(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 4 月 16 日 (16.04.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/073492 A1

(51) 国际专利分类号:
G06F 21/62 (2013.01) *G06K 9/62* (2006.01)

(21) 国际申请号: PCT/CN2018/122734

(22) 国际申请日: 2018 年 12 月 21 日 (21.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811187262.8 2018年10月12日 (12.10.2018) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 史光辉 (SHI, Guanghui); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 王涵 (WANG, Han); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 王建明 (WANG, Jianming); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳众鼎专利商标代理事务所（普通合伙) (SHENZHEN ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市龙岗

(54) **Title:** DATA SECURITY PROCESSING METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) **发明名称:** 数据安全处理方法、装置、计算机设备及存储介质

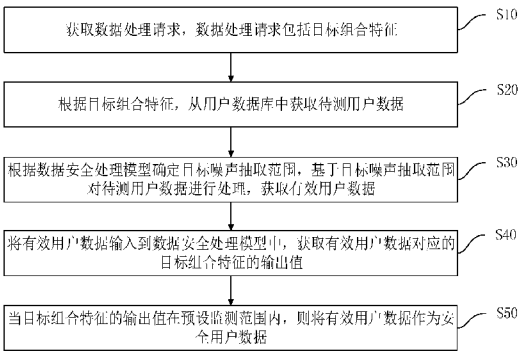

图 2

S10 Acquire a data processing request, the data processing request comprising a target combination feature
S20 Acquire user data to be tested from a user database according to the target combination feature
S30 Determine a target noise extraction range according to a data security processing model, and on the basis of the target noise extraction range, process said user data to obtain valid user data
S40 Input the valid user data into the data security processing model to obtain an output value of the target combination feature corresponding to the valid user data
S50 If the output value of the target combination feature is within a preset monitoring range, regard the valid user data as secure user data

(57) **Abstract:** Disclosed are a data security processing method and apparatus, a computer device and a storage medium. The method comprises: acquiring a data processing request, the data processing request comprising a target combination feature; determining a target noise extraction range according to a data security processing model, and on the basis of the target noise extraction range, processing user data to be tested to obtain valid user data; inputting the valid user data into the data security processing model to obtain an output value of the target combination feature corresponding to the valid user data; if the output value of the target combination feature is

WO 2020/073492 A1

[见续页]

区龙城街道中心城清林路546号城投商务中心4层/B, Guangdong 518172 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护)：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

within a preset monitoring range, regarding the valid user data as secure user data. The present method not only prevents a person using data from easily inferring personal privacy information of a user, but also maximizes the value of target data.

(57) 摘要：本申请公开了一种数据安全处理方法、装置、计算机设备及存储介质，该方法包括获取数据处理请求，数据处理请求包括目标组合特征；根据目标组合特征，从用户数据库中获取待测用户数据；根据数据安全处理模型确定目标噪声抽取范围，基于目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值；当目标组合特征的输出值在预设监测范围内，则将有效用户数据作为安全用户数据。该方法既保证数据的使用者不能轻易推断出用户的个人隐私信息，又能最大化程度地发挥目标数据的价值。

数据安全处理方法、装置、计算机设备及存储介质

本申请以 2018 年 10 月 12 日提交的申请号为 201811187262.8，名称为“数据安全处理方法、装置、计算机设备及存储介质”的中国发明专利申请为基础，并要求其优先权。

技术领域

本申请涉及数据安全技术领域，尤其涉及一种数据安全处理方法、装置、计算机设备及存储介质。

背景技术

近年来，随着信息技术的迅猛发展，大数据的应用越来越广泛，数据共享已成为社会发展的一种趋势。但数据在共享时需要符合特定的条件，保证不能包含个人身份的信息以及可轻易推断出个人隐私信息。如金融保险行业进行数据共享时，根据中国银行业监督管理委员会（简称银监会）和中国保险监督管理委员会（简称中国保监会）等规定，对表示用户身份的信息以及可轻易推断出用户身份的信息需要进行脱敏处理，既保证数据的使用者不能轻易推断出用户的个人隐私信息，又能最大化发挥数据的价值。当前按照规定对用户数据逐一进行脱敏处理，脱敏处理数据量大，不方便操作并且耗时耗力。

发明内容

本申请实施例提供一种数据安全处理方法、装置、计算机设备及存储介质，以解决用户数据脱敏处理数据量大，不方便操作并且耗时耗力的问题。

一种数据安全处理方法，包括：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

一种数据安全处理装置，包括：

数据处理请求获取模块，用于获取数据处理请求，所述数据处理请求包括目标组合特征；

待测用户数据获取模块，用于根据所述目标组合特征，从用户数据库中获取待测用户数据；

有效用户数据获取模块，用于根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

数据安全处理模块，用于将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

安全用户数据获取模块，用于当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器实现如下步骤：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

本申请的一个或多个实施例的细节在下面的附图及描述中提出。本申请的其他特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

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的描述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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请一实施例中数据安全处理方法的一应用场景图；

图 2 是本申请一实施例中数据安全处理方法的一流程图；

图 3 是图 2 中步骤 S30 的一具体流程图；

图 4 是图 2 中步骤 S30 之前的一流程图；

图 5 是图 4 中步骤 S301 的一具体流程图；

图 6 是本申请一实施例中数据安全处理装置的一示意图；

图 7 是本申请一实施例中计算机设备的一示意图。

具体实施方式

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清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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，都属于本申请保护的范围。

本申请提供的数据安全处理方法，可应用在如图 1 的应用环境中，其中，客户端通过网络与服务器进行通信。其中，客户端可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备。服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

在一实施例中，如图 2 所示，提供一种数据安全处理方法，以该方法应用在图 1 中的服务器为例进行说明，包括如下步骤：

S10：获取数据处理请求，数据处理请求包括目标组合特征。

本实施例中的数据处理请求指客户端发送给服务器用于处理用户数据的请求，该数据处理请求包括目标组合特征。该目标组合特征指用户根据需要在客户端填写的关于用户数据的特征，包括但不限于用户数据库中存储的用户的年龄、性别、是否有车、是否购买保险（包括但不限于车险、产险和寿险）等特征。可以理解地，该目标组合特征包括至少一个用户数据的特征。

S20：根据目标组合特征，从用户数据库中获取待测用户数据。

其中，用户数据库指用于存储用户数据的数据库。本实施例中的用户数据库中存储的用户数据包括但不限于用户的基本信息、购买行为和交易数据。其中，基本信息包括但不限于姓名、年龄（出生年月）、性别、籍贯、民族、教育背景和工作经历等。购买行为指用户在平台购买产品的行为，包括但不限于用户购买的理财产品和保险产品。交易数据指用户的投资、消费和转账的记录数据。

具体地，当用户在客户端设置好目标组合特征后，发送给服务器，服务器根据获取到的目标组合特征，从用户数据库中获取与目标组合特征对应的用户数据作为待测用户数据。其中，待测用户数据指需要进行测试的用户数据。

S30：根据数据安全处理模型确定目标噪声抽取范围，基于目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据。

为了保护用户数据的安全性，本实施例对待测用户数据添加非目标组合特征对应的用户数据，使得获取的有效用户数据不全为目标组合特征对应的用户数据，降低基本信息用户数据的准确性来保证用户数据的安全性。

