

业务任务执行方法、装置以及计算机可读存储介质

技术领域

本发明涉及能源技术领域，尤其涉及业务任务执行方法、装置以及计算机可读存储介质。

5

背景技术

作为一种新型机器学习理念，联合学习通过分布式训练及加密技术确保用户隐私数据得到最大限度的保护，以提升用户对人工智能技术的信任。在联合学习机制下，联合学习服务器初始化全局模型，下发给每个用户作为初始化模型，用户根据自身数据训练本地的局部模型，然后上传局部模型到联合学习服务器，联合学习服务器聚合局部模型再下发到每个用户作为初始化模型训练，如此迭代直至模型收敛，最后得到全局模型，通过联合每个用户数据信息，在数据不出本地的情况下提高全局模型的精度。

目前，首先通过联合学习得到目标用户的全局模型，然后根据目标用户的本地数据微调全局模型，得到适合目标用户的模型。

15 但是，微调全局模型需要利用目标用户的标签数据，然而在很多应用场景下，目标用户的标签数据难以获取，使得此类方式在使用时面临困难。

发明内容

本发明提供了一种业务任务执行方法、装置、计算机可读存储介质及电子设备，可在目标用户不具有标签的前提下，通过标签数据的权重，将非标签数据迁移到了标签数据上，确保能够实现目标用户的业务任务。

第一方面，本发明提供了一种业务任务执行方法，包括：
对目标用户的业务任务对应的多个非标签数据进行聚类，以确定至少两个聚类中心点；
根据所述至少两个聚类中心点以及所述多个非标签数据，确定联合用户的多个标签数据各自对应的权重，所述多个标签数据和所述业务任务对应；

25 根据所述联合用户的所述多个标签数据以及所述多个标签数据各自对应的权重，构建联合学习模型，所述联合学习模型用于执行所述目标用户的业务任务。

第二方面，本发明提供了一种业务任务执行装置，包括：
聚类模块，用于对目标用户的业务任务对应的多个非标签数据进行聚类，以确定至少两个聚类中心点；

30 权重确定模块，用于根据所述至少两个聚类中心点以及所述多个非标签数据，确定联合

用户的多个标签数据各自对应的权重，所述多个标签数据和所述业务任务对应；

构建模块，用于根据所述联合用户的所述多个标签数据以及所述多个标签数据各自对应的权重，构建联合学习模型，所述联合学习模型用于执行所述目标用户的业务任务。

第三方面，本发明提供了一种计算机可读存储介质，包括执行指令，当电子设备的处理器执行所述执行指令时，所述处理器执行如第一方面中任一所述的方法。

第四方面，本发明提供了一种电子设备，包括处理器以及存储有执行指令的存储器，当所述处理器执行所述存储器存储的所述执行指令时，所述处理器执行如第一方面中任一所述的方法。

本发明提供了一种业务任务执行方法、装置、计算机可读存储介质及电子设备，该方法通过对目标用户的业务任务对应的多个非标签数据进行聚类，以确定两个或多个聚类中心点，然后，根据两个或多个聚类中心点以及多个非标签数据，确定联合用户的多个标签数据各自对应的权重，多个标签数据和业务任务对应，之后，根据联合用户的多个标签数据以及多个标签数据各自对应的权重，构建联合学习模型，联合学习模型用于执行目标用户的业务任务。综上所述，通过本发明的技术方案，可在目标用户不具有标签的前提下，通过标签数据的权重，将非标签数据迁移到了标签数据上，确保能够实现目标用户的业务任务。

上述的非惯用的优选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

为了更清楚地说明本发明实施例或现有的技术方案，下面将对实施例或现有技术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发明中记载的一些实施例，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为本发明一实施例提供了一种业务任务执行方法的流程示意图；

图2为本发明一实施例提供的另一种业务任务执行方法的结构示意图；

图3为本发明一实施例提供了一种业务任务执行装置的结构示意图；

图4为本发明一实施例提供了一种电子设备的结构示意图。

具体实施方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合具体实施例及相应的附图对本发明的技术方案进行清楚、完整地描述。显然，所描述的实施例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做出创造性劳动

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。

如图 1 所示，为本发明实施例提供的一种业务任务执行方法。本发明实施例所提供的方法可应用在电子设备上，具体可以应用于服务器或一般计算机上。本实施例中，所述方法具体包括以下步骤：

- 5 步骤 101、对目标用户的业务任务对应的多个非标签数据进行聚类，以确定至少两个聚类中心点。

目标用户指的是具有业务需求的设备，可以是能源设备，比如燃气蒸汽锅炉、光伏电站、燃气内燃机、燃气轮机等。

- 10 业务任务指的是目标用户最终要实现的目标，比如，可以是故障预测、设备剩余使用寿命预测、变量预测等。

- 15 非标签数据指的是不具有标签的特征数据，特征数据是一个一维行向量，行向量包括多个特征各自对应的特征值，其中，特征指的是影响业务任务的影响因素，应当理解的，多个非标签数据具有顺序号，每个非标签数据对应的多个特征相同，且多个特征具有顺序号，在实际应用中，第*i*个非标签数据表示为 $[x_{i,1}, x_{i,2}, \dots, x_{i,j-1}, x_{i,j}]$ ，这里， $x_{i,j}$ 表示第*i*个非
15 标签数据中的第*j*个特征对应的特征值，其他数据项表示含义类似，这里不做过多赘述。

- 具体地，通过聚类算法对多个非标签数据进行聚类，从而确定两个或多个聚类中心点。其中，聚类算法可以是 k-means 聚类、层次聚类算法或密度聚类，优选 K-means 聚类。在一些可能的情况中，聚类中心点和多个非标签数据中的任意一个非标签数据不同，从而确保了数据安全。具体地，通过聚类算法对多个非标签数据进行聚类，以确定若干个聚类簇，针对
20 每个聚类簇，计算聚类簇中的多个非标签数据的均值，当计算得到的均值为非标签数据时，将该均值以及距离该均值最近的非标签数据之间的均值确定为聚类中心点。

