙箱的具体应用。

相较于已有技术文献，本文并未在技术原理和算法框架层面上进行创新，而是结合跨域授权运营中数据加工的场景，探讨了如何合理利用以上技术取得数据安全保护与加工效率之间的平衡，并给出具体的应用案例。

2 跨域数据授权运营平台设计

2.1 跨域数据授权运营全链路结构

跨域数据授权运营的链路结构可以分为数据授权、数据加工、产品交付审查、安全监控4个主要模块。数据授权模块负责审批运营单位加入平台；数据加工模块完成数据产品的生产；产品交付审查模块完成加工后数据产品审查，并交付数据使用方；安全监控模块对系统运营全流程进行监管。具体结构如图1所示。

数据授权主要包括3个流程：①授权运营单位入驻授权运营域，并完成建档；②授权运营单位发起授权数据申请；③公共数据主管部门会同数据提供单位完成审批。

数据加工是跨域数据授权运营链路中的核心环节，也是本文主要分析和研究的环节。数据加工包括4个部分：对申请获取的公共数据进行抽样脱敏处理；将脱敏后的数据分发至开发环境进行数据加工和模型训练；模型训练完成后，将数据模型转产至测试环境，使用抽样数据验证模型计算的可行性和计算结果的可用性；模型测试通过后，将数据模型转产到生产环境，执行计算任务，得到最终的计算结果。数据加工也是数据隐私信息容易泄露的环境，用于训练模型和计算的公共数据与社会数据会集中在加工过程中，造成隐私泄露的可能性。同时，根据文献[29]所述，数据加工形成的模型以及分析结果也包含了隐私敏感信息。

产品交付审查主要包括两部分：①公共数据主管部门进行审核；②将通过审核的数据产品交付指定的授权运营场景进一

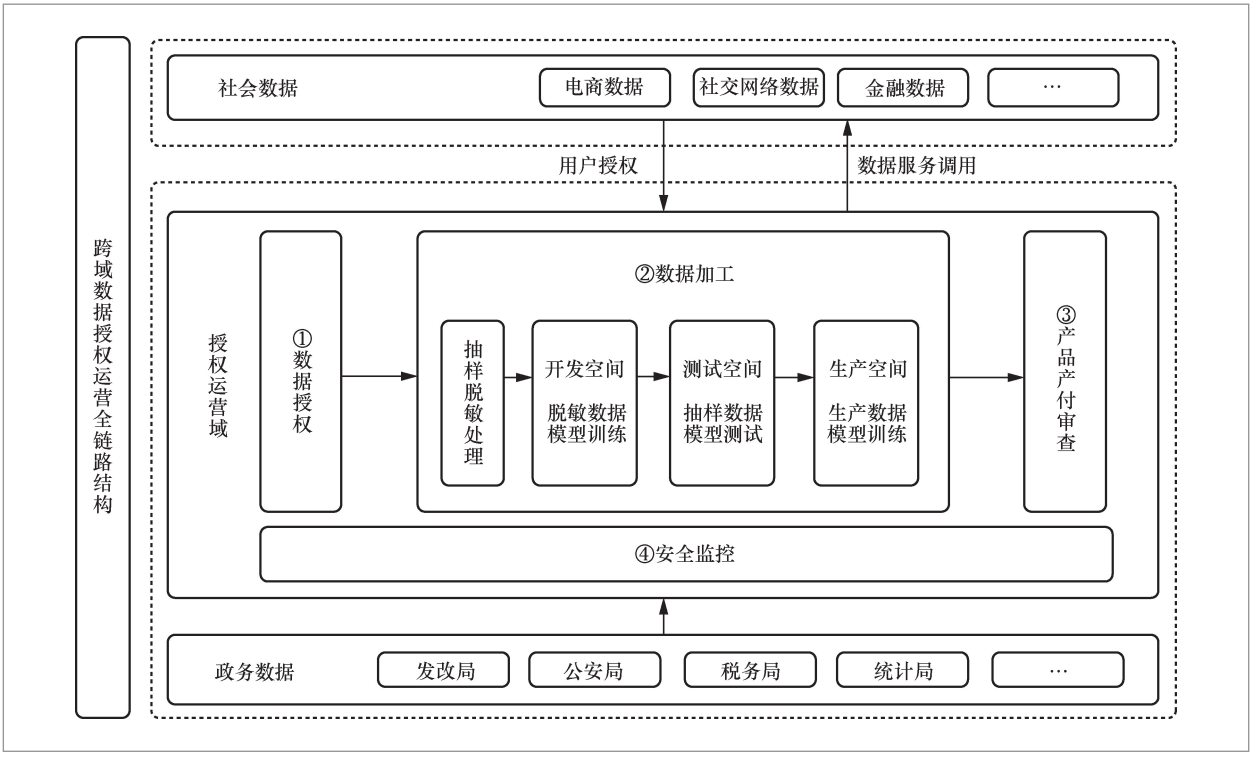

图1 跨域数据授权运营全链路结构

步利用。其中数据产品审核包括形式审查和技术审查两方面，形式审查内容包括数据产品的应用场景、数据服务、数据产品形态等，技术审查内容包括申报信息一致性审核等。

安全监控指对公共数据授权运营业务全过程进行安全监控与防范，覆盖数据申请授权、数据加工、数据产品交付的整个业务过程。其中具体包括业务流程安全、人员安全、环境安全、数据安全、应用安全、合规审计6个层面，对每个环节各个层面的安全问题设置对应的监控方式和防范措施。

