(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116451277 B
(45)授权公告日 2023.09.29

(21)申请号 202310718356.8

(22)申请日 2023.06.16

(65)同一申请的已公布的文献号
申请公布号 CN 116451277 A

(43)申请公布日 2023.07.18

(73)专利权人 中用科技有限公司
地址 230601 安徽省合肥市经济技术开发区宿松路3963号智能装备科技园E栋12层

(72)发明人 胡增 江大白 彭鹏

(74)专利代理机构 北京国源中科知识产权代理事务所(普通合伙) 16179
专利代理师 王少勇

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 21/60(2013.01)
G06F 16/2458(2019.01)
G06F 16/2455(2019.01)
G06F 16/27(2019.01)
G06N 3/098(2023.01)
H04L 9/00(2022.01)
H04L 67/104(2022.01)
H04L 67/12(2022.01)

(56)对比文件
US 2020193292 A1,2020.06.18
CN 115510494 A,2022.12.23
CN 116261717 A,2023.06.13
CN 105849749 A,2016.08.10
漆桂林;高桓;吴天星.知识图谱研究进展.情报工程.2017,(01),全文.

审查员 叶珊

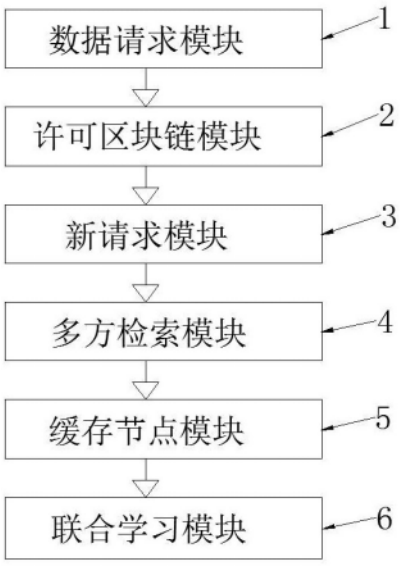
权利要求书3页 说明书8页 附图1页

(54)发明名称

一种基于共享情况下工业数据安全的保护系统及方法

(57)摘要

本发明公开了一种基于共享情况下工业数据安全的保护系统及方法,该基于共享情况下工业数据安全的保护系统包括:数据请求模块、许可区块链模块、新请求模块、多方检索模块、缓存节点模块及联合学习模块;其中,所述数据请求模块,用于接收来自数据请求者的数据共享请求,并将其记录在所述许可区块链模块中。本发明实现了CPU的高效利用,既实现了高效的数据访问也没有过多的消耗CPU,通过分布式多方共享数据,从而降低数据泄露的风险,使得数据所有者可以通过该架构进一步控制对共享数据的访问,并且将差异隐私集成到联邦学习中,进而可以进一步的保护数据隐私。



CN 116451277 B

1.一种基于共享情况下工业数据安全的保护系统,包括:数据请求模块、许可区块链模块、新请求模块、多方检索模块、缓存节点模块及联合学习模块;

其中,所述数据请求模块,用于接收来自数据请求者的数据共享请求,并将其记录在所述许可区块链模块中;

所述许可区块链模块,用于通过加密记录建立安全连接,使用许可的区块链来管理数据的可访问性和共享事件,并跟踪数据的使用情况;

所述新请求模块,用于接收管理后的数据请求者的共享请求,并将其转发给适当的超级节点进行处理;具体包括:

各方在许可区块链中注册并上传重新评估记录;

数据请求者向附近的超级节点SNreq启动包含一组查询Fx={f1,f2,…,fx}共享请求Req;

当数据请求者发起共享请求Req时,它将请求提交给其附近的超级节点SNreq;

SNreq首先搜索区块链以确定请求之前是否被处理过;

若有查找命中,则将之前计算的缓存全局数据模型直接返回给数据请求者;

否则,节点SNreq通过多方数据检索过程,并在区块链中查找相关节点;

所述多方检索模块,用于执行多方数据检索过程,根据注册记录查找相关方,并将共享请求转发给适当的超级节点进行处理;

所述缓存节点模块,用于将已经处理过的数据共享请求和共享结果存储在本地缓存中,并为后续请求提供结果;

所述联合学习模块,用于训练全局数据模型,并对共享请求提供预测和响应;

通过利用图来表示原始数据以供进一步处理,并保留更多的结构和上下文信息,具体包括:

加权图G={V,E}包括一组节点V和一组边$E \subseteq V \times V$;

每个节点n_i包含一个文本术语及其权重w_{ni}(n_i,w_{ni});

每个边e_{ij}连接节点n_i和n_j,其中边的权重是w_{eij};

通过权重矩阵A=[a_{ij}]来表示图,其中$a_{ij}=w_{ni}$;

如果i=j,则$a_{ij}=w_{eij}$;

如果i≠j,则利用术语频率逆文档频率来构建图,并将其序列化为序列向量;

通过将图序列化为线性向量,使用Jaccard相似性作为距离函数;

用k-均值算法对文档进行聚类,并根据文本相似性和参与的用户将数据集{D1,...,Dn}聚类为各种类别;

其特征在于:

只有委员会节点参与训练全局数据模型,而非委员会节点不参与训练全局数据模型;

用于训练全局数据模型,并对共享请求提供预测和响应包括:

选择一组具有专业知识和经验的实体或个人作为委员会节点;

委员会节点之间进行联合学习,共享知识和经验;

委员会节点使用联合学习得到的知识来训练全局数据模型;

当全局数据模型被训练并验证成功,委员会节点可以使用其处理查询请求;

所述委员会节点之间进行联合学习,共享知识和经验包括:

委员会节点Pi从数据请求者那里学习到一个本地全局数据模型mi；

根据委员会节点Pi的本地检索表将模型mi发送给其他相关的参与者；

经过训练的全局数据模型将返回给数据请求者，作为其数据共享请求的答案；

终端物联网设备输出和维护结构化数据包括非结构化数据及非结构化数据文本数据；

所述非结构化数据文本数据定义用于检索文本数据的两步距离度量学习方案，并量化指定数据的相似性。

2.根据权利要求1所述的一种基于共享情况下工业数据安全的保护系统，其特征在于：所述用于通过加密记录建立安全连接，使用许可的区块链来管理数据的可访问性和共享事件，并跟踪数据的使用情况包括：