步骤 102、根据所述至少两个聚类中心点以及所述多个非标签数据，确定联合用户的多个标签数据各自对应的权重，所述多个标签数据和所述业务任务对应。

- 25 本实施例通过两个或多个聚类中心点以及多个非标签数据，确定出联合用户的多个标签数据各自对应的权重，在目标用户不具有标签的前提下，将非标签数据迁移到了标签数据上，确保能够实现目标用户的业务任务。

- 可以理解的，标签数据指的是具有标签的特征数据，标签数据和非标签数据对应的多个特征相同，从而使得目标用户和联合用户之间能够进行横向联合学习。其中，标签和业务任务有关，比如，业务任务是故障预测，则标签可以是故障类型，业务任务是烟气含氧量预测，
30 则标签可以是烟气含氧量，业务任务是设备剩余使用寿命，则标签可以是设备剩余使用寿命，在实际应用中，第*i*个标签数据表示为 $[x_{i,1}, x_{i,2}, \dots, x_{i,j-1}, x_{i,j}, y_i]$ ，这里， $x_{i,j}$ 表示第*i*个标签

数据中的第 j 个特征对应的特征值， y_i 表征第 i 个标签数据对应的标签，其他数据项表示含义类似，这里不做过多赘述。

具体地，标签数据的权重指的是标签数据相对于非标签数据的重要程度，从而将非标签数据迁移到标签数据上。

5 在一些可行的实施例中，步骤 102 包括：

根据所述至少两个聚类中心点和所述多个非标签数据，确定所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度；

10 根据所述至少两个聚类中心点、所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度以及联合用户的多个标签数据，确定所述至少两个聚类中心点各自对应的相似度权重；

根据所述至少两个聚类中心点各自对应的相似度权重，确定所述联合用户的多个标签数据各自对应的权重。

本实施例中，通过确定聚类中心点与多个非标签数据之间的目标相似度，并基于聚类中心点和聚类中心点与多个非标签数据之间的目标相似度，确定聚类中心点对应的相似度权重，
15 基于所有聚类中心点各自对应的相似度权重，确定联合用户的多个标签数据各自对应的权重，并不涉及到非标签数据和标签数据之间的交互，确保了数据安全，同时，聚类中心点和目标相似度能够表征多个非标签数据之间的关系，从而确保多个标签数据各自对应的权重的参考价值，且得到的权重综合考虑到聚类中心点对应的相似度权重，具有相对较高的准确性。其中，聚类中心点对应的相似度权重指示了聚类中心点和联合用户的标签数据之间的相似度的
20 重要程度。

可选地，针对每个聚类中心点，基于多个非标签数据各自与聚类中心点之间的相似度，然后对多个非标签数据各自与聚类中心点之间的相似度进行平均，得到聚类中心点与多个非标签数据之间的目标相似度，即所述目标相似度基于对所述多个标签数据各自与所述聚类中心点之间的目标相似度进行平均得到。具体地，可通过任何现有技术中的相似度计算方法
25 确定聚类中心点和非标签数据之间的相似度，比如，可以通过计算聚类中心点和非标签数据之间的距离，并将该距离确定为相似度，也可以通过核函数，计算聚类中心点和标签数据之间的核函数值，并将核函数值确定为相似度，换言之，非标签数据与聚类中心点之间的相似度可以通过核函数计算，其中，核函数可以是现有技术中的任意一种核函数，比如，多项式核函数、线性核函数、径向基核函数、指数核函数，优选径向基核函数中的高斯核函数。具体
30 可通过如下第一公式计算聚类中心点与多个非标签数据之间的目标相似度；其中，所述第一公式包括：

$$\hat{h}_l = \frac{1}{n} \sum_{i=1}^n K(x_i, x_l)$$

其中, $\hat{h}_l$ 表征第 l 个聚类中心点与多个非标签数据的目标相似度; n 表征多个非标签数据的数据个数; x_i 表征第 i 个非标签数据; x_l 表征第 l 个聚类中心点; $K(\cdot)$ 表征核函数。应当理解的, 基于核函数 $K(\cdot)$ 所计算出来的核函数值理解为聚类中心点和非标签数据之间的相似度。优选高斯核函数。

- 5 可选地, 针对所述联合用户的每个所述标签数据, 根据所述至少两个聚类中心点各自对应的相似度权重, 对所述至少两个聚类中心点各自与所述标签数据之间的相似度进行加权求和, 以确定所述标签数据对应的权重。具体可通过如下第六公式计算标签数据的权重; 其中, 所述第六公式如下:

$$\hat{r}(x_j)_\alpha = \sum_{l=1}^k \hat{\theta}_l K(x_j, x_l)$$

- 10 其中, $\hat{r}(x_j)_\alpha$ 表征第 j 个标签数据的权重; x_j 表征第 j 个标签数据; x_l 表征第 l 个聚类中心点; $\hat{\theta}_l$ 表征第 l 个聚类中心点对应的相似度权重; k 表征聚类中心点的个数; $K(\cdot)$ 表征核函数。这里, k 个聚类中心点各自对应的相似度权重之和等于 1。

具体可通过如下两种实现方式实现根据所述至少两个聚类中心点、所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度以及联合用户的多个标签数据, 确定所述至少两个聚类中心点各自对应的相似度权重。

- 15 实现方式 1: 根据所述至少两个聚类中心点以及联合用户的多个标签数据, 确定所述至少两个聚类中心点各自与所述多个标签数据之间的参考相似度; 针对每个所述聚类中心点, 计算所述聚类中心点与所述多个非标签数据之间的目标相似度, 以及所述聚类中心点与所述多个标签数据之间的参考相似度的比值, 并将该比值确定为所述聚类中心点对应的相似度权重。需要说明的是, 目标相似度和参考相似度的计算方法相同, 区别仅仅在于目标相似度是
20 针对目标用户的非标签数据, 参考相似度是针对联合用户的标签数据。

