



(12) 发明专利申请

(10) 申请公布号 CN 113240461 A

(43) 申请公布日 2021.08.10

(21) 申请号 202110493396.8

(22) 申请日 2021.05.07

(71) 申请人 广州银行股份有限公司

地址 510620 广东省广州市天河区珠江东路30号

(72) 发明人 赵志东 阚建国 钟海 于晗宇 邓景熹 郑立志

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

代理人 麦小婵 郝传鑫

(51) Int. Cl.

G06Q 30/02 (2012.01)

G06Q 40/04 (2012.01)

G06N 20/00 (2019.01)

G06K 9/62 (2006.01)

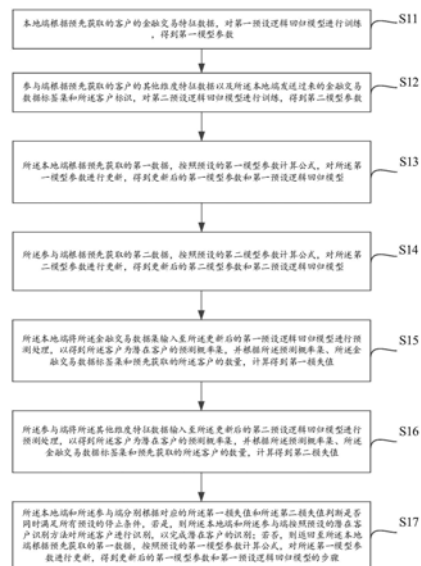
权利要求书4页 说明书14页 附图3页

(54) 发明名称

基于纵向联邦学习的潜在客户的识别方法、系统及介质

(57) 摘要

本发明公开了一种基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过本地端计算得到更新后的第一预设逻辑回归模型的第一损失值,通过参与端计算得到更新后的第二预设逻辑回归模型的第二损失值,最后本地端和参与端分别根据对应的第一损失值和第二损失值判断是否同时满足所有预设的停止条件,若是,则本地端和参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则重复执行指定的步骤并重新进行判断。本发明实施例提供的基于纵向联邦学习的潜在客户的识别方法、系统及介质,从而提高识别潜在客户的准确率。



1. 一种基于纵向联邦学习的潜在客户的识别方法,其特征在于,包括:

本地端根据预先获取的客户的金融交易特征数据,对第一预设逻辑回归模型进行训练,得到第一模型参数;其中,所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集;

参与端根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识,对第二预设逻辑回归模型进行训练,得到第二模型参数;

所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型;

所述参与端根据预先获取的第二数据,按照预设的第二模型参数计算公式,对所述第二模型参数进行更新,得到更新后的第二模型参数和第二预设逻辑回归模型;

所述本地端将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值;

所述参与端将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值;

所述本地端和所述参与端分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

2. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述客户的金融交易特征数据的获取方式具体包括:

获取客户的金融交易数据集和客户标识;

根据所述客户的金融交易数据集得到所述客户的金融交易数据标签集;

对所述客户的金融交易数据集中包含的所有缺失值数据采用负无穷数值进行补全,得到补全后的客户的金融交易数据集;

将所述补全后的客户的金融交易数据集中包含的所有日期型数据修改为对应的距离指定观察日的天数的格式的数据,得到修改后的金融交易数据集;

对所述修改后的金融交易数据集进行数据分箱处理,得到预处理后的金融交易数据集;

将所述预处理后的金融交易数据集按照所述客户标识进行划分,得到客户的金融交易特征数据。

3. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述客户的其他维度特征数据的获取方式具体为:

所述本地端对所述客户标识进行哈希加密,并将加密后的客户标识上传至所述联邦大数据集市;

所述联邦大数据集市将内部的加密信息与所述加密后的客户标识进行撞库匹配,得到客户的其他维度特征数据,并将所述其他维度特征数据发送至参与端。

4. 根据权利要求3所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述本地端获取第一数据和所述参与端获取第二数据的步骤具体为:

联邦参数仲裁服务器根据预先获取的随机安全数生成公钥-私钥对,并将所述公钥-私钥对中的公钥分别发送至所述本地端和所述参与端;

所述参与端根据所述公钥对所述第二模型参数和所述其他维度特征数据的乘积结果进行加密,得到第一加密数据,并将所述第一加密数据发送至所述本地端;

所述本地端根据所述公钥对所述第一模型参数和所述金融交易特征数据的乘积结果进行加密,得到第二加密数据;

所述本地端根据所述第一加密数据、所述第二加密数据和所述金融交易数据标签集,按照预设的梯度权重计算公式,得到所述本地端的第一预设逻辑回归模型的加密后的梯度权重,并将所述加密后的梯度权重发送至所述参与端;

所述本地端将所述加密后的梯度权重和所述金融交易特征数据的乘积结果与预先获取的第一随机数相加,得到第一组合数据,并将所述第一组合数据发送至所述联邦参数仲裁服务器;

所述参与端将所述加密后的梯度权重和所述其他维度特征数据的乘积结果与预先获取的第二随机数相加,得到第二组合数据,并将所述第二组合数据发送至所述联邦参数仲裁服务器;

所述联邦参数仲裁服务器根据所述公钥-私钥对中的私钥分别对所述第一组合数据和所述第二组合数据进行解密,并将解密后的第一组合数据发送至所述本地端,将解密后的第二组合数据发送至所述参与端;

所述本地端将所述解密后的第一组合数据中的随机数减去,得到第一数据;

所述参与端将所述解密后的第二组合数据中的随机数减去,得到第二数据。

5. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述预设的第一模型参数计算公式具体为:

$$\left(\frac{1}{2}Y(w')^T X - 1\right) \frac{1}{2} YX = gX$$

其中, $X$ 为所述金融交易特征数据, $Y$ 为所述金融交易数据标签集, $w'$ 为所述第一模型参数, $gX$ 为所述第一数据。

6. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述本地端将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值的计算公式具体为:

$$\text{Loss}_1 = -\frac{1}{n} \sum_{i=1}^n [y^i \ln p^i + (1 - y^i) \ln (1 - p^i)]$$

其中, $n$ 为所述客户的数量, $y^i$ 为第*i*个客户的是否为潜在客户的数值标签, $n$ 个客户的所述数值标签的集合即为所述金融交易数据标签集, $p^i$ 为第*i*个客户是否为潜在客户的预测概率, $n$ 个客户的所述预测概率的集合即为所述预测概率集。

7. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法,其特征在于,所述

预设的停止条件包括第一条件、第二条件和第三条件；

其中，所述第一条件具体为：

所述本地端连续k次满足本轮计算得到的所述第一损失值与相对于所述本轮的上一轮计算得到的第一损失值的差值的绝对值不大于预设的阈值；

所述第二条件具体为：

所述参与端连续k次满足本轮计算得到的所述第二损失值相对于所述本轮的上一轮计算得到的第二损失值的差值的绝对值不大于预设的阈值；

所述第三条件具体为：

所述本地端和所述参与端分别同时满足其对应的所述第一条件和所述第二条件。

8. 根据权利要求1所述的基于纵向联邦学习的潜在客户的识别方法，其特征在于，所述所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别，以完成潜在客户的识别，具体包括：

所述本地端将所述客户输入至所述更新后的第一预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的第一概率集，并将所述第一概率集加密后发送至所述联邦参数仲裁服务器；