其中，数据安全处理模型指预先训练好的用于确定目标噪声抽取范围的模型。目标噪

声抽取范围指用于抽取非目标组合特征对应的用户数据的范围。本实施例中的数据安全处理模型包括目标梯度提升决策树模型和目标逻辑回归模型。目标梯度提升决策树模型指预先训练好的满足用户要求的梯度提升决策树，梯度提升决策树（Gradient Boosting Decision Tree，GBDT）是一种迭代的决策树算法，该算法由多棵决策树组成，所有树的结论累加起来作为最终预测结果。目标逻辑回归模型指预先训练好的用于获取目标噪声抽取范围的逻辑回归模型。逻辑回归模型（Logistic Regression，LR）是一种减小预测范围，将预测值限定为[0,1]间的一种回归模型。

在获取待测用户数据后，根据安全处理模型确定目标噪声抽取范围，从用户数据库中抽取特定数量的非目标组合特征对应的用户数据，加入待测用户数据中，生成有效用户数据。有效用户数据指在待测用户数据中添加了非目标组合特征对应的用户数据后得到的数据。需要说明的是，本实施例中的特定数量指在目标噪声抽取范围内的数量。服务器根据目标噪声抽取范围抽取非目标组合特征对应的用户数据加入到待测用户数据中，使得后续步骤获取的安全用户数据既能保证用户数据的安全性，又满足对用户数据的价值的需求。

S40：将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值。

具体地，在获取有效用户数据后，还需要对有效用户数据进行进一步地验证处理，确定有效用户数据是否满足用户要求。因此，还需将获取的有效用户数据输入到预先训练好的数据安全处理模型中进行处理，获取有效用户数据在经过数据安全处理模型处理得到的输出值。该输出值是一个概率值，用于检测有效用户数据是否满足用户设定的要求。

S50：当目标组合特征的输出值在预设监测范围内，则将有效用户数据作为安全用户数据。

具体地，预设监测范围指预先设置的用于检测有效用户数据是否满足用户要求的范围。本实施中设置的预设监测范围为 70%-90%，既能保证用户数据的安全性，又满足对用户数据的价值需求。当通过数据安全处理模型的计算，获取的目标组合特征的输出值在预设监测范围内，则表示有效用户数据满足要求，将有效用户数据作为安全用户数据。

步骤 S10-步骤 S50，通过目标组合特征从用户数据库中获取待测用户数据，然后根据数据安全处理模型确定的目标噪声抽取范围，抽取特定数量的非目标组合特征对应的用户数据加入到待测用户数据中，以获取有效用户数据，降低有效用户数据的准确性来保证用户数据的安全性。在获取有效用户数据后，为了进一步确定有效用户数据是否满足要求，还需要将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值，当输出值在预设监测范围，则表示有效用户数据满足要求，将有效用户数据作为安全用户数据发送给需要的发送数据处理请求的客户端，既保证数据的使用者不能轻易推断出用户的个人隐私信息，又能最大化程度地发挥目标数据的价值。

在一实施例中，如图 3 所示，步骤 S30 中，基于目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据，具体包括如下步骤：

S31：基于目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据。

具体地，在数据安全处理模型确定目标噪声抽取范围后，服务器根据目标噪声抽取范围，从用户数据库选取特定数量的非目标组合特征对应的用户数据作为目标噪声数据，为后续步骤提供数据来源。

如根据数据安全处理模型确定目标噪声抽取范围为0.05-0.25，非目标组合特征对应的用户数据有 10 万个，服务器根据目标噪声抽取范围 5000-25000，从用户数据库选取5000-25000 中任意一个数字的非目标组合特征对应的用户数据作为目标噪声数据。

S32：将目标噪声数据加入到待测用户数据中，获取有效用户数据。

具体地，在获取目标噪声数据后，将目标噪声数据加入到待测用户数据中，获取有效用户数据，使得有效用户数据不全为目标组合特征对应的用户数据，降低有效用户数据的准确性，使得有效用户数据既可以满足用户对用户数据的需求，又能保证用户数据的安全性。

步骤 S31-步骤 S32，通过目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据，并将目标噪声数据加入到待测用户数据中，获取有效用户数据来保证用户数据的安全性。

在一实施例中，如图 4 所示，在步骤 S30 之前，即根据数据安全处理模型确定目标噪声抽取范围的步骤之前，数据安全处理方法还包括如下步骤：

S301：获取待训练数据，将待训练数据划分为训练集和测试集。

具体地，从样本数据库中获取待训练数据，并将待训练数据划分为训练集和测试集用于训练模型和测试模型。本实施例中需要训练的模型指原始梯度提升决策树模型。其中，样本数据库指用于存储待训练数据的数据库。待训练数据指用于训练原始梯度提升决策树模型的数据。训练集（training set)是用于训练原始梯度提升决策树模型的待训练数据的集合。测试集（test set)是用于测试训练好的原始梯度提升决策树模型是否准确的待训练数据的集合。

S302：初始化原始梯度提升决策树模型的模型参数，模型参数包括梯度提升决策树的最大深度和最大迭代次数。

具体地，在对原始梯度提升决策树模型进行训练之前，首先需要对原始梯度提升决策树模型中的模型参数进行初始化设置。本实施例中的模型参数包括原始梯度提升决策树的最大深度和最大迭代次数。根据实验数据表示，原始梯度提升决策树的最大深度设置为 3，最大迭代次数设置为 50，效果最好，本实施例中初始化设置其最大深度为 3，最大迭代次数为 50。对原始梯度提升决策树模型的模型参数进行初始化设置可以在后续训练原始梯度提升决策树模型时，缩短训练时间，提高识别准确率。

S303：将训练集对应的待训练数据输入到原始梯度提升决策树模型中，当原始梯度提升决策树模型中的训练深度达到最大深度且迭代次数达到最大迭代次数，则停止训练原始

梯度提升决策树模型，获取原始梯度提升决策树模型中每个决策树路径对应的原始组合特征。

具体地，在对原始梯度提升决策树模型进行初始化设置后，将训练集中的待训练数据输入到原始梯度提升决策树模型中，原始梯度提升决策树模型会在待训练数据对应的目标组合特征中选择一个特征作为第一个分叉点，获取训练集中待训练数据在该分叉点的残差，然后将目标组合特征中剩余的特征进行再次分叉，将第一个分叉点对应的残差作为第二个决策树输入，不断迭代，当原始梯度提升决策树模型中的训练深度达到最大深度且迭代次数达到最大迭代次数，则停止训练原始梯度提升决策树模型，获取原始梯度提升决策树模型中每个决策树路径对应的原始组合特征。其中，原始组合特征指训练集对应的待训练数据输入到原始梯度提升决策树模型中，每个决策树路径对应的组合特征。

S304：将原始组合特征输入到原始逻辑回归模型中，获取原始组合特征对应的输出值。

具体地，在获取原始组合特征之后，将原始组合特征输入到原始逻辑回归模型中，通过 Sigmoid 函数计算原始组合特征对应的概率，即原始组合特征对应的输出值。本实施例中的 Sigmoid 函数可表示为：$g_{(z)}=\left\{\begin{matrix}\frac{1}{1+e^{-\theta^{T}x}},q=1\\\frac{e^{-\theta T_{z}}}{1+e^{-\theta^{T}x}},q=0\end{matrix}\right\}$其中，$g_{(x)}$为原始组合特征发生的概率，$g_{(x)}\in(0,1)$，$x$为原始组合特征，$T$为用户根据实际设定的$\theta$变化快慢的参数，$\theta$为原始梯度提升决策树模型中每个用户数据的特征对应的权重的组合。$q=1$表示原始组合特征为目标组合特征成立，$q=0$表示原始组合特征为目标组合特征不成立，即原始组合特征为非目标组合特征。

