(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114282692 A
(43)申请公布日 2022.04.05

(21)申请号 202210217753.2

(22)申请日 2022.03.08

(71)申请人 富算科技(上海)有限公司
地址 200135 上海市浦东新区中国(上海)自由贸易试验区浦东大道1200号2层A区

(72)发明人 尤志强 卞阳

(74)专利代理机构 北京超凡宏宇专利代理事务所(特殊普通合伙) 11463
代理人 唐正瑜

(51)Int.Cl.
G06N 20/20(2019.01)

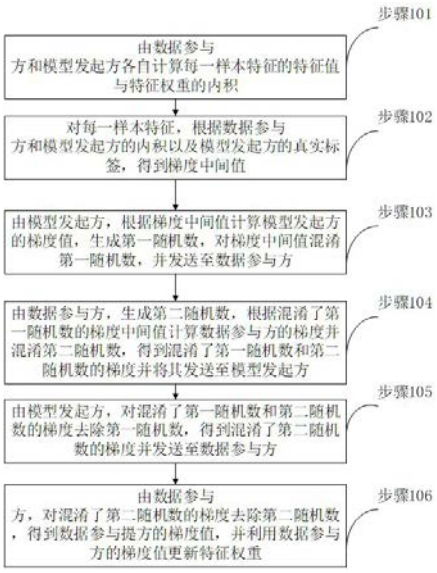
权利要求书3页 说明书10页 附图3页

(54)发明名称

一种纵向联邦学习的模型训练方法及系统

(57)摘要

本申请提供一种纵向联邦学习的模型训练方法及系统，在多个参与方的联邦学习过程中，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，模型发起方对梯度中间值混淆第一随机数，数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且降低了加密的耗时，提供了处理的效率。



CN 114282692 A

1.一种纵向联邦学习的模型训练方法,其特征在于,包括:

由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积;

对每一样本特征,根据所述数据参与方和所述模型发起方的内积以及所述模型发起方的真实标签,得到梯度中间值;

由所述模型发起方,根据梯度中间值计算所述模型发起方的梯度值,生成第一随机数,对梯度中间值混淆第一随机数,并发送至所述数据参与方;

由所述数据参与方,生成第二随机数,根据混淆了第一随机数的梯度中间值计算所述数据参与方的梯度并混淆第二随机数,得到混淆了第一随机数和第二随机数的梯度并将其发送至所述模型发起方;

由所述模型发起方,对混淆了第一随机数和第二随机数的梯度去除第一随机数,得到混淆了第二随机数的梯度并发送至所述数据参与方;以及

由所述数据参与方,对混淆了第二随机数的梯度去除第二随机数,得到所述数据参与方的梯度值,并利用所述数据参与方的梯度值更新特征权重。

2.如权利要求1所述的方法,其特征在于,还包括:

由所述模型发起方基于真实标签和预测值计算模型的损失值,根据损失值判断模型是否收敛:

若收敛,则确定模型训练完成;

若不收敛,则继续迭代更新。

3.如权利要求1所述的方法,其特征在于,所述梯度中间值由对每一样本特征,将所述数据参与方和所述模型发起方的内积相加,得到总内积值,使用预设函数对总内积值进行转换得到预测值,并将预测值与对应的所述模型发起方的真实标签做差的方式获得;

其中,所述预设函数包括sigmoid函数;所述预测值由$y=1/(1+e^{-Z})$计算;其中,Z为每条样本的总内积值。

4.如权利要求1所述的方法,其特征在于,所述混淆了第一随机数的梯度中间值由E_gradf _i = (y_hat_i - y_i)×R_ai计算;

其中,y_hat_i为预测值,y_i为真实标签,R_ai为第一随机数,i表示样本索引。

5.如权利要求4所述的方法,其特征在于,所述混淆了第一随机数和第二随机数的梯度由S_E_gradf _ij = E_gradf _ij×R_bi计算;

其中,E_gradf _ij = E_gradf _i×X_bij,X_bij为特征权重,j表示特征索引。

6.如权利要求5所述的方法,其特征在于,所述混淆了第二随机数的梯度由D_E_gradf _ij = S_E_gradf _ij / R_ai计算。

7.如权利要求6所述的方法,其特征在于,所述数据参与方的梯度值由gradf _ij = D_E_gradf _ij / R_bi计算。

8.如权利要求1所述的方法,其特征在于,在所述纵向联邦学习的模型训练中采用分批训练的方式。

9.如权利要求8所述的方法,其特征在于,得到所述梯度中间值之后,对所有梯度中间值进行过滤处理,得到处理后的梯度中间值;所述过滤处理包括:将梯度中间值绝对值大于或等于中间值阈值的梯度中间值保留,将梯度中间值绝对值小于中间值阈值的梯度中间值按采样比例进行采样。

10.如权利要求1所述的方法,其特征在于,利用梯度值更新特征权重,包括:

判断梯度值的绝对值是否大于梯度阈值,若是,则利用该梯度值更新特征权重;若否,则不更新特征权重。

11.一种纵向联邦学习的模型训练方法,其特征在于,应用于模型发起方,包括:

计算每一样本特征的特征值与特征权重的内积;

接收数据参与方发送的内积;

对每一样本特征,将所述数据参与方和所述模型发起方的内积相加,得到总内积值,使用预设函数对总内积值进行转换得到预测值;