- 实现方式 2: 根据所述至少两个聚类中心点和所述多个非标签数据, 确定所述至少两个聚类中心点中任意两个聚类中心点之间的初始相关性; 根据所述任意两个聚类中心点以及联合用户的多个标签数据, 确定所述任意两个聚类中心点之间的参考相关性; 根据所述任意两个聚类中心点之间的初始相关性以及参考相关性, 确定所述任意两个聚类中心点之间的目标相关性; 根据所述任意两个聚类中心点之间的目标相关性和所述至少两个聚类中心点各自与
25 所述多个非标签数据之间的目标相似度, 确定所述至少两个聚类中心点各自对应的相似度权重。

在实现方式 2 中,通过任意两个聚类中心对应应在联合用户的多个标签数据上的参考相关性,以及任意两个聚类中心对应应在目标用户的多个非标签数据上的初始相关性,确定出任意两个聚类中心对应的目标相关性,目标相关性用于表征目标用户和联合用户之间的数据相关程度,之后,基于任意两个聚类中心点之间的目标相关性和所有聚类中心点各自与联合用户的多个非标签数据之间的目标相似度,确定出所有的聚类中心点各自对应的相似度权重。可以理解的,得到的相似度权重综合考虑到聚类中心点,聚类中心点与多个非标签数据之间的目标相似度,任意两个聚类中心点之间的初始相关性以及参考相关性,具有相对较高的准确性。其中,两个聚类中心点之间的初始相关性指示了两个聚类中心点对应在目标用户的多个非标签数据上的相关程度,初始相关性越大,则说明两个聚类中心点对应在非标签数据上的相关性越大。两个聚类中心点之间的参考相关性指示了两个聚类中心点对应在联合用户的多个标签数据上的相关程度。

在实现方式 2 中,可选地,所述初始相关性基于目标概率分布权重,对所述多个非标签数据各自对应的目标相似度乘积值的平均值进行修正得到,所述目标相似度乘积值基于对所述任意两个聚类中心点各自与所述非标签数据之间的目标相似度进行相乘得到。在实际应用中,针对任意两个聚类中心点,计算任意两个聚类中心点各自与同一非标签数据之间的目标相似度,并对任意两个聚类中心点各自与同一非标签数据之间的目标相似度进行相乘,得到目标相似度乘积值,之后,得到多个非标签数据各自对应的目标相似度乘积值,对多个非标签数据各自对应的目标相似度乘积值进行平均,得到平均结果,基于目标概率分布权重,对平均结果进行修正,得到该聚类中心对应的初始相关性。具体可通过如下第二公式计算所述任意两个聚类中心点之间的初始相关性;其中,所述第二公式包括:

$$\hat{H}_{l,l'}^a = \frac{\alpha}{n} \sum_{i=1}^n K(x_i, x_l) K(x_i, x_{l'})$$

其中, $\hat{H}_{l,l'}^a$ 表征第 l 个聚类中心点和第 l' 个聚类中心点之间的初始相关性; n 表征各个所述非标签数据的数据个数; x_i 表征第 i 个所述非标签数据; x_l 表征第 l 个聚类中心点; $x_{l'}$ 表征第 l' 个聚类中心点; α 表征目标概率分布权重; $K(\cdot)$ 表征核函数。

对应的,所述参考相关性基于参考概率分布权重,对所述多个标签数据各自对应的参考相似度乘积值的平均值进行修正得到,所述参考相似度乘积值基于对所述任意两个聚类中心点各自与所述标签数据之间的参考相似度进行相乘得到。在实际应用中,通过如下第三公式计算所述任意两个聚类中心点之间的参考相关性;其中,所述第三公式包括:

$$\hat{H}_{l,l'}^b = \frac{1-\alpha}{m} \sum_{j=1}^m K(x_j, x_l) K(x_j, x_{l'})$$

其中， $\hat{H}_{l,l'}^b$ 表征第 l 个聚类中心点和第 l' 个聚类中心点之间的参考相关性； x_j 表征所述联合用户的第 j 个所述标签数据； m 表征所述联合用户的各个所述标签数据的数据个数； $1-\alpha$ 表征参考概率分布权重， α 表征目标概率分布权重。

应当理解的，目标概率分布权重指示了目标用户的多个非标签数据的概率分布的重要性，
5 作为一种可能的实现方式，可根据实际需求人为设定。作为另一种可能的情况，具体可通过如下方式确定目标概率分布权重：

获取所述多个非标签数据对应的多个验证数据以及预设概率分布权重；

根据所述预设概率分布权重以及所述多个验证数据，确定所述多个验证数据各自对应的验证权重；

10 根据所述多个验证数据各自对应的权重标签和所述多个验证数据各自对应的验证权重，确定所述预设概率分布权重对应的误差数据；

根据各个所述预设概率分布权重对应的误差数据，确定所述目标概率分布权重。

本实施例中，采用确定联合用户的多个标签数据各自对应的权重的相同方法，通过预设概率分布权重，确定多个验证数据各自对应的验证权重，然后，判断多个验证数据各自对应的
15 的权重标签和多个验证数据各自对应的验证权重，确定预设概率分布权重对应的误差数据，基于误差数据判断预设概率分布权重的准确性，将准确性最高的预设概率分布权重确定为目标概率分布权重，从而确保基于目标概率分布权重确定出的联合用户的多个标签数据各自对应的权重的准确性。这里，多个验证数据可以是目标用户的业务任务的其他非标签数据，也可以是联合用户的业务任务的多个标签数据，具体需要结合实际情况确定。误差数据可以是
20 多个验证数据各自对应的权重标签和验证权重之间的差值的标准差、方差等用于评价误差的参数，此处不做具体限定。应当理解的，确定验证数据的验证权重与确定联合用户的标签数据的权重的方法是相同的。

在实现方式2中，作为一种可能的情况，通过如下方式实现根据所述任意两个聚类中心点之间的目标相关性和所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相
25 似度，确定所述至少两个聚类中心点各自对应的相似度权重：

