

兼顾隐私保护和公平性的图神经网络的训练方法及装置

5 本申请要求于 2022 年 11 月 29 日提交中国国家知识产权局、申请号为 202211507949.1、申请名称为“兼顾隐私保护和公平性的图神经网络的训练方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

10 本说明书涉及图神经网络技术领域，尤其涉及一种兼顾隐私保护和公平性的图神经网络的训练方法及装置。

背景技术

可信 AI (Trustworthy AI) 是当今机器学习模型发展中的一个重要议题。随着模型能力的逐步提升，数据量的进一步增大，如何让模型在学习的过程当中能够不因为弱势群体占比小而产生歧视，催生了可信 AI 中的一个重要的分支-公平性问题。

15 目前，在解决机器学习模型（例如图神经网络）的公平性问题的方法中，部分方法中需要考虑关于人员的某些属性特征（例如：性别以及年龄等），以训练针对这些属性特征具有公平性的图神经网络（即公平图神经网络），这些属性特征一般具有隐私特性，易造成人员的隐私数据的泄露。那么如何提供一种兼顾隐私保护和公平性的图神经网络的训练方法成为亟待解决的问题。

20 发明内容

本说明书一个或多个实施例提供了一种兼顾隐私保护和公平性的图神经网络的训练方法及装置，以实现训练得到兼顾隐私保护和公平性的图神经网络。

根据第一方面，提供一种兼顾隐私保护和公平性的图神经网络的训练方法，包括：

25 利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，得到所述 N 个目标用户的用户表征；

至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失；

根据各预测损失，确定各目标用户对应的权重值，使得预测损失越大，所对应目标用户的权重值越大；

30 基于各目标用户的预测损失和权重值，确定总预测损失；

以最小化所述总预测损失为目标, 调整所述图神经网络的参数。

根据第二方面, 提供一种兼顾隐私保护和公平性的图神经网络的训练装置, 包括:

聚合模块, 配置为利用图神经网络, 对用户关系网络图中 N 个目标用户对应的节点进行表征聚合, 得到所述 N 个目标用户的用户表征;

5 第一确定模块, 配置为至少基于各目标用户的用户表征, 采用与目标业务相关的预设损失函数, 确定各目标用户对应的预测损失;

第二确定模块, 配置为根据各预测损失, 确定各目标用户对应的权重值, 使得预测损失越大, 所对应目标用户的权重值越大;

第三确定模块, 配置为基于各目标用户的预测损失和权重值, 确定总预测损失;

10 调整模块, 配置为以最小化所述总预测损失为目标, 调整所述图神经网络的参数。

根据第三方面, 提供一种计算机可读存储介质, 其上存储有计算机程序, 当所述计算机程序在计算机中执行时, 令计算机执行第一方面所述的方法。

根据第四方面, 提供一种计算设备, 包括存储器和处理器, 其中, 所述存储器中存储有可执行代码, 所述处理器执行所述可执行代码时, 实现第一方面所述的方法。

15 根据本说明书实施例提供的方法及装置, 利用图神经网络, 处理以用户为节点的用户关系网络图, 得到其中 N 个目标用户的用户表征, 之后, 至少基于各目标用户的用户表征, 采用与目标业务相关的预设损失函数, 确定出各目标用户对应的预测损失, 接着考虑到对弱势群体的公平性, 要求在网络模型训练过程中不能仅仅关注主流群体, 而需要同时保证在弱势群体中的网络模型性能得到可靠保证, 相应的, 根据各预测损失, 确定各目标用户对应的权重值, 使得预测损失越大, 所对应目标用户的权重值越  
20 大, 之后基于各目标用户的预测损失和权重值, 确定总预测损失; 以最小化总预测损失为目标, 调整图神经网络的参数。上述过程中, 预测损失越大, 所对应目标用户的权重值越大, 可以提高预测损失较大的目标用户(理论上属于弱势群体)在网络模型训练过程中的受关注度, 从而提高图神经网络对弱势群体的公平性。且该训练过程中,  
25 无需预先知晓各目标用户的隐私数据, 借鉴分布鲁棒优化思想, 训练图神经网络, 以保证图神经网络对弱势群体(预测损失大的目标用户)的表征聚合性能, 实现对用户隐私数据的保护和对弱势群体公平性的保证。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单的介绍。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。