终端物联网设备通过加密方式向许可区块链发送数据共享请求，以建立安全连接；

记录所有数据提供者的唯一身份以及其数据的配置；

许可区块链记录所有检索和数据共享请求，并由超级节点维护其加密记录；

超级节点使用计算和存储资源来维护许可区块链中的记录，并确保其安全性和可靠性；

当终端物联网设备需要检索相关数据时，它会向许可区块链发出检索请求；

许可区块链根据请求从其记录中检索数据，并将其返回给终端物联网设备；

若数据共享交易发生，许可区块链将记录这些交易，并跟踪数据的使用情况以进行审计；

将每个数据档案以交易的形式记录在区块链上，并由区块链节点通过Merkle树进行验证；

将每个数据共享事件作为交易存储在区块链中。

3.根据权利要求2所述的一种基于共享情况下工业数据安全的保护系统，其特征在于，

所述用于执行多方数据检索过程，根据注册记录查找相关方，并将共享请求转发给适当的超级节点进行处理包括：

根据共享请求中的查询条件和注册记录中的相关信息，确定需要搜索的相关方和数据源；

通过访问注册中心或查询数据库的方式，查找与查询条件匹配的相关方；

根据搜索结果选择适当的超级节点进行处理；

将共享请求转发给所选的超级节点进行处理；

等待超级节点返回共享结果，并将返回的共享结果存储在所述本地缓存中以供后续使用。

4.根据权利要求3所述的一种基于共享情况下工业数据安全的保护系统，其特征在于，

当所述全局数据模型被训练并验证成功，委员会节点可以使用其处理查询请求包括：

数据请求使用Req＝{f1,f2,…,fx}作为全局数据模型的输入；

通过输入得到相应的共享结果M(Req)；

全局数据模型能够接受查询集中的任何查询fx，并为查询提供结果M(fx)，且M对新查询进行预测。

5.一种基于共享情况下工业数据安全的保护方法，基于如权利要求1-4中任一项所述的基于共享情况下工业数据安全的保护系统实施，其特征在于，

该方法包括以下步骤：

接收来自数据请求者的数据共享请求，并将其记录在所述许可区块链模块中；

通过加密记录建立安全连接，使用许可的区块链来管理数据的可访问性和共享事件，并跟踪数据的使用情况；

接收管理后的数据请求者的共享请求，并将其转发给适当的超级节点进行处理；

执行多方数据检索过程，根据注册记录查找相关方，并将共享请求转发给适当的超级节点进行处理；

将已经处理过的数据共享请求和共享结果存储在本地缓存中，并为后续请求提供结果；

训练全局数据模型，并对共享请求提供预测和响应。

一种基于共享情况下工业数据安全的保护系统及方法

技术领域

[0001] 本发明涉及业数据安全隐私技术领域,具体来说,涉及一种基于共享情况下工业数据安全的保护系统及方法。

背景技术

[0002] 在工业互联网中,连接设备产生的数据量的快速增长为通过数据共享提高新兴应用的服务质量开辟了新的可能性。然而,安全和隐私问题(例如,数据泄露)是数据提供商在无线网络中共享数据的主要障碍。隐私数据的泄露可能会给提供者带来严重的财务问题。

[0003] 工业物联网(IIoT)范式中连接设备产生的数据量见证了工业4.0的巨大增长。随着数据带来的价值,随之而来的是对数据隐私的严重担忧。数据泄露可能发生在数据存储、数据传输和数据共享过程中,这可能会给数据所有者和提供商带来严重问题。在这方面,现有的工作主要集中在利用有关数据的聚合信息,而不破坏参与者的隐私。他们通过对原始数据的关键贡献进行一些修改来解决这个问题,例如k-匿名、l-多样性。但大多数方法都假设攻击者只有有限的背景知识,其中数据仍然容易受到基于算法的攻击或背景知识攻击。差异隐私提供了最可靠的隐私保障,通常认为这种保障足够强大,可以保护数据免受隐私攻击。在差异隐私的限制下,有的工作提出了一种机器学习差异隐私来发布数据结构,而不是直接发布查询和回复。

[0004] IIoT应用程序的数据可能包括敏感信息。在这方面,保护数据隐私是一个关键问题。有的工作提出了一种满足差异隐私的保护方法来保护位置数据隐私,而不会降低数据在IIoT中的效用。也有一些工作在探索使用区块链来增强IIoT中的数据安全性。有的将区块链集成到边缘智能中,用于IIoT中的资源分配。尽管这种组合很有希望,但机器学习方法还可以进一步改进。因此,一些工作利用Markov模型进行资源分配,该模型可以在不了解手头问题的情况下说明活动事务。在这些工作中,共识协议是实现所有参与节点之间共识的核心技术组成部分。在工作量证明(PoW)中,首先解决数学难题的矿工赢得了生成区块的权利。但是过于巨大的资源利用率是解决这些难题的必要要求,这限制了基于PoW的共识机制的适用性。

[0005] 最近,联合学习已经出现,允许多个数据所有者在不共享原始数据的情况下协作训练全局模型,同时尊重共享数据的隐私问题。有的工作提出了一种客户端差异隐私保护联合优化算法,以隐藏客户端在训练过程中的贡献。基于服务器聚合用户训练更新的分层架构,有的工作提出了一种基于联邦学习的主动内容缓存方案。

[0006] 然而,在大多数现有的数据共享方案中,集中式策展人的存在增加了数据泄露的风险,尤其是在分布式多方的应用中。主要有两个障碍:一是策展人可能会处理来自不同各方的大量聚合数据,包括一些未知的新数据;另一方面,这些各方都不完全信任他人(包括策展人),因此担心数据泄露。为此,协作数据共享在IIoT中的应用面临着几个挑战。因此,多个不可信方之间分布式数据共享的新协作机制适用于IIoT应用。

[0007] 针对相关技术中的问题,目前尚未提出有效的解决方案。

发明内容

[0008] 针对相关技术中的问题，本发明提出一种基于共享情况下工业数据安全的保护系统及方法，以克服现有相关技术所存在的上述技术问题。

[0009] 为此，本发明采用的具体技术方案如下：

[0010] 根据本发明的一个方面，提供了一种基于共享情况下工业数据安全的保护系统，该基于共享情况下工业数据安全的保护系统包括：数据请求模块、许可区块链模块、新请求模块、多方检索模块、缓存节点模块及联合学习模块；