在整个跨域数据授权运营平台的结构中，数据加工是其中较为复杂和重要的模块，也是数据安全问题集中体现的部分。因此，后文主要针对跨域数据加工过程中的具体问题进行分析讨论。

2.2 跨域数据加工的挑战

跨域数据加工过程面临的主要挑战包括两个方面：相较于单域数据处理的场景，如何保证在跨域场景下数据加工有较高的效率；在跨域数据加工的过程中，如何避免数据共享流通引起的隐私泄露等数据安全问题。

在加工效率方面，由于数据跨域存储，数据加工的工作量也比单域存储场景下要大。主要体现在两个方面：①由于数据量扩大，数据加工过程中的模型训练可能需要重复多次；②采用隐私保护技术需要对数据进行进一步处理，这会带来额外的处理时间，降低数据加工的效率。

在数据安全方面，跨域数据加工中存在的问题主要有3点：①数据安全的威胁更多来自内部而不是外部；②数据开发人员接触大量敏感数据，造成数据泄露；③企业存留原始数据变相用于其他未授权场景。这些挑战都要求跨域场景下的数据加工方式考虑隐私保护问题。

2.3 两种跨域数据加工模式

针对跨域数据加工中存在的效率问题和数据安全问题，可以采用集中式和隐私计算两种技术模式来解决。如图2所示，集中式加工模式适用于社会数据可出域的场景，将数据集中到授权运营平台统一加工处理，具有较高的加工效率。而基于隐私计算的加工模式适用于社会数据不可出域的场景，通过隐私计算技术进行数据联合加工，会造成额外的计算时间，一定程度上降低了加工效率。因此，在实际应用情境中，应当在算法执行效率和数据是否出域两方面进行权衡，以选择合适的跨域数据加工模式。在这两种情况下，为了加强对公共数据隐私的保护，与文献[30]所述场景不同，此处公共数据不出域。

2.3.1 集中式跨域数据加工模式

集中式跨域数据加工模式如图3所示，公共数据和社会数据集中发送到数据授权运营平台，在租户空间中，依照前述流程依次完成模型开发、模型验证和数据产品生成3个步骤，最终到产品交付审查模块进行审查。

在集中式数据加工模式下，由于采取了将数据集中至平台融合加工的方法，避免了在多个数据域分别加工造成的过高的时间成本，能够有效提高数据加工效率。

为了应对加工过程中可能出现的隐私泄露问题，集中式加工模式采取两方面的解决方案。一方面，将开发环境与生产环境分离，开发环境只提供脱敏抽样的数据，使开发人员不接触原始敏感数据，避