5 图 1 为本说明书披露的一个实施例的实施框架示意图；

图 2 为实施例提供的兼顾隐私保护和公平性的图神经网络的训练方法的一种流程示意图；

图 3 为实施例提供的用户关系网络图的一种示意图；

10 图 4 为实施例提供的兼顾隐私保护和公平性的图神经网络的训练装置的一种示意性框图。

### 具体实施方式

下面将结合附图，详细描述本说明书实施例的技术方案。

15 本说明书实施例披露一种兼顾隐私保护和公平性的图神经网络的训练方法及装置，下面首先对兼顾隐私保护和公平性的图神经网络的训练方法的应用场景和技术构思进行介绍，具体如下：

如前所述，在解决机器学习模型（例如图神经网络）的公平性问题的方法中，部分方法中需要考虑关于人员的某些属性特征（例如：性别以及年龄等），以训练针对这些属性特征具有公平性的图神经网络（即公平图神经网络），这些属性特征一般具有隐私特性，易造成人员的隐私数据的泄露。

20 鉴于此，发明人提出一种兼顾隐私保护和公平性的图神经网络的训练方法。首先，需要说明的是，本说明书实施例提供的该训练方法主要关注于 Rawlsian Max-Min（罗尔斯安 最大-最小）公平性，即要求在网络模型训练过程中，不能仅仅关注其在主流群体（即数量占比较多的用户群体）中的性能，还需要同时保证其在弱势群体（即数量占比较少的用户群体）中的性能，也就是保护弱势群体。

25 例如，在后续图 3 中所示的社交网络场景中，低互动的用户可以认为是社交网络场景下用户关系网络图中的弱势群体。又例如，在电商平台（或电子支付平台）的用户关系网络图中，所包含的年龄超过第一年龄阈值的用户占比低于预设比例阈值，则可以认为该年龄超过第一年龄阈值的用户群体为其中的弱势群体，等。

可以理解的，弱势群体在总体中一般是一个占比较低的群体。其可以体现为，在网络模型训练所需的样本集中为一个占比较低的子集。例如，在训练针对用户进行分类分析的网络模型的过程中，若年龄超过预设值的用户群体样本的占比较低，可称该用户群体为弱势群体。在一些实现中，未考虑公平性问题的网络模型的训练过程中，一般通过优化用户样本集中各样本的平均误差，来实现对网络模型的优化训练。而该过程中，容易忽视弱势群体的特征表达，使得弱势群体的特征表达在网络模型的优化训练过程中被主流群体遮盖，然后使得网络模型在弱势群体中的性能不够好。相应的，网络模型在弱势群体中的性能不够好可以表现为，其针对弱势群体的预测结果的准确性不够高；以及在其训练过程中，其针对弱势群体的预测损失较大。

10 在此基础上，为了提高图神经网络在弱势群体中的性能，实现对弱势群体的公平处理，在网络模型的训练过程，需要增加对弱势群体的关注，并且需要关注到对用户群体的隐私保护。相应的，图 1 示出根据一个实施例的兼顾隐私保护和公平性的图神经网络的训练场景的示意图。该场景示意图中，具体的，首先获取以用户为节点的用户关系网络图，其中的边表示用户直接的关联关系。该关联关系例如可以为社交关系、  
15 交易关系以及转账关系等等。利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，得到 N 个目标用户的用户表征；至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失。