S305：当原始组合特征对应的输出值在预设监测范围内，则将原始梯度提升决策树模型和原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型。

具体地，当原始组合特征对应的输出值在预设监测范围内，则表示原始梯度提升决策树模型和原始逻辑回归模型训练成功，将原始梯度提升决策树模型和原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型。

S306：采用测试集对应的待训练数据对目标梯度提升决策树模型和目标逻辑回归模型进行测试，若获取到的每一待训练数据对应的输出值均在预设监测范围内，则将目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

在确定原始梯度提升决策树模型和原始逻辑回归模型为目标梯度提升决策树模型和目标逻辑回归模型之后，为了防止过拟合，需要采用测试集中的待训练数据对目标梯度提升决策树模型和目标逻辑回归模型进行测试，验证目标梯度提升决策树模型和目标逻辑回归模型的准确性，若获取到的测试集中每一待训练数据对应的输出值均在预设监测范围

内，则标识将目标梯度提升决策树模型和目标逻辑回归模型训练的比较成功，满足要求，则将目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

步骤 S301-步骤 S306，通过训练集中的待训练数据训练原始梯度提升决策树模型，并将原始梯度提升决策树模型中每个决策树路径对应的原始组合特征输入到原始逻辑回归模型中，获取原始组合特征对应的输出值。若原始组合特征对应的输出值在预设监测范围内，则表示原始梯度提升决策树模型和原始逻辑回归模型训练成功。为了防止出现过拟合的情况，还需要将测试集对应的待训练数据输入到训练好的目标梯度提升决策树模型和目标逻辑回归模型中进行测试，若每一待训练数据对应的输出值均在预设监测范围内，则表示训练后的目标梯度提升决策树模型和目标逻辑回归模型符合需求，可以作为最终确定目标噪声抽取范围的数据安全处理模型，为后续步骤获取有效用户数据使用。

在一实施例中，如图 5 所示，步骤 S301，获取待训练数据，将待训练数据划分为训练集和测试集，具体包括如下步骤：

S3011：获取模型训练请求，模型训练请求包括训练组合特征。

其中，模型训练请求指客户端发送的用于进行模型训练的请求。训练组合特征指用户在客户端根据需要设置的用于训练模型的用户数据的特征。具体地，用户按照要求在客户端设置训练组合特征，然后点击发送的操作，服务器会获取到携带有训练组合特征的模型训练请求。

S3012：根据训练组合特征，从用户数据库中选取与训练组合特征匹配的训练用户数据和与训练组合特征不匹配的非训练用户数据。

具体地，在获取到模型训练请求后，服务器根据模型训练请求中包括的训练组合特征，从用户数据库中选取与训练组合特征匹配的训练用户数据和与训练组合特征不匹配的非训练用户数据。其中，训练用户数据指与训练组合特征匹配的用户数据。非训练用户数据指与训练组合特征不匹配的用户数据。如训练合特征为年龄 20-30 岁和购买产险，则根据训练特征获取年龄 20-30 岁和购买产险的用户数据作为训练用户数据，选取不同时包括年龄在 20-30 岁和购买产险的用户数据作为非训练用户数据。根据训练组合特征获取训练用户数据和非训练用户数据为后续获取原始正样本和原始负样本提供数据来源。

进一步地，为方便后续步骤对用户数据进行处理，在获取训练用户数据后，将训练组合特征作为训练用户数据的用户标签，非目标组合特征作为非训练用户数据的用户标签。如将训练合特征为年龄 20-30 岁和购买产险作为训练用户数据的用户标签，将年龄不在 20-30 岁，没有购买产险作为非训练用户数据的标签。

S3013：按照预设的正样本数量，从训练用户数据中选取对应的训练用户数据作为原始正样本。

本实施例中正样本数量指预先设置的用于获取原始正样本的数量。原始正样本指从训练用户数据中根据预设的正样本数量抽取的训练用户数据。按照预设的正样本数量，从训练用户数据中获取原始正样本，为后续步骤获取有效正样本和有效负样本提供数据来源。

S3014：按照正负样本比例，从非训练用户数据中选取对应的非训练用户数据作为原始负样本。

具体地，在获取原始正样本后，服务器根据预先设置的正负样本比例，从非训练用户数据中选取对应的非训练用户数据作为原始负样本。其中，正负样本比例指预先设置的用于根据正样本数量确定原始负样本的数量的比列。本实施例中的原始负样本的数量是根据正样本数量决定的，使得原始负样本的数量和正样本数量成一定比例，满足后续进行模型训练的要求。

S3015：根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将负噪声数据加入到原始负样本中，生成有效负样本。

其中，第一抽取范围指预先设置好的将原始正样本作为负噪声数据的比例范围。负噪声数据指将原始正样本中部分用户数据对应的用户标签进行人为进行修改，使得该部分用户数据携带的用户标签变为非目标用户特征对应的用户标签。如原始正样本中的用户数据携带的用户标签为年龄 20-30 岁和购买产险，人为将部分用户标签修改为非目标组合特征对应的用户标签（如将用户标签修改为 35-45 岁和购买产险，或者将用户标签修改为 20-30 岁和购买车险，或者修改为 30-40 岁和有车等任意非目标组合特征对应的用户标签）。

具体地，在获取原始正样本和原始负样本之后，服务器根据第一噪声抽取范围，从原始正样本中随机抽取特定数量的用户数据，人为将该部分用户数据对应的用户标签修改为非目标组合特征对应的用户标签，将该部分修改了用户标签的用户数据作为负噪声数据。可以理解地，该特定数量指第一噪声抽取范围内的任意数字。在获取负噪声数据后，将负噪声数据加入到原始负样本中，生成有效负样本，降低原始负样本的准确性，使得原始负样本中的用户数据不全为非目标组合特征对应的用户数据，达到降低原始负样本中用户数据的敏感性的目的。

例如，原始正样本的数量为 10 万个，第一抽取范围设置为 0.1-0.3，根据第一抽取范围，从 10 万个原始正样本中随机抽取 1 万-3 万的用户数据，将该部分用户数据携带的用户标签修改为非目标组合特征对应的用户标签，将该部分用户数据作为负噪声数据。

S3016：根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将正噪声数据加入到原始正样本中，生成有效正样本。

其中，第二抽取范围指预先设置好的将原始负样本作为正噪声数据的比例范围。正噪声数据指将原始负样本中部分用户数据对应的用户标签变为目标组合特征对应的用户标签。需要说明的是，由于正样本的数量与负样本的数量是按照预先设置好的正负样本比例选取的，因此，第一抽取范围和第二抽取范围也是不一致的，如将第一抽取范围设置为 0.1-0.3，则第二抽取范围设置为 0.05-0.2。

具体地，根据第二抽取范围，从原始负样本中随机抽取特定数量的用户数据，人为修改该部分用户数据对应的用户标签，将用户标签修改为目标组合特征对应的用户标签，将该部分修改了用户标签的原始负样本作为正噪声数据。可以理解地，该特定数量指第二噪

声抽取范围内的任意数字。获取正噪声数据后，将正噪声数据添加到原始正样本中，形成有效正样本。可以理解地，有效正样本指在原始正样本中加入负噪声数据后形成的正样本。

例如，原始负样本的数量为 100 万个，第二抽取范围设置为 0.05-0.2，根据第二抽取范围，从 100 万个原始负样本中随机抽取 5 万-20 万范围内任何一个数字对应的用户数据，然后将抽取的用户数据携带的用户标签人为修改为目标组合特征对应的用户标签。修改用户标签后，将修改用户标签的原始负样本作为正噪声数据，加入到原始正样本中，生成有效正样本。