对每一样本特征,将预测值与对应的所述模型发起方的真实标签做差,得到梯度中间值;

根据梯度中间值计算所述模型发起方的梯度值,利用所述模型发起方的梯度值更新样本特征的特征权重;生成第一随机数,对梯度中间值混淆第一随机数,向所述数据参与方发送混淆了第一随机数的梯度中间值;其中,第一随机数为不为0的实数;

接收所述数据参与方发送的混淆了第一随机数和第二随机数的梯度;其中,第二随机数为不为0的实数;以及

对混淆了第一随机数和第二随机数的梯度去除第一随机数,得到混淆了第二随机数的梯度,向所述数据参与方发送混淆了第二随机数的梯度。

12.一种纵向联邦学习的模型训练方法,其特征在于,应用于数据参与方,包括:

计算每一样本特征的特征值与特征权重的内积,向模型发起方发送所述数据参与方的内积;

接收所述模型发起方发送的混淆了第一随机数的梯度中间值;其中,第一随机数为不为0的实数;

生成第二随机数,根据混淆了第一随机数的梯度中间值计算所述数据参与方的梯度并混淆第二随机数,向所述模型发起方发送混淆了第一随机数和第二随机数的梯度;其中,第二随机数为不为0的实数;

接收所述模型发起方发送的混淆了第二随机数的梯度;以及

对混淆了第二随机数的梯度去除第二随机数,得到所述数据参与方的梯度值,并利用所述数据参与方的梯度值更新特征权重。

13.一种纵向联邦学习的模型训练系统,其特征在于,包括:

内积计算模块,用于由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积;

中间值计算模块,用于对每一样本特征,根据所述数据参与方和所述模型发起方的内积以及所述模型发起方的真实标签,得到梯度中间值;

一次混淆模块,用于由所述模型发起方,根据梯度中间值计算所述模型发起方的梯度值,生成第一随机数,对梯度中间值混淆第一随机数,并发送至所述数据参与方;

二次混淆模块,用于由所述数据参与方,生成第二随机数,根据混淆了第一随机数的梯度中间值计算所述数据参与方的梯度并混淆第二随机数,得到混淆了第一随机数和第二随机数的梯度并将其发送至所述模型发起方;

一次解混模块,用于由所述模型发起方,对混淆了第一随机数和第二随机数的梯度去

除第一随机数，得到混淆了第二随机数的梯度并发送至所述数据参与方；以及

二次解混模块，用于由所述数据参与方，对混淆了第二随机数的梯度去除第二随机数，得到所述数据参与方的梯度值，并利用所述数据参与方的梯度值更新特征权重。

14.一种纵向联邦学习的模型训练方法，其特征在于，包括：

由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；

对每一样本特征，根据所述数据参与方和所述模型发起方的内积以及所述数据参与方的真实标签，得到梯度中间值；

由所述数据参与方，根据梯度中间值计算所述数据参与方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至所述模型发起方；

由所述模型发起方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算所述模型发起方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至所述数据参与方；

由所述数据参与方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至所述模型发起方；以及

由所述模型发起方，对混淆了第二随机数的梯度去除第二随机数，得到所述模型发起方的梯度值，并利用所述模型发起方的梯度值更新特征权重。

一种纵向联邦学习的模型训练方法及系统

技术领域

[0001] 本申请涉及联邦学习技术领域，具体而言，涉及一种纵向联邦学习的模型训练方法及系统。

背景技术

[0002] 联邦学习作为一种数据安全计算的技术，在企业中逐步得到应用，其能够实现在原始数据不出门的前提下，让数据价值在各个机构之间进行流动，创造业务价值，比如应用在金融风控、广告推荐等领域。联邦学习是一种分布式计算架构，支持多方安全计算，根据不同的业务使用场景，主要包括纵向联邦学习、横向联邦学习以及联邦迁移算法三种类型。目前联邦学习已经可以支持多种机器学习算法。

[0003] 诸如逻辑回归算法(logistic regression)是一种经典的机器学习模型，适用于分类问题。因为其具有简单、快速、可解释强等特性，被广泛应用于金融风控等领域。企业实际业务一般要求在纵向联邦学习的场景下完成逻辑回归模型的训练和使用，比如银行与运营商之间联合建模评分卡场景。

[0004] 然而在纵向联邦学习场景中，已有的基于梯度下降优化算法的机器学习算法往往依赖一个可信赖的协调方，特别是针对逻辑回归算法。该协调方作为一种第三方角色，是独立于数据参与方的，对数据参与方与模型发起方之间进行相关中间结果处理以及通信。然而在现实场景中，特别是银行、运营商等对数据安全极其严格的机构，是不能接收这种依赖可信第三方的算法，因为难以找到被各个参与方认可的对象机构承担该协调者角色。因此，现有技术通常采用半同态加密技术，学习过程中需要涉及半同态加密、模型发起方与数据参与方之间的公钥通信和半同态解密等步骤，在大数据量下耗时较高，运算速度较慢。

发明内容

[0005] 本申请实施例的目的在于提供一种纵向联邦学习的模型训练方法及系统，用以解决现有技术的学习过程中需要涉及半同态加密、模型发起方与数据参与方之间的公钥通信和半同态解密等步骤，在大数据量下耗时较高，运算速度较慢的问题。

