面向LLM开放域问答中多方私有表格筛选:一种MPC可公开聚合审计与动态信誉的增强方法

胡 睿,吴 昊,潘宇轩,张 琳,刘 雨,朱孔林*
(北京邮电大学人工智能学院,北京 100876)

摘 要: 大语言模型(Large Language Model,LLM)驱动的开放域问答(Open-Domain Question Answering,ODAQ)系统,如GIST(Generating Identifiers and Selecting chunks for Tables)框架,在处理海量表格数据时展现出巨大潜力,受到了广泛关注. 然而,当ODQA系统需要整合多方私有表格数据进行Top-K候选筛选等环节时,传统方法需要访问全部原数据,这在数据隐私、计算透明度及参与方行为可信度方面面临挑战. 虽然现有研究采用零知识证明和基于权益的机制实现了公开可验证性,但在大规模场景下生成和验证单个证明的开销过高,而传统的基于权益的机制在公平性和对动态环境的适应性方面也存在局限性. 对此,本文基于多方安全计算(Multi-Party Computation,MPC)、可公开聚合审计与动态信誉机制,提出了一种面向LLM开放域问答中多方私有表格筛选的增强方法. 将Top-K多方私有表格筛选过程通过MPC完成,以保护多方私有数据隐私. 同时,引入高效的聚合审计机制,将零知识证明技术与随机抽样、聚合证明构造、基于时间窗口的批处理和错误定位相结合,确保评分与排序过程的正确性可以被批量、公开验证. 基于区块链的动态信誉反馈机制的集成也增强了系统的公平性,并约束了恶意行为. 实验评估表明,本文的Top-K候选筛选方法在保证隐私的同时与GIST原有筛选方法在结果上达到0.91的Top-50平均召回率和0.83的平均Jaccard指数,具有高度一致性,不会影响ODQA端到端任务性能. 同时,大规模任务下可公开审计的证明和验证效率均得到提升,与单独的证明相比节省了约87%的证明时间. 反馈机制的适应性和公平性也得到了增强.

关键词: 开放域问答;大语言模型;多方安全计算;可公开审计;零知识证明;区块链
基金项目: 国家重点研发计划(No.2023YFB2704500)
中图分类号: TP399 **文献标识码:** A **文章编号:** 0372-2112(2025)09-3089-14
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250451

Multi-Party Private Table Screening for LLM-Driven ODQA: An Enhanced Method with MPC, Publicly Aggregable Audit, and Dynamic Reputation

HU Rui, WU Hao, PAN Yu-xuan, ZHANG Lin, LIU Yu, ZHU Kong-lin*
(*Department of Artificial Intelligence*, *Beijing University of Posts and Telecommunications*, *Beijing* 100876, *China*)

Abstract: Large language model (LLM) driven open-domain question answering (ODQA) systems, exemplified by frameworks like GIST (Generating Identifiers and Selecting chunks for Tables), have garnered considerable research attention due to their significant potential in processing extensive tabular data. However, when such ODQA systems integrate data from multiple providers for Top-K candidate screening, traditional methods requiring access to raw data encounter substantial challenges concerning data privacy, computational transparency, and participant trustworthiness. While existing research employs zero-knowledge proofs and stake-based mechanisms to achieve public verifiability, the overhead of generating and verifying individual proofs in large-scale scenarios is often prohibitive. Moreover, conventional stake-based mechanisms exhibit limitations in fairness and adaptability within dynamic environments. This paper proposes an enhanced method for multi-party private table screening in LLM-driven ODQA, which integrates multi-party computation (MPC), a publicly aggregable audit mechanism, and a dynamic reputation system. This study adapt the Top-K multi-party private table screening process using MPC to ensure data privacy. Concurrently, an efficient aggregable audit mechanism is introduced;

收稿日期:2025-05-31;录用日期:2025-09-24;责任编辑:覃怀银
*通讯作者:朱孔林

this mechanism combines zero-knowledge proof techniques with random sampling, aggregate proof construction, time-window-based batching, and error localization, thereby enabling the public and batch-verified correctness of the scoring and ranking process. The integration of a blockchain-based dynamic reputation feedback mechanism further enhances system fairness and constrains malicious behavior. Experimental evaluations demonstrate that our Top-K candidate screening method, while preserving privacy, achieves high consistency with the original GIST screening approach, attaining a Top-50 average recall of 0.91 and an average Jaccard index of 0.83, thus indicating minimal impact on end-to-end ODQA task performance. Furthermore, the efficiency of publicly auditable proof generation and verification for large-scale tasks is significantly improved, saving approximately 87% of proof time compared to individual proofs. The adaptability and fairness of the feedback mechanism are also demonstrably enhanced.

Key words: open-domain question answering; large language models; secure multi-party computation; publicly auditable; zero-knowledge proofs; blockchain

Foundation Item(s): National Key Research and Development Program of China (No.2023YFB2704500)

1 引言

近年来,以大语言模型(Large Language Model,LLM)为核心驱动的人工智能技术取得了突破性进展,深刻改变了信息获取与知识服务的方式.特别是在开放域问答(Open-Domain Question Answering,ODQA)领域,大语言模型凭借其强大的自然语言理解与推理能力,使得从海量、多样化的数据(包括结构化表格与非结构化文本)中精准获取答案成为可能[1].例如,GIST(Generating Identifiers and Selecting chunks for Tables)[2]等前沿框架,通过充分挖掘LLM的上下文学习潜力,实现了在缺乏大规模标注数据的情况下对表格等复杂数据源的高效问答,极大地提升了信息检索的便捷性和智能化水平.

然而,随着大语言模型在互联网及各类应用中的深度融合与广泛部署,一系列新的挑战也日益凸显.其中极为关键的是,当ODQA系统(如GIST)需要整合来自多个不同提供方的表格等数据,其核心处理流程涉及多方计算实体的协作时,系统的透明度、计算结果的内在正确性、用户数据的隐私安全以及参与各方的行为可信度等问题变得尤为关键[3].在众多实际应用场景中,表格数据往往承载着商业机密、个人隐私等高度敏感的信息[4],数据持有者通常不愿让中心化的问答系统或其他任何第三方直接大量地访问与问题回答实际并不相关的数据.这种固有的数据私密性需求与充分利用多方数据以提升问答服务质量的目标之间,形成了日益尖锐的矛盾.因此,相关研究考虑引入多方计算的场景,而不是中心化地完成数据筛选与判断.

目前,虽然部分研究尝试采用零知识证明等技术手段增强多方计算过程的公开可验证性[5],但在处理大规模多方计算时,传统方法面临着证明生成与验证开销过大、在动态和复杂环境中适应性不足等瓶颈.与此同时,在多个不同提供方提供数据的情况下,由于计算本身固有的"黑箱"特性,其内部决策逻辑难以完全透明化和被追溯[6].计算过程筛选的是LLM赖以决策的数据来源,即从海量潜在表格中初步筛选候选集的过程,如果缺乏可信保障,那么最终输出答案的可靠性将受到严重质疑,甚至可能引发数据误用、偏见放大等一系列的潜在风险[7,8].因此,在大模型与互联网深度融合的时代背景下,如何构建一个既能充分发挥LLM的强大问答功能,又能有效保障在多方数据协作场景下的隐私安全、计算正确、行为可信且过程可审计的开放域问答系统,是急需解决的关键问题.

针对上述挑战,本研究设计并实现了一种多方数据问答的可公开聚合审计与动态信誉增强机制.本文在ODQA任务下基于大语言模型的GIST框架,聚焦于其中数据访问最为密集、隐私泄露风险最高的初始候选集筛选环节,即从海量多方私有表格中遴选Top-K候选表格的过程.在该环节引入多方安全计算(Multi-Party Computation,MPC)、高效的聚合审计机制以及动态的信誉反馈系统,以期在有效保护数据隐私的前提下,确保大规模分布式计算的可审计性与正确性,并促进参与生态的健康公平发展.这对于推动大模型技术在金融、医疗等关键敏感领域的安全合规应用,以及保障网络空间中数据要素的安全、有序与高效利用,具有重要意义.

本文的主要贡献包括以下几点.

(1)提出了一种针对LLM驱动的开放域问答系统中多方私有数据初步筛选环节的可信增强方法.该方法创新性地将MPC应用于GIST等系统在处理海量多方私有表格时的Top-K候选筛选过程,在保护数据提供方隐私的同时,为后续LLM的精细化处理提供可靠的数据基础.

(2)设计并实现了高效的可公开聚合审计机制,以保障多方表格评分计算的可验证性与透明度.该机制显著减少了审计大规模多方计算流程所需的证明数量,降低了计算复杂度.通过结合随机抽样、聚合验证、基于时间窗口的批处理和错误定位,本文的方法能够

确保评分过程的正确性被高效、公开地验证,从而提升了整个问答系统数据处理的可信度.