之后，为了提高图神经网络对弱势群体的关注，实现对弱势群体的保护，相应的，可以根据各目标用户对应的预测损失，确定各目标用户对应的权重值，使得预测损失  
20 越大，所对应目标用户的权重值越大。可以理解的是，结合前述的未考虑公平性问题的网络模型，其在弱势群体中的性能不够好，可以表现为：在网络模型的训练过程中，网络模型针对弱势群体的预测损失较大。鉴于此，可以通过权重值的设置，来增大图神经网络对弱势群体（预测损失大的目标用户）的关注。具体的，预测损失越大，所对应目标用户的权重值越大，即对相应的目标用户的关注度越大。并且训练过程中，  
25 无需提前预知用户关系网络图中哪些用户为弱势群体，即无需提前知晓用户群体的隐私数据，而是基于训练过程中图神经网络在目标业务任务下对目标用户的性能表现，来预估属于弱势群体的目标用户，其中，预测损失越大，所对应目标用户属于弱势群体的可能性越大，相应的，越需要关注该类目标用户，也就是该目标用户的权重值越大。通过前述方式，可以增加图神经网络在目标业务任务下对弱势群体的关注度，以  
30 在训练过程中提高图神经网络在目标业务任务下对弱势群体的保护能力，且可以实现

对用户群体的隐私数据的保护。

接着，基于各目标用户的预测损失和权重值，确定总预测损失，具体的可以是：计算各目标用户的预测损失和权重值的乘积的和值，将该和值确定为总预测损失。之后以最大化总预测损失为目标，调整图神经网络的参数。

5       上述过程中，预测损失越大，所对应目标用户的权重值越大，可以提高预测损失较大的目标用户（理论上属于弱势群体）在图神经网络的训练过程中的受关注度，从而提高图神经网络对弱势群体的公平性。且该训练过程中，无需预先知晓各目标用户的隐私数据，基于分布鲁棒优化思想，构建各目标用户对应的预测损失的权重值的最差情况下的分布，进而得到在该最差情况下的分布下的最优的解，即以最小化总预测  
10   损失为目标，训练图神经网络，以保证图神经网络对弱势群体（预测损失大的目标用户）的表征聚合性能，实现对用户隐私数据的保护和对弱势群体公平性的保证。

下面结合具体实施例，对本说明书提供的兼顾隐私保护和公平性的图神经网络的训练方法及装置进行详细阐述。

图 2 示出了本说明书一个实施例中兼顾隐私保护和公平性的图神经网络的训练方法  
15   的流程图。该方法可以通过任何具有计算、处理能力的装置、设备、平台、设备集群等来实现。在训练过程中，如图 2 所示，所述方法包括如下步骤 S210- S250：

首先在步骤 S210，利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，得到 N 个目标用户的用户表征。本步骤中，该用户关系网络图可以是针对目标平台的用户及其之间的关联关系所构建的，其中，各节点对应目标平台的  
20   各用户，边表示用户之间的关联关系。在一种情况中，该目标平台例如可以是电商平台、电子支付平台、金融平台或者社交平台等。在一个示例中，该目标平台为电商平台的情况下，用户关系网络图中各节点对应电商平台的各用户，边表示的关联关系可以为电商平台的各用户之间的交易关系。在又一个示例中，该目标平台为电子支付平台（或者金融平台）的情况下，用户关系网络图中各节点对应电子支付平台的各用户，  
25   边表示的关联关系可以为电商平台的各用户之间的转账关系（或者借贷关系）。在又一个示例中，该目标平台为社交平台的情况下，用户关系网络图中各节点对应社交平台的各用户，边表示的关联关系可以为社交平台的各用户之间的社交互动关系。

在步骤 S210，可以预先根据目标业务的业务需求，随机从用户关系网络图中确定出 N 个目标用户。一种情况中，目标业务为分类业务（例如预测用户分类）或者回归

业务（预测用户指标值）的情况下，各目标用户为具有与目标业务对应的标签数据的用户。另一种情况中，目标业务为自编码业务的情况下，目标用户可以为用户关系网络图中任意的用户。