在原始正样本中加入正噪声数据，使得原始正样本不只是仅包括真正的目标组合特征对应的用户数据，还包括一些非目标组合特征的正噪声数据，使得原始正样本中的用户数据不是 100%真实的训练用户数据，可以避免输出真实的用户数据，也不用对用户数据中的一些敏感数据进行逐一脱敏处理，保证了用户数据的安全性。

S3017：将有效正样本和有效负样本作为待训练数据，并存储在样本数据库中。

具体地，在获取有效正样本和有效负样本后，为了方便后续训练原始梯度提升决策树模型和验证训练好的目标梯度提升决策树模型是否准确，将有效正样本和有效负样本作为待训练数据存储在样本数据库中。

步骤 S3011-步骤 S3017，通过对原始正样本加入正噪声数据获取有效正样本，对原始负样本加入负噪声数据获取有效负样本，强行改变原始正样本和原始负样本的准确性，使得生成的有效正样本和有效负样本中的用户数据不全为含有目标组合特征的用户数据或者非目标组合特征的用户数据，避免输出真实的用户数据，也不用对用户数据中的一些敏感数据进行逐一脱敏处理，保证了用户数据的安全性。

在一实施例中，在步骤 S304，将原始组合特征输入到原始逻辑回归模型中，获取原始组合特征对应的输出值的步骤之后，该数据安全处理方法还包括：当原始组合特征对应的输出值低于预设监测范围，则减小第一噪声抽取范围和第二噪声抽取范围，并提高正负样本比例；当原始组合特征对应的输出值高于预设监测范围，则增大第一噪声抽取范围和第二噪声抽取范围，并降低正负样本比例。

具体地，当训练集对应的待训练数据经过原始梯度提升决策树模型和原始逻辑回归模型的训练与计算，得到的原始组合特征对应的输出值不在预设监测范围内，则表示第一噪声抽取范围、第二噪声抽取范围和正负样本比例设置的不准确，需要动态调整第一噪声抽取范围、第二噪声抽取范围和正负样本比例。

进一步地，当原始组合特征对应的输出值低于预设监测范围，则表示负样本数量过多，应该提高正负样本比例，减少负样本数量。此时，第一噪声抽取范围和第二噪声抽取范围设置的偏高，减小第一噪声抽取范围和第二噪声抽取范围的设置。当原始组合特征对应的输出值高于预设监测范围，则表示负样本数量过少，应该降低正负样本比例，增加负样本数量。此时，第一噪声抽取范围和第二噪声抽取范围设置的偏低，应该提高第一噪声抽取范围和第二噪声抽取范围的设置。

本申请提供的数据安全处理方法，通过目标组合特征从用户数据库中获取待测用户数据，然后根据数据安全处理模型确定的目标噪声抽取范围，抽取特定数量的非目标组合特征对应的用户数据加入到待测用户数据中，降低待测用户数据的准确性来保证用户数据的安全性。该目标噪声抽取范围是根据训练好的数据安全处理模型确定的，使得根据目标噪声抽取范围获取的有效用户数据更加准确，满足用户要求，既保证用户数据的安全又能使用户数据发挥数据的价值。获取有效用户数据后，为了进一步确定有效用户数据是否满足要求，还需要将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值，当输出值在预设监测范围，则表示有效用户数据满足要求，将有效用户数据作为安全用户数据发送给需要的用户。当原始组合特征对应的输出值低于预设监测范围，则减小第一噪声抽取范围和第二噪声抽取范围，并提高正负样本比例；当原始组合特征对应的输出值高于预设监测范围，则增大第一噪声抽取范围和第二噪声抽取范围，并降低正负样本比例，使得获取安全用户数据更加接近用户的要求。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

在一实施例中，提供一种数据安全处理装置，该数据安全处理装置与上述实施例中数据安全处理方法一一对应。如图 6 所示，该数据安全处理装置包括数据处理请求获取模块 10、待测用户数据获取模块 20、有效用户数据获取模块 30、数据安全处理模块 40 和安全用户数据获取模块 50。各功能模块详细说明如下：

数据处理请求获取模块 10，用于获取数据处理请求，数据处理请求包括目标组合特征。

待测用户数据获取模块 20，用于根据目标组合特征，从用户数据库中获取待测用户数据。

有效用户数据获取模块 30，用于根据数据安全处理模型确定目标噪声抽取范围，基于目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据。

数据安全处理模块 40，用于将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值。

安全用户数据获取模块 50，用于当目标组合特征的输出值在预设监测范围内，则将有效用户数据作为安全用户数据。

进一步地，有效用户数据获取模块 30 包括目标噪声数据获取单元和有效用户数据获取单元。

目标噪声数据获取单元，用于基于目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据。

有效用户数据获取单元，用于将目标噪声数据加入到待测用户数据中，获取有效用户数据。

进一步地，在有效用户数据获取模块之前，数据安全处理装置还包括：

待训练数据处理单元，用于获取待训练数据，将待训练数据划分为训练集和测试集。

模型参数初始化单元，用于初始化原始梯度提升决策树模型的模型参数，模型参数包括梯度提升决策树的最大深度和最大迭代次数。

模型训练单元，用于将训练集对应的待训练数据输入到原始梯度提升决策树模型中，当原始梯度提升决策树模型中的训练深度达到最大深度且迭代次数达到最大迭代次数，则停止训练原始梯度提升决策树模型，获取原始梯度提升决策树模型中每个决策树路径对应的原始组合特征。

输出值计算单元，用于将原始组合特征输入到原始逻辑回归模型中，获取原始组合特征对应的输出值。

模型确定单元，用于当原始组合特征对应的输出值在预设监测范围内，则将原始梯度提升决策树模型和原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型。

模型测试单元，用于采用测试集对应的待训练数据对目标梯度提升决策树模型和目标逻辑回归模型进行测试，若获取到的每一待训练数据对应的输出值均在预设监测范围内，则将目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

进一步地，待训练数据处理单元包括模型训练请求获取子单元、用户数据选取子单元、原始正样本获取子单元、原始负样本获取子单元、有效负样本获取子单元、有效正样本获取子单元和待训练数据生成子单元。

模型训练请求获取子单元，用于获取模型训练请求，模型训练请求包括训练组合特征。

用户数据选取子单元，用于根据训练组合特征，从用户数据库中选取与训练组合特征匹配的训练用户数据和与训练组合特征不匹配的非训练用户数据。

原始正样本获取子单元，用于按照预设的正样本数量，从训练用户数据中选取对应的训练用户数据作为原始正样本。

原始负样本获取子单元，用于按照正负样本比例，从非训练用户数据中选取对应的非训练用户数据作为原始负样本。

有效负样本获取子单元，用于根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将负噪声数据加入到原始负样本中，生成有效负样本。

有效正样本获取子单元，用于根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将正噪声数据加入到原始正样本中，生成有效正样本。

待训练数据生成子单元，用于将有效正样本和有效负样本作为待训练数据存储在样本数据库中。

进一步地，数据安全处理方法还用于当原始组合特征对应的输出值低于预设监测范围，则减小第一噪声抽取范围和第二噪声抽取范围，并提高正负样本比例；当原始组合特征对应的输出值高于预设监测范围，则增大第一噪声抽取范围和第二噪声抽取范围，并降低正负样本比例。