根据所述任意两个聚类中心点之间的目标相关性，确定所述至少两个聚类中心点对应的目标相关性矩阵；根据所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度，确定目标相似度向量；根据正则化参数以及单位矩阵，对所述目标相关性矩阵进行修正，

以确定修正相关性矩阵；根据所述修正相关性矩阵和所述目标相似度向量，确定相似度权重向量，所述相似度权重向量包括所述至少两个聚类中心点各自对应的相似度权重。

可以理解的，为了防止过拟合，采用正则化参数和单位矩阵对相关性矩阵进行了修正，确定了修正相关性矩阵，然后根据修正相关性矩阵和相似度向量，确定相似度权重向量，从而得到各个聚类中心点各自对应的相似度权重。

具体地，对正则化参数和相似度向量相乘后得到的结果，与相关性矩阵相加，得到修正相关性矩阵，然后修正相关性矩阵的倒数和相似度向量相乘得到相似度权重向量。在实际应用中，通过如下第四公式计算所述修正相关性矩阵；其中，所述第四公式包括：

$$\hat{\mathbf{H}}' = \hat{\mathbf{H}} + \lambda \mathbf{I}_n$$

其中， $\hat{\mathbf{H}}'$ 表征所述修正相关性矩阵； $\hat{\mathbf{H}}$ 表征所述相关性矩阵； λ 表征所述正则化参数； $\mathbf{I}_n$ 表征所述单位矩阵。

通过如下第五公式计算相似度权重向量；其中，所述第五公式包括：

$$\hat{\theta} = \hat{\mathbf{H}}'^{-1} \hat{h}$$

其中， $\hat{\theta}$ 表征相似度权重向量； $\hat{\mathbf{H}}'$ 表征所述修正相关性矩阵； $\hat{h}$ 表征所述目标相似度向量。

需要说明的是，目标相关性矩阵中的聚类中心点的编号和目标相似度向量中的聚类中心点的编号相同，应当理解的，聚类中心点的编号大小指示了聚类中心点的顺序。

具体地，目标相关性矩阵中的矩阵元素综合考虑两个聚类中心点之间的初始相关性以及参考相关性，确保了相关性矩阵的参考价值。具体可通过如下2种实现方式确定目标相关性。

实现方式1，所述目标相关性基于对所述任意两个聚类中心点之间的所述初始相关性以及所述参考相关性相加得到。具体地，对两个或多个聚类中心点进行编号，构建二维矩阵，将任意两个聚类中心点之间的初始相关性和参考相关性放入二维矩阵中成为元素，计算任意两个聚类中心点之间的初始相关性和参考相关性之和，以得到目标相关性矩阵。应当理解的，不同联合用户各自计算任意两个聚类中心点的目标相关性。

应当理解的，本实施例中的核心思路是计算目标用户的概率分布 $p(x)$ 和每个联合用户的概率分布 $q(x)$ 的概率分布比率 $w(x)$ ，从而为联合用户的标签数据设定权重，多个联合用户的计算过程相同，这里以一个联合用户为例进行说明，假设多个非标签数据表示为 $\{x_i\}_{i=1}^n$ ，其中， n 表示非标签数据的数据个数，联合用户的多个标签数据表示为 $\{x_j\}_{j=1}^m$ ，其中， m 表示标签数据的数据个数。

令 $w(x) = \hat{f}(x)_\alpha$ ， $\hat{f}(x)_\alpha = p(x)/(\alpha p(x) + (1 - \alpha)q(x))$ ，基于数据和几个聚类点之间的

相似度的线性组合的思路构建回归模型，另 $\hat{f}(x)_\alpha = \sum_{l=1}^k \hat{\theta}_l K(x, x_l)$ ，其中， $K(x, x_l)$ 表征核函数，然后最小化损失函数 $\hat{\theta} = \operatorname{argmin}[\frac{1}{2}\theta^T \hat{H}\theta - \hat{h}^T \theta + \frac{\lambda}{2}\theta^T \theta]$ ，其中， θ 具有解析解， $\hat{\theta} = (\hat{H} + \lambda I_n)^{-1} \hat{h}$ ，其中， λ 表征正则化参数； I_n 表征单位矩阵， $\hat{h}$ 表征向量； $\hat{H}$ 表征矩阵。 $\hat{H}$ 中的每个元素表示如下：

$$\hat{H}_{l,l'} = \frac{\alpha}{n} \sum_{i=1}^n K(x_i, x_l) K(x_i, x_{l'}) + \frac{1-\alpha}{m} \sum_{j=1}^m K(x_j, x_l) K(x_j, x_{l'})$$

5 其中， $\hat{H}_{l,l'}$ 表征第 l 行和第 l' 列相交位置的矩阵元素； n 表征各个所述非标签数据的数据个数； x_i 表征第 i 个所述非标签数据； x_l 表征第 l 个聚类簇的聚类中心点； $x_{l'}$ 表征第 l' 个聚类簇的聚类中心点； α 表征目标概率分布权重； $K(\cdot)$ 表征核函数； x_j 表征所述联合用户的第 j 个所述标签数据； m 表征所述联合用户的各个所述标签数据的数据个数； $1-\alpha$ 表征参考概率分布权重。

10 $\hat{h}$ 中的向量元素表示如下：

$$\hat{h}_l = \frac{1}{n} \sum_{i=1}^n K(x_i, x_l)$$

其中， $\hat{h}_l$ 表征第 l 个元素； n 表征各个所述非标签数据的数据个数； x_i 表征第 i 个所述非标签数据； x_l 表征第 l 个聚类簇的聚类中心点； $K(\cdot)$ 表征核函数。