[0011] 其中，所述数据请求模块，用于接收来自数据请求者的数据共享请求，并将其记录在所述许可区块链模块中；

[0012] 所述许可区块链模块，用于通过加密记录建立安全连接，使用许可的区块链来管理数据的可访问性和共享事件，并跟踪数据的使用情况；

[0013] 所述新请求模块，用于接收管理后的数据请求者的共享请求，并将其转发给适当的超级节点进行处理；

[0014] 所述多方检索模块，用于执行多方数据检索过程，根据注册记录查找相关方，并将共享请求转发给适当的超级节点进行处理；

[0015] 所述缓存节点模块，用于将已经处理过的数据共享请求和共享结果存储在本地缓存中，并为后续请求提供结果；

[0016] 所述联合学习模块，用于训练全局数据模型，并对共享请求提供预测和响应。

[0017] 可选地，所述用于通过加密记录建立安全连接，使用许可的区块链来管理数据的可访问性和共享事件，并跟踪数据的使用情况包括：

[0018] 终端物联网设备通过加密方式向许可区块链发送数据共享请求，以建立安全连接；

[0019] 记录所有数据提供者的唯一身份以及其数据的配置；

[0020] 许可区块链记录所有检索和数据共享请求，并由超级节点维护其加密记录；

[0021] 超级节点使用计算和存储资源来维护许可区块链中的记录，并确保其安全性和可靠性；

[0022] 当终端物联网设备需要检索相关数据时，它会向许可区块链发出检索请求；

[0023] 许可区块链根据请求从其记录中检索数据，并将其返回给终端物联网设备；

[0024] 若数据共享交易发生，许可区块链将记录这些交易，并跟踪数据的使用情况以进行审计；

[0025] 将每个数据档案以交易的形式记录在区块链上，并由区块链节点通过Merkle树进行验证；

[0026] 将每个数据共享事件作为交易存储在区块链中。

[0027] 可选地，所述终端物联网设备输出和维护结构化数据包括非结构化数据及非结构化数据文本数据；

[0028] 所述非结构化数据文本数据定义用于检索文本数据的两步距离度量学习方案，并量化指定数据的相似性；

[0029] 通过利用图来表示原始数据以供进一步处理，并保留更多的结构和上下文信息。

[0030] 可选地，所述通过利用图来表示原始数据以供进一步处理，并保留更多的结构和

上下文信息包括：

[0031] 加权图G={V,E}包括一组节点V和一组边$E \subseteq V \times V$；

[0032] 每个节点n_i包含一个文本术语及其权重w_{ni}(n_i,w_{ni})；

[0033] 每个边e_{ij}连接节点n_i和n_j,其中边的权重是w_{eij}；

[0034] 通过权重矩阵A=[a_{ij}]来表示图,其中a_{ij}=w_{ni}；

[0035] 如果i=j,则a_{ij}=w_{eij}；

[0036] 如果i≠j,则利用术语频率逆文档频率来构建图,并将其序列化为序列向量；

[0037] 通过将图序列化为线性向量,使用Jaccard相似性作为距离函数；

[0038] 用k-均值算法对文档进行聚类,并根据文本相似性和参与的用户将数据集{D_1,...,D_n}聚类为各种类别。

[0039] 可选地,所述用于接收管理后的数据请求者的共享请求,并将其转发给适当的超级节点进行处理包括：

[0040] 各方在许可区块链中注册并上传重新评估记录；

[0041] 数据请求者向附近的超级节点SN_{req}启动包含一组查询F_x={f_1,f_2,…,f_x}共享请求Req；

[0042] 当数据请求者发起共享请求Req时,它将请求提交给其附近的超级节点SN_{req}；

[0043] SN_{req}首先搜索区块链以确定请求之前是否被处理过；

[0044] 若有查找命中,则将之前计算的缓存全局数据模型直接返回给数据请求者；

[0045] 否则,节点SN_{req}通过多方数据检索过程,并在区块链中查找相关节点。

[0046] 可选地,所述用于执行多方数据检索过程,根据注册记录查找相关方,并将共享请求转发给适当的超级节点进行处理包括：

[0047] 根据共享请求中的查询条件和注册记录中的相关信息,确定需要搜索的相关方和数据源；

[0048] 通过访问注册中心或查询数据库的方式,查找与查询条件匹配的相关方；

[0049] 根据搜索结果选择适当的超级节点进行处理；

[0050] 将共享请求转发给所选的超级节点进行处理；

[0051] 等待超级节点返回共享结果,并将返回的共享结果存储在所述本地缓存中以供后续使用。