所述参与端将所述客户输入至所述更新后的第二预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的第二概率集，并将所述第二概率集加密后发送至所述联邦参数仲裁服务器；

所述联邦参数仲裁服务器分别对加密后的所述第一概率集和加密后的所述第二概率集进行解密，并根据解密后的第一概率集和解密后的第二概率集，按照预设的概率计算方法，得到所述客户为潜在客户的综合概率集；

所述联邦参数仲裁服务器将所述潜在客户的综合概率集中的每一概率值与预设的潜在客户阈值进行逐一比较，得到每一概率比较结果，并根据所述每一概率比较结果，按照预设的概率判断方法，得到所述客户的潜在客户综合标签集，并将所述潜在客户综合标签集分别发送至所述本地端和所述参与端，以完成潜在客户的识别。

9. 一种基于纵向联邦学习的潜在客户的识别系统，其特征在于，包括本地端和参与端；其中：

所述本地端，用于根据预先获取的客户的金融交易特征数据，对第一预设逻辑回归模型进行训练，得到第一模型参数；其中，所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集；

所述参与端，用于根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识，对第二预设逻辑回归模型进行训练，得到第二模型参数；

所述本地端，还用于根据预先获取的第一数据，按照预设的第一模型参数计算公式，对所述第一模型参数进行更新，得到更新后的第一模型参数和第一预设逻辑回归模型；

所述参与端，还用于根据预先获取的第二数据，按照预设的第二模型参数计算公式，对所述第二模型参数进行更新，得到更新后的第二模型参数和第二预设逻辑回归模型；

所述本地端，还用于将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理，以得到所述客户为潜在客户的预测概率集，并根据所述预测概率集、所述

金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值;

所述参与端,还用于将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值;

所述本地端和所述参与端,还用于分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如权利要求1至8中任意一项所述的基于纵向联邦学习的潜在客户的识别方法。

## 基于纵向联邦学习的潜在客户的识别方法、系统及介质

### 技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种基于纵向联邦学习的潜在客户的识别方法、系统及介质。

### 背景技术

[0002] 当前随着生活科技化、数字化程度越来越高,公众在吃、穿、住、行、用等各方面产生了大量隐私数据,这些全方位的数据给人工智能、区块链、云计算、大数据等新兴技术的创新应用提供了肥沃土壤。与此同时,如何利用新兴技术安全高效地挖掘用户数据并创造商业价值,也越来越成为银行等金融机构关注的重点。由于客户在柜台、手机客户端、网上客户端等渠道办理开户、转账、存取款、交易等金融业务时会产生大量的金融交易数据,这些数据潜藏着客户基础信息、持有资产、投资倾向等,如果可以挖掘利用这些数据,就能精确的触达客户的潜在需求,识别哪些客户为潜在理财推荐、保险推荐、贷款推荐的目标客户。但是,本发明人发现,现有技术一般使用单一模型对潜在客户进行识别,导致识别潜在客户的预测效果不佳。

### 发明内容

[0003] 本发明提供一种基于纵向联邦学习的潜在客户的识别方法、系统及介质,以解决现有技术采用单一模型识别潜在客户的预测效果不佳的问题,通过采用基于纵向联邦学习的联合建模方法对潜在客户进行识别,从而提高识别潜在客户的准确率。

[0004] 为实现上述目的,本发明实施例提供了一种基于纵向联邦学习的潜在客户的识别方法,包括以下步骤:

[0005] 本地端根据预先获取的客户的金融交易特征数据,对第一预设逻辑回归模型进行训练,得到第一模型参数;其中,所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集;

[0006] 参与端根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识,对第二预设逻辑回归模型进行训练,得到第二模型参数;

[0007] 所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型;

[0008] 所述参与端根据预先获取的第二数据,按照预设的第二模型参数计算公式,对所述第二模型参数进行更新,得到更新后的第二模型参数和第二预设逻辑回归模型;

[0009] 所述本地端将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值;

[0010] 所述参与端将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述

金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值;

[0011] 所述本地端和所述参与端分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

[0012] 进一步的,所述客户的金融交易特征数据的获取方式具体包括:

[0013] 获取客户的金融交易数据集和客户标识;

[0014] 根据所述客户的金融交易数据集得到所述客户的金融交易数据标签集;

[0015] 对所述客户的金融交易数据集中包含的所有缺失值数据采用负无穷数值进行补全,得到补全后的客户的金融交易数据集;

[0016] 将所述补全后的客户的金融交易数据集中包含的所有日期型数据修改为对应的距离指定观察日的天数的格式的数据,得到修改后的金融交易数据集;

[0017] 对所述修改后的金融交易数据集进行数据分箱处理,得到预处理后的金融交易数据集;

[0018] 将所述预处理后的金融交易数据集按照所述客户标识进行划分,得到客户的金融交易特征数据。

[0019] 进一步的,所述客户的其他维度特征数据的获取方式具体为:

[0020] 所述本地端对所述客户标识进行哈希加密,并将加密后的客户标识上传至所述联邦大数据集市;

[0021] 所述联邦大数据集市将内部的加密信息与所述加密后的客户标识进行撞库匹配,得到客户的其他维度特征数据,并将所述其他维度特征数据发送至参与端。

[0022] 进一步的,所述本地端获取第一数据和所述参与端获取第二数据的步骤具体为:

[0023] 联邦参数仲裁服务器根据预先获取的随机安全数生成公钥-私钥对,并将所述公钥-私钥对中的公钥分别发送至所述本地端和所述参与端;

[0024] 所述参与端根据所述公钥对所述第二模型参数和所述其他维度特征数据的乘积结果进行加密,得到第一加密数据,并将所述第一加密数据发送至所述本地端;

[0025] 所述本地端根据所述公钥对所述第一模型参数和所述金融交易特征数据的乘积结果进行加密,得到第二加密数据;

[0026] 所述本地端根据所述第一加密数据、所述第二加密数据和所述金融交易数据标签集,按照预设的梯度权重计算公式,得到所述本地端的第一预设逻辑回归模型的加密后的梯度权重,并将所述加密后的梯度权重发送至所述参与端;

[0027] 所述本地端将所述加密后的梯度权重和所述金融交易特征数据的乘积结果与预先获取的第一随机数相加,得到第一组合数据,并将所述第一组合数据发送至所述联邦参数仲裁服务器;

[0028] 所述参与端将所述加密后的梯度权重和所述其他维度特征数据的乘积结果与预先获取的第二随机数相加,得到第二组合数据,并将所述第二组合数据发送至所述联邦参数仲裁服务器;

[0029] 所述联邦参数仲裁服务器根据所述公钥-私钥对中的私钥分别对所述第一组合数

据和所述第二组合数据进行解密,并将解密后的第一组合数据发送至所述本地端,将解密后的第二组合数据发送至所述参与端;

[0030] 所述本地端将所述解密后的第一组合数据中的随机数减去,得到第一数据;

[0031] 所述参与端将所述解密后的第二组合数据中的随机数减去,得到第二数据。

[0032] 进一步的,所述预设的第一模型参数计算公式具体为:

$$[0033] \quad \left(\frac{1}{2}Y(w')^T X - 1\right) \frac{1}{2}YX = gX$$

[0034] 其中,X为所述金融交易特征数据,Y为所述金融交易数据标签集,w'为所述第一模型参数,gX为所述第一数据。

[0035] 进一步的,所述本地端将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值的计算公式具体为:

$$[0036] \quad \text{Loss}_1 = -\frac{1}{n} \sum_{i=1}^n [y^i \ln p^i + (1 - y^i) \ln (1 - p^i)]$$