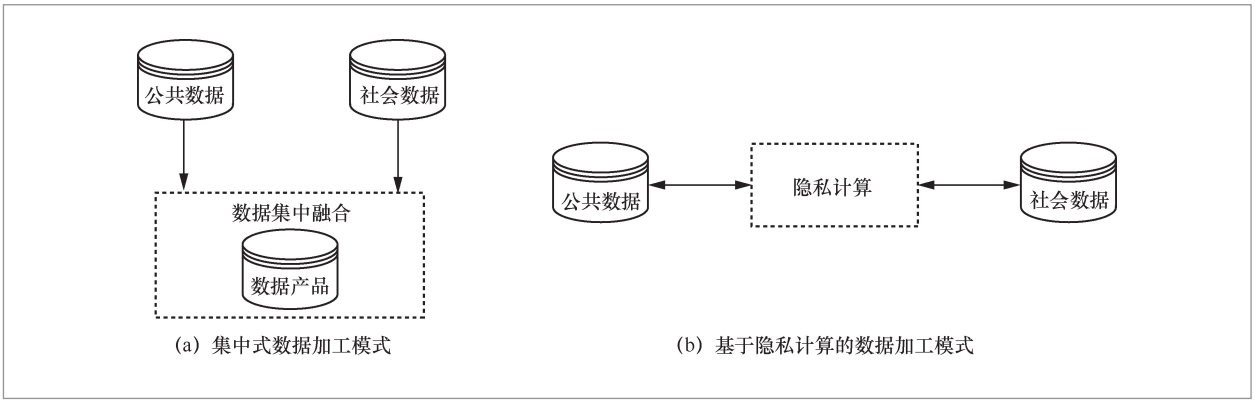

图 2 两种跨域数据加工模式

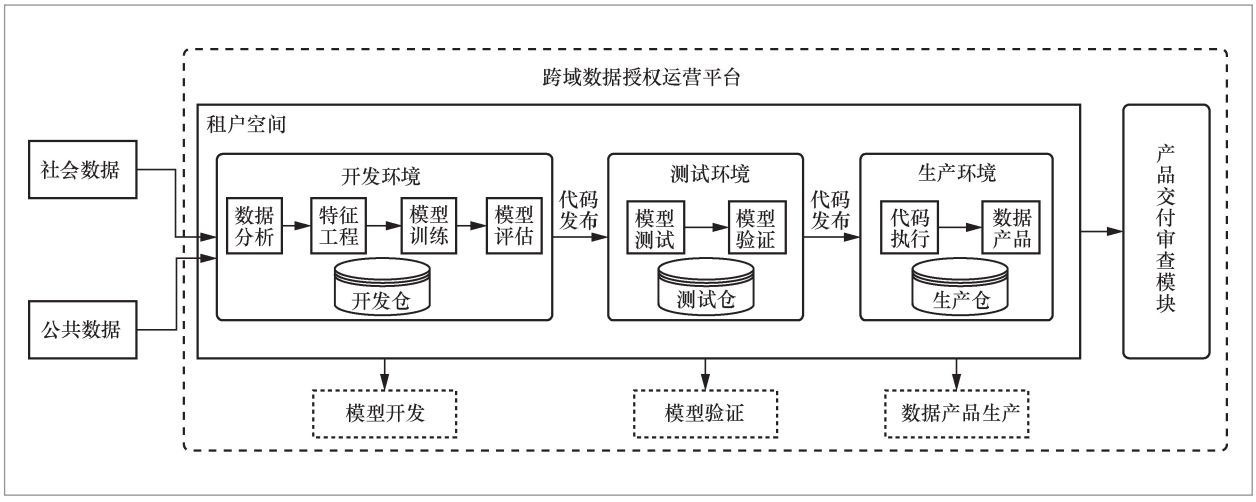

图 3 集中式跨域数据加工模式

免开发过程中的数据泄露；另一方面，建设统一的数据沙箱，严格限制数据的导出，避免数据流向其他使用场景。

2.3.2 基于隐私计算的跨域数据加工模式

基于隐私计算的跨域数据加工模式在计算过程中利用密码学原理，使原始明文数据不可见；同时保证原始数据不出域，避免数据被滥用，并在数据通信过程中采用加密通信的方式，避免隐私信息在域之间流通。由于采用联合加工方式，基于隐私计算的跨域加工模式也能够在一定程度上达到并行处理和提高数据加工效率的目的。相比于集中式的加工模式，隐私计算加工模式虽然在效率上有一定降低，但强化了对数据隐私的保护力度。

基于隐私计算的跨域加工流程如**图4**所示，授权运营单位包含一个隐私计算节点，创建了联合查询、联合分析和联合统计等基于数据产品的服务接口。公共数据授权运营平台包含本地隐私计算节点和隐私计算服务平台，其中隐私计算服务平台负责调度所有隐私计算节点。社会数据经由隐私计算节点加密传输至授权运营平台