在实现方式1中，进一步的，还包括：

15 根据所述多个非标签数据、所述至少两个聚类中心点以及所述至少两个聚类中心点各自对应的相似度权重，确定所述目标用户和所述联合用户之间的数据分布相似度；

根据各个所述联合用户各自与所述目标用户之间的数据分布相似度，确定各个所述联合用户各自的重要性；

根据各个所述联合用户各自的重要性，调整所述联合学习模型。

20 具体地，通过如下第七公式计算目标用户和联合用户之间的数据分布相似度；其中，第七公式如下：

$$\hat{L}_s = \frac{1}{n} \sum_{i=1}^n \log[\hat{\theta}_{l,s} K(x_i, x_l)]$$

其中， $\hat{L}_s$ 表征目标用户和第 s 个联合用户的数据分布相似度， $\hat{\theta}_{l,s}$ 表示第 s 个联合用户的第 l 个聚类中心点的相似度权重。

具体地，通过如下第八公式计算联合用户的重要性：

$$Score_s = \frac{e^{-|\hat{L}_s|}}{\sum_{s=1}^N e^{-|\hat{L}_s|}}$$

其中, $Score_s$ 表征第 s 个联合用户; N 表征联合用户的用户数量。

在实际应用中, 获取目标用户将采用联合学习模型进行预测的预测值和预测值对应的真实值, 当预测值和真实值之间的误差较大时, 比如, 大于预设阈值时, 此时, 可以基于目标用户和联合用户之间的数据分布相似度, 确定联合用户的重要性, 将重要性较低的联合用户
5 进行删除, 保留重要性较高的用户, 并通过重要性较高的联合用户进行联合学习, 对联合学习模型进行修正, 得到准确性更高的联合学习模型。还可以基于联合用户的重要性, 对联合用户进行奖励, 使得重要性较高的联合用户提供更多的标签数据, 从而对联合学习模型进行修正, 得到准确性更高的联合学习模型。

实现方式 2, 基于目标用户的初始相关性和不同联合用户各自的参考相关性进行融合,
10 得到共享的目标相关性, 换言之, 不同联合用户共享任意两个聚类中心点之间的目标相关性。换言之, 各个所述联合用户共享所述任意两个聚类中心点之间的目标相关性; 所述目标相关性基于所述任意两个聚类中心点之间的所述初始相关性以及各个所述联合用户各自的所述任意两个聚类中心点之间的参考相关性确定。

具体地, 针对所有聚类中心点中的任意两个聚类中心点, 目标相关性可以是各个联合用户
15 各自的任意两个聚类中心点之间的参考相关性的均值, 与任意两个聚类中心点之间的初始相关性之和, 本实施例对如何得到目标相关性不做具体限定, 任何基于任意两个聚类中心点之间的初始相关性以及各个联合用户各自的任意两个聚类中心点之间的参考相关性确定的相关性即可。

步骤 103、根据各个所述联合用户各自的所述多个标签数据以及所述多个标签数据各自
20 对应的权重, 构建联合学习模型, 所述联合学习模型用于执行所述目标用户的业务任务。

具体地, 针对每个联合用户, 根据联合用户的多个标签数据及其各自对应的权重进行初始模型的训练, 得到联合用户的局部模型, 之后, 将各个联合用户各自的局部模型发送到目标用户, 目标用户对各个联合用户各自的局部模型聚合后得到更新后的模型, 将更新后的模型再下发到每个联合用户作为初始化模型进行训练, 如此迭代直至模型收敛, 最后得到联合
25 学习模型。得到的联合学习模型用于执行业务任务, 比如, 当业务任务是故障类型预测时, 联合学习模型用于预测目标用户的故障类型。

应当理解的, 联合用户的多个标签数据各自对应的权重用于调整模型中的模型参数, 从而使得调整后的模型能够反映出目标用户的非标签数据和业务任务之间的联系, 确保联合学习模型的模型精度。在实际应用中, 具体可通过如下实现方式确定联合用户的局部模型:

A1、根据将标签数据中的多个特征数据分别代入初始模型中的预测结果以及标签数据中的多个特征数据各自对应的标签，确定标签数据对应的第一误差，对多个标签数据各自对应的第一误差以及权重进行相乘计算，以确定多个标签数据各自对应的第二误差；

- A2、判断是否满足迭代次数或者多个标签数据各自对应的第二误差是否满足预设条件，
5 如果是，则将初始模型确定为局部模型，如果否，则执行 A3；

A3、根据多个标签数据各自对应的第二误差，对初始模型中的模型参数进行调整，以确定调整后的模型参数，并将初始模型中的模型参数替换为调整后的模型参数，执行 A1。

- 需要说明的是，各个联合用户各自的多个标签数据分布在物联网中的不同的节点，共享数据会产生数据安全问题，通过节点中的非共享数据以及非共享数据的权重进行联合学习，
10 进而得到节点的局部模型，实现将非共享数据迁移到目标用户上，使得节点之间不存在数据共享，避免了直接共享数据带来的数据安全问题。其中，节点能进行数据处理以及数据交互，包括但不限于边缘服务器、边缘网关以及边缘控制器中的任意一种或多种。目标用户和联合用户之间的数据交互仅仅涉及到目标相似度、初始相关性以及聚类中心点，并不涉及到非标签数据的交互。

- 15 作为一种可能的情况，联合用户和目标用户之间的数据分布相似度不小于预设阈值。这里，数据分布相似度可以基于上述第七公式进行计算。

- 通过以上技术方案可知，本实施例存在的有益效果是：对目标用户的业务任务对应的多个非标签数据进行聚类，确定聚类中心点，并确定聚类中心点与多个非标签数据，确定联合用户的标签数据的权重，将非标签数据迁移到了标签数据上，实现了数据迁移，确保了数据
20 量，之后，根据联合用户的多个标签数据以及多个标签数据各自对应的权重，构建联合学习模型，联合学习模型用于执行目标用户的业务任务，在目标用户缺少标签的前提下，能够完成目标用户的业务任务。