[0037] 其中,n为所述客户的数量, $y^i$ 为第i个客户的是否为潜在客户的数值标签,n个客户的所述数值标签的集合即为所述金融交易数据标签集, $p^i$ 为第i个客户是否为潜在客户的预测概率,n个客户的所述预测概率的集合即为所述预测概率集。

[0038] 进一步的,所述预设的停止条件包括第一条件、第二条件和第三条件;

[0039] 其中,所述第一条件具体为:

[0040] 所述本地端连续k次满足本轮计算得到的所述第一损失值与相对于所述本轮的上一轮计算得到的第一损失值的差值的绝对值不大于预设的阈值;

[0041] 所述第二条件具体为:

[0042] 所述参与端连续k次满足本轮计算得到的所述第二损失值相对于所述本轮的上一轮计算得到的第二损失值的差值的绝对值不大于预设的阈值;

[0043] 所述第三条件具体为:

[0044] 所述本地端和所述参与端分别同时满足其对应的所述第一条件和所述第二条件。

[0045] 进一步的,所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别,具体包括:

[0046] 所述本地端将所述客户输入至所述更新后的第一预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的所述第一概率集,并将所述第一概率集加密后发送至所述联邦参数仲裁服务器;

[0047] 所述参与端将所述客户输入至所述更新后的第二预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的第二概率集,并将所述第二概率集加密后发送至所述联邦参数仲裁服务器;

[0048] 所述联邦参数仲裁服务器分别对加密后的所述第一概率集和加密后的所述第二概率集进行解密,并根据解密后的所述第一概率集和解密后的所述第二概率集,按照预设的概率计算方法,得到所述客户为潜在客户的综合概率集;

[0049] 所述联邦参数仲裁服务器将所述潜在客户的综合概率集中的每一概率值与预设的潜在客户阈值进行逐一比较,得到每一概率比较结果,并根据所述每一概率比较结果,按照预设的概率判断方法,得到所述客户的潜在客户综合标签集,并将所述潜在客户综合标签集分别发送至所述本地端和所述参与端,以完成潜在客户的识别。

[0050] 本发明另一实施例对应提供了一种基于纵向联邦学习的潜在客户的识别系统,其特征在于,包括本地端和参与端;其中:

[0051] 所述本地端,用于根据预先获取的客户的金融交易特征数据,对第一预设逻辑回归模型进行训练,得到第一模型参数;其中,所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集;

[0052] 所述参与端,用于根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识,对第二预设逻辑回归模型进行训练,得到第二模型参数;

[0053] 所述本地端,还用于根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型;

[0054] 所述参与端,还用于根据预先获取的第二数据,按照预设的第二模型参数计算公式,对所述第二模型参数进行更新,得到更新后的第二模型参数和第二预设逻辑回归模型;

[0055] 所述本地端,还用于将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值;

[0056] 所述参与端,还用于将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值;

[0057] 所述本地端和所述参与端,还用于分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

[0058] 本发明另一实施例对应提供了一种计算机可读存储介质,其特征在于,所述计算机可读存储介质包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如上述发明实施例所述的基于纵向联邦学习的潜在客户的识别方法。

[0059] 与现有技术相比,本发明实施例公开的一种基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过本地端得到训练完成后的第一预设逻辑回归模型的第一模型参数,并按照预设的第一模型参数计算公式,更新第一模型参数,然后根据更新后的第一模型参数计算得到更新后的所述第一预设逻辑回归模型的第一损失值,同时,通过参与端得到训练完成后的第二预设逻辑回归模型的第二模型参数,并按照预设的第二模型参数计算公式,更新第二模型参数,然后根据更新后的第二模型参数计算得到更新后的第二预设逻辑回归模型的第二损失值,最后本地端和参与端分别根据对应的第一损失值和第二损失值判断是否同时满足所有预设的停止条件,若是,则本地端和参与端按照预设的潜在客户识别

方法对所述客户进行识别,以完成潜在客户的识别;若否,则重复执行指定的步骤并重新进行判断。本发明实施例提供的基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过采用基于纵向联邦学习的联合建模方法对潜在客户进行识别,能够在保证客户的隐私数据安全可控的前提下,克服各公司之间客户数据种类单一和不均衡的问题,实现各公司之间的客户数据共享,从而提高识别潜在客户的准确率,精准触达不同业务场景下的潜在客户。

### 附图说明

[0060] 图1是本发明实施例提供的一种基于纵向联邦学习的潜在客户的识别方法的流程图示意图;

[0061] 图2是本发明实施例提供的一种基于纵向联邦学习的潜在客户的识别方法的另一流程图示意图;

[0062] 图3是本发明实施例提供的一种潜在客户的识别系统的结构示意图。

### 具体实施方式

[0063] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0064] 在本申请的描述中,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。在本申请的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0065] 参见图1,是本发明实施例提供的一种基于纵向联邦学习的潜在客户的识别方法的流程图示意图,所述方法包括步骤S11至步骤S17:

[0066] S11、本地端根据预先获取的客户的金融交易特征数据,对第一预设逻辑回归模型进行训练,得到第一模型参数;其中,所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集。

[0067] S12、参与端根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识,对第二预设逻辑回归模型进行训练,得到第二模型参数。

[0068] S13、所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型。

[0069] S14、所述参与端根据预先获取的第二数据,按照预设的第二模型参数计算公式,对所述第二模型参数进行更新,得到更新后的第二模型参数和第二预设逻辑回归模型。

[0070] S15、所述本地端将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值。

[0071] S16、所述参与端将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、

所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值。

[0072] S17、所述本地端和所述参与端分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端和所述参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

[0073] 与现有技术相比,本发明实施例公开的一种基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过本地端得到训练完成后的第一预设逻辑回归模型的第一模型参数,并按照预设的第一模型参数计算公式,更新第一模型参数,然后根据更新后的第一模型参数计算得到更新后的所述第一预设逻辑回归模型的第一损失值,同时,通过参与端得到训练完成后的第二预设逻辑回归模型的第二模型参数,并按照预设的第二模型参数计算公式,更新第二模型参数,然后根据更新后的第二模型参数计算得到更新后的第二预设逻辑回归模型的第二损失值,最后本地端和参与端分别根据对应的第一损失值和第二损失值判断是否同时满足所有预设的停止条件,若是,则本地端和参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则重复执行指定的步骤并重新进行判断。本发明实施例提供的基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过采用基于纵向联邦学习的联合建模方法对潜在客户进行识别,能够在保证客户的隐私数据安全可控的前提下,克服各公司之间客户数据种类单一和不均衡的问题,实现各公司之间的客户数据共享,从而提高识别潜在客户的准确率,精准触达不同业务场景下的潜在客户。