确定出 N 个目标用户之后，在一个实施例中，可以将该用户关系网络图输入图神经网络，利用图神经网络的 K 个聚合层，至少根据该 N 个目标用户各自对应的 K 跳邻居节点集，分别对用户关系网络图中该 N 个目标用户对应的节点进行 K 级表征聚合，得到 N 个目标用户的用户表征。N 和 K 均为预先设置的数值，为了训练所得性能较好的图神经网络，N 越大越好。K 可以根据实际需求（例如图神经网络的聚合层的个数）进行设置，例如设置为 2。目标用户的用户表征中可以聚合有目标用户自身的特征数据，以及其 K 跳邻居节点集中各节点的特征数据。

考虑到用户关系网络图的整体数据量较大，为了节省计算资源消耗，在又一个实施例中，在步骤 S210，可以包括：在用户关系网络图中，分别以各目标用户对应的节点为中心节点，确定该中心节点的 K 跳邻居节点集，该中心节点及其 K 跳邻居节点集构成一个样本子图；将各样本子图输入图神经网络，对其中的中心节点进行表征聚合。各样本子图中包括中心节点及中心节点的 K 跳邻居节点集，以及各节点之间的边。将各样本子图输入图神经网络之后，可以利用图神经网络的 K 个聚合层，根据各样本子图中的节点的特征数据，对其中的中心节点进行 K 级表征聚合。一种实现中，可以通过 AGL 系统实现对样本子图的采样过程。

在一种情况中，可能存在与目标用户存在关联关系的用户数量较少的情况。例如在社交网络场景中，有部分低互动的用户，其用户关系网络图的局部示意图可以如图 3 所示，其中，低互动用户对应的节点比较孤立，一般存在于比较特殊的子图中，例如，低互动用户对应的节点所在子图的节点数量较少（例如低于预设数量，例如为 3，又例如节点没有邻居节点）。相应的，若该类用户（例如没有邻居的用户）被确定为目标用户，则其样本子图可以仅包括该目标用户对应的节点。

在聚合得到 N 个目标用户的用户表征之后，在步骤 S220，至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失。

在一个实施例中，该目标业务可以为预测用户分类的业务、预测用户指标值的业务或者自编码业务，不同的目标业务可以对应不同的预设损失函数。例如：该目标业务为预测用户分类的业务的情况下，该预设损失函数可以为交叉熵损失函数，该目标

业务为预测用户指标值的业务的情况下,该预设损失函数可以为均方差 MSE 损失函数,该目标业务为自编码业务的情况下,该预设损失函数可以为用于构建无监督任务中特征重建损失的损失函数。

5 在一个实施例中,目标业务为预测用户分类的业务或者预测用户指标值的业务的情况下,各目标用户具有与目标业务对应的标签数据;相应的,在步骤 S220,具体可以包括:利用与目标业务相关的预测网络,对各目标用户的用户表征进行处理,得到各目标用户对应的预测结果;将标签数据和预测结果输入预设损失函数,得到对应的预测损失。其中,目标业务为预测用户分类的业务的情况下,该预测网络为用户分类网络;该目标业务为预测用户指标值的业务的情况下,该预测网络为用户指标预测网络。  
10 络。

具体的,得到 N 个目标用户的用户表征之后,将各目标用户的用户表征输入该预测网络,利用该预测网络对各目标用户的用户表征进行处理,得到各目标用户对应的预测结果,将各目标用户对应的标签数据和预测结果分别输入预设损失函数,得到各目标用户对应的预测损失。

15 在又一个实施例中,目标业务为自编码业务的情况下,在步骤 S220,具体可以包括:利用与目标业务相关的解码网络,处理各目标用户的用户表征,确定出各目标用户的重构特征数据;基于各目标用户的重构特征数据和各目标用户对应的原始特征数据,采用预设损失函数,计算得到各目标用户的预测损失。本步骤中,将各目标用户的用户表征分别输入解码网络,以利用解码网络,处理各目标用户的用户表征,得到  
20 各目标用户的重构特征数据,之后,基于各目标用户的重构特征数据和各目标用户对应的原始特征数据,采用预设损失函数,计算得到各目标用户的预测损失。具体的,可以是:计算各目标用户的重构特征数据和原始特征数据之间的特征差异,基于各目标用户对应的特征差异确定各目标用户的预测损失。一种实现中,该原始特征数据可以包括所对应目标用户的基本属性数据以及与关联关系相关的特征数据。