(3)设计并实现了基于区块链和智能合约的动态信誉反馈机制,促进协作生态的公平与可持续发展,从而增强系统的适应性和公平性. 该机制将聚合审计的验证结果直接转化为对数据提供方信誉的实时调整与反馈,通过有效的奖惩措施激励诚实行为,约束潜在的恶意操作,为构建一个公平、透明且可持续的多方协作问答生态系统提供了关键的技术支撑.

2　相关工作

2.1　大语言模型与开放域问答

大语言模型已显著提升了问答系统的能力[9],研究者探索了多种利用LLM进行信息检索、证据抽取和答案生成的方法,比如引入用于文档检索的URL(Uniform Resource Locator)标识符[10]、引导LLM逐步生成SQL(Structured Query Language)[11]等. 然而,这些研究均未关注LLM在开放域任务中的应用. 当前,ODQA的主流技术范式是"检索器-阅读器"(Retriever-Reader)架构. 该架构首先由检索器从大规模语料库中快速召回与问题相关的片段,然后由阅读器对这些文档进行深度理解和分析,最终抽取出精确答案. 在检索器实现上,存在两种主要技术路线:稀疏检索与密集检索. 稀疏检索以TF-IDF、BM25等传统信息检索方法为代表,其优势在于计算速度快,但在处理词汇不匹配的语义相关问题时表现不佳. 密集检索以DPR(Dense Passage Retrieval)为代表[12],它利用预训练语言模型将问题和文档编码为低维稠密向量,通过向量相似度匹配,能更准确地捕捉语义相关性,显著提升召回的准确率. 此外,为进一步优化系统性能,部分工作也探索了检索器与阅读器之间的强化学习、记忆与迭代交互等机制[13],以动态优化检索结果. GIST框架则在"检索器-阅读器"架构的基础上针对开放域问答,引入筛选、分解、证据选择等子环节,通过LLM生成标识符进行表格检索,并利用LLM选择证据块,展示了LLM在处理表格数据问答任务中的潜力. 尽管这些研究在提升问答效果方面取得了显著提升,但在涉及多方私有数据源协作时,如何在保护数据隐私的前提下有效利用LLM进行Top-K候选筛选等初步数据处理环节,仍是一个亟待解决的问题.

2.2　MPC在数据处理与隐私保护中的研究

MPC作为一种关键的隐私保护技术,允许各方在不泄露各自私有输入的情况下共同完成计算任务. 近年来,MPC在隐私保护机器学习、数据分析和数据库查询等领域得到了广泛应用. 例如,有研究将MPC应用于联邦学习中,以保护各参与方本地模型的梯度信息[14]. 另外一些工作则探索了MPC在执行SQL类查询[15]或进行统计分析[16]时的隐私保护能力. MPC为多方数据协作提供了理论基础,但将其应用于LLM驱动的ODQA系统中,特别是针对Top-K表格筛选这类需要平衡隐私保护和可公开并与LLM模块交互的特定环节,结合后续公开可审计的设计,仍需要进一步探索.

2.3　计算过程的公开可审计研究

确保计算过程的正确性和可追溯性对于构建可信系统至关重要,尤其是在应对与公共审计相关的挑战方面. Zhu等人[17]提出的用于双方计算的公开验证方案,融合了区块链和零知识证明,以确保计算的透明性和正确性. 然而,将该方法扩展到大规模多方计算系统则仍然不可行. 类似地,Yang等人[5]、Cordi等人[18]和Kanjalkar等人[19]提出的结合区块链的去中心化多方计算(MPC)框架,将零知识证明与密码学原语结合使用. 这些框架实现了隐私保护和透明计算,尤其是在工业物联网环境中. 然而,随着参与者数量的增加,交互复杂性和运营成本显著增加,如何高效实现可公开审计具有较大的探索价值.

2.4　基于区块链的激励机制研究

在多方协作系统中,建立有效的激励机制以促进诚实参与、约束恶意行为是维持系统稳定和公平的关键. Seo等人[20]和Jiang等人[21]在MPC和区块链系统中引入了基于存款的激励方案,对恶意行为进行了约束. 然而,对固定存款金额的依赖导致其对动态环境的适应性有限,从而导致奖惩结构缺乏灵活性. Jin等人[22]提出了一个基于区块链的MPC框架,该框架能够对整个计算过程进行公开审计,同时融入了作弊者检测机制,对不诚实的参与者施加经济惩罚,并通过补偿激励诚实的行为. 尽管此类机制有其优势,但对虚拟代币的广泛依赖会在不同的监管环境下产生合规风险,并引发对系统性问题的担忧,包括资源丰富的参与者的主导地位以及新市场参与者的进入壁垒.

3　相关技术基础

3.1　GIST:基于大语言模型的开放域表格问答

GIST是一个利用LLM进行ODQA的框架[2],其特点在于无需对特定数据集进行大量微调即可实现有效的问答. GIST的核心流程主要包括以下几个关键步骤.

(1)标识符生成. 针对用户提出的自然语言问题,GIST首先借助LLM的理解能力,从中提取生成一组能够精确捕捉问题核心语义的离散标识符.

(2)表格检索. 利用上一步生成的标识符,GIST在表格语料库中进行检索,以召回与问题最相关的候选表格集合. 此过程涉及语义相似度匹配和重排序等技术.

(3)证据选择. 为减少后续LLM推理的负担并排

除无关信息,GIST设计了一个选择器模块. 该模块将候选表格分割成更小的文本块(chunks),并再次利用LLM的判断能力,从这些文本块中筛选出与问题强相关的子证据.

(4)答案生成. GIST的阅读器模块将筛选出的子证据和原始问题一同提交给LLM,由LLM进行最终的综合推理并生成自然语言形式的答案.

GIST框架通过巧妙地利用LLM在不同阶段的能力,有效地应对了开放域表格问答的挑战. 然而,当GIST需要处理来自多个不愿直接共享其私有表格数据的提供方的数据时,其原有的数据处理流程(特别是在表格检索阶段)将面临隐私保护和协作可信度的挑战,这正是本文所要解决的核心问题之一.

3.2 基于秘密共享的多方安全计算

基于秘密共享的多方安全计算协议使多方能够基于各自的私有输入联合计算函数,而无需彼此透露输入. 这类协议包含许多经典且广泛使用的方案,例如GMW协议、SPDZ[6]及其各种扩展. 这些协议通常在有限域F_p上运行,其中p是足够大的素数.

秘密共享和乘法三元组. 每个私有值x被拆分为个份额$[x]$,分配给n个参与者,这样就没有参与者能够得知x的值. 为了实现高效的乘法,该协议使用预处理的乘法三元组(a,b,c),其中a、b是F_p中的随机元素,且$c=a\times b$. 每个值都是各方共享的秘密.

安全乘法. 给定秘密共享的输入x和y,各方使用乘法三元组$d=x-a,e=y-b$,共同计算$z=x\times y$. d和e的值被重建并显示出来,由于a和b的随机性,这样不会泄露x或y的信息. 然后,乘积计算如下:$z=c+d\times b+e\times a+d\times e$. 重复此过程,可以安全地计算复杂函数,而不会泄露任何一方的隐私输入.

3.3 Pedersen承诺

Pedersen承诺是一种具有完美隐藏性和计算绑定性的密码学承诺方案[23]. 对一个值x和一个随机选择的致盲因子r,其承诺形式为$\mathrm{Com}(x;r)=g^x h^r$,其中$g$和$h$是一个循环群的两个生成元,且$h$相对于$g$的离散对数是未知的.

该承诺的关键特性在于其同态性,两个承诺的乘积等于其对应值之和的承诺,即$\mathrm{Com}(x_1,r_1)\cdot\mathrm{Com}(x_2,r_2)=\mathrm{Com}(x_1+x_2,r_1+r_2)$. 这一特性使得Pedersen承诺非常适用于构建聚合证明,因为它允许将对多个值的承诺聚合成对这些值某种线性组合的承诺,而无需暴露原始值.

3.4 零知识证明

零知识证明(Zero-Knowledge Proofs,ZKPs)允许证明者向验证者证明其拥有某个知识或某个声明为真,而在此过程中不泄露任何关于该知识或声明的具体信息(除了其真实性本身). 在本文中主要关注非交互式零知识证明(Non-Interactive Zero-Knowledge proofs,NIZKs),它消除了证明者与验证者之间多轮交互的需求,通常通过Fiat-Shamir启发式等技术将交互式证明转换为一次性消息[24].

验证者收到一个声明S和对应的NIZK证明π后,可以通过一个公开的验证算法$\mathrm{Verify}(S,\pi)$来判断证明的有效性. 若验证通过,则验证者确信声明S为真. NIZKs的安全性依赖于成熟的密码学假设,即离散对数问题(Discrete Logarithm Problem,DLP)和计算性Diffie-Hellman(Computational Diffie-Hellman,CDH)假设.

4 系统框架与核心方法

针对LLM驱动的ODQA系统在集成多方私有数据时面临的隐私、可信与效率挑战,本节将概述本研究提出的核心方法. 该方法是一个集隐私保护、可公开审计与动态激励于一体的综合性框架,旨在为多方协作环境下的Top-K表格筛选提供解决方案.