[0074] 需要说明的是,所述第一预设逻辑回归模型和所述第二预设逻辑回归模型为同一逻辑回归模型,只是分别被本地端和参与端使用。逻辑回归模型为一种广义线性的机器学习模型,用于分类模型,常用于二分类模型。因其简单、可并行化、可解释的特性,在工业届被广泛接受。逻辑回归模型核心假设是数据服从某一种分布,然后使用极大似然估计做参数的估计算法。其中,第一预设逻辑回归模型的表达式为:

$$[0075] \quad LR_{\text{model}} = \frac{1}{1 + e^{-(w^T x + b)}}$$

[0076] 所述第一预设逻辑回归模型的损失函数的表达式为:

$$[0077] \quad L(w, x, y) = \log(1 + e^{-yw^T x})$$

[0078] 其中, $w$ 为所述预设回归模型的参数, $w^T$ 为所述参数的转置, $x$ 为特征数据, $y$ 为所述特征数据对应的标签, $b$ 为偏置参数。

[0079] 需要说明的是,金融交易数据标签集是根据客户在银行内部是否进行过特定的交易来进行定义的,若客户在银行内部进行过特定的交易,则将所述客户的金融交易数据对应的标签设置为1,反之,设置为-1,并根据每一客户的金融交易数据构成的金融交易数据集得到每一客户对应的标签,即得到金融交易数据标签集。

[0080] 作为其中一个可选的实施方式,在步骤S12中,本地端是通过虚拟网络专用信道(以下简称专线)将金融交易数据标签集传送至参与端的,但是本地端并未告知具体标签的含义,参与端只能接受到单纯的(-1,1)数值数据。

[0081] 值得说明的是,为解决当前各商业公司之间数据不能互通进行联合建模的问题,

缓解“数据不能出本地”的瓶颈,本实施例通过在保证客户隐私数据安全可控、不泄露出本地的前提下共享数据,打破“数据孤岛”的困境,克服各银行业金融机构中客户数据种类单一、不均衡的现状,让不同银行业金融机构、公司之间的客户数据共享,实现利用多种、多维数据联合建模,精准触达不同金融业务场景下的潜在客户。

[0082] 作为上述方案的改进,所述客户的金融交易特征数据的获取方式具体包括:

[0083] 获取客户的金融交易数据集和客户标识;

[0084] 根据所述客户的金融交易数据集得到所述客户的金融交易数据标签集;

[0085] 对所述客户的金融交易数据集中包含的所有缺失值数据采用负无穷数值进行补全,得到补全后的客户的金融交易数据集;

[0086] 将所述补全后的客户的金融交易数据集中包含的所有日期型数据修改为对应的距离指定观察日的天数的格式的数据,得到修改后的金融交易数据集;

[0087] 对所述修改后的金融交易数据集进行数据分箱处理,得到预处理后的金融交易数据集;

[0088] 将所述预处理后的金融交易数据集按照所述客户标识进行划分,得到客户的金融交易特征数据。

[0089] 需要说明的是,通过结构化查询语句从全体存量客户群体中筛查出一批目标客户群,得到目标客群的金融交易数据,其中,所述金融交易数据 $[(X, Y)]$ 就包括客户的金融交易数据集 $X$ 和金融交易数据标签集 $Y$ 。

[0090] 示例性的, $X = [x_{11}, x_{12}, \dots, x_{21}, x_{22}, \dots, x_{ij}, \dots]$ ,其中, $x_{ij}$ 表示第 $i$ 个客户的第 $j$ 维金融交易特征, $i$ 的取值范围为 $[1, 66, 274]$ , $j$ 的取值范围为 $[1, 65]$ , $Y = [y_1, y_2, \dots, y, \dots]$ ,其中, $y_i$ 表示第 $i$ 个客户是否为潜在客户的数值标签,该标签是根据机构内存有客户是否有购买(理财、保险)或持有(贷款)金融产品等的历史数据进行标记,若有则标记为1(是),反之为-1(否)。

[0091] 作为其中一个可选的实施方式,对所述修改后的金融交易数据集进行数据分箱处理,得到预处理后的金融交易数据集,具体为对所述修改后的金融交易数据集进行数据分箱处理,将连续数据进行离散化分段处理并进行分类,其中,数据分箱处理的原则如下:基于数据连续分布形态,可将特征等频分箱为5类;基于数据零/非零分布形态,可将特征分为2类;基于数据负/零/正分布形态,可将特征分为3类;基于数据无/有数值形态,可将特征分为2类。

[0092] 示例性的,将所述预处理后的金融交易数据集 $[(X, Y)]$ 按照所述客户标识进行划分,得到客户的金融交易特征数据 $[(UID, X, Y)]$ ,即为 $[(X, Y)] \Rightarrow [(UID, X, Y)]$ ,其中,UID是所述客户的标识,对于每一客户是唯一存在的,一般是指客户的手机号, $UID = [uid_1, uid_2, \dots, uid_i, \dots]$ ,其中, $uid_i$ 表示第 $i$ 个客户的唯一标识。

[0093] 值得说明的是,通过上述的数据分箱处理的规则,可以得到预处理后的金融交易数据集,其中所述数据分箱处理能够使及其学习模型训练计算的迭代过程更加稳定,并降低模拟过拟合的风险。

[0094] 作为上述方案的改进,所述客户的其他维度特征数据的获取方式具体为:

[0095] 所述本地端对所述客户标识进行哈希加密,并将加密后的客户标识上传至所述联邦大数据集市;

[0096] 所述联邦大数据集市将内部的加密信息与所述加密后的客户标识进行撞库匹配，得到客户的其他维度特征数据，并将所述其他维度特征数据发送至参与端。

[0097] 需要说明的是，联邦大数据集市也称为潜在特征联邦大数据集市 (Potential Feature Federal Big Data Mart, PFDM)，是由各联邦参与端联合建成的一个保护隐私的包含各类客户潜在特征的大数据合作平台，以联邦学习为基础设施，各方共同贡献出自身脱敏的客户潜在特征数据，所有数据按种类可以分为若干类，其中每一类数据的特征维度均在几千维左右，所有数据经过清洗与治理，以客户唯一标识ID为索引。各联邦方均可在联邦大数据集市上通过比对自身客户唯一标识ID与数据集市的客户唯一标识ID，即撞库匹配得到对应的数据。数据种类包括但不限于设备类(设备属性、APP安装与使用偏好)、基础类(基础属性)、金融类(财富属性、电商属性、支付属性)、活动类(到访偏好、旅行偏好)。根据联邦学习的设计原理，各方只能从潜在特征联邦大数据集市中获得数值型数据，无法得知具体获得哪些种类数据。

[0098] 需要说明的是，联邦大数据集市的内部的加密信息是指对客户手机号进行哈希加密后的哈希值，联邦大数据集市将所述哈希值与本地端上传的加密后的客户标识值进行对比，撞库匹配即得到同一批客户的其他维度特征数据，即： $\text{Hash}(\text{UID}) \xrightarrow{\text{PFDM}} Z$ ，有 $Z = [z_{11}, z_{12}, \dots, z_{21}, z_{22}, \dots, z_{ik}, \dots]$ ，其中 $z_{ik}$ 表示第 $i$ 个客户的第 $k$ 维其他潜在特征， $k$ 的取值范围为 $[1, 3000]$ 。

[0099] 作为上述方案的改进，所述本地端获取第一数据和所述参与端获取第二数据的步骤具体为：

[0100] 联邦参数仲裁服务器根据预先获取的随机安全数 $\lambda$ 生成公钥-私钥对 $(p_k, s_k)$ ，并将所述公钥-私钥对中的公钥 $p_k$ 分别发送至所述本地端和所述参与端；