25 需要理解的,本说明书实施例提供的该方法主要关注于 Rawlsian Max-Min (罗尔斯安 最大-最小) 公平性,即要求在网络模型训练过程中,不能仅仅关注其在主流群体(即数量占比较多的用户群体)中的性能,还需要同时保证其在弱势群体(即数量占比较少的用户群体)中的性能,也就是保护弱势群体。

为此,借鉴分布鲁棒优化思想,认为各目标用户对应的预测损失(即各目标用户)

存在分布漂移情况，之后通过对各预测损失赋予权重值（即赋权），使得赋权后的各预测损失形成一个最差情况的数据分布（即预测损失越大，所对应目标用户的权重值越大，且预测损失与对应的权重值乘积之和最大）。之后针对该最差情况的数据分布训练图神经网络，训练目标是，使得图神经网络在赋权后的各预测损失形成的最差情况的数据分布下，达到最好的性能。如此，在不需要预先获知用户群体的隐私数据（即关注隐私保护）的前提下，训练得到能够保护弱势群体（即实现公平性）的图神经网络。

具体的，在步骤 S230，根据各预测损失，确定各目标用户对应的权重值，使得预测损失越大，所对应目标用户的权重值越大。可以理解的是，各目标用户对应的预测损失，可以在一定程度上指示出图神经网络在目标业务任务下对目标用户的表征能力（即性能）的优劣，其中，目标用户对应的预测损失越大，可以认为图神经网络在目标业务任务下针对目标用户的性能越差。对于预测损失越大的目标用户（即弱势群体），为其赋予越大的权重值，使得图神经网络越关注该类目标用户，以提高图神经网络对该类用户（弱势群体）的公平性，提高在目标业务任务下对弱势群体的性能。

其中，各目标用户对应的权重值的取值范围均为[0,1)，且各目标用户对应的权重值的和为 1。在一种情况中，目标用户对应的预测损失低于预设损失值时，可以设置该目标用户对应的权重值为 0。

在一个实施例中，在步骤 S230，具体可以包括：以各预测损失与其对应的权重值的乘积之和最大化为目标，在预设约束条件下，确定各权重值，其中，该预设约束条件包括：该权重值形成的实际分布与预设先验分布之间的距离不超过扰动半径。其中，该距离可以指该权重值形成的实际分布与预设先验分布之间的 f 散度距离或 wasserstein 距离或 CVaR 值。一种实现中，该预设先验分布可以为均匀分布。

其中，可以通过如下公式，表示确定各目标用户的权重值的过程：

$$Q^* = \operatorname{argmax}_{Q: D_{fk}(Q||\widehat{P}_N) \leq \rho} \left\{ \sum_{i=1}^N q_i l(\theta; X_i) \right\};$$

其中，Q 表示各目标用户的权重值形成的实际分布， $\widehat{P}_N$  表示预设先验分布， $\rho$  表示扰动半径， $D_{fk}(Q||\widehat{P}_N) \leq \rho$  表示实际分布与预设先验分布之间的 f 散度距离不超过（小于等于）扰动半径； $q_i$  表示第 i 个目标用户的权重值， $l(\theta; X_i)$  表示第 i 个目标用户的预测损失，其中  $X_i$  表示第 i 个目标用户的原始特征数据， $\theta$  表示图神经网络（以

及预测网络或者解码网络)的参数。因此,求和符号所得的结果为,各目标用户的预测损失与其对应的权重值的乘积之和。 $Q^*$ 表示所得到的各目标用户的权重值形成的最优的实际分布,即上述乘积之和达到最大。

5 各预测损失与其对应的权重值的乘积之和最大化,对应于赋权后的各预测损失达到分布漂移情况下的最差情况的数据分布。相应的,使得图神经网络(以及预测网络或解码网络)更加关注在该最差情况的数据分布下的性能(worst-case performance),以实现关于分布漂移情况下的鲁棒性,这样可以提高图神经网络的公平性和隐私保护性能,同时也可以提升图神经网络(以及预测网络或解码网络)的尾部性能(tail performance)。