本研究提出的方法是作为GIST这类先进的ODQA框架在处理多方私有数据场景下的一个可信协作与隐私安全增强层,系统涉及技术模块的角色定位如下.

GIST框架是核心的问答引擎. 它负责处理高级的语义理解任务,如解析用户问题、从筛选后的证据中综合推理并生成最终答案.

本研究提出的增强方法是介于GIST框架与多方数据源之间的可信协作与安全增强层. 它专注于多方私有数据的安全协作问题,确保在调用多方私有数据进行Top-K候选筛选时,过程是隐私保护、计算正确且参与方行为可信的.

Hyperledger Fabric区块链是本文增强方法中的“可信基础设施”. 选用区块链技术,借助其不可篡改、去中心化和公开透明的特性,为本研究设计的“可公开聚合审计”和“动态信誉反馈”机制提供了技术实现平台,是确保审计结果和信誉记录公信力的基石.

为了将这一核心方法付诸实践,本文设计了一个由3类关键实体构成的系统架构,并规划了其工作流程,以此作为该方法的具体实现蓝图. 该架构的组成实体与工作流程的各个阶段,分别对应了本研究方法中的不同角色和核心机制.

4.1 系统实体

本文定义的系统架构由3类实体组成:数据提供方、数据用户和基于区块链的审计者.

数据提供方(Data Providers,DPs). 每个数据提供方拥有并控制其私有表格数据集. DPs是GIST系统进行Top-K候选表格筛选时的数据来源,并作为核心参与

者加入相关的MPC协议中. 在此过程中,DPs负责对其私有数据进行秘密共享、执行必要的本地计算,并为后续的审计生成相应的密码学证明.

数据用户(Data Users,DUs). 数据用户是问答服务的最终使用者,他们向GIST系统提交自然语言形式的问题,并期望获得基于经过可靠筛选的表格数据、由LLM生成的答案. 在特定场景下,DUs自身也可能扮演数据提供方的角色,参与到多方计算过程中.

审计者(Auditor). 审计者被设计为一个基于区块链的去中心化实体,其功能通过预设的智能合约自动执行. 其核心职责包括收集由DPs(或其他计算参与方)提交的密码学承诺(如Pedersen承诺)和有效性证明(如ZKPs),对这些证明执行公开、透明的验证程序,并以不可篡改的方式记录审计结果及各参与方的动态信誉评分.

4.2　工作流程

本研究的方法通过一个包含3个阶段的标准化工作流程来实现,如图1所示. 这3个阶段清晰地串联起了本研究方法的各个技术环节. 图1中的具体符号和算法将在第4.4节、第4.5节、第4.6节中详细介绍.

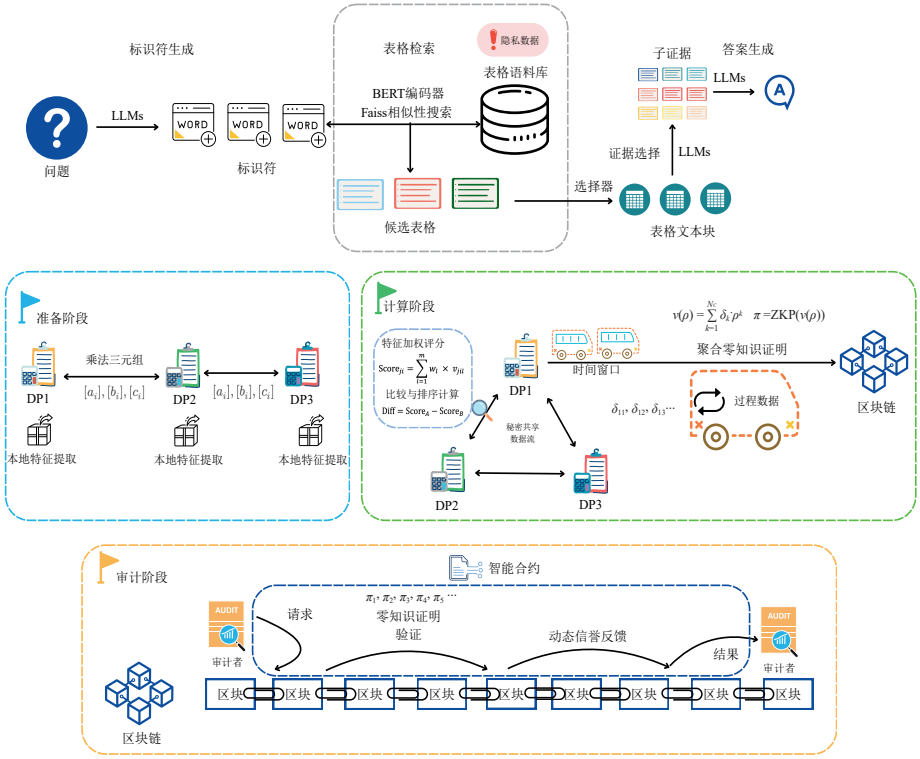


图1　基于MPC与聚合审计的多方私有表格筛选系统框架

准备阶段. 数据提供者本地提取特征,并生成后续MPC协议所需的密码学材料(如乘法三元组),分发给其他计算参与方.

计算阶段. 数据提供者执行MPC协议,协同完成特征加权评分、比较与排序等计算任务,即Top-K多方私有表格筛选过程. 同时,为保证计算的正确性和可被追溯,各方还需生成相应的密码学承诺与证明,并提交至区块链.

审计阶段. 审计者对计算过程的有效性证明进行公开验证,并根据审计结果,通过智能合约自动更新所有数据提供者的信誉评分.

这3个阶段的实现,分别依赖于本研究方法中的核心机制. 计算阶段的核心是“基于MPC的隐私保护筛选机制”(详见第4.4节),审计阶段的前半部分对应“可公开的聚合审计机制”(详见第4.5节),后半部分则对应“动态信誉反馈机制”(详见第4.6节). 通过这样的设计,本文将方法落地为具体的、紧密连接的技术实现.

4.3　安全目标与威胁模型

本方法致力于实现数据隐私性、计算正确性、公开可审计性及参与方行为的问责与激励. 本文假设一个标准的对抗模型,其中审计者和GIST核心模块在执行其既定功能时被认为是可信的,能够正确执行验证和信誉更新,但所有其他参与者可能会任意偏离协议. 数据提供方和数据用户可能会采取恶意行为,包括提交错误的份额、生成无效证明、与其他方合谋或试图推断他人的隐私输入. 本文不假设数据提供者或用户中存

在诚实的大多数. 本方法旨在通过聚合证明与激励机制,抵抗来自DPs和DUs的恶意行为及潜在信息的泄露风险,确保Top-K表格筛选过程的安全与可信.

4.4 面向Top-K私有表格筛选的MPC增强机制

传统的开放域问答系统,如GIST,在处理公开可用的表格数据时展现出卓越的性能. 然而,在许多现实应用中,有价值的表格数据往往分散在不同的DPs手中,并且由于包含商业机密或个人隐私,这些DPs不愿将其全部原始数据直接共享给中心化的问答系统或任何其他第三方. 这就对GIST这类系统在利用多方私有数据时提出了严峻的挑战,特别是在需要从所有潜在的私有表格中初步筛选出最相关候选的Top-K环节. 该环节因需要对最广泛的数据进行评估,是整个问答流程中数据访问最为密集、潜在隐私泄露风险最高的瓶颈. 为实现隐私保护下的Top-K表格筛选,本研究设计了一个包含基于MPC的表格相关性评分和基于MPC的Top-K选择两个核心阶段的Top-K私有表格筛选机制.

此阶段旨在为每个DP_j所拥有的每一个私有表格T_{ji},计算一个GIST系统生成的公开查询标识符Q_{ids}的相关性评分Score_{ji},整个计算过程严格保护T_{ji}的隐私.

4.4.1 基于MPC的表格相关性评分

参数初始化与本地特征提取. GIST的查询处理模块首先将用户问题转换为公开的查询标识符Q_{ids}. GIST的LLM模块根据Q_{ids}动态生成一组公开的特征权重$\boldsymbol{W}=\{w_1, w_2, \cdots, w_m\}$,这些权重反映了不同特征对于当前查询的重要性,并被广播给所有参与的DPs. 随后,每个DP_j对其本地持有的私有表格T_{ji}进行预处理. 针对一组预定义的、公开的打分函数$H=\{h_1, h_2, h_3\}$,DP_j在本地计算出其表格T_{ji}对应的特征向量$\boldsymbol{v}_{ji}=\{v_{ji1}, v_{ji2}, v_{ji3}\}$,其中$v_{jil}=h_l(T_{ji}, Q_{\mathrm{ids}})$. 对于打分函数的设计,本文与GIST原有筛选逻辑中考虑的因素保持一致,以确保筛选结果的有效性,具体包括以下几点.

h_1:语义相似度得分. DP_j本地计算其表格T_{ji}的标题和元数据的嵌入向量,并与公开的Q_{ids}的嵌入向量计算相似度,得到v_{ji1}.

h_2:关键词匹配度. DP_j本地统计其表格T_{ji}的文本描述列中包含Q_{ids}内关键词的归一化频次,得到v_{ji2}.

h_3:数值/条件满足度. DP_j本地判断其表格T_{ji}的数值是否满足Q_{ids}中隐含的数值范围约束或特定条件,输出一个布尔值(0或1)作为v_{ji3}.