[0101] 所述参与端根据所述公钥 $p_k$ 对所述第二模型参数和所述其他维度特征数据 $Z$ 的乘积 $w''Z$ 进行加密，得到第一加密数据 $[[w''Z]]$ ，并将所述第一加密数据发送至所述本地端；

[0102] 所述本地端根据所述公钥 $p_k$ 对所述第一模型参数 $w'$ 和所述金融交易特征数据 $X$ 的乘积 $w'X$ 进行加密，得到第二加密数据 $[[w'X]]$ ；

[0103] 所述本地端根据所述第一加密数据 $[[w''Z]]$ 、第二加密数据 $[[w'X]]$ 和所述金融交易数据标签集 $Y$ ，按照预设的梯度权重计算公式，得到所述本地端的第一预设逻辑回归模型的加密后的梯度权重 $[[g]]$ ，并将所述加密后的梯度权重发送至所述参与端；

[0104] 所述本地端将所述加密后的梯度权重 $[[g]]$ 和所述金融交易特征数据 $X$ 的乘积 $[[gX]]$ 与预先获取的第一随机数 $\sigma_1$ 相加，得到第一组合数据 $([[gX]] + \sigma_1)$ ，并将所述第一组合数据发送至所述联邦参数仲裁服务器；

[0105] 所述参与端将所述加密后的梯度权重 $[[g]]$ 和所述其他维度特征数据 $Z$ 的乘积 $[[g]]Z$ 与预先获取的第二随机数 $\sigma_2$ 相加，得到第二组合数据 $([[gZ]] + \sigma_2)$ ，并将所述第二组合数据发送至所述联邦参数仲裁服务器；

[0106] 所述联邦参数仲裁服务器根据所述公钥-私钥对中的私钥 $s_k$ 分别对所述第一组合数据和所述第二组合数据进行解密，并将解密后的第一组合数据 $(gX + \sigma_1)$ 发送至所述本地端，将解密后的第二组合数据 $(gZ + \sigma_2)$ 发送至所述参与端；

[0107] 所述本地端将所述解密后的第一组合数据 $(gX + \sigma_1)$ 中的随机数 $\sigma_1$ 减去，得到第一数

据 $gX$ ;

[0108] 所述参与端将所述解密后的第二组合数据( $gZ+\sigma_2$ )中的随机数 $\sigma_2$ 减去,得到第二数据 $gZ$ 。

[0109] 需要说明的,上述通过公钥-私钥对对数据进行加密是采用的同态加密算法,同态加密算法是指满足密文同态运算性质的加密算法,即明文数据经过同态加密之后得到密文数据,对密文数据进行特定方式计算后得到的密文计算结果,再经过对应的同态解密后的明文,等同于明文数值直接以相同方式计算得出的明文结果,可实现数据的“可算不可见”。支持密文以任意形式计算的算法称为全同态加密(Fully Homomorphic Encryption,FHE),仅支持加法、乘法、有限次加法、有限次乘法的算法称为半同态加密或部分加密(Partially Homomorphic Encryption,PHE)。

[0110] 其中,需要特别指出,半同态加密算法(PHE)具有以下特性: $PHE_{EN}(*,p_k)=[[*]]$ , $[[u]]+[[v]]=[[u+v]]$ , $n[[u]]=[[nu]]$ 。其中, $PHE_{EN}(*,p_k)=[[*]]$ 表示任一明文数据\*和公钥 $p_k$ (Public Key),经过半同态加密算法 $PHE_{EN}$ 计算,可得到明文数据的加密形式 $[[*]]$ ;  $[[u]]+[[v]]=[[u+v]]$ 表示半同态加密密文数据支持有限次加法; $u[[v]]=[[uv]]$ 表示半同态加密密文数据支持有限次乘法。另外,对于半同态加密后的数据可以进行半同态解密,即 $PHE_{DE}([[*]],s_k)=*$ ,表示加密数据 $[[*]]$ 和私钥 $s_k$ (Secret Key),经过半同态解密算法 $PEH_{DE}$ 计算,可得到加密数据的明文形式。

[0111] 作为其中一个可选的实施方式,所述按照预设的梯度权重计算方法,得到所述本地端的第一预设逻辑回归模型的加密后的梯度权重的计算公式具体为:

$$[0112] \quad [[g]] = \left( \frac{1}{2} Y([[w''Z]] + [[w'X]] - 1) \right) \frac{1}{2} Y$$

[0113] 其中, $[[g]]$ 为所述加密后的梯度权重, $[[w''Z]]$ 为第一加密数据, $[[w'X]]$ 为第二加密数据, $Y$ 为所述金融交易数据标签集, $w'$ 为所述第一模型参数, $w''$ 为所述第二模型参数, $X$ 为所述金融交易特征数据, $Z$ 为所述其他维度特征数据。

[0114] 需要说明的是,第一随机数和第二随机数是由随机数生成程序产生的随机数。

[0115] 作为上述方案的改进,所述所述本地端根据预先获取的第一数据,按照预设的第一模型参数计算方法,更新所述第一模型参数,得到更新后的第一模型参数的表达式具体为:

$$[0116] \quad \left( \frac{1}{2} Y(w')^T X - 1 \right) \frac{1}{2} YX = gX$$

[0117] 其中, $X$ 为所述金融交易特征数据, $Y$ 为所述金融交易数据标签集, $w'$ 为所述第一模型参数, $gX$ 为所述第一数据。

[0118] 作为上述方案的改进,所述所述参与端根据预先获取的第二数据,按照预设的第二模型参数计算方法,更新所述第二模型参数,得到更新后的第二模型参数的表达式具体为:

$$[0119] \quad \left( \frac{1}{2} Y(w'')^T Z - 1 \right) \frac{1}{2} YZ = gZ$$

[0120] 其中, $Z$ 为所述其他维度特征数据, $Y$ 为所述金融交易数据标签集, $w''$ 为所述第一模型参数, $gZ$ 为所述第二数据。

[0121] 作为上述方案的改进,所述步骤S15具体为:

[0122] 本地端将金融交易数据集输入至更新后的第一预设逻辑回归模型进行预测处理,得到客户是否为潜在客户的预测概率集,包括每一客户是潜在客户和不是潜在客户的预测概率,然后根据所述预测概率集、金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值。

[0123] 示例性的,本地端通过模型参数计算模型损失值的公式如下:

$$[0124] \quad \text{Loss}_1 = -\frac{1}{n} \sum_{i=1}^n [y^i \ln p^i + (1 - y^i) \ln (1 - p^i)]$$

[0125] 其中,n为所述客户的数量, $y^i$ 代表第i个客户的是否为潜在客户的数值标签,n个客户的数值标签的集合即为金融交易数据标签集, $p^i$ 为第i个客户是否为潜在客户的预测概率,n个客户的预测概率的集合即为所述预测概率集。

[0126] 作为上述方案的改进,所述步骤S16具体为:

[0127] 参与端将其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,得到客户是否为潜在客户的预测概率集,包括每一客户是潜在客户和不是潜在客户的预测概率,然后根据所述预测概率集、金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值。

[0128] 示例性的,参与端通过模型参数计算模型损失值的公式如下:

$$[0129] \quad \text{Loss}_2 = -\frac{1}{n} \sum_{i=1}^n [y^i \ln q^i + (1 - y^i) \ln (1 - q^i)]$$