图 1 所示仅为本发明所述方法的基础实施例，在其基础上进行一定的优化和拓展，还能够得到所述方法的其他优选实施例。

- 25 如图 2 所示，为本发明所述业务任务执行方法的另一个具体实施例。本实施例在前述实施例的基础上，结合应用场景进行了更加具体的描述。

本实施例所结合的具体场景为：目标用户的多个非标签数据表示为 $\{x_i\}_{i=1}^n$ ，其中， n 表示非标签数据的数据个数，联合用户的多个标签数据表示为 $\{x_j\}_{j=1}^m$ ，其中， m 表示标签数据的数据个数。多个联合用户的计算过程相同，这里仅以一个联合用户为例进行说明。

- 30 所述方法具体包括以下步骤：

步骤 201、对目标用户的业务任务对应的多个非标签数据进行聚类，以确定至少两个聚类中心点。

目标用户通过 K-means 聚类算法对多个非标签数据进行聚类，得到 k 个聚类簇以及每个聚类簇的聚类中心点，每个聚类中心点均与非标签数据不同，确保数据安全与隐私。

- 5 步骤 202、根据所述至少两个聚类中心点以及所述多个非标签数据，确定所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度以及所述至少两个聚类中心点中的任意两个聚类中心点之间的初始相关性。

目标用户通过上述第一公式 $\hat{h}_l = \frac{1}{n} \sum_{i=1}^n K(x_i, x_l)$ 计算聚类中心点与多个非标签数据之间的目标相似度，得到 k 个聚类中心点各自对应的目标相似度，k 个目标相似度分别表示为

- 10 $\hat{h}_1$ 、 $\hat{h}_2$ 、...、 $\hat{h}_l$ 、...、 $\hat{h}_k$ 。其中， $K(\cdot)$ 为高斯核函数。

目标用户通过上述第二公式 $\hat{H}_{l,l'}^a = \frac{\alpha}{n} \sum_{i=1}^n K(x_i, x_l) K(x_i, x_{l'})$ 计算任意两个聚类中心点之间的初始相关性，得到 k^2 个初始相关性，通过如下表格 1 表示：

	1	2	...	l'	...	k
1	$\hat{H}_{1,1}^a$	$\hat{H}_{1,2}^a$	...	$\hat{H}_{1,l'}^a$	...	$\hat{H}_{1,k}^a$
2	$\hat{H}_{2,1}^a$	$\hat{H}_{2,2}^a$	...	$\hat{H}_{2,l'}^a$	...	$\hat{H}_{2,k}^a$
...	...	...	...	...	...	...
l	$\hat{H}_{l,1}^a$	$\hat{H}_{l,2}^a$	...	$\hat{H}_{l,l'}^a$	...	$\hat{H}_{l,k}^a$
...	...	...	...	...	...	...
k	$\hat{H}_{k,1}^a$	$\hat{H}_{k,2}^a$	...	$\hat{H}_{k,l'}^a$	...	$\hat{H}_{k,k}^a$

表 1

- 15 步骤 203、根据所述任意两个聚类中心点以及联合用户的多个标签数据，确定所述任意两个聚类中心点之间的参考相关性；根据所述任意两个聚类中心点之间的初始相关性以及参考相关性，确定所述任意两个聚类中心点之间的目标相关性。

目标用户将 k 个聚类中心点各自对应的目标相似度和表 1 中 k^2 个初始相关性发给联合用户，联合用户通过上述第三公式 $\hat{H}_{l,l'}^b = \frac{1-\alpha}{m} \sum_{j=1}^m K(x_j, x_l) K(x_j, x_{l'})$ 计算任意两个聚类中心点之间的参考相关性，得到 k^2 个参考相关性，通过如下表格 2 表示：

	1	2	...	l'	...	k
--	---	---	-----	------	-----	-----

1	$\hat{H}_{1,1}^b$	$\hat{H}_{1,2}^b$	...	$\hat{H}_{1,l'}^b$	...	$\hat{H}_{1,k}^b$
2	$\hat{H}_{2,1}^b$	$\hat{H}_{2,2}^b$	...	$\hat{H}_{2,l'}^b$	...	$\hat{H}_{2,k}^b$
...	...	...	...	...	...	...
l	$\hat{H}_{l,1}^b$	$\hat{H}_{l,2}^b$	...	$\hat{H}_{l,l'}^b$	...	$\hat{H}_{l,k}^b$
...	...	...	...	...	...	...
k	$\hat{H}_{k,1}^b$	$\hat{H}_{k,2}^b$	...	$\hat{H}_{k,l'}^b$	...	$\hat{H}_{k,k}^b$

表 2

作为一种可能的情况，每个联合用户各自计算任意两个聚类中心点的目标相关性，针对每个联合用户，任意两个聚类中心点的目标相关性为任意两个聚类中心点的初始相关性以及参考相关性之和，通过如下表格 3 表示 k^2 个目标相关性：

	1	2	...	l'	...	k
1	$\hat{H}_{1,1}^a + \hat{H}_{1,1}^b$	$\hat{H}_{1,2}^a + \hat{H}_{1,2}^b$	...	$\hat{H}_{1,l'}^a + \hat{H}_{1,l'}^b$	...	$\hat{H}_{1,k}^a + \hat{H}_{1,k}^b$
2	$\hat{H}_{2,1}^a + \hat{H}_{2,1}^b$	$\hat{H}_{2,2}^a + \hat{H}_{2,2}^b$	...	$\hat{H}_{2,l'}^a + \hat{H}_{2,l'}^b$	...	$\hat{H}_{2,k}^a + \hat{H}_{2,k}^b$
...	...	...	...	...	...	...
l	$\hat{H}_{l,1}^a + \hat{H}_{l,1}^b$	$\hat{H}_{l,2}^a + \hat{H}_{l,2}^b$	...	$\hat{H}_{l,l'}^a + \hat{H}_{l,l'}^b$	...	$\hat{H}_{l,k}^a + \hat{H}_{l,k}^b$
...	...	...	...	...	...	...
k	$\hat{H}_{k,1}^a + \hat{H}_{k,1}^b$	$\hat{H}_{k,2}^a + \hat{H}_{k,2}^b$	...	$\hat{H}_{k,l'}^a + \hat{H}_{k,l'}^b$	...	$\hat{H}_{k,k}^a + \hat{H}_{k,k}^b$