[0052] 可选地,用于训练全局数据模型,并对共享请求提供预测和响应包括：

[0053] 选择一组具有专业知识和经验的实体或个人作为委员会节点；

[0054] 委员会节点之间进行联合学习,共享知识和经验；

[0055] 委员会节点使用联合学习得到的知识来训练全局数据模型；

[0056] 当全局数据模型被训练并验证成功,委员会节点可以使用其处理查询请求。

[0057] 可选地,所述委员会节点之间进行联合学习,共享知识和经验包括：

[0058] 委员会节点P_i从数据请求者那里学习到一个本地全局数据模型m_i；

[0059] 根据委员会节点P_i的本地检索表将模型m_i发送给其他相关的参与者；

[0060] 经过训练的全局数据模型将返回给数据请求者,作为其数据共享请求的答案。

[0061] 可选地,当所述全局数据模型被训练并验证成功,委员会节点可以使用其处理查

询请求包括：

[0062] 数据请求使用Req＝{f_1,f_2,…,f_x}作为全局数据模型的输入；

[0063] 通过输入得到相应的共享结果M(Req)；

[0064] 全局数据模型能够接受查询集中的任何查询f_x,并为查询提供结果M(f_x),且M对新查询进行预测。

[0065] 根据本发明的另一个方面,还提供了一种基于共享情况下工业数据安全的保护方法,该方法包括以下步骤：

[0066] 接收来自数据请求者的数据共享请求,并将其记录在所述许可区块链模块中；

[0067] 通过加密记录建立安全连接,使用许可的区块链来管理数据的可访问性和共享事件,并跟踪数据的使用情况；

[0068] 接收管理后的数据请求者的共享请求,并将其转发给适当的超级节点进行处理；

[0069] 执行多方数据检索过程,根据注册记录查找相关方,并将共享请求转发给适当的超级节点进行处理；

[0070] 将已经处理过的数据共享请求和共享结果存储在本地缓存中,并为后续请求提供结果；

[0071] 训练全局数据模型,并对共享请求提供预测和响应。

[0072] 本发明的有益效果为：

[0073] 本发明实现了CPU的高效利用,既实现了高效的数据访问也没有过多的消耗CPU,通过利用联邦学习来构建数据模型,并共享数据模型而不是原始数据,将数据共享问题转化为机器学习问题,并提出了一种新的区块链授权协同架构,通过分布式多方共享数据,从而降低数据泄露的风险,使得数据所有者可以通过该架构进一步控制对共享数据的访问,并且将差异隐私集成到联邦学习中,进而可以进一步的保护数据隐私。

附图说明

[0074] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0075] 图1是根据本发明实施例的一种基于共享情况下工业数据安全的保护系统的原理框图。

[0076] 图中：

[0077] 1、数据请求模块;2、许可区块链模块;3、新请求模块;4、多方检索模块;5、缓存节点模块;6、联合学习模块。

具体实施方式

[0078] 为进一步说明各实施例,本发明提供有附图,这些附图为本发明揭露内容的一部分,其主要用以说明实施例,并可配合说明书的相关描述来解释实施例的运作原理,配合参考这些内容,本领域普通技术人员应能理解其他可能的实施方式以及本发明的优点,图中的组件并未按比例绘制,而类似的组件符号通常用来表示类似的组件。

[0079] 根据本发明的实施例,提供了一种基于共享情况下工业数据安全的保护系统及方法。

[0080] 现结合附图和具体实施方式对本发明进一步说明,如图1所示,根据本发明实施例的基于共享情况下工业数据安全的保护系统,该基于共享情况下工业数据安全的保护系统包括:数据请求模块1、许可区块链模块2、新请求模块3、多方检索模块4、缓存节点模块5及联合学习模块6;

[0081] 其中,所述数据请求模块1,用于接收来自数据请求者的数据共享请求,并将其记录在所述许可区块链模块2中;

[0082] 所述许可区块链模块2,用于通过加密记录建立安全连接,使用许可的区块链来管理数据的可访问性和共享事件,并跟踪数据的使用情况;

[0083] 所述新请求模块3,用于接收管理后的数据请求者的共享请求,并将其转发给适当的超级节点进行处理;

[0084] 所述多方检索模块4,用于执行多方数据检索过程,根据注册记录查找相关方,并将共享请求转发给适当的超级节点进行处理;

[0085] 所述缓存节点模块5,用于将已经处理过的数据共享请求和共享结果存储在本地缓存中,并为后续请求提供结果;