[0130] 其中,n为所述客户的数量, $y^i$ 代表第i个客户的是否为潜在客户的数值标签,n个客户的数值标签的集合即为金融交易数据标签集, $q^i$ 为第i个客户是否为潜在客户的预测概率,n个客户的预测概率的集合即为所述预测概率集。

[0131] 作为上述方案的改进,本地端保存原始逻辑回归模型的参数和每轮联邦参数仲裁服务器通过专线传送来的最新逻辑回归模型参数,预先设置停止标志 $\text{STOP}_1=0$ 。当本地端通过模型参数 $w'_t$ 计算出来的损失值和上一次接收的逻辑回归模型参数 $w'_{t-1}$ 计算出来的损失值的差值绝对值在设定阈值以下,且已经连续出现了n次,即将停止标志设为1,再加密成 $[[\text{STOP}_1]]$ ,通过专线传送给联邦参数仲裁服务器。

[0132] 同时,参与端保存原始逻辑回归模型的参数和每轮联邦参数仲裁服务器通过专线传送来的最新逻辑回归模型参数,预先设置停止标志 $\text{STOP}_2=0$ 。当参与端通过模型参数 $w''_t$ 计算出来的损失值和上一次接收的逻辑回归模型参数 $w''_{t-1}$ 计算出来的损失值的差值绝对值在设置阈值之下,且已经连续出现了n次,即将停止标志设为1。再经过半同态算法加密成 $[[\text{STOP}_1]]$ ,通过专线传送给联邦参数仲裁服务器。

[0133] 需要说明的是,由于同态加密算法对逻辑回归模型损失函数和梯度函数的限制,在本地端和与参与方是通过使用泰勒公式对上述逻辑回归模型的损失函数和梯度权重的计算公式进行近似展开的。

[0134] 联邦参数仲裁服务器在接收到本地端和参与端通过专线传送过来的各自的停止标志后,经过半同态解密为 $\text{STOP}_1$ 、 $\text{STOP}_2$ ,当且仅当识别 $\text{STOP}_1$ 、 $\text{STOP}_2$ 都为1时,满足预设停止条件。

[0135] 联邦参数仲裁服务器发送全体停止标志 $\text{STOP}$ 给本地端和参与端,表示本地端和参

与端已联合建模训练完成。

[0136] 作为上述方案的改进,所述预设的停止条件包括第一条件、第二条件和第三条件;

[0137] 其中,所述第一条件具体为:

[0138] 所述本地端连续k次满足本轮计算得到的所述第一损失值与相对于所述本轮的上一轮计算得到的第一损失值的差值的绝对值不大于预设的阈值;

[0139] 所述第二条件具体为:

[0140] 所述参与端连续k次满足本轮计算得到的所述第二损失值相对于所述本轮的上一轮计算得到的第二损失值的差值的绝对值不大于预设的阈值;

[0141] 所述第三条件具体为:

[0142] 所述本地端和所述参与端分别同时满足其对应的所述第一条件和所述第二条件。

[0143] 需要说明的是,在执行完所述步骤S16后,本地端、参与端和联邦仲裁服务器三方需要不断从所述步骤S13开始循环执行,直到满足预设的停止条件,则表示本地端和参与端已联合建模训练完成。

[0144] 需要说明的是,所述预设的停止条件主要是指本地端和参与端同时出现连续k次的第(t-1)轮和第t轮逻辑回归模型的损失值的差值绝对值小于或等于预设的阈值。

[0145] 示例性的,k一般取值为15~20,所述预设阈值一般为0.005。

[0146] 需要说明的是,本地端保存的原始逻辑回归模型的参数为相对于本轮的上一轮所更新的参数,若不满足预设的停止条件,则继续进行迭代循环,不断更新模型参数,并根据上一轮更新的模型所计算出的第一损失值与本轮更新后的模型计算出的第二损失值进行比较,若两者的差值小于预设阈值,且连续循环n次的本轮与上一轮的损失值差值均小于预设阈值,则停止循环。参与段与本地端的原理相同,在此不作赘述。

[0147] 值得说明的是,本实施例通过设置停止条件来使得本地端和参与端对各自的逻辑回归模型不断进行学习训练,直至得到使模型收敛的最优参数,从而从中获得逻辑回归模型学习客户潜在的交易规律,并利用纵向联邦学习技术保证在客户的隐私数据安全可控的前提下,克服各公司之间客户数据种类单一和不均衡的问题,实现各公司之间的客户数据共享,从而提高识别潜在客户的准确率,精准触达不同业务场景下的潜在客户。

[0148] 作为上述方案的改进,由于当前联合建模训练已经完成,本地端和参与端分别拥有联合模型的一部分,任一方如果要预测潜在客户的需求,将按照以下方式进行,即通过以下步骤完成潜在客户的识别,具体包括:

[0149] 所述本地端将所述客户输入至所述更新后的第一预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的第一概率集P,并将所述第一概率集P加密后发送至所述联邦参数仲裁服务器;其中, $P = [p_1, p_2, \dots, p_m, \dots]$ , $p_m$ 为第m个客户在本地端被预测成潜在客户的概率;

[0150] 所述参与端将所述客户输入至所述更新后的第二预设逻辑回归模型进行预测处理以得到所述客户为潜在客户的第二概率集Q,并将所述第二概率集Q加密后发送至所述联邦参数仲裁服务器;其中, $Q = [q_1, q_2, \dots, q_m, \dots]$ , $q_m$ 为第m个客户在参与端被预测成潜在客户的概率;

[0151] 所述联邦参数仲裁服务器分别对加密后的所述第一概率集[[P]]和加密后的所述第二概率集[[Q]]进行解密,得到解密后的第一概率集P和解密后的第二概率集Q,并根据所

述解密后的第一概率集P和所述解密后的第二概率集Q,按照预设的概率计算方法,得到所述客户为潜在客户的综合概率集F;

[0152] 所述联邦参数仲裁服务器将所述潜在客户的综合概率集F中的每一概率值与预设的潜在客户阈值 $\tau$ 进行逐一比较,得到每一概率比较结果,并根据所述每一概率比较结果,按照预设的概率判断方法,得到所述客户的潜在客户综合标签集,并将所述潜在客户综合标签集分别发送至所述本地端和所述参与端,以完成潜在客户的识别。

[0153] 作为其中一个可选的实施方式,所述根据所述解密后的第一概率集和所述解密后的第二概率集,按照预设的概率计算方法,得到所述客户为潜在客户的综合概率集的计算公式具体为:

$$[0154] \quad F = \frac{\alpha \times \text{Norm}(P) + \beta \times \text{Norm}(Q) + \varepsilon}{2}$$

[0155] 其中, $\alpha$ 为本地端的自有模型权重因子, $\beta$ 为参与端的自有模型权重因子,将在训练中计算得出,Norm()为数值标准化算法,将本地端和参与端得出的客户是否为潜在客户的概率进行按比例缩放,并统一计量单位, $\varepsilon$ 为光滑因子,防止被除数为过小趋近于0。

[0156] 则 $F = [f_1, f_2, \dots, f_m, \dots]$ ,其中 $f_m$ 为第m个客户是否在潜在客户的综合概率。

[0157] 作为其中一个可选的实施方式,所述按照预设的概率判断方法,得到所述客户的潜在客户综合标签集,具体为:

[0158] 如果潜在客户的综合概率集F中的概率值大于预设的潜在客户阈值 $\tau$ ,则赋值为1,反之则赋值为-1(否),从而得到潜在客户综合标签集 $F_{\text{set}}$ 。

[0159] 示例性的,预设的潜在客户阈值 $\tau$ 的取值范围为0.01~0.05,可以是根据对准确性的实际需求进行设定,在此不作限制。

[0160] 值得说明的是,参见图2,是本发明实施例提供的一种基于纵向联邦学习的潜在客户的识别方法的另一流程示意图,本实施例通过纵向联邦学习技术,建立适用于各个银行业金融机构、商业公司之间的识别潜在客户的联合模型,包括但不限于理财推荐、贷款推荐、保险推荐等金融业务中识别潜在(需求)客户的场景,在保证各参与端客户隐私特征数据被合法合规保护的前提下,丰富客户多维特征数据,实现联合建模,提高潜在客户识别模型的预测准确性。

[0161] 参见图3,是本发明实施例提供的一种潜在客户的识别系统的结构示意图,所述系统30包括本地端31和参与端32;其中:

[0162] 所述本地端31,用于根据预先获取的客户的金融交易特征数据,对第一预设逻辑回归模型进行训练,得到第一模型参数;其中,所述客户的金融交易特征数据包括客户标识、客户的金融交易数据集和客户的金融交易数据标签集;

[0163] 所述参与端32,用于根据预先获取的客户的其他维度特征数据以及所述本地端发送过来的金融交易数据标签集和所述客户标识,对第二预设逻辑回归模型进行训练,得到第二模型参数;

[0164] 所述本地端31,还用于根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型;

[0165] 所述参与端32,还用于根据预先获取的第二数据,按照预设的第二模型参数计算

公式,对所述第二模型参数进行更新,得到更新后的第二模型参数和第二预设逻辑回归模型;

[0166] 所述本地端31,还用于将所述金融交易数据集输入至所述更新后的第一预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第一损失值;

[0167] 所述参与端32,还用于将所述其他维度特征数据输入至所述更新后的第二预设逻辑回归模型进行预测处理,以得到所述客户为潜在客户的预测概率集,并根据所述预测概率集、所述金融交易数据标签集和预先获取的所述客户的数量,计算得到第二损失值;

[0168] 所述本地端31和所述参与端32,还用于分别根据对应的所述第一损失值和所述第二损失值判断是否同时满足所有预设的停止条件,若是,则所述本地端31和所述参与端32按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则返回至所述本地端31根据预先获取的第一数据,按照预设的第一模型参数计算公式,对所述第一模型参数进行更新,得到更新后的第一模型参数和第一预设逻辑回归模型的步骤。

[0169] 与现有技术相比,本发明实施例公开的一种基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过本地端得到训练完成后的第一预设逻辑回归模型的第一模型参数,并按照预设的第一模型参数计算公式,更新第一模型参数,然后根据更新后的第一模型参数计算得到更新后的所述第一预设逻辑回归模型的第一损失值,同时,通过参与端得到训练完成后的第二预设逻辑回归模型的第二模型参数,并按照预设的第二模型参数计算公式,更新第二模型参数,然后根据更新后的第二模型参数计算得到更新后的第二预设逻辑回归模型的第二损失值,最后本地端和参与端分别根据对应的第一损失值和第二损失值判断是否同时满足所有预设的停止条件,若是,则本地端和参与端按照预设的潜在客户识别方法对所述客户进行识别,以完成潜在客户的识别;若否,则重复执行指定的步骤并重新进行判断。本发明实施例提供的基于纵向联邦学习的潜在客户的识别方法、系统及介质,通过采用基于纵向联邦学习的联合建模方法对潜在客户进行识别,能够在保证客户的隐私数据安全可控的前提下,克服各公司之间客户数据种类单一和不均衡的问题,实现各公司之间的客户数据共享,从而提高识别潜在客户的准确率,精准触达不同业务场景下的潜在客户。

[0170] 作为其中一个可选的实施方式,所述系统30还包括联邦参数仲裁服务器33;

[0171] 所述联邦参数仲裁服务器33,用于根据预先获取的随机安全数生成公钥-私钥对,并将所述公钥分别发送至所述本地端和所述参与端;

[0172] 所述联邦参数仲裁服务器33,还用于根据所述私钥分别对所述第一组合数据和所述第二组合数据进行解密,并将解密后的第一组合数据发送至所述本地端,将解密后的第二组合数据发送至所述参与端;

[0173] 所述联邦参数仲裁服务器33,还用于分别对加密后的所述第一概率集和加密后的所述第二概率集进行解密,得到解密后的第一概率集和解密后的第二概率集,并根据所述解密后的第一概率集和所述解密后的第二概率集,按照预设的概率计算方法,得到所述客户为潜在客户的综合概率集;

[0174] 所述联邦参数仲裁服务器33,还用于将所述潜在客户的综合概率集中的每一概率值与预设的潜在客户阈值进行逐一比较,得到每一概率比较结果,并根据所述每一概率比较结果,按照预设的概率判断方法,得到所述客户的潜在客户综合标签集,并将所述潜在客

户综合标签集分别发送至所述本地端和所述参与端,以完成潜在客户的识别。

[0175] 本发明实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质包括存储的计算机程序,其中,在所述计算机程序运行时控制所述计算机可读存储介质所在设备执行如上述所述的基于纵向联邦学习的潜在客户的识别方法。