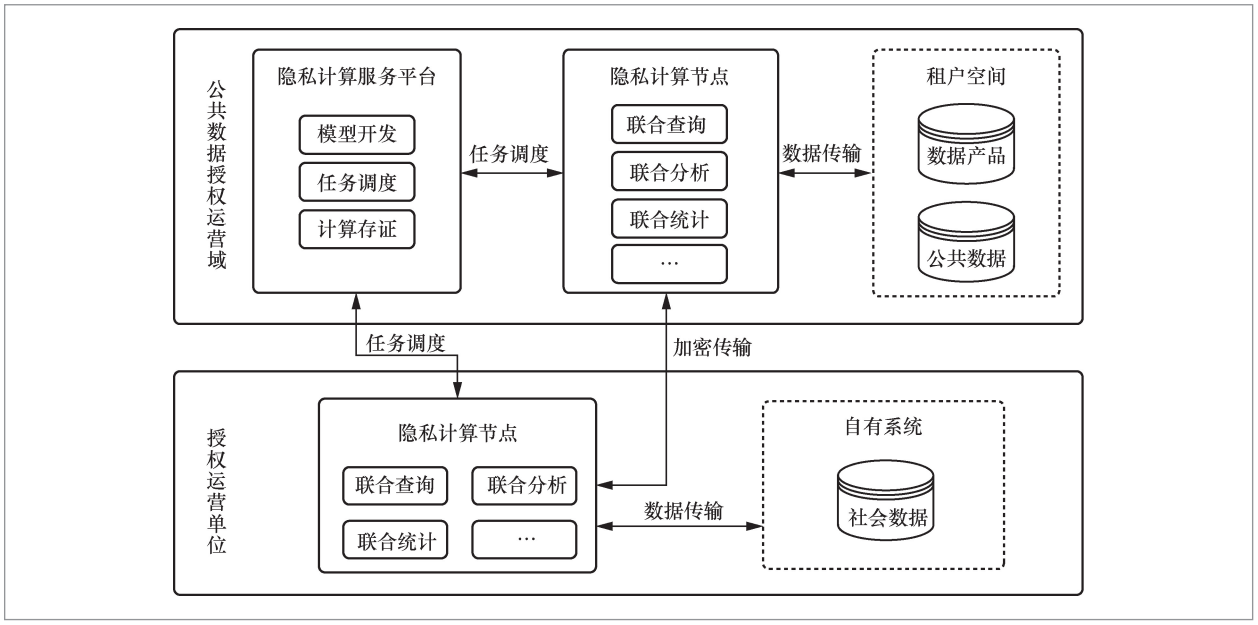

图 4 基于隐私计算跨域数据加工模式

中的隐私计算节点，由隐私计算服务平台调度计算任务，加工数据产品并传输至租户空间。

隐私计算跨域数据加工模式主要采取联邦学习和安全多方计算两种技术实现。

在**图4**中，公共数据授权运营平台中的隐私计算服务平台可以作为联邦学习的中央服务器，隐私计算节点可以作为不同的客户端，从而构建联邦学习框架。如**图5**所示，基于联邦学习的跨域数据加工过程大致包括以下几个步骤：

- 隐私计算服务平台初始化模型参数，并发布模型训练任务到各个域的隐私计算节点；
- 隐私计算节点读取源数据；
- 隐私计算节点导入计算模块；
- 根据训练任务调用隐私计算节点本地的算法库，进行本地模型训练；
- 本地模型训练的参数更新值通过差分隐私、同态加密、秘密分享等隐私保护技术进一步处理后，传输至隐私计算服务平台；
- 隐私计算服务平台执行模型聚合算法，生成全局模型参数；
- 新的全局模型分发至各个隐私计算节点；
- 上述步骤执行多轮后模型收敛，最终得出模型训练结果，并传输至计算节点的前置结果区。

与联邦学习不同的是，在安全多方计算框架中，隐私计算服务平台只负责调度任务和监控隐私计算子节点计算，并不直接参与模型计算过程。如**图6**所示，基于安全多方计算的跨域数据加工流程大概包括以下步骤：

- 隐私计算服务平台发起数据加工或计算任务；
- 隐私计算节点读取源数据；
- 隐私计算节点导入计算模块；
- 不同域隐私计算节点之间借助差分隐私、同态加密和秘密分享等隐私保护技术处理数据，并在域间相互传输，进行样本对齐；
- 不同域隐私计算节点之间借助加密