[0086] 所述联合学习模块6,用于训练全局数据模型,并对共享请求提供预测和响应。

[0087] 具体的,全局数据模型通过联合学习,允许多个数据所有者在不共享原始数据的情况下协作训练全局模型,同时尊重共享数据的隐私问题,这个全局模型是每个任务中都不一样的。

[0088] 在一个实施例中,所述用于通过加密记录建立安全连接,使用许可的区块链来管理数据的可访问性和共享事件,并跟踪数据的使用情况包括:

[0089] 终端物联网设备通过加密方式向许可区块链发送数据共享请求,以建立安全连接;

[0090] 记录所有数据提供者的唯一身份以及其数据的配置;

[0091] 许可区块链记录所有检索和数据共享请求,并由超级节点维护其加密记录;

[0092] 超级节点使用计算和存储资源来维护许可区块链中的记录,并确保其安全性和可靠性;

[0093] 当终端物联网设备需要检索相关数据时,它会向许可区块链发出检索请求;

[0094] 许可区块链根据请求从其记录中检索数据,并将其返回给终端物联网设备;

[0095] 若数据共享交易发生,许可区块链将记录这些交易,并跟踪数据的使用情况以进行审计;

[0096] 将每个数据档案以交易的形式记录在区块链上,并由区块链节点通过Merkle树进行验证;

[0097] 将每个数据共享事件作为交易存储在区块链中。

[0098] 具体的,区块链上相关参与者对一个数据共享请求的检索是该模型需要解决的一个基本问题。由于有许多参与者,那些拥有与请求相关数据的人应该参与数据共享,以提高响应结果的准确性。尽管如此,检索过程不应该破坏每个参与者的隐私。需要一种分布式检索方案来快速定位分布在参与者之间的请求数据,参与者可以协同响应请求。

[0099] 因此设计了区块链中的第二方检索机制。所有参与者根据其数据类别被划分为不同的社区，也就是说，一个社区的成员持有相似类别的数据。每个社区都维护一个本地可检索的log(n)记录，这些记录指向log(n)个不同的社区。对于社区中的每个节点，它存储所有社区成员的ID，以及其最接近(在数据类别中最相关)的log社区的log((n))节点。通过这种方式，最相关的参与者将在本地检索表Pi上进行本地检索。

[0100] 从每个参与者的数据中提取一个关键字列表，作为哈希值形式的代表特征。此外，由于IIoT设备的通信资源有限，在检索过程中还需要考虑两个节点之间的物理距离。然后，基于Jaccard距离计算它们关键项之间的逻辑距离。每个参与者(设备)的ID根据逻辑距离生成。也就是说，两个节点的相对关系越大，它们的公共ID前缀就越长。

[0101] 当用户向其附近的节点P_i提交数据共享请求时，与P_i处于同一社区的所有节点将请求发送到其本地路由表中具有一定距离的节点来启动检索过程。这个过程将递归地实现，直到遍历相关距离内的所有节点。在检索结束时，我们获得了请求的相关子集节点$P_s \subseteq P$，这些节点也是委员会节点，用于运行共识流程来批准数据共享结果。