10 在一个实施例中,前述的扰动半径根据预设的用户关系网络图中弱势群体用户的占比而确定。在一种实现方式中,预设的用户关系网络图中弱势群体用户的占比 $\alpha$ 的取值范围可以为(0, 0.5),一种情况中, $\alpha$ 可以取[0.1, 0.3]。在一种实现中,可以通过如下公式,确定扰动半径 $\rho$ ,其中,扰动半径 $\rho = (1/\alpha - 1)^2$ 。

15 确定出各目标用户的权重值之后,在步骤S240,基于各目标用户的预测损失和权重值,确定总预测损失。在一个实施例中,在步骤S240,具体可以包括:计算各目标用户的预测损失及对应的权重值的乘积的和值,将该和值作为总预测损失。这样,计算所得的总预测损失可以更好的关注弱势群体(即预测损失大的目标用户)。接着,在步骤S250,以最小化总预测损失为目标,调整图神经网络的参数。本步骤中,基于总预测损失,利用反向传播算法,确定图神经网络的参数梯度。利用所确定的模型参数梯度以及图神经网络的参数的当前取值,确定图神经网络的参数的更新值。进而基于更新值,调整图神经网络的参数。其中,确定图神经网络的参数梯度是以最小化总预测损失为目标得到的。

25 在一个实施例中,目标业务为预测用户分类的业务或者预测用户指标值的业务的情况下,图神经网络之后还连接有与目标业务相关的预测网络(为用户分类网络或者用户指标值预测网络)在步骤S250,可以具体包括:以最小化总预测损失为目标,调整图神经网络和预测网络的参数。

在又一个实施例中,目标业务为自编码业务的情况下,图神经网络(即编码网络)之后还连接有与目标业务相关的解码网络,用于对各目标用户的用户表征进行解码,

得到各目标用户的重构特征数据。相应的，在步骤 S250，还可以具体包括：以最小化总预测损失为目标，调整图神经网络和解码网络的参数。

上述步骤 S210~S250 为一次迭代训练过程。为了训练得到更好的图神经网络（以及  
5 与目标业务相关的预测网络或解码网络），可以多次迭代执行上述过程。也就是在步骤 S250 之后基于图神经网络（以及与目标业务相关的预测网络或解码网络）的参数  
的更新值，返回执行步骤 S210。上述迭代训练过程的停止条件可以包括，迭代训练次数  
达到预设次数阈值，或者迭代训练时长达到预设时长，或者总预测损失小于设定的损  
失阈值等等。

本实施例中，预测损失越大，所对应目标用户的权重值越大，可以提高预测损失  
10 较大的目标用户（理论上属于弱势群体）在图神经网络的训练过程中的受关注度，从而  
提高图神经网络对弱势群体的公平性。且该训练过程中，无需预先知晓各目标用户的  
隐私数据，基于分布鲁棒优化思想，构建各目标用户对应的预测损失的权重值的最  
差情况下的分布，进而得到在该最差情况下的分布下的最优解，即以最小化总预测  
损失为目标，训练图神经网络，以保证图神经网络对弱势群体（预测损失大的目标用  
15 户）的表征聚合性能，实现对用户隐私数据的保护和对弱势群体公平性的保证。

并且，本实施例中，可以认为在图神经网络模型的训练过程中，以松耦合的形式，  
在总预测损失计算过程中，嵌入一个用于计算 DRO（分布鲁棒优化）权重值的计算单  
元，以此来使得训练所得的图神经网络兼顾隐私保护和公平性。