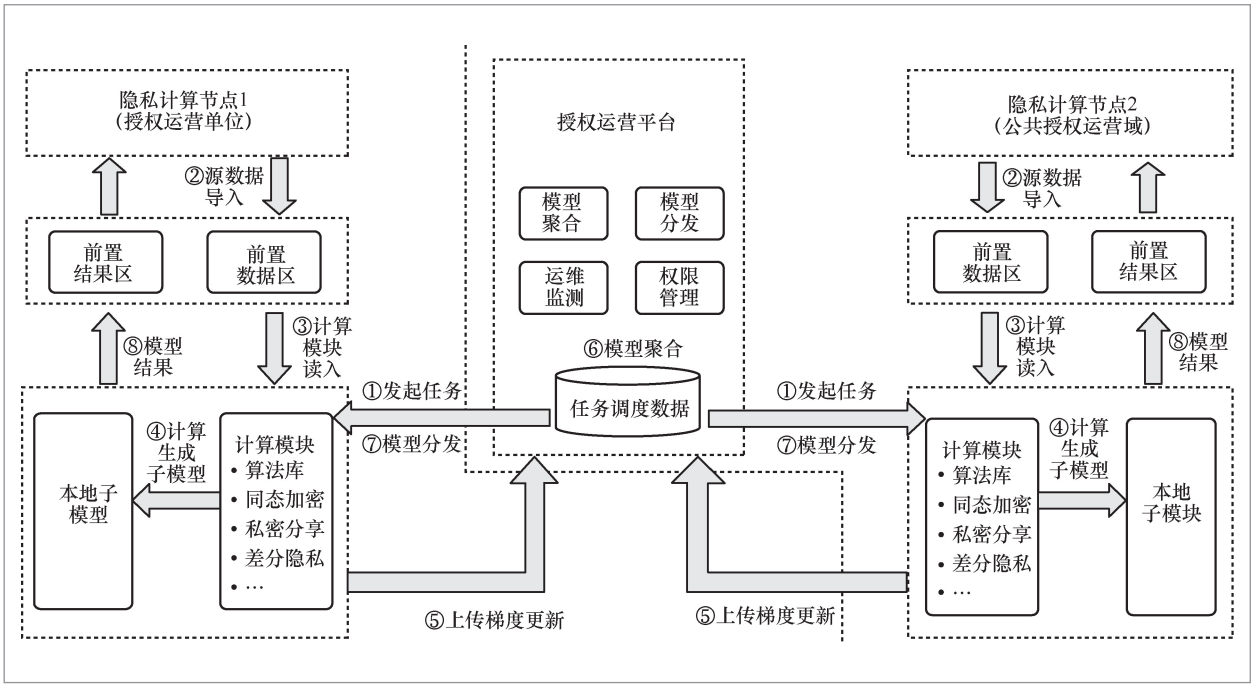

图 5 基于联邦学习的跨域数据加工流程

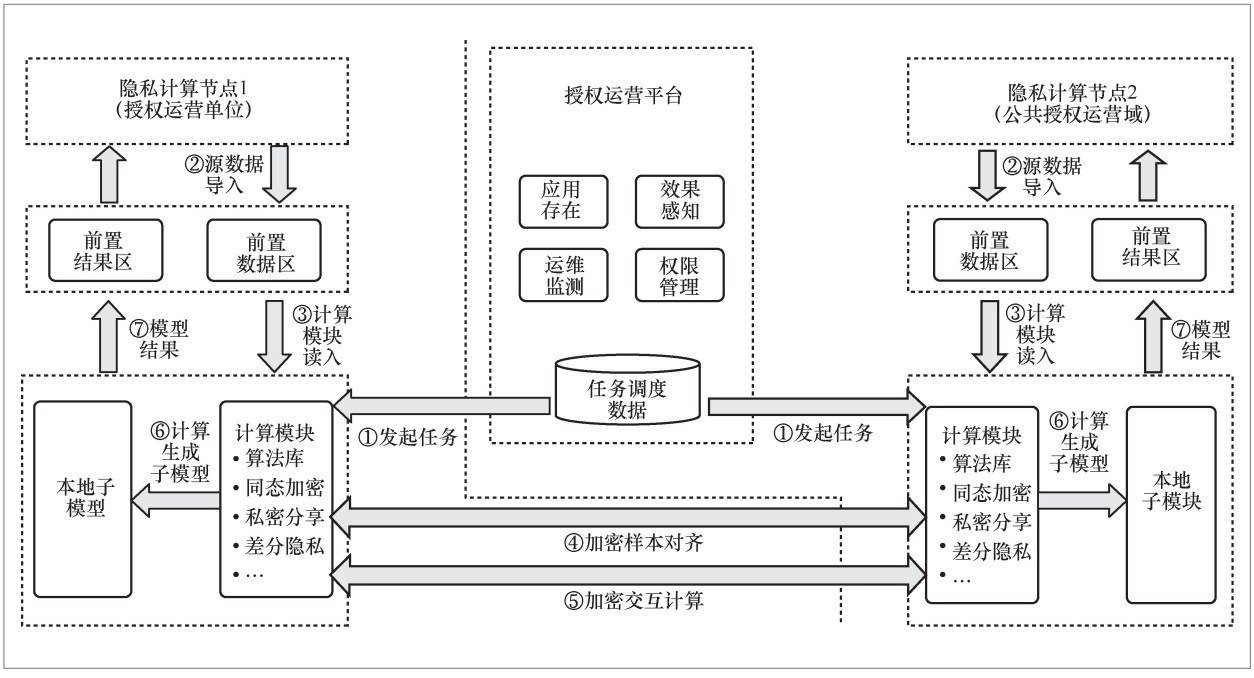

图 6 基于 MPC 的跨域数据加工流程

技术，共享计算过程数据，实现加密交互计算；

- 隐私计算节点计算生成本地子模型或运算结果；
- 最终产生出模型训练结果或计算结果传输至隐私计算节点的前置结果区。

2.3.3 两种加工模式的对比分析

由于集中式数据加工模式和基于隐私计算的数据加工模式采用的技术不同，两种加工模式具有不同的特点，具体概括见**表1**。

从数据安全性角度来讲，集中式数据加工模式由于数据集中统一存储，会带来数据安全问题；而基于隐私计算的加工模式采用了一系列隐私保护技术，隐私数据泄露的可能性大大降低，对数据安全性的保障高于集中式加工方式。

从效率角度来讲，隐私计算技术需要对数据进行加密等一系列处理操作，增加了数据加工的时间成本，因此加工效率要低于集中式加工模式。例如采用同态加密技术实现机器学习模型，相比于明文上的处理，计算时间普遍延长。比如在Graepel T等人[31]的实验中，线性分类器在密文上的训练和分类时间大约比明文上慢5~6个数量级，在明文上模型训练时长为2.3450×10^{-5} s，而同样条件下在密文上训练的时间可达6.221 s。