[0006] 本申请实施例提供的一种纵向联邦学习的模型训练方法，包括：

由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；

对每一样本特征，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值；

由模型发起方，根据梯度中间值计算模型发起方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至数据参与方；

由数据参与方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至模型发起方；

由模型发起方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到

混淆了第二随机数的梯度并发送至数据参与方;以及

由数据参与方,对混淆了第二随机数的梯度去除第二随机数,得到数据参与方的梯度值,并利用数据参与方的梯度值更新特征权重。

[0007] 上述技术方案中,在多个参与方的联邦学习过程中,根据数据参与方和模型发起方的内积以及模型发起方的真实标签,得到梯度中间值,模型发起方对梯度中间值混淆第一随机数,数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数,之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数,使得模型发起方和数据参与方都有各自的梯度值,可以对各自特征权重进行更新,由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信,不受计算次数的限制,不存在精度丢失的问题,能够支持更复杂的联邦学习模型学习需求,并且,降低了加密的耗时,减小通信量,提升模型训练的性能,满足了大规模数据场景下的高性能计算的要求。

[0008] 在一些可选的实施方式中,还包括:

由模型发起方基于真实标签和预测值计算模型的损失值,根据损失值判断模型是否收敛:

若收敛,则确定模型训练完成;

若不收敛,则继续迭代更新。

[0009] 在一些可选的实施方式中,梯度中间值由对每一样本特征,将数据参与方和模型发起方的内积相加,得到总内积值,使用预设函数对总内积值进行转换得到预测值,并将预测值与对应的模型发起方的真实标签做差的方式获得;

其中,预设函数包括sigmoid函数;预测值由$y=1/(1+e^{-Z})$计算;其中,Z为每条样本的总内积值。

[0010] 在一些可选的实施方式中,混淆了第一随机数的梯度中间值由E_gradf _i = (y_hat_i - y_i)×R_ai计算;

其中,y_hat_i为预测值,y_i为真实标签,R_ai为第一随机数,i表示样本索引。

[0011] 在一些可选的实施方式中,混淆了第一随机数和第二随机数的梯度由S_E_gradf _ij = E_gradf _ij×R_bi计算;

其中,E_gradf _ij = E_gradf _i×X_bij,X_bij为特征权重,j表示特征索引。

[0012] 在一些可选的实施方式中,混淆了第二随机数的梯度由D_E_gradf _ij = S_E_gradf _ij / R_ai计算。

[0013] 在一些可选的实施方式中,数据参与方的梯度值由gradf _ij = D_E_gradf _ij / R_bi计算。

[0014] 在一些可选的实施方式中,在纵向联邦学习的模型训练中采用分批训练的方式。

[0015] 上述技术方案中,纵向联邦学习的模型训练采用小批次的随机梯度下降算法,分批训练做聚合能够提高安全性,适用于一些对安全要求更高的场景下。

[0016] 在一些可选的实施方式中,得到梯度中间值之后,对所有梯度中间值进行过滤处理,得到处理后的梯度中间值;过滤处理包括:将梯度中间值绝对值大于或等于中间值阈值的梯度中间值保留,将梯度中间值绝对值小于中间值阈值的梯度中间值按采样比例进行采样。

[0017] 上述技术方案中,模型发起方会将每一个批次全部样本特征的梯度中间值信息传

给数据参与方，在数据量较大的情况下，其通信量会比较可观，在通信上存在大量的性能耗损。因此，对梯度中间值进行过滤处理，将梯度中间值绝对值小于中间值阈值的梯度中间值按采样比例进行采样，将梯度中间值绝对值大于或等于中间值阈值的梯度中间值保留，减小了通信量，提升了模型训练的性能和效率。

[0018] 在一些可选的实施方式中，利用梯度值更新特征权重，包括：

判断梯度值的绝对值是否大于梯度阈值，若是，则利用该梯度值更新特征权重；若否，则不更新特征权重。

[0019] 上述技术方案中，当梯度值绝对值很小时，梯度值对参数的更新操作，不能带来模型训练效果上的增益，因此，只对梯度值绝对值大于梯度阈值的梯度值更新特征权重，可以在减少通信量和计算量的同时，不影响模型的整体训练效果，进而达到提升模型训练性能提升目的。

[0020] 本申请实施例提供的一种纵向联邦学习的模型训练方法，应用于模型发起方，包括：

计算每一样本特征的特征值与特征权重的内积；

接收数据参与方发送的内积；

对每一样本特征，将数据参与方和模型发起方的内积相加，得到总内积值，使用预设函数对总内积值进行转换得到预测值；

对每一样本特征，将预测值与对应的模型发起方的真实标签做差，得到梯度中间值；

根据梯度中间值计算模型发起方的梯度值，利用模型发起方的梯度值更新样本特征的特征权重；生成第一随机数，对梯度中间值混淆第一随机数，向数据参与方发送混淆了第一随机数的梯度中间值；其中，第一随机数为不为0的实数；

接收数据参与方发送的混淆了第一随机数和第二随机数的梯度；其中，第二随机数为不为0的实数；以及

对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度，向数据参与方发送混淆了第二随机数的梯度。

[0021] 上述技术方案中，在多个参与方的联邦学习过程中，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，模型发起方对梯度中间值混淆第一随机数，数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0022] 本申请实施例提供的一种纵向联邦学习的模型训练方法，应用于数据参与方，包括：