本实施例可以实现对工业级大图上的兼顾隐私保护和公平性的图神经网络的训练，  
20 可用于可信 AI 的图学习实践中。

以各预测损失与其对应的权重值的乘积之和最大化为目标，确定各目标用户对应  
的权重值，以得到赋权之后的各预测损失的最差情况的数据分布，之后以最小化总预  
测损失（各预测损失与其对应的权重值的乘积之和）为目标，训练图神经网络（以及  
预测网络或解码网络），得到训练完成的图神经网络，实现得到在所述的最差情况  
25 的数据分布下的最优解，相应的图神经网络的鲁棒性在这个最差情况的数据分布下可  
以得到保证，即保证图神经网络在弱势群体下的性能。在存在弱势群体的用户关系网  
络图中，该图神经网络兼顾隐私保护和公平性的性能可以得到很好的表现。

上述内容对本说明书的特定实施例进行了描述，其他实施例在所附权利要求书的  
范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中

的顺序来执行，并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要按照示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的，或者可能是有利的。

相应于上述方法实施例，本说明书实施例，提供了一种兼顾隐私保护和公平性的图神经网络的训练装置 400，其示意性框图如图 4 所示，包括：

聚合模块 410，配置为利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，得到所述 N 个目标用户的用户表征；

第一确定模块 420，配置为至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失；

10 第二确定模块 430，配置为根据各预测损失，确定各目标用户对应的权重值，使得预测损失越大，所对应目标用户的权重值越大；

第三确定模块 440，配置为基于各目标用户的预测损失和权重值，确定总预测损失；

15 调整模块 450，配置为以最小化所述总预测损失为目标，调整所述图神经网络的参数。

在一种可选地实施方式中，各目标用户具有与所述目标业务对应的标签数据；

所述第一确定模块 420，具体配置为利用与所述目标业务相关的预测网络，对各目标用户的用户表征进行处理，得到各目标用户对应的预测结果；

将所述标签数据和预测结果输入所述预设损失函数，得到对应的预测损失。

20 在一种可选地实施方式中，所述调整模块 450，具体配置为以最小化所述总预测损失为目标，调整所述图神经网络和所述预测网络的参数。

在一种可选地实施方式中，所述第一确定模块 420，具体配置为利用与所述目标业务相关的解码网络，处理各目标用户的用户表征，确定出各目标用户的重构特征数据；

25 基于各目标用户的重构特征数据和各目标用户对应的原始特征数据，采用所述预设损失函数，计算得到各目标用户的预测损失。

在一种可选地实施方式中，所述目标业务为如下业务中的一种：预测用户分类、

预测用户指标值、自编码业务。

在一种可选地实施方式中，所述第二确定模块 430，配置为以各预测损失与其对应的权重值的乘积之和最大化为目标，在预设约束条件下，确定各权重值，其中，所述预设约束条件包括：所述权重值形成的实际分布与预设先验分布之间的距离不超过  
5 扰动半径。

在一种可选地实施方式中，所述预设先验分布为均匀分布。

在一种可选地实施方式中，所述扰动半径根据预设的所述用户关系网络图中弱势群体用户的占比而确定。

在一种可选地实施方式中，所述第三确定模块 440，配置为计算各目标用户的预测损失及对应的权重值的乘积的和值，作为总预测损失。  
10

在一种可选地实施方式中，所述聚合模块 410，配置为在所述用户关系网络图中，分别以各目标用户对应的节点为中心节点，确定该中心节点的 K 跳邻居节点集，该中心节点及其 K 跳邻居节点集构成一个样本子图；

将各样本子图输入所述图神经网络，对其中的中心节点进行表征聚合。

上述装置实施例与方法实施例相对应，具体说明可以参见方法实施例部分的描述，此处不再赘述。装置实施例是基于对应的方法实施例得到，与对应的方法实施例具有同样的技术效果，具体说明可参见对应的方法实施例。  
15

本说明书实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行本说明书所提供的所述兼顾隐私保护和公平性的图神经网络的训练方法。  
20

本说明书实施例还提供了一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现本说明书所提供的所述兼顾隐私保护和公平性的图神经网络的训练方法。

本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于存储介质和计算设备实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。  
25