特征值的秘密共享. 完成本地特征提取后,每个DP_j将其计算出的私有特征值$\boldsymbol{v}_{ji}$转换为秘密共享形式. 采用Shamir(K,N)门限秘密共享方案(其中,N是参与MPC的DPs总数,k是重构秘密所需的最小份额数),DP_j将$\boldsymbol{v}_{ji}$分割成N个份额,并将每个份额安全地分发给对应的DP. 至此,每个DP都持有了其他所有DPs(包括自己)的所有表格的所有特征值的一个份额.

MPC安全加权评分. 所有DPs共同为每个表格T_{ji}计算其加权总分Score_{ji}. 计算公式为

$$\mathrm{Score}_{ji}=\sum_{l=1}^{m} w_l \times v_{jil}$$

完成此阶段后,每个表格T_{ji}的最终相关性得分Score_{ji}仍以秘密共享的形式分布在所有参与的DPs之间,没有任何一个DP能够单独获知其他DP表格的准确得分.

4.4.2 MPC安全Top-K选择

在所有表格的秘密共享得分Score_{ji}计算完毕后,DPs继续通过MPC协议协同选出得分最高的Top-K个表格,同样不泄露任何单个表格的具体得分.

安全多方比较与排序. DPs共同逐个计算Score_{ji}的差值$\mathrm{Diff}=\mathrm{Score}_A-\mathrm{Score}_B$,这将确定Diff的符号位. 通过一系列这样的安全比较,可以实现对所有秘密共享得分Score_{ji}的完全排序,能够识别出Top-K个表格.

Top-K结果的确定与输出. 排序完成后,得分最高的K个秘密共享值所对应的表格索引即被确定下来. 这些索引以秘密共享的形式由DPs共同持有,并在需要时共同解密给GIST的LLM模块.

通过上述MPC赋能的筛选机制,GIST系统能够在不直接访问任何DP原始私有表格数据的前提下,获得一个与当前用户查询最相关的Top-K候选表格索引列表.

4.5 基于聚合证明的可公开审计机制

为确保在GIST框架下,通过MPC进行的Top-K多方私有表格筛选过程的计算正确性与DPs的行为诚实性,本研究引入了一种高效的公开可审计机制. 该机制的核心在于聚合证明技术,旨在显著降低验证大规模MPC计算的开销,同时保障审计的公开透明与结果的不可否认.

4.5.1 随机抽样和聚合证明

为确保MPC化Top-K表格筛选中核心评分计算$\left(\mathrm{Score}_{ji}=\sum_{l=1}^{m} w_l \times v_{jil}\right)$忠实于DPs所承诺的输入及公开算法,本框架集成了一种高效的公开可审计机制. 该机制采用聚合证明技术,以低开销实现大规模MPC计算的公开透明验证.

在本框架的MPC化表格相关性评分阶段,每个DP_j首先在本地为其私有表格T_{ji}计算出一组特征值$v_{jil}(l=1, 2, \cdots, m)$. 随后,$\mathrm{DP}_j$对这些私有特征值$v_{jil}$生成Pedersen承诺$C_{jil}=\mathrm{Com}(v_{jil}, r_{jil})$,并将这些承诺公开发布至区块链. 之后,DPs共同参与MPC协议,计算每个表格的加权总分Score_{ji},其对应的明文期望值为

$$\text{Score}_{ji}=\sum_{l=1}^{m}w_l\times v_{jil}$$

其中,w_l是公开的权重.

审计的目标是验证MPC协议计算出的最终秘密共享结果Score_{ji}(记其在后续零知识证明内部处理的对应值为 S_{ji})与基于DPs公开承诺的输入v_{jil}和公开算法计算出的期望值是否一致.

定义一个误差项δ_{ji}来表示这种一致性的偏差:

$$\delta_{ji}=S_{ji}-\sum_{l=1}^{m}w_l\times v_{jil}$$

在理想情况下,如果MPC协议被正确执行,并且所有DPs都忠实地使用了其承诺的v_{jil}值参与计算,那么所有δ_{ji}均应为0. 若MPC计算过程存在错误,或者有DP在MPC中使用了与其承诺不符的输入值,则至少有一个对应的$\delta_{ji}\neq 0$.

传统的审计方法需要对每一个δ_{ji}进行独立验证,这在涉及大量表格和多个DPs的场景下,证明生成和验证的开销将是巨大的. 为解决此问题,本研究采用聚合验证的思想. 审计者不再逐一验证每个$\delta_{ji}=0$,而是通过一个随机挑战将所有验证条件聚合成单个等式. 具体而言,审计者选择一个足够大的有限域F_ρ(其中ρ是一个大素数),并从中随机选取一个挑战值$\rho\in F_\rho$. 然后,审计者构建如下的聚合验证表达式(此处将索引定义为k,假设总共有N_c个需要审计的表格评分计算,每个计算对应一个δ_k):

$$V(\rho)=\sum_{k=1}^{N_c}\delta_k\times\rho^k$$

若所有计算均忠实于承诺的输入和公开算法,即所有$\delta_k=0$,则显然$V(\rho)=0$. 若存在至少一个计算使得$\delta_k\neq 0$,那么$V(\rho)$就构成了一个ρ关于次数最高为N_c的非平凡多项式. 根据Schwartz-Zippel引理,在一个大有限域中,一个非零多项式随机取值为0的概率极小,不超过N_c/ρ. 当ρ远大于N_c时,这个概率可以忽略不计,从而保证了审计的可靠性.

Pedersen承诺. 参与MPC计算的DPs需要对每一个独立的误差项δ_k生成Pedersen承诺. 由于Pedersen承诺的同态性,$\text{Com}(x;\ r)=g^x h^r$,可以计算出聚合承诺:

$$C_V=\text{Com}\left(\delta_k;\ r_k\right)^{\rho^k}=\text{Com}\left(\sum_{k=1}^{N_c}\delta_k\times\rho^k;\ \sum_{k=1}^{N_c}r_k\times\rho^k\right)$$

零知识证明构建. DPs为聚合误差值$V(\rho)$的正确性构建一个零知识证明π:

$$\pi=\text{ZKP}\left(V(\rho)\right)$$

该证明采用非交互式零知识证明(NIZKs)方案,通过Fiat-Shamir启发式算法生成. 此证明能够使验证者信服$V(\rho)$确实为0,而无需获知任何关于个体δ_k或其依赖的私有数据(如v_{jil}和S_{ji})的信息.

验证. 审计者从区块链中检索π及其承诺C_V,并通过计算以下公式验证该证明:

$$\text{VerifyZKP}(C_V,\pi)\rightarrow\{\text{accept, rejection}\}$$

如果被接受,则计算被视为正确;否则,审计者触发错误定位程序.

4.5.2　时间窗口与错误定位机制

为进一步降低与区块链交互的频率和链上存储开销,本文引入时间窗口机制. 在一个预设的时间窗口T内发生的所有表格评分计算(可能涉及多个用户查询或一批表格的处理)可以被聚合成一个批次,仅为这整个批次的计算生成一个聚合零知识证明,这种批处理方式显著提升了系统的整体效率和可扩展性:

$$V_T(\rho)=\sum_{i\in T}\delta_i\times\rho^i,\pi=\text{ZKP}\left(V(\rho)\right)$$

当聚合证明验证失败时,表明在该时间窗口或批次内至少存在一个计算错误或承诺与计算不一致的情况. 此时,需要启动错误定位机制. 本文采用基于二分法的批次错误定位方法:审计者请求将验证失败的批次递归地划分为更小的子批次,并对每个子批次重新进行聚合验证. 对于一个包含n_j个操作的子批次j,计算其子聚合验证$V_{\text{subj}}(\rho)$并验证对应的证明π_{subj}. 这一过程持续进行,直至定位到一个或多个包含错误的计算单元和具体的DP. 这种分层验证方法在保持隐私性的同时,有效地平衡了审计成本与错误追溯的精确度,为后续的信誉调整提供了依据. 审计的可能结果(通过、失败并定位错误、超时)及其对信誉的影响可参考图2.

4.6　动态信誉反馈机制

为有效激励DPs在集成MPC的Top-K表格筛选过程中的诚实行为,并对潜在的计算错误或恶意操作施加约束,本框架设计并实现了一个基于区块链的动态信誉反馈机制. 该机制根据公开可审计的计算验证结果,持续评估并调整各参与方的信誉得分,从而提升系统的整体鲁棒性、公平性和参与意愿. 本节的符号说明如表1所示.