计算每一样本特征的特征值与特征权重的内积，向模型发起方发送数据参与方的内积；

接收模型发起方发送的混淆了第一随机数的梯度中间值；其中，第一随机数为不

为0的实数；

生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，向模型发起方发送混淆了第一随机数和第二随机数的梯度；其中，第二随机数为不为0的实数；

接收模型发起方发送的混淆了第二随机数的梯度；以及

对混淆了第二随机数的梯度去除第二随机数，得到数据参与方的梯度值，并利用数据参与方的梯度值更新特征权重。

[0023] 上述技术方案中，在多个参与方的联邦学习过程中，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，模型发起方对梯度中间值混淆第一随机数，数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0024] 本申请实施例提供的一种纵向联邦学习的模型训练系统，包括：

内积计算模块，用于由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；

中间值计算模块，用于对每一样本特征，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值；

一次混淆模块，用于由模型发起方，根据梯度中间值计算模型发起方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至数据参与方；

二次混淆模块，用于由数据参与方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至模型发起方；

一次解混模块，用于由模型发起方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至数据参与方；以及

二次解混模块，用于由数据参与方，对混淆了第二随机数的梯度去除第二随机数，得到数据参与方的梯度值，并利用数据参与方的梯度值更新特征权重。

[0025] 上述技术方案中，在多个参与方的联邦学习过程中，利用内积计算模块和中间值计算模块，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，再通过一次混淆模块、二次混淆模块、一次解混模块和二次解混模块，采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0026] 本申请实施例提供的一种纵向联邦学习的模型训练方法，包括：

由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；

对每一样本特征，根据数据参与方和模型发起方的内积以及数据参与方的真实标签，得到梯度中间值；

由数据参与方，根据梯度中间值计算数据参与方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至模型发起方；

由模型发起方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算模型发起方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至数据参与方；

由数据参与方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至模型发起方；以及

由模型发起方，对混淆了第二随机数的梯度去除第二随机数，得到模型发起方的梯度值，并利用模型发起方的梯度值更新特征权重。

[0027] 上述技术方案中，在多个参与方的联邦学习过程中，数据参与方具有真实标签，由数据参与方计算得到梯度中间值并混淆第一随机数，模型发起方对混淆了第一随机数的梯度中间值计算模型发起方的梯度并混淆第二随机数，之后依次由数据参与方去除第一随机数、模型发起方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

附图说明

[0028] 为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本申请的某些实施例，因此不应被看作是对范围的限定，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相关的附图。

[0029] 图1为本申请实施例提供的一种纵向联邦学习的模型训练方法步骤流程图；

图2为本申请实施例提供的一种纵向联邦学习的模型训练方法工作流程图；

图3为本申请实施例提供的一种纵向联邦学习的模型训练系统的功能模块图。

[0030] 图标：1-内积计算模块，2-中间值计算模块，3-一次混淆模块，4-二次混淆模块，5-一次解混模块，6-二次解混模块。

具体实施方式

[0031] 下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行描述。

[0032] 在纵向联邦学习场景中，已有的基于梯度下降优化算法的机器学习算法往往依赖一个可信赖的协调方，特别是针对逻辑回归算法。该协调方作为一种第三方角色，是独立于数据参与方的，对数据参与方与模型发起方之间进行相关中间结果处理以及通信。然而在现实场景中，特别是银行、运营商等对数据安全极其严格的机构，是不能接收这种依赖可信第三方的算法，因为难以找到被各个参与方认可的对象机构承担该协调者角色。因此，现有技术通常采用半同态加密技术，学习过程中需要涉及半同态加密、模型发起方与数据参与方之间的公钥通信和半同态解密等步骤，在大数据量下耗时较高，运算速度较慢。

[0033] 在半同态加密过程中，从明文空间向密文空间中不存在完美的同态映射，因此同

态加密会存在不同程度的噪声，精确性会受到影响；其次，同态加密的发挥性能所产生的开销很大，对计算资源的要求也很高，这也是制约其大规模使用的最主要因素。另外，由于半同态加密使用的密钥长度较长，加密后的密文较大，在大批量数据的通信场景，会显著降低性能，特别是在大规模数据联邦学习场景下，制约算法训练预测的整体性能。

[0034] 因此，本申请实施例提供一种纵向联邦学习的模型训练方法及系统，采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。下面详细阐述：

请参照图1，图1为本申请实施例提供的一种纵向联邦学习的模型训练方法步骤流程图，包括：

步骤101、由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；

步骤102、对每一样本特征，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值；

步骤103、由模型发起方，根据梯度中间值计算模型发起方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至数据参与方；

步骤104、由数据参与方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至模型发起方；

步骤105、由模型发起方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至数据参与方；

步骤106、由数据参与方，对混淆了第二随机数的梯度去除第二随机数，得到数据参与方的梯度值，并利用数据参与方的梯度值更新特征权重。

[0035] 其中，联邦学习的参与方包括模型发起方和数据参与方，且满足以下条件：

(1)模型发起方，需要具有标签数据及部分特征数据。

[0036] (2)数据参与方仅具有部分特征数据。

[0037] (3)各参与方所持有的特征数据数量都分别需要大于3个，不允许存在单特征或者无特征。之所以控制数量，是为了避免信息泄漏风险。

[0038] (4)当然标签数据不一定在模型发起方，也可以在数据参与方，那么相应的流程需要做一定的调整，具有标签数据的一方需要计算梯度中间值信息以及加解密动作。本申请的一个或多个实施例中均以标签数据在模型发起方为例进行阐述。