关于数据安全处理装置的具体限定可以参见上文中对于数据安全处理方法的限定，在

此不再赘述。上述数据安全处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 7 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储数据安全处理方法涉及的数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种数据安全处理方法。

在一个实施例中，提供了一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，处理器执行计算机可读指令时实现上述数据安全处理方法的步骤，如图 2 所示的步骤 S10-步骤 S50。或者图 3 至图 5 所示的步骤。为避免重复，这里不再赘述。或者，处理器执行计算机可读指令时实现上述数据安全处理装置的步骤，如图 6 所示的数据处理请求获取模块 10、待测用户数据获取模块 20、有效用户数据获取模块 30、数据安全处理模块 40 和安全用户数据获取模块 50。为避免重复，这里不再赘述。

在一个实施例中，提供了一个或多个存储有计算机可读指令的非易失性可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现上述实施例中数据安全处理方法的步骤，如图 2 所示的步骤 S10-步骤 S50。或者图 3 至图 5 所示的步骤。为避免重复，这里不再赘述。或者，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现上述实施例中数据安全处理装置的步骤，如图 6 所示的数据处理请求获取模块 10、待测用户数据获取模块 20、有效用户数据获取模块 30、数据安全处理模块 40 和安全用户数据获取模块 50。为避免重复，这里不再赘述。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink） DRAM

（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种数据安全处理方法，其特征在于，包括：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

2. 如权利要求 1 所述的数据安全处理方法，其特征在于，所述基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据，包括：

基于所述目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据；

将所述目标噪声数据加入到待测用户数据中，获取有效用户数据。

3. 如权利要求 1 所述的数据安全处理方法，其特征在于，在所述根据数据安全处理模型确定目标噪声抽取范围的步骤之前，所述数据安全处理方法还包括：

获取待训练数据，将所述待训练数据划分为训练集和测试集；

初始化原始梯度提升决策树模型的模型参数，所述模型参数包括梯度提升决策树的最大深度和最大迭代次数；

将训练集对应的待训练数据输入到所述原始梯度提升决策树模型中，当所述原始梯度提升决策树模型中的训练深度达到所述最大深度且迭代次数达到最大迭代次数，则停止训练所述原始梯度提升决策树模型，获取所述原始梯度提升决策树模型中每个决策树路径对应的原始组合特征；

将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值；

当所述原始组合特征对应的输出值在所述预设监测范围内，则将所述原始梯度提升决策树模型和所述原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型；

采用测试集对应的待训练数据对所述目标梯度提升决策树模型和所述目标逻辑回归模型进行测试，若获取到的每一所述待训练数据对应的输出值均在所述预设监测范围内，则将所述目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

4. 如权利要求 3 所述的数据安全处理方法，其特征在于，所述获取待训练数据，将所述待训练数据划分为训练集和测试集，包括：

获取模型训练请求，所述模型训练请求包括训练组合特征；

根据所述训练组合特征，从用户数据库中选取与所述训练组合特征匹配的训练用户数据和与所述训练组合特征不匹配的非训练用户数据；

按照预设的正样本数量，从所述训练用户数据中选取对应的训练用户数据作为原始正样本；

按照正负样本比例，从所述非训练用户数据中选取对应的非训练用户数据作为原始负样本；

根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将所述负噪声数据加入到原始负样本中，生成有效负样本；

根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将所述正噪声数据加入到原始正样本中，生成有效正样本；

将所述有效正样本和所述有效负样本作为待训练数据存储在样本数据库中。

5. 如权利要求 4 所述的数据安全处理方法，其特征在于，在所述将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值的步骤之后，所述数据安全处理方法还包括：

当所述原始组合特征对应的输出值低于所述预设监测范围，则减小所述第一噪声抽取范围和所述第二噪声抽取范围，并提高所述正负样本比例；当所述原始组合特征对应的输出值高于所述预设监测范围，则增大所述第一噪声抽取范围和所述第二噪声抽取范围，并降低所述正负样本比例。

6. 一种数据安全处理装置，其特征在于，包括：

数据处理请求获取模块，用于获取数据处理请求，所述数据处理请求包括目标组合特征；

待测用户数据获取模块，用于根据所述目标组合特征，从用户数据库中获取待测用户数据；

有效用户数据获取模块，用于根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

数据安全处理模块，用于将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

安全用户数据获取模块，用于当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

7. 如权利要求 6 所述的数据安全处理装置，其特征在于，在有效用户数据获取模块之前，所述数据安全处理装置还包括：

待训练数据处理单元，用于获取待训练数据，将所述待训练数据划分为训练集和测试集；

模型参数初始化单元，用于初始化原始梯度提升决策树模型的模型参数，所述模型参数包括梯度提升决策树的最大深度和最大迭代次数；

模型训练单元，用于将训练集对应的待训练数据输入到所述原始梯度提升决策树模型中，当所述原始梯度提升决策树模型中的训练深度达到所述最大深度且迭代次数达到最大迭代次数，则停止训练所述原始梯度提升决策树模型，获取所述所述原始梯度提升决策树模型中每个决策树路径对应的原始组合特征；

输出值计算单元，用于将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值；

模型确定单元，用于当所述原始组合特征对应的输出值在所述预设监测范围内，则将所述原始梯度提升决策树模型和所述原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型；

模型测试单元，用于采用测试集对应的待训练数据对所述目标梯度提升决策树模型和所述目标逻辑回归模型进行测试，若获取到的每一所述待训练数据对应的输出值均在所述预设监测范围内，则将所述目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

8. 如权利要求 7 所述的数据安全处理装置，其特征在于，所述待训练数据处理单元，包括：

模型训练请求获取子单元，用于获取模型训练请求，所述模型训练请求包括训练组合特征；

用户数据选取子单元，用于根据所述训练组合特征，从用户数据库中选取与所述训练组合特征匹配的安全用户数据和与所述训练组合特征不匹配的非安全用户数据；

原始正样本获取子单元，用于按照预设的正样本数量，从所述安全用户数据中选取对应的安全用户数据作为原始正样本；

原始负样本获取子单元，用于按照正负样本比例，从所述非安全用户数据中选取对应的非安全用户数据作为原始负样本；

有效负样本获取子单元，用于根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将所述负噪声数据加入到原始负样本中，生成有效负样本；

有效正样本获取子单元，用于根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将所述正噪声数据加入到原始正样本中，生成有效正样本；

待训练数据生成子单元，用于将所述有效正样本和所述有效负样本作为待训练数据存储在样本数据库中。

9. 如权利要求 6 所述的数据安全处理装置，其特征在于，所述有效用户数据获取模块，包括：

目标噪声数据获取单元，用于基于所述目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据；

有效用户数据获取单元，用于将所述目标噪声数据加入到待测用户数据中，获取有效用户数据。

10. 如权利要求 8 所述的数据安全处理装置，其特征在于，所述数据安全处理装置还包括：

当所述原始组合特征对应的输出值低于所述预设监测范围，则减小所述第一噪声抽取范围和所述第二噪声抽取范围，并提高所述正负样本比例；当所述原始组合特征对应的输出值高于所述预设监测范围，则增大所述第一噪声抽取范围和所述第二噪声抽取范围，并降低所述正负样本比例。

11. 一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

12. 如权利要求 11 所述的计算机设备，其特征在于，所述基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据，包括：

基于所述目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据；

将所述目标噪声数据加入到待测用户数据中，获取有效用户数据。

13. 如权利要求 11 所述的计算机设备，其特征在于，在所述根据数据安全处理模型确定目标噪声抽取范围的步骤之前，所述处理器执行所述计算机可读指令时还实现如下步骤：