本领域技术人员应该可以意识到，在上述一个或多个示例中，本发明实施例所描

述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

- 5 以上所述的具体实施方式，对本发明实施例的目的、技术方案和有益效果进行了进一步的详细说明。所应理解的是，以上所述仅为本发明实施例的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的技术方案的基础之上所做的任何修改、等同替换、改进等，均应包括在本发明的保护范围之内。

## 权 利 要 求 书

1、一种兼顾隐私保护和公平性的图神经网络的训练方法，包括：

利用图神经网络，对用户关系网络图中N个目标用户对应的节点进行表征聚合，得到所述N个目标用户的用户表征；

至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失，所述预测损失用于确定对应目标用户属于弱势群体的概率，所述预测损失越大，所对应目标用户属于弱势群体的概率越大；

根据各预测损失，确定各目标用户对应的权重值，使得所述概率越大，所对应目标用户的权重值越大；

基于各目标用户的预测损失和权重值，确定总预测损失；

以最小化所述总预测损失为目标，调整所述图神经网络的参数。

2、如权利要求1所述的方法，其中，各目标用户具有与所述目标业务对应的标签数据；

所述确定各目标用户对应的预测损失，包括：

利用与所述目标业务相关的预测网络，对各目标用户的用户表征进行处理，得到各目标用户对应的预测结果；

将所述标签数据和预测结果输入所述预设损失函数，得到对应的预测损失。

3、如权利要求2所述的方法，其中，所述调整所述图神经网络的参数，包括：

以最小化所述总预测损失为目标，调整所述图神经网络和所述预测网络的参数。

4、如权利要求1所述的方法，其中，所述确定各目标用户对应的预测损失，包括：

利用与所述目标业务相关的解码网络，处理各目标用户的用户表征，确定出各目标用户的重构特征数据；

基于各目标用户的重构特征数据和各目标用户对应的原始特征数据，采用所述预设损失函数，计算得到各目标用户的预测损失。

5、如权利要求1所述的方法，其中，所述目标业务为如下业务中的一种：预测用

户分类、预测用户指标值、自编码业务。

6、如权利要求 1-5 任一项所述的方法，其中，所述确定各目标用户对应的权重值，包括：

以各预测损失与其对应的权重值的乘积之和最大化为目标，在预设约束条件下，确定各权重值，其中，所述预设约束条件包括：所述权重值形成的实际分布与预设先验分布之间的距离不超过扰动半径。

7、如权利要求 6 所述的方法，其中，所述预设先验分布为均匀分布。

8、如权利要求 6 所述的方法，其中，所述扰动半径根据预设的所述用户关系网络图中弱势群体用户的占比而确定。

9、如权利要求 1-5 任一项所述的方法，其中，所述确定总预测损失，包括：

计算各目标用户的预测损失及对应的权重值的乘积的和值，作为总预测损失。

10、如权利要求 1-5 任一项所述的方法，其中，所述利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，包括：

在所述用户关系网络图中，分别以各目标用户对应的节点为中心节点，确定该中心节点的 K 跳邻居节点集，该中心节点及其 K 跳邻居节点集构成一个样本子图；

将各样本子图输入所述图神经网络，对其中的中心节点进行表征聚合。

11、一种兼顾隐私保护和公平性的图神经网络的训练装置，包括：

聚合模块，配置为利用图神经网络，对用户关系网络图中 N 个目标用户对应的节点进行表征聚合，得到所述 N 个目标用户的用户表征；

第一确定模块，配置为至少基于各目标用户的用户表征，采用与目标业务相关的预设损失函数，确定各目标用户对应的预测损失，所述预测损失用于确定对应目标用户属于弱势群体的概率，所述预测损失越大，所对应目标用户属于弱势群体的概率越大；

第二确定模块，配置为根据各预测损失，确定各目标用户对应的权重值，使得所述概率越大，所对应目标用户的权重值越大；

第三确定模块，配置为基于各目标用户的预测损失和权重值，确定总预测损失；

调整模块，配置为以最小化所述总预测损失为目标，调整所述图神经网络的参数。

12、一种计算设备，包括存储器和处理器，其中，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-10 中任一项所述的方法。