4.6.1　信誉调整计算

信誉调整的核心在于将审计阶段输出的验证结果(如计算正确、计算错误、超时未响应)量化为对参与方信誉值的具体影响. 首先定义一个基础信誉调整值$\Delta R_i'$,它根据审计结果采用分段函数进行计算:

$$\Delta R_i'=\begin{cases}\gamma\times T_i\ , & \text{审计通过}\\ -\lambda\times T_i\times\dfrac{e_i}{n_i}\ , & \text{审计失败}\\ -\theta\times T_i\ , & \text{审计超时}\end{cases}$$

若审计通过,参与方获得正向信誉激励,鼓励其持续的诚实行为. 若审计失败,则根据错误操作的比例

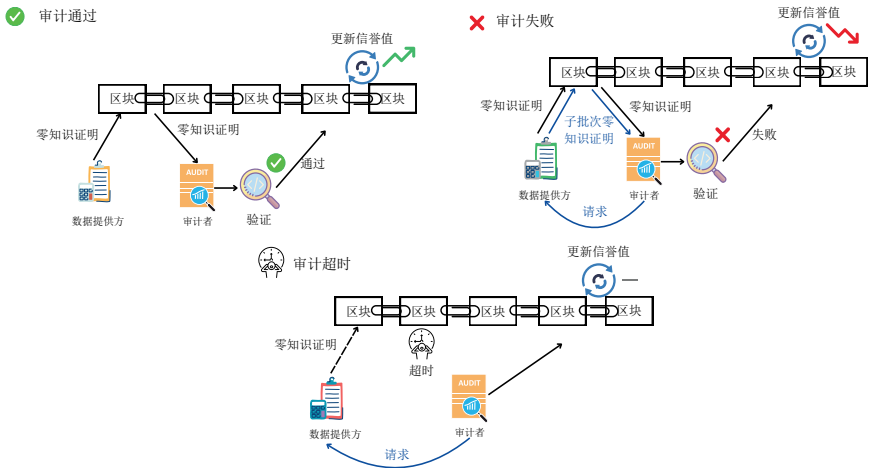


图2 审计结果情况示意图

表1 动态信誉反馈机制的符号和定义

符号	定义
R_i^{old}	参与者i的历史信誉(更新前)
R_i^{new}	参与者i的更新信誉(更新后)
n_i	参与者i在当前窗口中执行的操作数
e_i	参与者i的错误操作数
T_i	分配给参与者i任务的重要性权重
γ	奖励系数(正向激励)
λ	惩罚系数(惩罚错误)
θ	超时惩罚因子$(0<\theta<1)$
$\bar{R}$	所有参与者的平均信誉
η_1,η_2	性能比较的敏感度系数
β	信誉平滑因子$(0\leqslant\beta\leqslant 1)$
$\Delta R_i'$	参与者i的基础信誉调整
ΔR_i	参与者i的最终信誉调整

$\frac{e_i}{n_i}$施加相应惩罚,确保惩罚的公平性与对等性. 若审计超时,则施加一个相对温和的惩罚,以区分主观恶意与客观故障.

4.6.2 基于整体表现的信誉调整因子

为更全面地反映参与方的相对表现,并避免个体在孤立事件中信誉受到过度影响,本文引入一个基于全局平均信誉的性能调整因子F_i. 该因子将参与方i的历史信誉R_i^{old}与所有参与方的平均信誉$\bar{R}$进行比较:

$$F_i=\begin{cases}1+\eta_1\left(\dfrac{\bar{R}-R_i^{\text{old}}}{\bar{R}+1}\right), & \text{审计通过}\\ 1+\eta_2\left(\dfrac{R_i^{\text{old}}-\bar{R}}{\bar{R}+1}\right), & \text{审计失败/超时}\end{cases}$$

最终的信誉调整值ΔR_i由基础信誉调整$\Delta R_i'$与性能调整因子F_i共同决定:

$$\Delta R_i=\Delta R_i'\times F_i$$

4.6.3 带平滑的信誉更新

为防止参与方信誉因短期行为波动而产生剧烈变化,保持信誉系统的长期稳定性与可信度,本文采用指数平滑的方式更新参与方的信誉. 参与方i的新信誉值R_i^{new}由其旧有信誉R_i^{old}和最终信誉调整值ΔR_i加权计算得出:

$$R_i^{\text{new}}=(1-\beta)\times R_i^{\text{old}}+\beta\times\Delta R_i$$

本文提供了完整的动态信誉反馈机制算法,如算法1所示.

算法1 动态信誉反馈算法

输入: $R_i^{\text{old}},n_i,e_i,T_i,\gamma,\lambda,\theta,\eta_1,\eta_2,\beta$

输出: R_i^{new}

计算 $\bar{R}=\dfrac{1}{M}\sum_{i=1}^{M}R_i^{\text{old}}$

FOR 每个参与者i DO

  IF 审计通过 THEN

    $\Delta R_i'=\gamma\times T_i$;

    $F_i=1+\eta_1\left(\dfrac{\bar{R}-R_i^{\text{old}}}{\bar{R}+1}\right)$;

  ELSE IF 审计失败 THEN

    $\Delta R_i'=-\lambda\times T_i\times\dfrac{e_i}{n_i}$;

    $F_i=1+\eta_1\left(\dfrac{R_i^{\text{old}}-\bar{R}}{\bar{R}+1}\right)$;

  ELSE IF 审计超时 THEN

    $\Delta R_i'=-\theta\times T_i$;

    $F_i=1+\eta_1\left(\dfrac{R_i^{\text{old}}-\bar{R}}{\bar{R}+1}\right)$; $\Delta R_i=\Delta R_i'\times F_i$

  $R_i^{\text{new}}=(1-\beta)\times R_i^{\text{old}}+\beta\times\Delta R_i$

RETURN R_i^{new}

4.7 安全性分析

本研究提出的增强方法通过采用面向Top-K私有表格筛选的MPC增强机制、基于聚合证明的可公开审

计机制和动态信誉反馈机制构建了一个多层次的安全保障体系. 本节将依据第4.3节设定的安全目标与威胁模型,从数据隐私性、计算可信性与系统鲁棒性3个层面,对增强方法的安全性进行深入剖析.

本文的增强方法的核心安全保障是数据隐私性,其实现主要依赖于基于秘密共享的多方安全计算协议. 在Top-K筛选流程启动时,各DP仅将本地提取的私有特征向量转换为秘密份额后参与后续计算. 这一机制确保了任何原始表格数据及其衍生特征值自始至终不离开数据持有方的本地环境,并且任何计算中间值(如相关性得分)均以秘密共享形式存在. 因此,单个参与方或任何未达到预设门限的共谋方联盟,均无法重构出任何关于其他方私有输入的有效信息,从而为多方协作提供了坚实的隐私保护基础.

在保障数据隐私的基础上,增强方法通过可公开聚合审计机制确保了计算过程的可信性,即可验证的正确性、可审计性与不可否认性. 该机制利用区块链作为不可篡改的公共账本,要求所有参与方首先以Pedersen承诺的形式,将其用于计算的私有输入的密码学承诺公开发布. MPC计算完成后,各方并非对每个计算独立生成证明,而是协同生成一个聚合的非交互式零知识证明. 该聚合证明引入了一个由审计者提供的公共随机挑战值ρ,将计算的误差项δ构造成一个关于ρ的多项式$V(\rho)$. 各方需证明该多项式在挑战值ρ下的结果为0. 这一设计的安全性基于Schwartz-Zippel引理:若计算过程中存在至少一个错误(即至少有一个$\delta=0$),那么$V(\rho)$是一个非零多项式,其在一个从足够大有限域中随机选取的点ρ上求值为0的概率可以忽略不计. 这种聚合形式能够有效抵御共谋攻击,即使多个恶意参与方合谋,试图构造一组精心设计的错误,使得它们聚合后的δ为0,也几乎不可能在不知道随机挑战值ρ的情况下,使这个加权多项式$V(\rho)$的结果也为0. 这使得任何偏离协议的计算行为,无论是源于单方错误还是多方共谋,都将以极高的概率被检测出来. 零知识证明的可靠性进一步确保了恶意方无法伪造一个有效的证明. 同时,所有密码学证据均锚定在区块链上,构成了公开、透明且不可抵赖的审计轨迹,解决了分布式计算中的信任问题.

最后,本方法通过基于审计结果的动态信誉反馈机制,确保了系统的鲁棒性与对恶意行为的抵抗能力. 该机制与公开审计环节紧密耦合,将计算验证的结果转化为对参与方信誉的直接影响. 当聚合证明验证失败时,错误定位程序将追溯至具体的恶意或故障参与方. 随后,部署在区块链上的智能合约将自动执行预设的奖惩逻辑,对不诚实行为(如提交与承诺不符的输入)或拒绝协作的行为施加信誉惩罚. 这种将密码学证明与奖惩机制相结合的设计,有效地将参与方的行为约束在协议规范之内,形成强大的博弈威慑,从而抑制潜在的恶意企图,保障了整个协作生态系统的长期稳定与公平.