3 跨域数据授权运营应用案例

针对第2节提出的跨域数据加工技术，本节结合面向跨域数据加工的应用场景加以阐述。

3.1 场景描述

评估企业经营状况是一项具有重要意义的工作，对企业数据的有效分析有助于政府部门开展相关工作。本文列举一例具体场景：通过分析企业纳税状态，了解企业经营状况，主动发现纳税水平达到升级规上企业标准的小规模企业，并协助有关部门及时升规。

在这一场景下，数据建模分析的参与方包括两个部门，即税务局和统计局。税务局为数源部门，提供的数据条目见**表2**。统计局为数据需求部门，使用的数据包括企业名单和纳税规模界限。在数据分析的过程中，省统计局将每月正常纳税的企业和达到一定纳税规模的企业数据进行加密建模分析，输出企业特征标记。该场景符合跨域数据加工的设定，依据前文提出的方法，具体的数据加工流程在第3.2节给出。

表1 两种加工模式的差异对比

数据加工模式	安全性（数据是否可出域）	加工效率
集中式	数据可（加密）出域	较高
基于隐私计算	数据不可出域	较低

3.2 实施方案

在该案例中，对企业纳税状况的评估可以采取基于隐私计算的跨域数据加工模式。具体而言，省税务局和省统计局的私有数据向省大数据局开放授权，省大数据局以此建立授权运营平台，主要用于调度跨域数据联合分析任务、检测平台运行过程以及授权账号申请等。省统计局和省税务局分别建立域内隐私计算节点以及相应的数据存储管理架构。在三方建立稳定网络连接后，具体的跨域数据分析过程可以分为以下几个步骤：

- 由统计局发起计算任务；
- 大数据局调度任务，并下发给两个计算节点；
- 统计局及税务局读取数据至隐私计算节点；
- 统计局加密企业统一社会信用代码并发送至税务局；
- 税务局基于加密结果进行密文计算；

表 2 税务局提供的数据字段

序号	需求部门	数据名称	核心字段	数源部门
1	省统计局	增值税一般纳税人申报信息	纳税人名称 纳税人识别号 统一社会信用代码 所属行来 营业地址 纳税所在地（至区县级） 是否每月持续纳税	浙江省税务局
2		小规模纳税人申报信息	纳税人名称 纳税人识别号 统一社会信用代码 所属行业 营业地址 纳税所在地（至区县级） 是否达到纳税规模界限	

- 统计局获得密文计算结果并解密；
- 统计局应用计算结果。

基于前文提出的隐私计算跨域数据加工模式，可以达成两个方面的效果：①实现税务数据和统计数据联合、高效建模的目的，能够为统计部门评估企业经营状况提供有效数据支持；②在数据加工和传输过程中，采用安全多方计算等隐私计算技术，保证税务数据不出域。同时，如前文所述，能够在加工效率和数据安全性上得到兼顾。上述应用自上线以来，已对全省500余万家企业完成计算，其中计算出达到纳规标准的企业约有10万家，确保了公共数据的安全流通。

4 总结与展望

随着大数据时代的到来，数据成为重要的生产要素和战略资源，数据的价值和使用率不断提升。然而，数据的生产和存储在不同的数据域中，这造成了“数据孤岛”的存在，对推动数据流通、发掘数据价值和进一步融合大数据提供服务造成了很大的阻碍。在国家层面相关文件的指导下，部分省份提出了建立公共数据授权运营平台的方案，借助平台完成数据授权、加工、交付运营的全过程。在这一过程中，跨域场景下的数据加工面临着效率和安全两方面的挑战。本文结合集中式和隐私计算两种数据加工模式阐述了应对上述挑战的技术路线。集中式数据加工模式能够将公共数据和社会数据集中在加工平台中，适用于社会数据可出域的场景，借助数据脱敏和数据沙箱技术保护数据隐私；基于隐私计算的跨域数据加工模式可实现“数据不出域”，采用密码学和加密通信技术避免传输和加工中数据隐私的泄露。而后，文章给出了跨域数据加工模式应用于企业税收数据联合分析评估的案例，阐述了基于隐私计算的跨域数据加工流程。

在确保跨域数据授权运营过程中数据高效加工以及隐私信息保护的基础上，还要考虑多个授权运营平台联合运行的问题。这类问题主要分为两个部分，一是数据授权运营平台架构的兼容性，二是各级政府数据授权运营平台之间的协调关系。由于平台建设初期不同部门采用的第三方构建方案不同，在后续的平台联合运营中需要进一步开发以解决系统间差异引

起的问题；同时，上下级政府之间应当确立平台建设的统一标准，避免因为协调不当造成的冗余开发，促进数据授权运营高效开展。

参考文献:

[1] 袁博，闫树. 数据要素市场化配置上升为国家战略[J]. 互联网天地，2020(6)：34-37.
YUAN B，YAN S. Market-based allocation of data elements rises to national strategy[J]. Internet World，2020(6)：34-37.