[0039] (5)特征中的值不允许全0，或者全1。

[0040] 本申请实施例中，在多个参与方的联邦学习过程中，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，模型发起方对梯度中间值混淆第一随机数，数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，

不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0041] 在一些可选的实施方式中，还包括：由模型发起方基于真实标签和预测值计算模型的损失值，根据损失值判断模型是否收敛：若收敛，则确定模型训练完成；若不收敛，则继续迭代更新。

[0042] 在一些可选的实施方式中，梯度中间值由对每一样本特征，将数据参与方和模型发起方的内积相加，得到总内积值，使用预设函数对总内积值进行转换得到预测值，并将预测值与对应的模型发起方的真实标签做差的方式获得；其中，预设函数包括sigmoid函数；预测值由$y=1/(1+e^{-Z})$计算；其中，Z为每条样本的总内积值。

[0043] 在一些可选的实施方式中，混淆了第一随机数的梯度中间值由E_gradf _i = (y_hat_i - y_i)×R_ai计算；其中，y_hat_i为预测值，y_i为真实标签，R_ai为第一随机数，i表示样本索引。

[0044] 在一些可选的实施方式中，混淆了第一随机数和第二随机数的梯度由S_E_gradf _ij = E_gradf _ij×R_bi计算；其中，E_gradf _ij = E_gradf _i×X_bij，X_bij为特征权重，j表示特征索引。

[0045] 在一些可选的实施方式中，混淆了第二随机数的梯度由D_E_gradf _ij = S_E_gradf _ij / R_ai计算。

[0046] 在一些可选的实施方式中，数据参与方的梯度值由gradf _ij = D_E_gradf _ij / R_bi计算。

[0047] 本申请实施例中，纵向联邦学习的模型训练采用小批次的随机梯度下降算法，分批训练做聚合能够提高安全性，适用于一些对安全要求更高的场景下。

[0048] 在一些可选的实施方式中，得到梯度中间值之后，对所有梯度中间值进行过滤处理，得到处理后的梯度中间值；过滤处理包括：将梯度中间值绝对值大于或等于中间值阈值的梯度中间值保留，将梯度中间值绝对值小于中间值阈值的梯度中间值按采样比例进行采样。

[0049] 模型发起方(带标签数据)会将每一个批次全部样本的梯度中间值信息传给数据参与方。在数据量较大的情况下，其通信量会比较可观，在通信上存在大量的性能耗损。通过分析可以发现，当梯度中间值很小的情况下，在参数迭代上，其更新变化的幅度并没有很大，即表示当梯度中间值很小的时候，并不能给予模型更新足够的信息量，因此这类样本可以进行减少，且不会影响整体模型的训练效果，反而可以提升模型训练的性能效率。即，由于梯度中间值的计算是真实y与预测值y'的差，也就是梯度中间值越大，表示真实值y与预测值y'差距越大，表示该样本的梯度中间值可以提供更多的信息量，用于模型参数的更新。

[0050] 因此，模型发起方，首先对梯度中间值按照值的大小进行降序排列，值越大排在越靠前，其目的是为了对齐样本，此处可以采用其他排列方式。定义两个可调参数，一个是中间值阈值v，另一个是采样比例p。我们将梯度中间值绝对值大于或等于中间值阈值v的样本，全部保留。另外，将梯度中间值绝对值小于v的梯度中间值，按采样比例p进行采样。那么最后保留的样本梯度中间值信息，主要是由两部分构成：a. 梯度中间值绝对值大于或等于v；b. 梯度中间值绝对值小于v的部分中占比p的采样部分。下面举例说明：假如某个批次的样本有M条，那么梯度中间值|y-y'|同样量级为M。首先进行排序，保留梯度中间值绝对值大

于或等于v的梯度中间值，且保留对应的样本id或者索引idx，假设共有N条。然后在梯度中间值绝对值小于v的部分，即M-N量级中，随机抽样p比例的量级，得到第二部分需要保留的梯度中间值，即(M-N)×p，同时保留该抽样得到的样本梯度中间值对应的样本id或者索引idx。那么总体剩下的量级为D = N + (M-N)×p。将D与对应的样本id或者索引idx发送给数据参与方，过滤出对应的样本信息，进行后续的梯度计算。在实践过程中，可以发现随着迭代次数的增加，D的量级相比原先整体M的量级，要小的多，在不影响模型训练效果的前提下，大幅减小通信量，达到提升训练性能提升的目的。

[0051] 本申请实施例中，模型发起方会将每一个批次全部样本特征的梯度中间值信息传给数据参与方，在数据量较大的情况下，其通信量会比较可观，在通信上存在大量的性能耗损。因此，对梯度中间值进行过滤处理，将梯度中间值绝对值小于中间值阈值的梯度中间值按采样比例进行采样，将梯度中间值绝对值大于或等于中间值阈值的梯度中间值保留，减小了通信量，提升了模型训练的性能和效率。