获取待训练数据，将所述待训练数据划分为训练集和测试集；

初始化原始梯度提升决策树模型的模型参数，所述模型参数包括梯度提升决策树的最大深度和最大迭代次数；

将训练集对应的待训练数据输入到所述原始梯度提升决策树模型中，当所述原始梯度提升决策树模型中的训练深度达到所述最大深度且迭代次数达到最大迭代次数，则停止训练所述原始梯度提升决策树模型，获取所述原始梯度提升决策树模型中每个决策树路径对应的原始组合特征；

将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值；

当所述原始组合特征对应的输出值在所述预设监测范围内，则将所述原始梯度提升决

策树模型和所述原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型；

采用测试集对应的待训练数据对所述目标梯度提升决策树模型和所述目标逻辑回归模型进行测试，若获取到的每一所述待训练数据对应的输出值均在所述预设监测范围内，则将所述目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

14. 如权利要求 13 所述的计算机设备，其特征在于，所述获取待训练数据，将所述待训练数据划分为训练集和测试集，包括：

获取模型训练请求，所述模型训练请求包括训练组合特征；

根据所述训练组合特征，从用户数据库中选取与所述训练组合特征匹配的训练用户数据和与所述训练组合特征不匹配的非训练用户数据；

按照预设的正样本数量，从所述训练用户数据中选取对应的训练用户数据作为原始正样本；

按照正负样本比例，从所述非训练用户数据中选取对应的非训练用户数据作为原始负样本；

根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将所述负噪声数据加入到原始负样本中，生成有效负样本；

根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将所述正噪声数据加入到原始正样本中，生成有效正样本；

将所述有效正样本和所述有效负样本作为待训练数据存储在样本数据库中。

15. 如权利要求 14 所述的计算机设备，其特征在于，在所述将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值的步骤之后，所述处理器执行所述计算机可读指令时还实现如下步骤：

当所述原始组合特征对应的输出值低于所述预设监测范围，则减小所述第一噪声抽取范围和所述第二噪声抽取范围，并提高所述正负样本比例；当所述原始组合特征对应的输出值高于所述预设监测范围，则增大所述第一噪声抽取范围和所述第二噪声抽取范围，并降低所述正负样本比例。

16. 一个或多个存储有计算机可读指令的非易失性可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器实现如下步骤：

获取数据处理请求，所述数据处理请求包括目标组合特征；

根据所述目标组合特征，从用户数据库中获取待测用户数据；

根据数据安全处理模型确定目标噪声抽取范围，基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据；

将所述有效用户数据输入到所述数据安全处理模型中，获取所述有效用户数据对应的目标组合特征的输出值；

当所述目标组合特征的输出值在预设监测范围内，则将所述有效用户数据作为安全用户数据。

17. 如权利要求 16 所述的非易失性可读存储介质，其特征在于，所述基于所述目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据，包括：

基于所述目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据；

将所述目标噪声数据加入到待测用户数据中，获取有效用户数据。

18. 如权利要求 16 所述的非易失性可读存储介质，其特征在于，在所述根据数据安全处理模型确定目标噪声抽取范围的步骤之前，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

获取待训练数据，将所述待训练数据划分为训练集和测试集；

初始化原始梯度提升决策树模型的模型参数，所述模型参数包括梯度提升决策树的最大深度和最大迭代次数；

将训练集对应的待训练数据输入到所述原始梯度提升决策树模型中，当所述原始梯度提升决策树模型中的训练深度达到所述最大深度且迭代次数达到最大迭代次数，则停止训练所述原始梯度提升决策树模型，获取所述原始梯度提升决策树模型中每个决策树路径对应的原始组合特征；

将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值；

当所述原始组合特征对应的输出值在所述预设监测范围内，则将所述原始梯度提升决策树模型和所述原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型；

采用测试集对应的待训练数据对所述目标梯度提升决策树模型和所述目标逻辑回归模型进行测试，若获取到的每一所述待训练数据对应的输出值均在所述预设监测范围内，则将所述目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型。

19. 如权利要求 18 所述的非易失性可读存储介质，其特征在于，所述获取待训练数据，将所述待训练数据划分为训练集和测试集，包括：

获取模型训练请求，所述模型训练请求包括训练组合特征；

根据所述训练组合特征，从用户数据库中选取与所述训练组合特征匹配的训练用户数据和与所述训练组合特征不匹配的非训练用户数据；

按照预设的正样本数量，从所述训练用户数据中选取对应的训练用户数据作为原始正样本；

按照正负样本比例，从所述非训练用户数据中选取对应的非训练用户数据作为原始负样本；

根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将所述负噪声数据加入到原始负样本中，生成有效负样本；

根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将所述正噪声数据加入到原始正样本中，生成有效正样本；

将所述有效正样本和所述有效负样本作为待训练数据存储在样本数据库中。

20. 如权利要求 19 所述的非易失性可读存储介质，其特征在于，在所述将所述原始组合特征输入到原始逻辑回归模型中，获取所述原始组合特征对应的输出值的步骤之后，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

当所述原始组合特征对应的输出值低于所述预设监测范围，则减小所述第一噪声抽取范围和所述第二噪声抽取范围，并提高所述正负样本比例；当所述原始组合特征对应的输出值高于所述预设监测范围，则增大所述第一噪声抽取范围和所述第二噪声抽取范围，并降低所述正负样本比例。

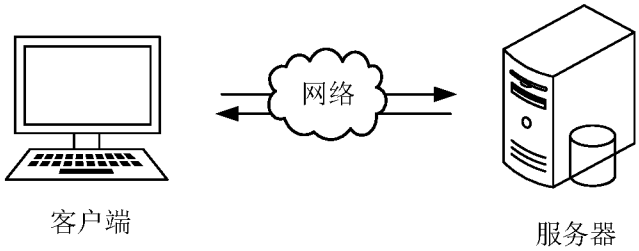
网络
客户端
服务器

图 1

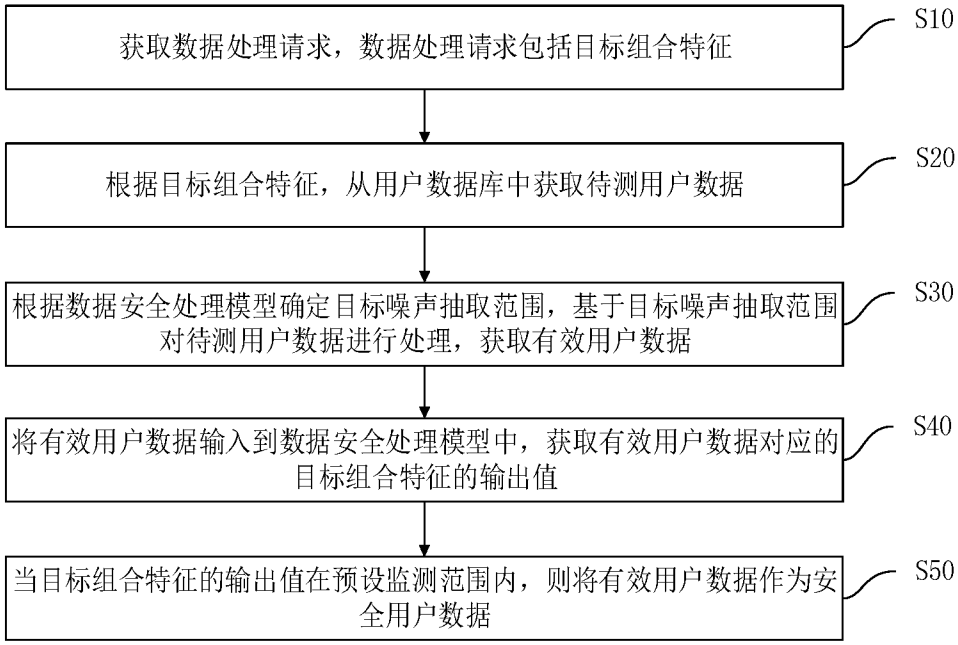
获取数据处理请求，数据处理请求包括目标组合特征
S10
根据目标组合特征，从用户数据库中获取待测用户数据
S20
根据数据安全处理模型确定目标噪声抽取范围，基于目标噪声抽取范围对待测用户数据进行处理，获取有效用户数据
S30
将有效用户数据输入到数据安全处理模型中，获取有效用户数据对应的目标组合特征的输出值
S40
当目标组合特征的输出值在预设监测范围内，则将有效用户数据作为安全用户数据
S50