[2] 郭志斌，马书惠. 主流公有云提供商产品体系研究[J]. 邮电设计技术，2015(7)：16-21.
GUO Z B，MA S H. Research on the product system of mainstream public cloud providers[J]. Post & Telecom Design Technology，2015(7)：16-21.

[3] 邓磊，吴健，张昌利，等. 电子政务中跨域可信数据交换模型设计与实现[J]. 计算机工程，2007，280(12)：4-9.
DENG L，WU J，ZHANG C L，et al. Design and Implementation of cross-domain trusted data exchange model in e-government[J]. Computer Engineering，2007，280(12)：4-9.

[4] 张彬，徐建民，吴姣. 跨域推荐中的知识融合研究进展[J]. 现代情报，2023，43(3)：157-166.
ZHANG B，XU J M，WU J. Advances in knowledge fusion research in cross-domain recommendation[J]. Modern Intelligence，2023，43(3)：157-166.

[5] 卢红建，王晗，马玮. 跨区域医疗信息共享与业务协同标准建设探索[J]. 中国卫生信息管理杂志，2021，18(2)：229-234，267.
LU H J，WANG H，MA W. Exploring the construction of cross-regional medical information sharing and business collaboration standards[J]. Chinese Journal of Health Information Management，2021，18(2)：229-234，267.

[6] 董晶，张天龙. 智慧城市跨域数据治理体系与平台研究[J]. 电子元器件与信息技术，2021，5(1)：53-54.
DONG J，ZHANG T L. Research on cross-domain data governance system and platform for smart cities[J]. Electronic Components and Information Technology，2021，5(1)：53-54.

[7] 王蒙蒙，朱婉婷. 面向联合作战的跨域数据安全互联方法[J]. 中国电子科学研究院学报，2020，15(5)：442-448.
WANG M M，ZHU W T. A secure interconnection method for cross-domain data for joint warfare [J]. Journal of the Chinese Academy of Electronic Science，2020，15(5)：442-448.

[8] 冯绮航. 考虑属性加密的物联网隐私数据跨域安全共享模型[J]. 现代电子技术，2023，46(1)：91-95.
FENG Q H. A cross-domain secure sharing model for IoT privacy data considering attribute encryption[J]. Modern Electronics Technology，2023，46(1)：91-95.

[9] 潘雪，袁凌云，黄敏敏. 主从链下的物联网隐私数据跨域安全共享模型[J]. 计算机应用研究，2022，39(11)：3238-3243.
PAN X，YUAN L Y，HUANG M M. A cross-domain secure sharing model for IoT privacy data under master-slave chain[J]. Computer Application Research，2022，39(11)：3238-3243.

[10] 陈嘉嫚. 基于区块链的跨域数据共享技术研究[D]. 成都：电子科技大学，2021.
CHEN J M. Research on cross-domain data sharing technology based on blockchain[D]. Chengdu：University of Electronic Science and Technology，2021.

[11] 郑佳伟. 基于区块链的物联网数据跨域受控共享平台设计与实现[D]. 西安：西安电子科技大学，2020.
ZHENG J W. Design and Implementation

of a blockchain-based cross-domain controlled sharing platform for IoT data[D]. Xi'an: Xi'an University of Electronic Science and Technology, 2020.

[12] 高丰. 厘清公共数据授权运营: 定位与内涵[J]. 大数据, 2023, 9(2): 16-32.
GAO F. Clarifying the authorized operation of public data: positioning and connotation[J]. Big Data Research, 2023, 9(2): 16-32.

[13] 林镇阳, 侯智军, 赵蓉, 等. 数据要素生态系统视角下数据运营平台的服务类型与监管体系构建[J]. 电子政务, 2022, 236(8): 89-99.
LIN Z Y, HOU Z J, ZHAO R, et al. Service types and regulatory system construction of data operation platform in the perspective of data element ecosystem[J]. E-Government, 2022, 236(8): 89-99.

[14] 胡业飞, 陈美欣, 张怡梦. 价值共创与数据安全的兼顾: 基于联邦学习的政府数据授权运营模式研究[J]. 电子政务, 2022, 238(10): 2-19.
HU Y F, CHEN M X, ZHANG Y M. Balancing value co-creation and data security: a federal learning-based operational model for government data empowerment[J]. E-Government, 2022, 238(10): 2-19.

[15] 冯洋. 公共数据授权运营的行政许可属性与制度建构方向[J]. 电子政务, 2023(6): 77-87.
FENG Y. The administrative license attributes of public data authorization operation and the direction of institutional construction[J]. E-Government, 2023(6): 77-87.

[16] 童楠楠, 杨铭鑫, 莫心瑶, 等. 数据财政: 新时期推动公共数据授权运营利益分配的模式框架[J]. 电子政务, 2023, 241(1): 23-35.
TONG N N, YANG M X, MO X Y, et al. Data finance: a model framework for promoting benefit sharing of public data empowerment operations in the new era[J]. E-Government, 2023, 241(1): 23-35.