5　实验分析与评估

5.1　Top-K表格筛选一致性评估

本文设置了实验以验证本方法中采用MPC进行Top-K私有表格筛选的可行性,评估其筛选结果与GIST框架在处理公开数据时原有筛选方法(下称“基准筛选方法”)的一致性,以证明不会对开放域问答任务的端到端性能产生影响. 其中,MPC加权评分与安全选择部分是在单机完成的模拟实验,不考虑通信开销. 实验设置如表2所示.

表2　Top-K表格筛选一致性实验设置

参数	描述
GIST方法	BERT编码器 Faiss相似性搜索
本文的方法	嵌入编码特征、BM25特征、数值布尔特征 MPC加权评分和安全选择
硬件参数	8核,16 GB Apple M3芯片
数据集	NQ-TABLES(100个问题) OTT-QA(100个问题)
评估指标	Top-50召回率,平均Jaccard指数
Top-50召回率	对每个问题,本文的方法与GIST方法选出的Top-50表格交集大小/50
Jaccard指数	对每个问题,本文的方法与GIST方法选出的Top-50表格交集大小/并集大小

结果如表3所示,在GIST方法评估所使用的两个数据集NQ-TABLES[4]和OTT-QA[25]上,对于大多数问题查询,两种方法选择的候选表格高度重合. 这为后续GIST的LLM模块提供了质量稳定且可靠的输入,从而可以说明在不显著影响开放域问答任务端到端性能的前提下,实现了对多方私有数据的安全利用.

表3　Top-K表格筛选一致性实验结果

评估指标/数据集	NQ-TABLES	OTT-QA
平均Top-50召回率	0.92	0.90
平均Jaccard指数	0.85	0.81

5.2　Top-K表格筛选性能评估

为继续深入探究增强方法中采用MPC进行Top-K私有表格筛选在多方场景下的实用性,本文设计并实现了针对Top-K私有表格筛选中关键操作的性能评估实验. 基于SPDZ协议的核心思想,在本地回环中通过多个进程模拟多个参与方,对Top-K私有表格筛选这一代表性瓶颈操作进行了量化测试. 其中包含了完整的Shamir秘密分享、安全乘法协议与安全比较协议. 实验设置如表4所示.

表4 Top-K表格筛选性能试验设置

参数	描述1	描述2
MPC计算参与方数	{2, 3, 5, 10, 15}	5
数据量	{2, 3, 5, 10, 15}	{25, 50, 100, 500, 1 000}
硬件参数	8核,16 GB Apple M3 芯片	8核,16 GB Apple M3 芯片
数据集	NQ-TABLES	NQ-TABLES
评估指标	平均耗时(s),通信轮次(轮)	平均耗时(s),通信轮次(轮)

结果如图3所示,采用MPC进行Top-K表格筛选是存在一定基线开销的,为满足安全与隐私需求,MPC在计算Diff的最高有效位时需要执行较多计算次数并进行通信.在真实的多方网络环境中,通信开销是影响MPC协议性能的主要瓶颈.根据对广泛使用的MPC框架(如SPDZ)在广域网(Wide Area Network,WAN)环境下的性能评测[26],执行一次涉及多方通信和大量比较操作的安全计算,其延迟通常在数秒至数分钟的范围内,其中主要取决于参与方数量与通信条件.这也与本文的实验结果相匹配.本实验的环境更接近数据中心或局域网(Local Area Network,LAN)环境,在较优的网络环境下,15个参与方参与的情况需约5 s,在WAN或互联网(WAN)环境中延迟会有明显提升,以约30 ms作为平均值,随着参与方的增加(2~15方),延迟会在2~28 s不等.同时,我们也注意到当需要筛选的数据总量从25增加到5 000时,平均耗时从约0.79 s上升至约15.36 s.通信轮次的显著增加也再次说明了性能瓶颈主要取决于通信条件,大量通信导致耗时增加.这也是因为在Top-K表格筛选环节所有表格的特征都需要进行安全排序.因此,无论是增加参与方数量(横向扩展),还是增加单个参与方的表格数量(纵向扩展),最终都会导致Top-K表格筛选阶段的规模变大,其共同测试了Top-K表格筛选模块的处理能力.

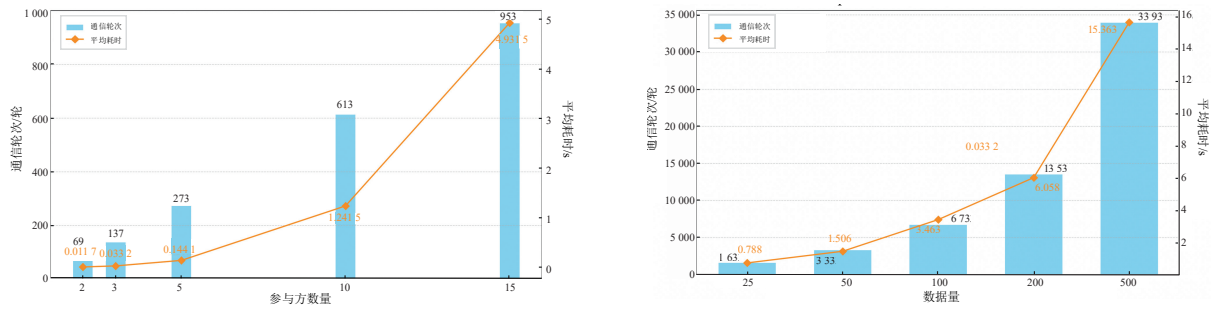


图3 Top-K表格筛选性能测试分析图

在完成实验评估时,针对海量数据也需要将此环节的性能开销置于整个ODQA框架下进行评估.正如同在GIST框架中进行的说明,Top-K表格筛选本身就是应对海量表格数据的第一道关口,其核心价值在于数据削减.本方案通过在这一阶段付出可控的隐私成本,可以将一个潜在规模极其庞大的私有表格库,高效地缩减为一个规模可控的候选集.这极大地减轻了后续的LLM精排和答案生成模块的处理负担,从而保障了整个端到端问答流程在面对海量数据时的可行性.

总体来看,通过本实验研究量化了隐私成本,这是为保证隐私和可信的核心目标所需要付出的性能成本,这也为在实际部署时进行成本效益分析提供了数据依据,需要考虑到任务执行涉及的参与方数量与数据量.ODQA任务不是一个强实时任务,在控制数据参与方和数据量的情况下,一定的性能成本是可接受的.

5.3 聚合审计机制性能评估

本文比较了所提出的基于聚合证明的可公开审计机制与个体审计机制[22]在证明生成时间方面的性能差异.实验设置如表5所示.

如图4所示,对于所提出的聚合证明机制,在时间窗口大小固定为50的情况下,证明生成时间保持稳定

表5 聚合审计机制性能实验设置

参数	描述
证明系统	Groth16
证明电路	Circom
挑战参数	随机选择
时间窗口大小	50(固定)
运算操作数	{50, 100, 200, 250, 500}
证明生成机制	聚合证明 VS 独立证明
时间计算方法	基于窗口的证明时间总和

在每个窗口255~272 ms之间. 在不同任务大小之间,每个窗口的生成时间差异较小,总生成时间随窗口数量的增加而线性增加. 相比之下,传统的个体审计机制的总生成时间随单个计算(如乘法)数量的增加而线性增加. 随着操作数量的增加,这种性能下降变得更加明显.

本文同时对时间窗口大小进行了实验,虽然较大的窗口有助于减少证明数量,但累积的错误需要频繁进行批量验证,从而增加证明生成时间. 经过实验,窗口大小为150被认为是最有效的配置,它在最小化证明数量和减少过多的批量验证开销之间取得了平衡.

结果表明,所提出的基于随机抽样的审计机制显著减少了证明生成和验证时间,尤其是在大规模多方安全计算系统中. 它在保持可公开审计和隐私保护的同时,实现了更好的可扩展性和更高的效率.