[0102] 具体的，P_s表示节点集合，P表示某个具体的节点。

[0103] 在一个实施例中，所述终端物联网设备输出和维护结构化数据包括非结构化数据及非结构化数据文本数据；

[0104] 所述非结构化数据文本数据定义用于检索文本数据的两步距离度量学习方案，并量化指定数据的相似性；

[0105] 通过利用图来表示原始数据以供进一步处理，并保留更多的结构和上下文信息。

[0106] 在一个实施例中，所述通过利用图来表示原始数据以供进一步处理，并保留更多的结构和上下文信息包括：

[0107] 加权图G={V,E}包括一组节点V和一组边$E \subseteq V \times V$；

[0108] 每个节点n_i包含一个文本术语及其权重$w_{ni}(n_i, w_{ni})$；

[0109] 每个边e_{ij}连接节点n_i和n_j，其中边的权重是w_{eij}；

[0110] 通过权重矩阵$A=[a_{ij}]$来表示图，其中$a_{ij}=w_{ni}$；

[0111] 如果i=j，则$a_{ij}=w_{eij}$；

[0112] 如果$i \neq j$，则利用术语频率逆文档频率来构建图，并将其序列化为序列向量；

[0113] 通过将图序列化为线性向量，使用Jaccard相似性作为距离函数；

[0114] 用k-均值算法对文档进行聚类，并根据文本相似性和参与的用户将数据集$\{D_1, \ldots, D_n\}$聚类为各种类别。

[0115] 具体的，将图合并为全局图$G=G_1 \cup G_2, \cdots \cup G_n$对于全局图G＝{V,E}，使用k代表顶点的数量；所以节点的归一化属性的大小将为k，边的归一化属性大小将为$k \times (k-1)/2$；因此，归一化向量$S_{eq}=V \cup E=\{V_1, \cdots, V_k\} \cup \{E_1, E_2, \cdots, E_{k(k-2)/2}\}$。利用Jaccard相似性作为距离函数，用k-均值算法对文档进行聚类；在归一化加权图和定义的距离度量的帮助下，根据文本相似性将数据集$\{D_1, \cdots, D_n\}$聚类为各种类别；根据数据将参与的用户分为不同的组。

[0116] 在一个实施例中，所述用于接收管理后的数据请求者的共享请求，并将其转发给适当的超级节点进行处理包括：

[0117] 各方在许可区块链中注册并上传重新评估记录；

[0118] 数据请求者向附近的超级节点SN_{req}启动包含一组查询$F_x=\{f_1,f_2,\cdots,f_x\}$共享请求Req；

[0119] 当数据请求者发起共享请求Req时，它将请求提交给其附近的超级节点SN_{req}；

[0120] SN_{req}首先搜索区块链以确定请求之前是否被处理过；

[0121] 若有查找命中，则将之前计算的缓存全局数据模型直接返回给数据请求者；

[0122] 否则，节点SN_{req}通过多方数据检索过程，并在区块链中查找相关节点。

[0123] 在一个实施例中，所述用于执行多方数据检索过程，根据注册记录查找相关方，并将共享请求转发给适当的超级节点进行处理包括：

[0124] 根据共享请求中的查询条件和注册记录中的相关信息，确定需要搜索的相关方和数据源；

[0125] 通过访问注册中心或查询数据库的方式，查找与查询条件匹配的相关方；

[0126] 根据搜索结果选择适当的超级节点进行处理；

[0127] 将共享请求转发给所选的超级节点进行处理；

[0128] 等待超级节点返回共享结果，并将返回的共享结果存储在所述本地缓存中以供后续使用。

[0129] 在一个实施例中，用于训练全局数据模型，并对共享请求提供预测和响应包括：

[0130] 选择一组具有专业知识和经验的实体或个人作为委员会节点；

[0131] 委员会节点之间进行联合学习，共享知识和经验；

[0132] 委员会节点使用联合学习得到的知识来训练全局数据模型；

[0133] 当全局数据模型被训练并验证成功，委员会节点可以使用其处理查询请求。

[0134] 在一个实施例中，所述委员会节点之间进行联合学习，共享知识和经验包括：

[0135] 委员会节点P_i从数据请求者那里学习到一个本地全局数据模型m_i；

[0136] 根据委员会节点P_i的本地检索表将全局数据模型m_i发送给其他相关的参与者；

[0137] 经过训练的全局数据模型将返回给数据请求者，作为其数据共享请求的答案。

[0138] 在一个实施例中，当所述全局数据模型被训练并验证成功，委员会节点可以使用其处理查询请求包括：

[0139] 数据请求使用$Req=\{f_1,f_2,\cdots,f_x\}$作为全局数据模型的输入；

[0140] 通过输入得到相应的共享结果M(Req)；

[0141] 全局数据模型能够接受查询集中的任何查询f_x，并为查询提供结果$M(f_x)$，且M对新查询进行预测。

[0142] 根据本发明的另一个实施例，还提供了一种基于共享情况下工业数据安全的保护方法，该方法包括以下步骤：

[0143] 接收来自数据请求者的数据共享请求，并将其记录在所述许可区块链模块2中；

[0144] 通过加密记录建立安全连接，使用许可的区块链来管理数据的可访问性和共享事件，并跟踪数据的使用情况；

[0145] 接收管理后的数据请求者的共享请求，并将其转发给适当的超级节点进行处理；

[0146] 执行多方数据检索过程，根据注册记录查找相关方，并将共享请求转发给适当的超级节点进行处理；

[0147] 将已经处理过的数据共享请求和共享结果存储在本地缓存中，并为后续请求提供结果；

[0148] 训练全局数据模型，并对共享请求提供预测和响应。

[0149] 综上所述，借助于本发明的上述技术方案，通过分布式多方共享数据，从而降低数据泄露的风险，使得数据所有者可以通过该架构进一步控制对共享数据的访问，并且将差异隐私集成到联邦学习中，进而可以进一步的保护数据隐私。