[17] 黄丽华, 杜万里, 吴蔽余. 基于数据要素流通价值链的数据产权结构性分置[J]. 大数据, 2023, 9(2): 5-15.
HUANG L H, DU W L, WU B Y. A structural division of data property rights based on the value chain of data elements circulation[J]. Big Data Research, 2023, 9(2): 5-15.

[18] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[J]. Artificial Intelligence and Statistics, PMLR, 2017: 1273-1282.

[19] MIHIR B, HOANG V T, ROGAWAY P. Foundations of garbled circuits[C]// ACM Conference on Computer and Communications Security. New York: ACM Press, 2012: 784-796.

[20] OGBURN M, TURNER C, DAHAL P. Homomorphic encryption[J]. Computer Science, 2013: 502-509.

[21] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 308-318.

[22] LIU Z Q, WANG Y X, SMOLA A. Fast differentially private matrix factorization[C]//Proceedings of the 9th ACM Conference on Recommender Systems. New York: ACM Press, 2015: 171-178.

[23] 唐迪, 顾健, 张凯悦, 等. 数据脱敏技术发展趋势[J]. 保密科学技术, 2021, 127(4): 4-11.
TANG D, GU J, ZHANG K Y, et al. Development trend of data desensitization technology[J]. Confidentiality Science and Technology, 2021, 127(4): 4-11.

[24] SWEENEY L. k-anonymous: a model

for protecting privacy[J]. Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[25] MACHANAVAJJHALA A, GEHRKE J, KIFER D, et al. L-diversity: privacy beyond k-anonymity[C]//Proceedings of 22nd International Conference on Data Engineering (ICDE'06). Piscataway: IEEE Press, 2006: 24.

[26] LI N H, LI T C, VENKATASUBRAMANIAN S. t-closeness: privacy beyond K-anonymity and L-diversity[C]//Proceedings of IEEE 23rd International Conference on Data Engineering. Piscataway: IEEE Press, 2007: 106-115.

[27] 佟玲玲, 李鹏霄, 段东圣, 等. 面向异构大数据环境的数据脱敏模型[J]. 北京航空航天大学学报, 2022, 48(2): 249-257.
TONG L L, LI P X, DUAN S D, et al. A data desensitization model for heterogeneous big data environment[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(2): 249-257.

[28] 王忠春, 陈庆荣, 刘婷. 大数据下新型安全沙箱技术运用分析与研究[J]. 网络空间安全, 2022, 13(6): 89-97.
WANG Z C, CHEN Q R, LIU T. Analysis and research on the use of new security sandbox technology under big data[J]. Cyberspace Security, 2022, 13(6): 89-97.

[29] 彭钺峰, 赵波, 刘会, 等. 针对机器学习的成员推断攻击综述[J]. 计算机科学, 2023, 50(3): 351-359.
PENG Y F, ZHAO B, LIU H, et al. A review of membership inference attacks against machine learning[J]. Computer Science, 2023, 50(3): 351-359.

[30] 叶兵, 宋从雅, 赵银银. 公共数据的金融共享应用[J]. 中国金融, 2022, 985(19): 73-74.
YE B, SONG C Y, ZHAO Y Y. Financial sharing applications of public data[J]. China Finance, 2022, 985(19): 73-74.

[31] GRAEPEL T, LAUTER K, NAEHRIG M. ML confidential: machine learning on encrypted data[C]//Proceeding of International Conference on Information Security and Cryptology. Heidelberg: Springer, 2013: 1-21.

作者简介

张纪林(1980-), 男, 博士, 浙江省大数据发展管理局数据资源处副处长, 曾任杭州电子科技大学网络空间安全学院副院长、教授, 主要研究方向为政务智能化、海量数据处理、数据安全。

顾小卫(1980-), 男, 博士, 浙江省委网信办网络数据与技术处处长, 曾任浙江理工大学国际处副处长、信息学院教授, 主要研究方向为信息安全、数据安全和电子技术等。

张亦钊（1998– ），男，浙江大学计算机科学与技术学院硕士生，主要研究方向为隐私计算。

郑小林（1977– ），男，博士，浙江大学计算机科学与技术学院教授、博士生导师，浙江大学人工智能研究所副所长，斯坦福大学访问学者，IEEE Senior Member，中国计算机学会杰出会员，主要研究方向为人工智能、隐私计算、智能电商、金融智能等。

陈超超（1988– ），男，博士，浙江大学计算机科学与技术学院特聘研究员，浙江大学现代服务创新实验室副主任，曾任蚂蚁集团高级算法专家，主要研究方向为隐私保护机器学习、分布式机器学习、图机器学习和推荐系统等。

收稿日期： 2023-02-28
通信作者： 张纪林，jilin.zhang@hdu.edu.cn
基金项目： 国家自然科学基金资助项目（No.62172362）
Foundation Item: The National Natural Science Foundation of China (No.62172362)