[0052] 在一些可选的实施方式中，利用梯度值更新特征权重，包括：判断梯度值的绝对值是否大于梯度阈值，若是，则利用该梯度值更新特征权重；若否，则不更新特征权重。

[0053] 本申请实施例中，当梯度值绝对值很小时，梯度值对参数的更新操作，不能带来模型训练效果上的增益，因此，只对梯度值绝对值大于梯度阈值的梯度值更新特征权重，可以在减少通信量和计算量的同时，不影响模型的整体训练效果，进而达到提升模型训练性能提升目的。

[0054] 请参照图2，图2为本申请实施例提供的一种纵向联邦学习的模型训练方法工作流程图。其中，应用于模型发起方，包括：计算每一样本特征的特征值与特征权重的内积；接收数据参与方发送的内积；对每一样本特征，将数据参与方和模型发起方的内积相加，得到总内积值，使用预设函数对总内积值进行转换得到预测值；对每一样本特征，将预测值与对应的模型发起方的真实标签做差，得到梯度中间值；根据梯度中间值计算模型发起方的梯度值，利用模型发起方的梯度值更新样本特征的特征权重；生成第一随机数，对梯度中间值混淆第一随机数，向数据参与方发送混淆了第一随机数的梯度中间值；其中，第一随机数为不为0的实数；接收数据参与方发送的混淆了第一随机数和第二随机数的梯度；其中，第二随机数为不为0的实数；以及对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度，向数据参与方发送混淆了第二随机数的梯度。

[0055] 本申请实施例提供的一种纵向联邦学习的模型训练方法，数据参与方的工作流程，包括：计算每一样本特征的特征值与特征权重的内积，向模型发起方发送数据参与方的内积；接收模型发起方发送的混淆了第一随机数的梯度中间值；其中，第一随机数为不为0的实数；生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，向模型发起方发送混淆了第一随机数和第二随机数的梯度；其中，第二随机数为不为0的实数；接收模型发起方发送的混淆了第二随机数的梯度；以及对混淆了第二随机数的梯度去除第二随机数，得到数据参与方的梯度值，并利用数据参与方的梯度值更新特征权重。

[0056] 本申请实施例中，在多个参与方的联邦学习过程中，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，模型发起方对梯度中间值混淆第一随机数，数据参与方对混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第

二随机数，之后依次由模型发起方去除第一随机数、数据参与方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0057] 请参照图3，图3为本申请实施例提供的一种纵向联邦学习的模型训练系统的功能模块图，包括内积计算模块1、中间值计算模块2、一次混淆模块3、二次混淆模块4、一次解混模块5和二次解混模块6。

[0058] 其中，内积计算模块1，用于由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；中间值计算模块2，用于对每一样本特征，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值；一次混淆模块3，用于由模型发起方，根据梯度中间值计算模型发起方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至数据参与方；二次混淆模块4，用于由数据参与方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至模型发起方；一次解混模块5，用于由模型发起方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至数据参与方；二次解混模块6，用于由数据参与方，对混淆了第二随机数的梯度去除第二随机数，得到数据参与方的梯度值，并利用数据参与方的梯度值更新特征权重。

[0059] 本申请实施例中，在多个参与方的联邦学习过程中，利用内积计算模块1和中间值计算模块2，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值，再通过一次混淆模块3、二次混淆模块4、一次解混模块5和二次解混模块6，采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0060] 本申请实施例提供的一种纵向联邦学习的模型训练方法，包括：由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积；对每一样本特征，根据数据参与方和模型发起方的内积以及数据参与方的真实标签，得到梯度中间值；由数据参与方，根据梯度中间值计算数据参与方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至模型发起方；由模型发起方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算模型发起方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至数据参与方；由数据参与方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至模型发起方；以及由模型发起方，对混淆了第二随机数的梯度去除第二随机数，得到模型发起方的梯度值，并利用模型发起方的梯度值更新特征权重。

[0061] 本申请实施例中，在多个参与方的联邦学习过程中，数据参与方具有真实标签，由数据参与方计算得到梯度中间值并混淆第一随机数，模型发起方对混淆了第一随机数的梯度中间值计算模型发起方的梯度并混淆第二随机数，之后依次由数据参与方去除第一随机

数、模型发起方去除第二随机数，使得模型发起方和数据参与方都有各自的梯度值，可以对各自特征权重进行更新，由于采用随机数混淆的方式实现数据参与方和模型发起方对梯度信息的加密通信，不受计算次数的限制，不存在精度丢失的问题，能够支持更复杂的联邦学习模型学习需求，并且，降低了加密的耗时，减小通信量，提升模型训练的性能，满足了大规模数据场景下的高性能计算的要求。

[0062] 同样的，本申请实施例之前的多个实施例的方案在适应性调整执行主体，包括模型发起方和数据参与方需进行转换，调整后的多个实施例同样适用于本申请实施例的方案中。例如：

在一些实施方式中，还包括：由数据参与方基于真实标签和预测值计算模型的损失值，根据损失值判断模型是否收敛：若收敛，则确定模型训练完成；若不收敛，则继续迭代更新。

[0063] 在一些实施方式中，梯度中间值由对每一样本特征，将数据参与方和模型发起方的内积相加，得到总内积值，使用预设函数对总内积值进行转换得到预测值，并将预测值与对应的数据参与方的真实标签做差的方式获得；其中，预设函数包括sigmoid函数；预测值由$y=1/(1+e^{-Z})$计算；其中，Z为每条样本的总内积值。