图 2

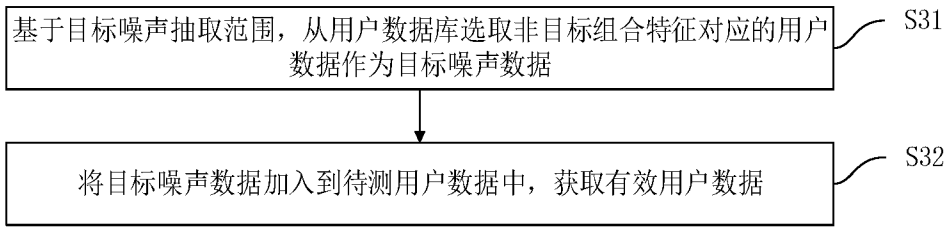
基于目标噪声抽取范围，从用户数据库选取非目标组合特征对应的用户数据作为目标噪声数据
S31
将目标噪声数据加入到待测用户数据中，获取有效用户数据
S32

图 3

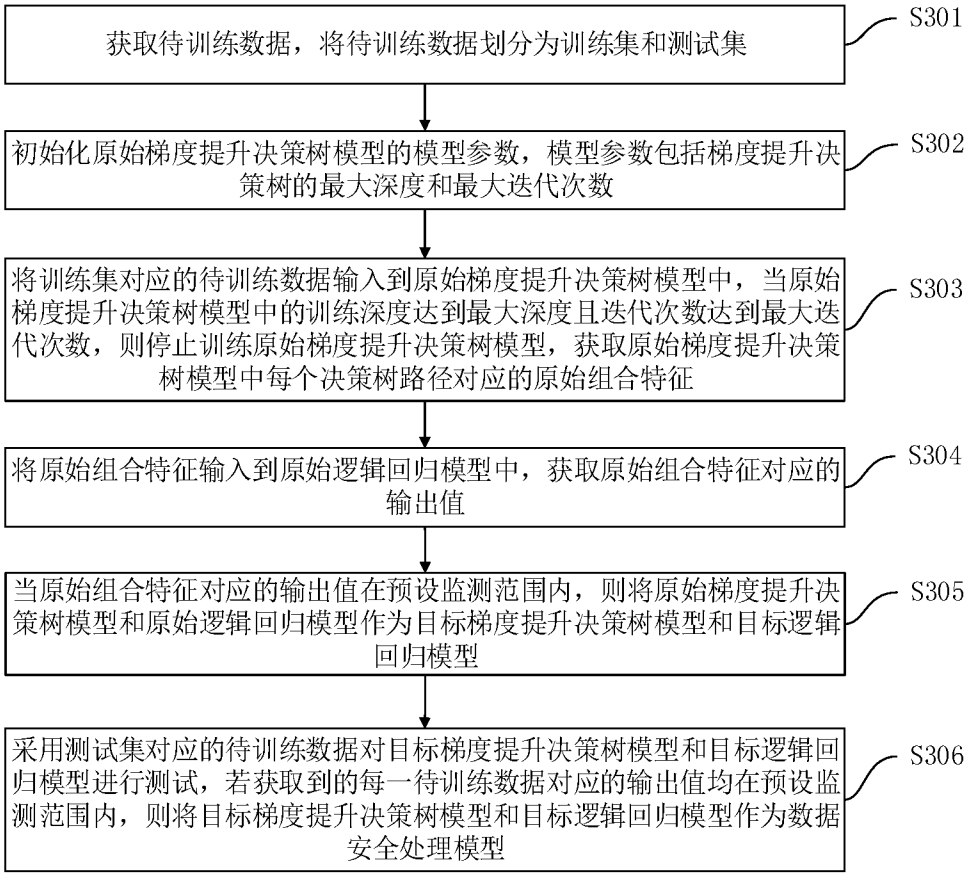
获取待训练数据，将待训练数据划分为训练集和测试集
S301
初始化原始梯度提升决策树模型的模型参数，模型参数包括梯度提升决策树的最大深度和最大迭代次数
S302
将训练集对应的待训练数据输入到原始梯度提升决策树模型中，当原始梯度提升决策树模型中的训练深度达到最大深度且迭代次数达到最大迭代次数，则停止训练原始梯度提升决策树模型，获取原始梯度提升决策树模型中每个决策树路径对应的原始组合特征
S303
将原始组合特征输入到原始逻辑回归模型中，获取原始组合特征对应的输出值
S304
当原始组合特征对应的输出值在预设监测范围内，则将原始梯度提升决策树模型和原始逻辑回归模型作为目标梯度提升决策树模型和目标逻辑回归模型
S305
采用测试集对应的待训练数据对目标梯度提升决策树模型和目标逻辑回归模型进行测试，若获取到的每一待训练数据对应的输出值均在预设监测范围内，则将目标梯度提升决策树模型和目标逻辑回归模型作为数据安全处理模型
S306