表 3

5

作为一种可能的情况，每个联合用户共享任意两个聚类中心点之间的目标相关性。针对任意两个聚类中心点，任意两个聚类中心点之间的目标相关性为对所有联合用户各自的任意两个聚类中心点之间的参考相关性平均后，加上任意两个聚类中心点之间的初始相关性的求和结果。比如，联合用户有 N 个，第 i 个联合用户的任意两个聚类中心点之间的参考相关性表示为 $\hat{H}_{l,l'}^{ib}$ ，则任意两个聚类中心点之间的目标相关性为 $\hat{H}_{l,l'}^a + \frac{1}{N} \sum_{i=1}^N \hat{H}_{l,l'}^{ib}$ 。

10

步骤 204、根据所述任意两个聚类中心点之间的目标相关性，确定所述至少两个聚类中心点对应的目标相关性矩阵；根据所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度，确定目标相似度向量；根据正则化参数以及单位矩阵，对所述目标相关性矩阵进行修正，以确定修正相关性矩阵。

通过上述第四公式 $\hat{\mathbf{H}}' = \hat{\mathbf{H}} + \lambda I_n$ 计算修正相关性矩阵。

步骤 205、根据所述修正相关性矩阵和所述目标相似度向量，确定相似度权重向量，所述相似度权重向量包括所述至少两个聚类中心点各自对应的相似度权重。

通过上述第五公式 $\hat{\theta} = \hat{\mathbf{H}}'^{-1} \hat{h}$ 计算相似度权重向量。

5 步骤 206、针对所述联合用户的每个所述标签数据，根据所述至少两个聚类中心点各自对应的相似度权重，对所述至少两个聚类中心点各自与所述标签数据之间的相似性进行加权求和，以确定所述标签数据对应的权重。

通过上述第六公式 $\hat{r}(x_j)_\alpha = \sum_{i=1}^k \hat{\theta}_i K(x_j, x_i)$ 计算每个标签数据对应的权重。

10 步骤 207、根据所述多个非标签数据、所述至少两个聚类中心点以及所述至少两个聚类中心点各自对应的相似度权重，确定所述目标用户和所述联合用户之间的数据分布相似度。

通过上述第七公式 $\hat{L}_s = \frac{1}{n} \sum_{i=1}^n \log[\hat{\theta}_{i,s} K(x_i, x_i)]$ 计算联合用户和目标用户之间的数据分布相似度。

15 步骤 208、将满足联合学习条件的数据分布相似度对应的联合用户作为目标联合用户，根据所述目标联合用户的所述多个标签数据以及所述多个标签数据各自对应的权重，构建联合学习模型。

通过以上技术方案可知，本实施例存在的有益效果是：对目标用户的业务任务对应的多个非标签数据进行聚类，确定聚类中心点，并确定聚类中心点与多个非标签数据之间的目标相似度以及任意两个聚类中心点之间的初始相关性，从而得到非标签数据的描述信息，确保数据隐私和安全，根据聚类中心点、聚类中心点与多个非标签数据之间的目标相似度、任意两个聚类中心点之间的初始相关性、任意两个聚类中心点之间的参考相关性，确定所有聚类中心点各自对应的相似度权重，根据所有聚类中心点各自对应的相似度权重，对标签数据与所有的聚类中心点之间的相似度进行加权，确定标签数据对应的权重，将非标签数据迁移到了标签数据上，实现了数据迁移，确保了数据量，之后，基于联合用户和目标用户之间的数据分布相似度，对联合用户进行选择，基于数据分布相似度较高的联合用户的多个标签数据以及多个标签数据各自对应的权重，构建联合学习模型，联合学习模型用于执行目标用户的业务任务，在目标用户缺少标签的前提下，能够完成目标用户的业务任务，同时确保模型精度。

基于与本发明方法实施例相同的构思，请参考图 3，本发明实施例还提供了一种业务任务执行装置，包括：

30 聚类模块 301，用于对目标用户的业务任务对应的多个非标签数据进行聚类，以确定至

少两个聚类中心点；

权重确定模块 302，用于根据所述至少两个聚类中心点以及所述多个非标签数据，确定联合用户的多个标签数据各自对应的权重，所述多个标签数据和所述业务任务对应；

构建模块 303，用于根据所述联合用户的所述多个标签数据以及所述多个标签数据各自
5 对应的权重，构建联合学习模型，所述联合学习模型用于执行所述目标用户的业务任务。

本发明一个实施例中，所述权重确定模块 302，包括：相似度确定单元、第一权重确定单元以及第二权重确定单元；其中，

所述相似度确定单元，用于根据所述至少两个聚类中心点和所述多个非标签数据，确定所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度；

10 所述第一权重确定单元，用于根据所述至少两个聚类中心点、所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度以及联合用户的多个标签数据，确定所述至少两个聚类中心点各自对应的相似度权重；

所述第二权重确定单元，用于根据所述至少两个聚类中心点各自对应的相似度权重，确定所述联合用户的多个标签数据各自对应的权重。

15 本发明一个实施例中，还包括：相关性确定模块；

所述相关性确定模块，用于根据所述至少两个聚类中心点中任意两个聚类中心点以及所述多个非标签数据，确定所述任意两个聚类中心点之间的初始相关性；

所述第一权重确定单元，包括：第一相关性确定子单元、第二相关性确定子单元以及第一权重确定子单元；其中，

20 所述第一相关性确定子单元，用于根据所述任意两个聚类中心点以及联合用户的多个标签数据，确定所述任意两个聚类中心点之间的参考相关性；

所述第二相关性确定子单元，用于根据所述任意两个聚类中心点之间的初始相关性以及参考相关性，确定所述任意两个聚类中心点之间的目标相关性；

25 所述第一权重确定子单元，用于根据所述任意两个聚类中心点之间的目标相关性和所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度，确定所述至少两个聚类中心点各自对应的相似度权重。