[0064] 在本申请所提供的实施例中，应该理解到，所揭露装置和方法，可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，又例如，多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0065] 另外，作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0066] 再者，在本申请各个实施例中的各功能模块可以集成在一起形成一个独立的部分，也可以是各个模块单独存在，也可以两个或两个以上模块集成形成一个独立的部分。

[0067] 在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。

[0068] 以上所述仅为本申请的实施例而已，并不用于限制本申请的保护范围，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

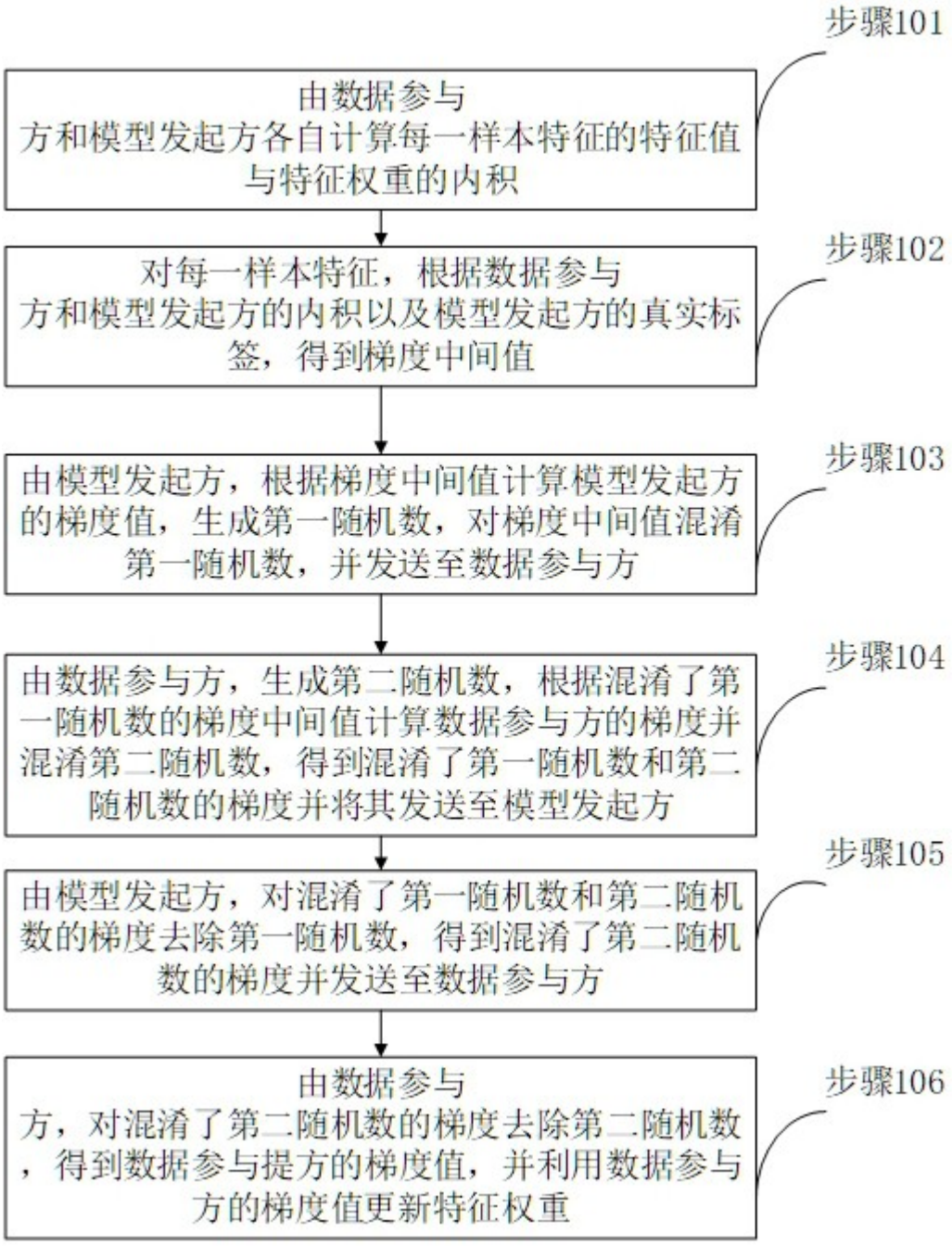

步骤101
由数据参与方和模型发起方各自计算每一样本特征的特征值与特征权重的内积
步骤102
对每一样本特征，根据数据参与方和模型发起方的内积以及模型发起方的真实标签，得到梯度中间值
步骤103
由模型发起方，根据梯度中间值计算模型发起方的梯度值，生成第一随机数，对梯度中间值混淆第一随机数，并发送至数据参与方
步骤104
由数据参与方，生成第二随机数，根据混淆了第一随机数的梯度中间值计算数据参与方的梯度并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度并将其发送至模型发起方
步骤105
由模型发起方，对混淆了第一随机数和第二随机数的梯度去除第一随机数，得到混淆了第二随机数的梯度并发送至数据参与方
步骤106
由数据参与方，对混淆了第二随机数的梯度去除第二随机数，得到数据参与提方的梯度值，并利用数据参与方的梯度值更新特征权重