图 4

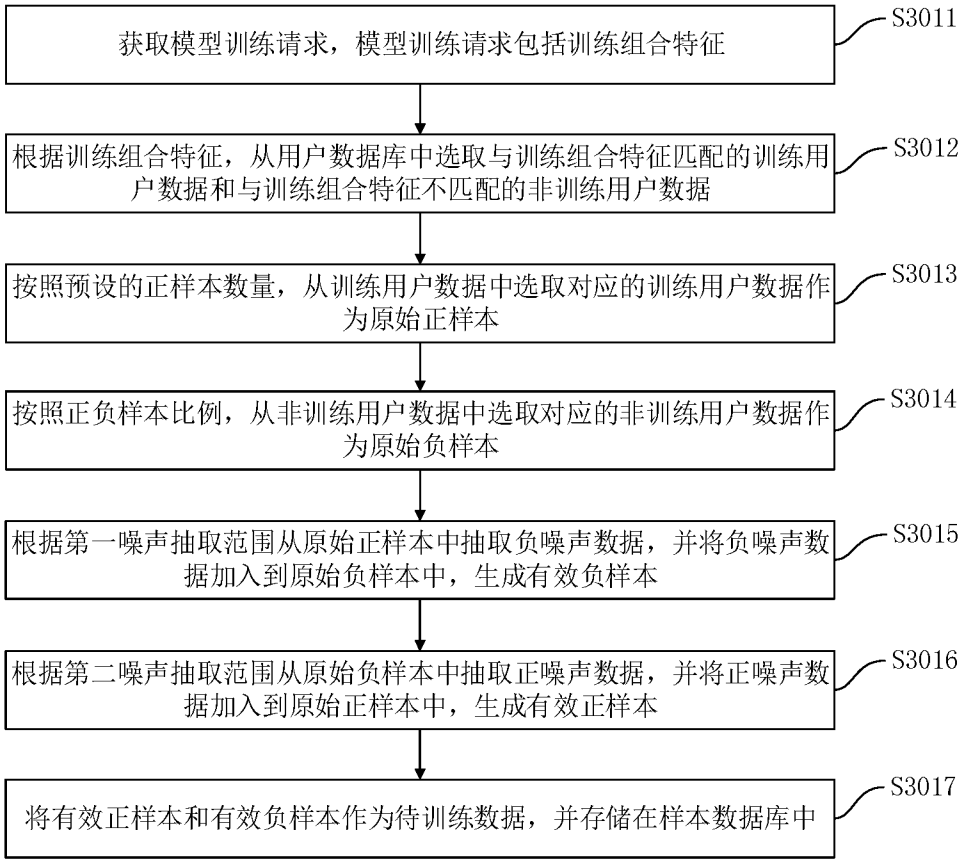
获取模型训练请求，模型训练请求包括训练组合特征
S3011
根据训练组合特征，从用户数据库中选取与训练组合特征匹配的训练用户数据和与训练组合特征不匹配的非训练用户数据
S3012
按照预设的正样本数量，从训练用户数据中选取对应的训练用户数据作为原始正样本
S3013
按照正负样本比例，从非训练用户数据中选取对应的非训练用户数据作为原始负样本
S3014
根据第一噪声抽取范围从原始正样本中抽取负噪声数据，并将负噪声数据加入到原始负样本中，生成有效负样本
S3015
根据第二噪声抽取范围从原始负样本中抽取正噪声数据，并将正噪声数据加入到原始正样本中，生成有效正样本
S3016
将有效正样本和有效负样本作为待训练数据，并存储在样本数据库中
S3017

图 5

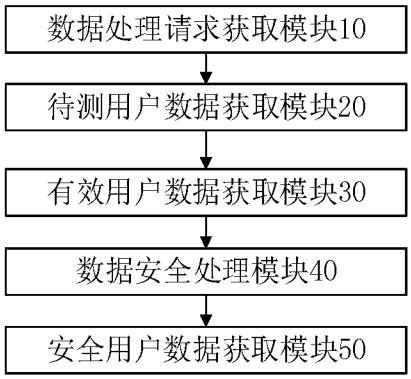
数据处理请求获取模块10
待测用户数据获取模块20
有效用户数据获取模块30
数据安全处理模块40
安全用户数据获取模块50

图 6

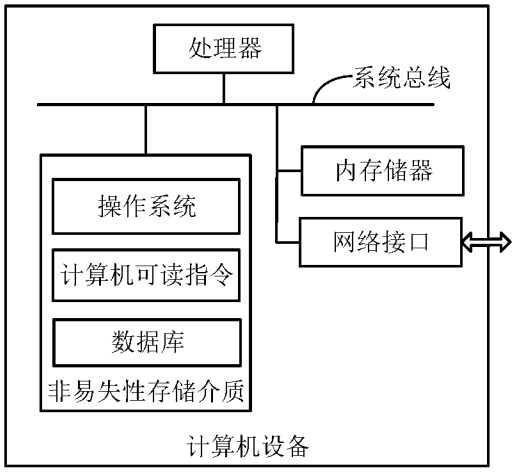
处理器
系统总线
内存储器
网络接口
操作系统
计算机可读指令
数据库
非易失性存储介质
计算机设备

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/122734

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/62(2013.01)i; G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNKI, CNTXT, VEN: 数据, 信息, 隐私, 保护, 加噪, 脱敏, 噪声, 加入, 添加, 扰动, 干扰, 决策树, 逻辑回归, data, information, privacy, intimacy, intimity, preserv+, protect+, noise, add+, increas+, disturb+, decision 1w tree, logistic, regression

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 108537055 A (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 14 September 2018 (2018-09-14) description, paragraphs 7-67	1-20
Y	李睿琪 (LI, Ruiqi). "针对匿名电信客户数据的流失预测模型 (Churn Prediction Models for Anonymous Telecom Customer Dataset)" *中国优秀硕士学位论文全文数据库经济与管理科学辑 (Non-official translation: Economics and Management, China Master's Theses Full-Text Database),* No. no. 1, 15 January 2018 (2018-01-15), ISSN: 1674-0246, sections 2.1 and 4.3	1-20
A	CN 102549614 A (MICROSOFT CORPORATION) 04 July 2012 (2012-07-04) entire document	1-20
A	CN 106339714 A (SHANGHAI JIAO TONG UNIVERSITY) 18 January 2017 (2017-01-18) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 July 2019	**29 July 2019**
Name and mailing address of the ISA/CN **State Intellectual Property Office of the P. R. China** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/122734

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 108537055 A	14 September 2018	None	
CN 102549614 A	04 July 2012	EP 2486536 A2	15 August 2012
		US 8312273 B2	13 November 2012
		WO 2011044232 A2	14 April 2011
		JP 2013507687 A	04 March 2013
		WO 2011044232 A3	30 June 2011
		KR 2012009259 A	21 August 2012
		JP 5662459 B2	28 January 2015
		KR 101735136 B1	12 May 2017
		EP 2486536 A4	26 November 2014
		CN 102549614 B	23 November 2016
		US 2011083013 A1	07 April 2011
		HK 1167502 A0	30 November 2012
CN 106339714 A	18 January 2017	None	

国际检索报告

国际申请号
PCT/CN2018/122734

A. 主题的分类

G06F 21/62(2013.01)i; G06K 9/62(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNABS,CNKI,CNTXT,VEN:数据，信息，隐私，保护，加噪，脱敏，噪声，加入，添加，扰动，干扰，决策树，逻辑回归，data，information，privacy，intimacy，intimity，preserv+，protect+，noise，add+，increas+，disturb+，decision 1w tree，logistic，regression

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN 108537055 A (南京邮电大学) 2018年 9月 14日 (2018-09-14) 说明书第7-67段	1-20
Y	李睿琪. "针对匿名电信客户数据的流失预测模型" 中国优秀硕士学位论文全文数据库经济与管理科学辑, 第1期，2018年 1月 15日 (2018-01-15), ISSN：1674-0246, 第2.1、4.3小节	1-20
A	CN 102549614 A (微软公司) 2012年 7月 4日 (2012-07-04) 全文	1-20
A	CN 106339714 A (上海交通大学) 2017年 1月 18日 (2017-01-18) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 7月 12日	2019年 7月 29日

ISA/CN的名称和邮寄地址	受权官员
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	孙泽竑
传真号 (86-10)62019451	电话号码 62089929

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/122734

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	108537055	A	2018年 9月 14日	无			
CN	102549614	A	2012年 7月 4日	EP	2486536	A2	2012年 8月 15日
				US	8312273	B2	2012年 11月 13日
				WO	2011044232	A2	2011年 4月 14日
				JP	2013507687	A	2013年 3月 4日
				WO	2011044232	A3	2011年 6月 30日
				KR	2012009259	A	2012年 8月 21日
				JP	5662459	B2	2015年 1月 28日
				KR	101735136	B1	2017年 5月 12日
				EP	2486536	A4	2014年 11月 26日
				CN	102549614	B	2016年 11月 23日
				US	2011083013	A1	2011年 4月 7日
				HK	1167502	A0	2012年 11月 30日
CN	106339714	A	2017年 1月 18日	无			