[0150] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

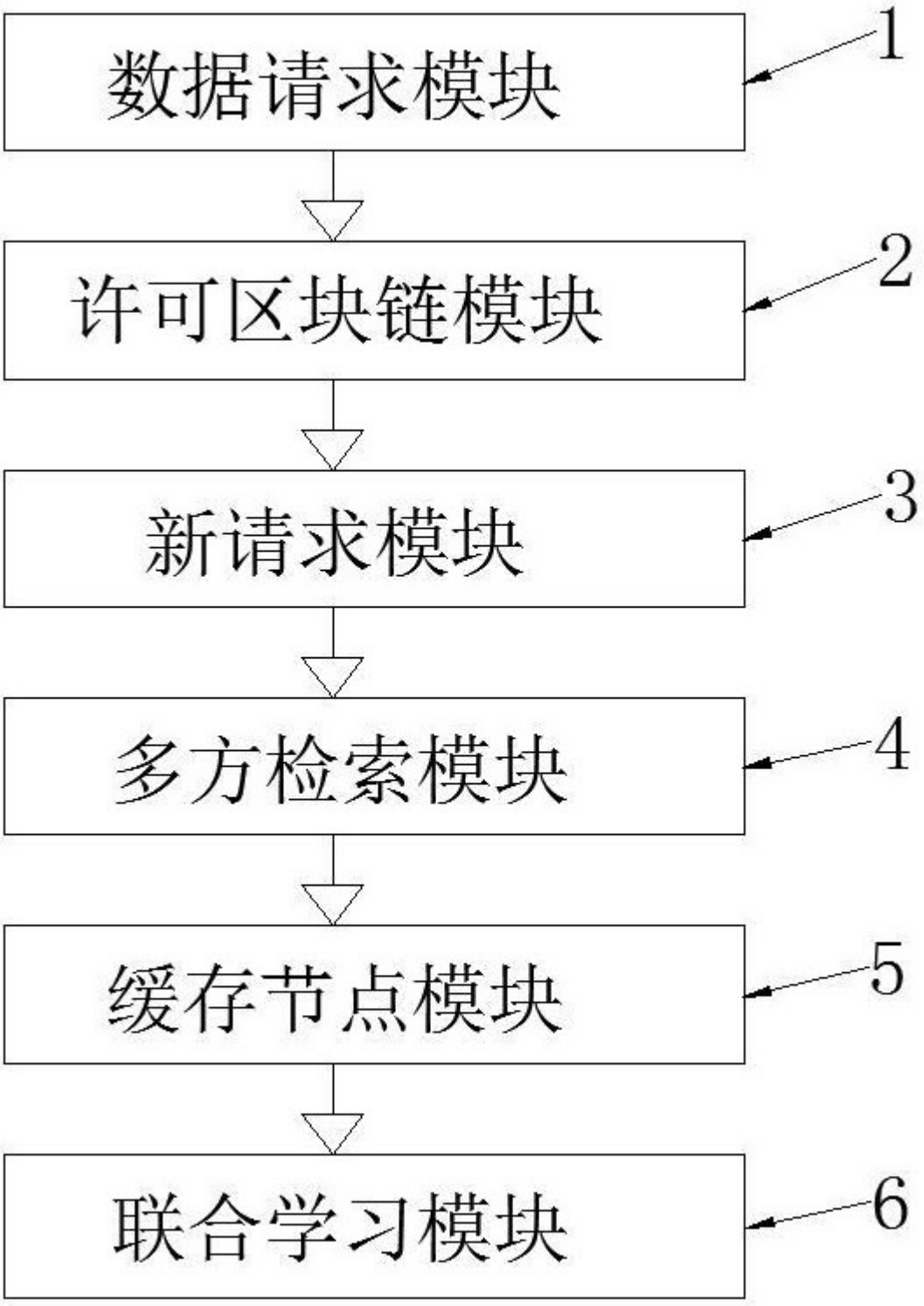


图 1