图1

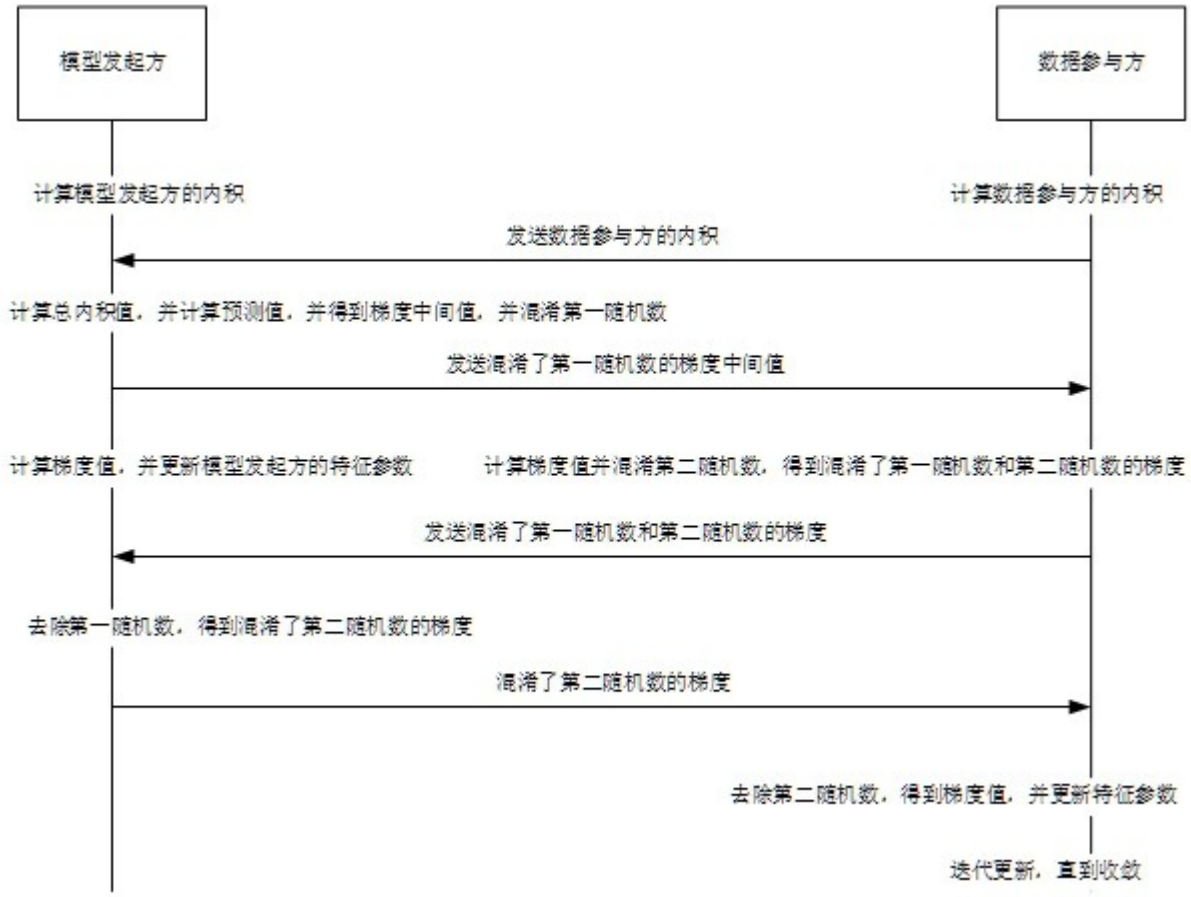

模型发起方
数据参与方
计算模型发起方的内积
计算数据参与方的内积
发送数据参与方的内积
计算总内积值，并计算预测值，并得到梯度中间值，并混淆第一随机数
发送混淆了第一随机数的梯度中间值
计算梯度值，并更新模型发起方的特征参数
计算梯度值并混淆第二随机数，得到混淆了第一随机数和第二随机数的梯度
发送混淆了第一随机数和第二随机数的梯度
去除第一随机数，得到混淆了第二随机数的梯度
混淆了第二随机数的梯度
去除第二随机数，得到梯度值，并更新特征参数
迭代更新，直到收敛

图2

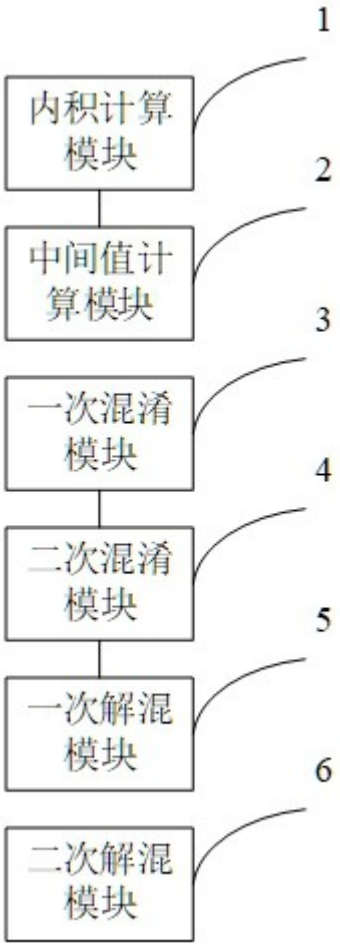


图3