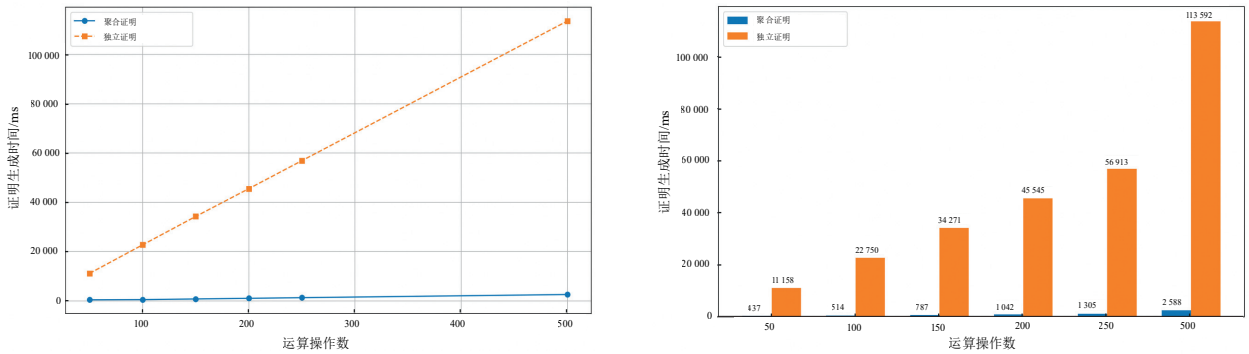


图4　聚合证明VS独立证明时间性能分析图

5.4　智能合约与动态信誉反馈机制评估

5.4.1　智能合约吞吐量测试

实验设置如表6所示.

表6　智能合约与信誉反馈机制评估实验设置

参数	描述
硬件参数	8核,16 GB Apple M3 芯片
实现语言	Python、Go
区块链平台	Hyperledger Fabric
性能指标	吞吐量、吞吐量变化率
并发请求数量	{250, 500, 750, 1 000, 1 250, 1 500}
吞吐量定义	每秒交易数(TPS)
吞吐量变化率定义	级别k的吞吐量/级别$k-1$的吞吐量

结果如图5所示,吞吐量随着并发度的增加略有下降,但仍在可接受的范围内,平均约为66笔/s. 吞吐量变化率始终接近1,最低为0.94,表明合约在高并发度下性能稳定.

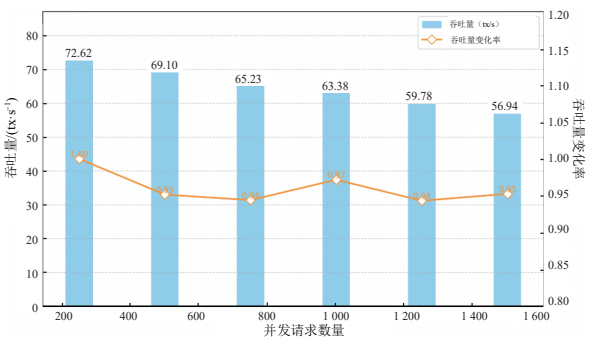


图5　不同并发请求数量下信誉更新智能合约的吞吐量分析图

5.4.2　追赶测试

本文模拟了3位行为各异的数据提供者,以评估信誉反馈机制的追赶能力,实验设置如表7所示.

表7　追赶测试实验设置

参数	描述
核心参数	$\gamma=0.2, \lambda=0.4, \theta=0.1,$ $\beta=0.6, \eta=0.5$
任务重要性权重	$T_i \in \{1, 2, 3, 4, 5\}$
初始信誉	$R_{P1}=50$, $R_{P2}=45$, $R_{P3}=50$
特殊条件	P2在第7个审核窗口加入

P1:早加入者,表现平平,经常性不参与数据提供.

P2:晚加入者,表现优秀,积极参与到各类问题的数据提供中.

P3:早加入者,行为恶意.

如图6所示,尽管P2加入较晚,但由于积极主动地参与任务,其信誉迅速提升. 反馈机制对审计失败施加了严厉的惩罚,对超时施加了较温和的惩罚,同时有效地奖励了准确执行任务的行为. 因此,信誉反馈机制能够成功激励诚实和积极主动的行为,从而保障系统的长期稳健性和公平性.

5.5　端到端整体性能评估

基于以上实验部分中本文对MPC核心组件、聚合审计、信誉机制智能合约的实验数据,本节将主要针对整个增强方法的端到端性能进行评估. 系统的总时延主要由两部分构成:相对固定的LLM调用开销和审计开销、随参与方数量变化的MPC计算开销. 为精确评

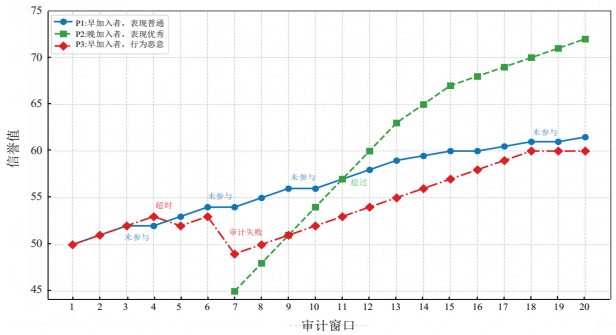


图6 追赶测试中各参与者信誉动态情况图

估总时延，本文将用户感知的问答流程分解为以下4个主要阶段.

$T_{\text{llm\_retrieval}}$（LLM检索增强耗时）. 此阶段包含GIST框架中进行标识符生成（Identifier Generation）和证据块选择（Selector）所需的两次LLM API调用.

$T_{\text{llm\_reader}}$（LLM答案生成耗时）. 此阶段为GIST框架最终生成答案所需的LLM API调用，通常因处理的上下文更长而耗时更多.

T_{mpc}（采用MPC进行Top-K私有表格筛选）. 其耗时采用第5.2节的实验数据.

$T_{\text{audit\_reputation}}$（链上审计与信誉更新耗时）. 此阶段包含聚合审计和动态信誉阶段所需耗时，它在答案生成后异步触发，完成聚合证明并提交至区块链，由智能合约完成验证和信誉更新. 其耗时采用第5.3节和第5.4节的实验数据.

在GIST的论文中讨论了不同LLM（GPT-3.5、GPT-4、LLAMA）对性能的影响，其中主要受API调用影响，使用本地部署的模型（如LLAMA）则可以消除账户响应时间的干扰来减少问答延迟，同时本文限制了LLM的深度思考. 在第5.2节的相同硬件环境下，针对NQ-TABLES数据集中的问题样本，对LLM调用耗时进行了测量，得到的平均值分别为 $T_{\text{llm\_retrieval}} = 1.8$ s，$T_{\text{llm\_reader}} = 2.5$ s. 本文取针对NQ-TABLES数据集中的问题样本所需要的平均证明生成与验证时间以及更新信誉合约所需执行时间 $T_{\text{audit\_reputation}} = 2.1$ s. 将上述测量值进行合并，可以得到端到端性能预估，相对固定的LLM调用开销和审计开销约为6.4 s，故端到端性能主要仍旧取决于随参与方数量、数据量变化的MPC计算开销，而该计算开销主要取决于参与方间的通信开销.

其中，可公开聚合审计机制是一项关键的性能优化. 传统的可信计算方案若要实现公开可审计，通常需要对每一次计算生成独立的零知识证明. 如本文第5.3节的性能评估所示，本方案的聚合审计机制通过时间窗口处理和聚合验证，与生成独立证明相比，节省了约87%的证明时间. 同时，第5.4节的测试也表明，信誉更新智能合约具有高吞吐量和稳定性，确保了该环节不会成为系统瓶颈. 因此，本方案在引入可信保障的同时，已解决了大规模审计所带来的部分性能挑战.

总体来看，本文的增强方案仍旧保有可行性，尤其是在参与方数量较少、通信环境较优的场景下，系统的端到端响应时间能够控制在10 s以内，可以做到实时的问答. 在参与方增加、通信环境恶化的情况下，通过异步问答的模式，这一性能水平对于许多专业领域的问答应用，如企业内网知识库、金融投研、医疗辅助诊断等，是完全可以接受的，因为在这些场景下，数据的安全性、隐私性与结果的可追溯性远比实时响应更为重要. 具体来看，本文的增强方法在实际应用中的可接受性，可以从以下3个层面进行分析.

首先，本文的增强方案的首要目标并非追求极致的实时响应，而是在多方协作场景下，为因数据主权、商业机密或个人隐私法规（如《中华人民共和国数据安全法》）而无法流通的数据提供一套安全可信的解决方案. 在金融、医疗、政务等高度敏感的领域，直接共享原始数据进行中心化计算往往是不可行的. 因此，数秒的延迟是为实现数据隐私保护、计算过程可审计与结果可追溯性这些核心价值而付出的一定的隐私成本，它使得过去无法实现的安全数据协作成为可能.

其次，本文的增强方案的典型应用并非面向公众的高并发、强实时问答，而是服务于专业领域的深度知识探索，例如企业内网的联合知识库查询、跨机构的金融投研数据分析、多中心医疗数据的辅助诊断等. 在这些场景中，用户对答案的准确性、可靠性和数据来源的可公开性、可追踪性要求极高，远超对响应速度的需求. 对于一个需要严谨推理和证据支撑的复杂问题，用户心理预期与执行一次复杂的数据库查询或数据分析任务类似，数十秒的响应时间在专业工作流中是完全可以接受的.

最后，系统交互模式本身也存在一定的工程优化潜力. 在现实系统部署中，可通过引入异步问答机制来优化用户体验. 在用户提交问题后，系统可在后台执行MPC筛选和LLM生成任务，并在计算完成后，通过消息推送、邮件通知或页面通知等方式将最终答案呈现给用户. 这种交互模式将前端用户感知与后端实际计算耗时完全解耦，从而有效规避了MPC带来的延迟对用户体验的直接影响.