[0176] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

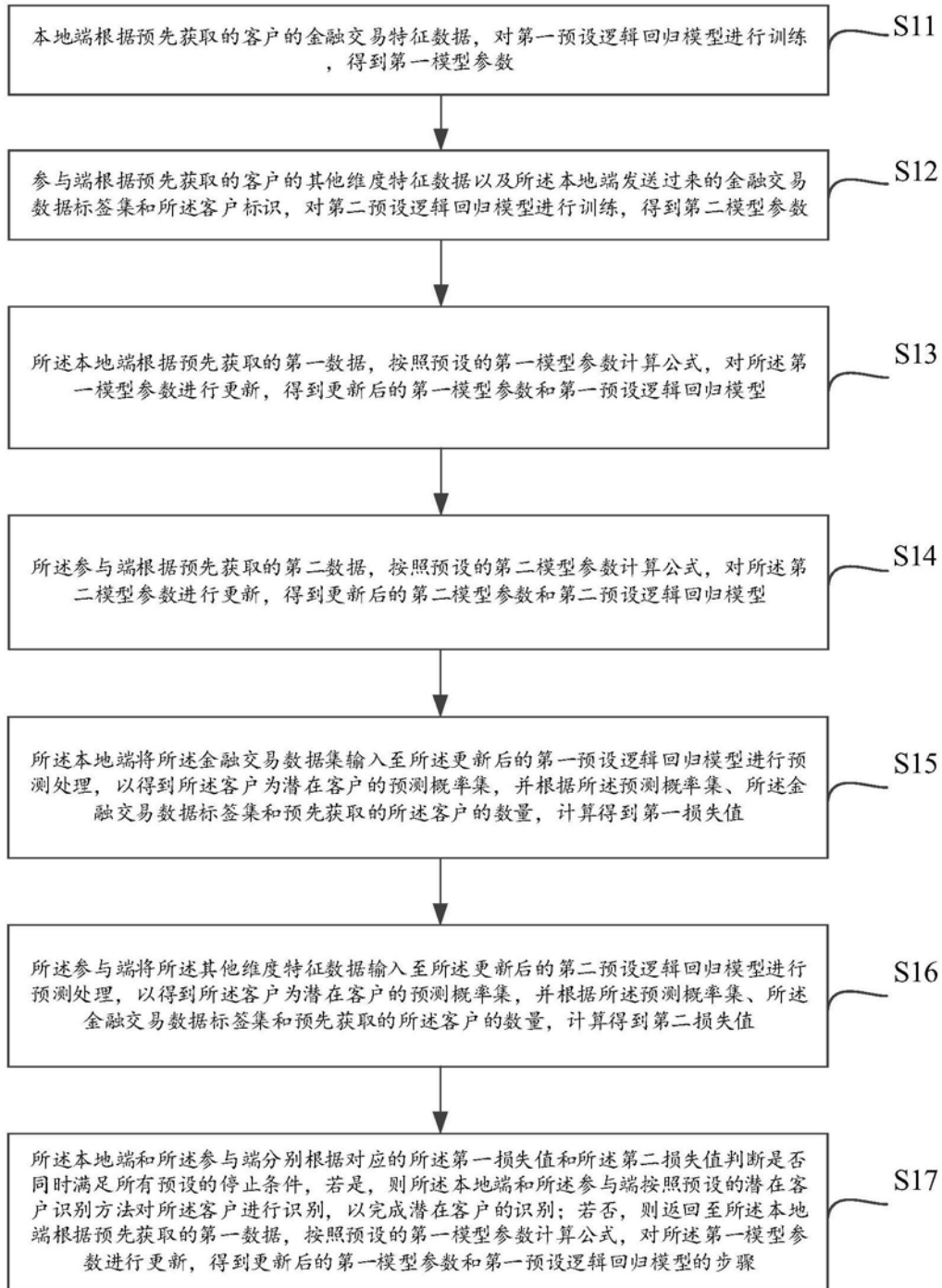


图1

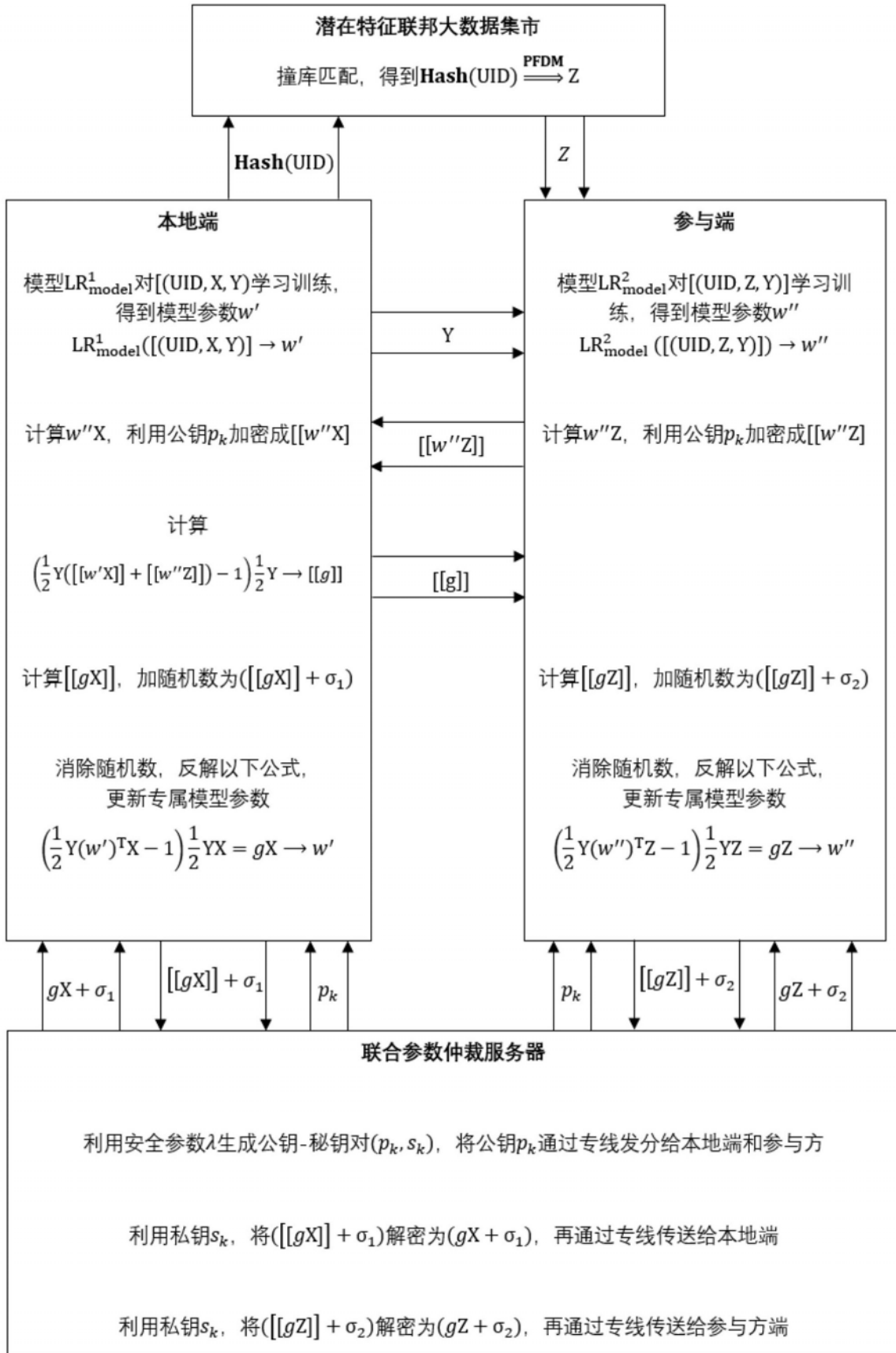


图2

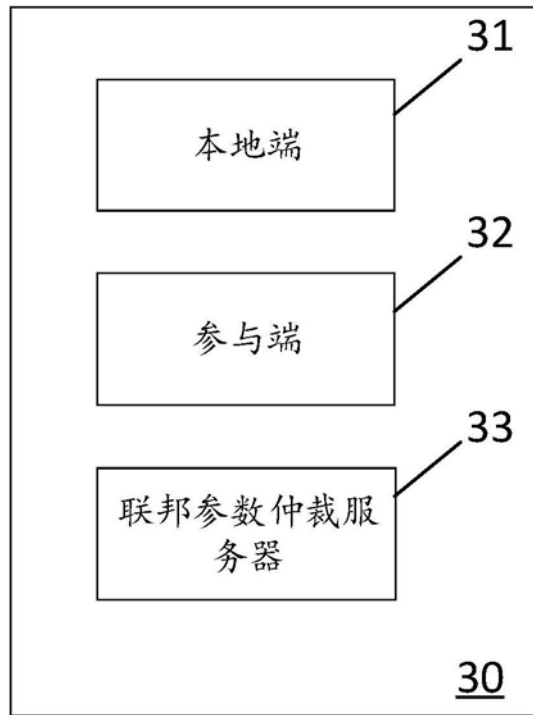


图3