综上所述，本文的增强方法在性能上的考量是多方面的. 一方面，本文接受了为实现核心隐私目标所必需的MPC性能开销，这是为安全隐私保障能力所付出的一定的隐私成本. 另一方面，本文创新性地设计了高效的可公开聚合审计等机制，以解决部分可信保障环节的性能瓶颈. 结合其目标应用场景的特点、价值权衡以及可能的灵活交互设计，本文的增强方法在现实世

界中具有部署可行性与应用价值.

6　结论与未来工作

本研究针对LLM驱动的ODQA系统在处理多方私有表格数据,特别是在Top-K候选筛选环节所面临的隐私、可信与效率挑战,提出了一种集MPC、可公开聚合审计及动态信誉机制于一体的增强方法. 该方法通过基于MPC的Top-K表格筛选机制保护了数据隐私,利用聚合审计高效验证了计算正确性,并通过动态信誉系统激励了诚实参与方. 实验评估验证了本文方法在筛选结果一致性、审计效率和信誉机制有效性方面的优势,证明了其在保障LLM多方协作问答安全可信方面的应用潜力. 未来研究可进一步深化本文方法与LLM核心模块的融合,探索端到端的可信问答链路. 同时,可在不牺牲隐私与安全的前提下,通过引入更先进的MPC协议来进一步优化性能,以支持更快、更大规模的协作.

参考文献

[1] JIN N Z, SIEBERT J, LI D F, et al. A survey on table question answering: Recent advances[M]//Knowledge Graph and Semantic Computing: Knowledge Graph Empowers the Digital Economy. Singapore: Springer Nature Singapore, 2022: 174-186.

[2] LIANG X Y, HU R, LIU Y, et al. Open-domain question answering over tables with large language models[C]//Advanced Intelligent Computing Technology and Applications. Singapore: Springer, 2024: 347-358.

[3] HARMELINK R, JOOSTEN R, TOPAN E, et al. Data: To share or not to share? A Semi-Systematic Literature Review in (rational) data sharing in inter-organizational systems[J]. Discover Data, 2024, 2(1): 13.

[4] HERZIG J, MÜLLER T, KRICHENE S, et al. Open domain question answering over tables via dense retrieval[EB/OL]. (2021-06-09)[2025-05-21]. https://arXiv.org/abs/2103.12011.

[5] YANG Y H, WU J, LONG C N, et al. Blockchain-enabled multiparty computation for privacy preserving and public audit in industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 9259-9267.

[6] KELLER M. MP-SPDZ: A versatile framework for multi-party computation[C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2020: 1575-1590.

[7] LAKHANPAL S, KUMAR S V, RAJ V H, et al. Machine learning-enhanced blockchain solutions for sustainable energy consumption and carbon footprint tracking[C]//2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0. Piscataway: IEEE, 2024: 1-6.

[8] TAMMINA M R, POSINASETTY B, NAIR P S, et al. Machine learning enabled healthcare balancing patient privacy and data utility[C]//2024 Ninth International Conference on Science Technology Engineering and Mathematics. Piscataway: IEEE, 2024: 1-6.

[9] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and open problems in federated learning[J]. Foundations and Trends in Machine Learning, 2021, 14(1/2): 1-210.

[10] ZIEMS N, YU W H, ZHANG Z H, et al. Large language models are built-in autoregressive search engines[EB/OL]. (2023-05-16)[2025-05-21]. https://arXiv.org/abs/2305.09612.

[11] CHENG Z J, XIE T B, SHI P, et al. Binding language models in symbolic languages[EB/OL]. (2023-03-01)[2025-05-21]. https://arXiv.org/abs/2210.02875.

[12] LIU Y, HASHIMOTO K, ZHOU Y B, et al. Dense hierarchical retrieval for open-domain question answering[EB/OL]. (2021-10-28)[2025-05-21]. https://arXiv.org/abs/2110.15439.

[13] WU Y X, ZHAO Y, HU B T, et al. An efficient memory-augmented transformer for knowledge-intensive NLP tasks[EB/OL]. (2022-10-30)[2025-05-21]. https://arXiv.org/abs/2210.16773.

[14] DUA D, GUPTA S, SINGH S, et al. Successive prompting for decomposing complex questions[EB/OL]. (2022-12-08)[2025-05-21]. https://arXiv.org/abs/2212.04092.

[15] DESHMUKH P, CARTER B. Secure multi-party computation protocols for privacy-preserving data analysis[J]. International Journal of Recent Advances in Engineering and Technology, 2023, 12(2): 1-6.

[16] LAVIN R, LIU X K, MOHANTY H, et al. A survey on the applications of zero-knowledge proofs[EB/OL]. (2024-08-01)[2025-05-21]. https://arXiv.org/abs/2408.00243.

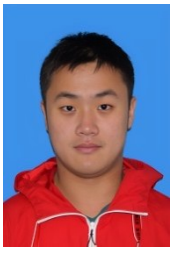
[17] ZHU R Y, DING C C, HUANG Y. Efficient publicly verifiable 2PC over a blockchain with applications to financially-secure computations[C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019: 633-650.

[18] CORDI C, FRANK M P, GABERT K, et al. Auditable, available and resilient private computation on the blockchain via MPC[M]//Cyber Security, Cryptology, and Machine Learning. Cham: Springer International Publishing, 2022: 281-299.

[19] KANJALKAR S, ZHANG Y, GANDLUR S, et al. Publicly Auditable MPC-as-a-Service with succinct verification and universal setup[C]//2021 IEEE European Symposium on Security and Privacy Workshops. Piscataway: IEEE, 2021: 386-411.
[20] SEO M. Fair and secure multi-party computation with cheater detection[J]. Cryptography, 2021, 5(3): 19.
[21] JIANG Y B, ZHOU Y, FENG T. A blockchain-based secure multi-party computation scheme with multi-key fully homomorphic proxy re-encryption[J]. Information, 2022, 13(10): 481.
[22] JIN S, LI Y, CHEN X, et al. Blockchain based publicly auditable multi-party computation with cheater detection[C]//Information and Communications Security. Singapore: Springer, 2023: 608-626.
[23] PEDERSEN T P. Non-interactive and information-theoretic secure verifiable secret sharing[C]//Advances in Cryptology-CRYPTO'91. Berlin: Springer, 1992: 129-140.
[24] WENG C K, YANG K, KATZ J, et al. Wolverine: Fast, scalable, and communication-efficient zero-knowledge proofs for Boolean and arithmetic circuits[C]//2021 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2021: 1074-1091.
[25] CHEN W H, CHANG M W, SCHLINGER E, et al. Open question answering over tables and text[EB/OL]. (2021-02-10)[2025-05-21]. https://arXiv.org/abs/2010.10439.
[26] WANG H Z, YANG Y J, WANG E, et al. Bilateral privacy-preserving worker selection in spatial crowdsourcing[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(3): 2533-2546.

作者简介

胡 睿 男,2001年8月出生于北京市. 现为北京邮电大学人工智能专业硕士研究生. 主要研究方向为区块链、隐私保护和数字资产流通.
E-mail: hurui@bupt.edu.cn



吴 昊 男,1996年10月出生于北京市. 现为北京邮电大学人工智能专业博士研究生. 主要研究方向为区块链、共识算法和多方安全计算.
E-mail: wuhaodoc@bupt.edu.cn

潘宇轩 男,1997年12月出生于广东省广州市. 现为北京邮电大学信息与通信工程专业博士研究生. 主要研究方向为沉浸式多媒体传输与处理、数字图像三维建模.
E-mail: panyx@bupt.edu.cn

张 琳 男,1974年7月出生于山东省济南市. 现为北京市大数据中心主任,北京邮电大学兼职教授、博士生导师. 主要研究方向为大数据与人工智能、网络信号处理等. 曾多次荣获中国电子学会科技进步奖、北京市高校优秀教学成果奖.
E-mail: zhanglin@bupt.edu.cn

刘 雨 女,1978年10月出生于山东省莱阳市. 现为北京邮电大学人工智能学院副教授、博士生导师. 主要研究方向为图像处理、通信网理论和天地一体化信息网络等. 2013年获北京市教育工会组织的北京市高校教学基本功比赛二等奖,入选首批北京高等学校"青年英才计划". 发表学术论文90余篇.
E-mail: liuy@bupt.edu.cn

朱孔林 男,1985年11月出生于山东省临沂市. 现为北京邮电大学人工智能学院教授、博士生导师. 主要研究方向为车联网、边缘计算、隐私计算和数字资产流通等. 主持国家重点研发计划青年科学家项目、国家自然科学基金、装备发展教育部联合基金、北京市自然科学基金等项目20余项,发表学术论文60余篇.
E-mail: klzhu@bupt.edu.cn