在一个实施例中，所述第二权重确定单元，包括：第二权重确定子单元；其中，

所述第二权重确定子单元，用于针对所述联合用户的每个所述标签数据，根据所述至少两个聚类中心点各自对应的相似度权重，对所述至少两个聚类中心点各自与所述标签数据之
30 间的相似度进行加权求和，以确定所述标签数据对应的权重。

在一个实施例中，还包括：相似度计算模块、重要性计算模块以及调整模块；其中，

所述相似度计算模块，用于根据所述多个非标签数据、所述至少两个聚类中心点以及所述至少两个聚类中心点各自对应的相似度权重，确定所述目标用户和所述联合用户之间的数据分布相似度；

- 所述重要性计算模块，用于根据各个所述联合用户各自与所述目标用户之间的数据分布相似度，确定各个所述联合用户各自的重要性；

所述调整模块，用于根据各个所述联合用户各自的重要性，调整所述联合学习模型。

在一个实施例中，所述第一权重确定子单元，用于执行如下步骤：

根据所述任意两个聚类中心点之间的目标相关性，确定所述至少两个聚类中心点对应的目标相关性矩阵；

- 10 根据所述至少两个聚类中心点各自与所述多个非标签数据之间的目标相似度，确定目标相似度向量；

根据正则化参数以及单位矩阵，对所述目标相关性矩阵进行修正，以确定修正相关性矩阵；

- 15 根据所述修正相关性矩阵和所述目标相似度向量，确定相似度权重向量，所述相似度权重向量包括所述至少两个聚类中心点各自对应的相似度权重。

在一个实施例中，所述修正相关性矩阵基于对所述目标相关性矩阵以及所述正则化参数与所述单位矩阵相乘的结果进行求和得到；

所述相似度权重向量基于对所述修正相关性矩阵的倒数与所述相似度向量相乘得到；

- 20 所述目标相关性基于对所述任意两个聚类中心点之间的所述初始相关性以及所述参考相关性相加得到；

所述目标相似度基于对所述多个标签数据各自与所述聚类中心点之间的目标相似度进行平均得到；

- 25 所述初始相关性基于目标概率分布权重，对所述多个非标签数据各自对应的目标相似度乘积值的平均值进行修正得到，所述目标相似度乘积值基于对所述任意两个聚类中心点各自与所述非标签数据之间的目标相似度进行相乘得到；

所述参考相关性基于参考概率分布权重，对所述多个标签数据各自对应的参考相似度乘积值的平均值进行修正得到，所述参考相似度乘积值基于对所述任意两个聚类中心点各自与所述标签数据之间的参考相似度进行相乘得到；

- 30 其中，所述目标概率分布权重和所述参考概率分布权重之和等于 1，所述参考相似度和所述目标相似度基于相同的核函数计算得到。

在一个实施例中，各个所述联合用户共享所述任意两个聚类中心点之间的目标相关性；

所述目标相关性基于所述任意两个聚类中心点之间的所述初始相关性以及各个所述联合用户各自的所述任意两个聚类中心点之间的参考相关性确定。

在一个实施例中，所述聚类中心点和所述多个非标签数据中的任意一个非标签数据不同。

图 4 是本发明实施例提供的一种电子设备的结构示意图。在硬件层面，该电子设备包括
5 处理器 401 以及存储有执行指令的存储器 402，可选地还包括内部总线 403 及网络接口 404。
其中，存储器 402 可能包含内存 4021，例如高速随机存取存储器（Random-Access Memory，
RAM），也可能还包括非易失性存储器 4022（non-volatile memory），例如至少 1 个磁盘存
储器等；处理器 401、网络接口 404 和存储器 402 可以通过内部总线 403 相互连接，该内部
总线 403 可以是 ISA（Industry Standard Architecture，工业标准体系结构）总线、PCI（Peripheral
10 Component Interconnect，外设部件互连标准）总线或 EISA（Extended Industry Standard
Architecture，扩展工业标准结构）总线等；内部总线 403 可以分为地址总线、数据总线、控
制总线等，为便于表示，图 4 中仅用一个双向箭头表示，但并不表示仅有一根总线或一种类
型的总线。当然，该电子设备还可能包括其他业务所需要的硬件。当处理器 401 执行存储器
402 存储的执行指令时，处理器 401 执行本发明任意一个实施例中的方法，并至少用于执行
15 如图 1 或图 2 所示的方法。

在一种可能实现的方式中，处理器从非易失性存储器中读取对应的执行指令到内存中然
后运行，也可从其它设备上获取相应的执行指令，以在逻辑层面上形成一种业务任务执行装
置。处理器执行存储器所存放的执行指令，以通过执行的执行指令实现本发明任一实施例
中提供的一种业务任务执行方法。

20 处理器可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各
步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可
以是通用处理器，包括中央处理器（Central Processing Unit，CPU）、网络处理器（Network
Processor，NP）等；还可以是数字信号处理器（Digital Signal Processor，DSP）、专用集成
电路（Application Specific Integrated Circuit，ASIC）、现场可编程门阵列（Field-Programmable
25 Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。
可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微
处理器或者该处理器也可以是任何常规的处理器等。

本发明实施例还提供了一种计算机可读存储介质，包括执行指令，当电子设备的处理器
执行执行指令时，所述处理器执行本发明任意一个实施例中提供的方法。该电子设备具体可
30 以是如图 4 所示的电子设备；执行指令是一种业务任务执行装置所对应计算机程序。

本领域内的技术人员应明白，本发明的实施例可提供为方法或计算机程序产品。因此，

本发明可采用完全硬件实施例、完全软件实施例，或软件和硬件相结合的形式。

本发明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，对于装置实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

5 还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素
10 的过程、方法、商品或者设备中还存在另外的相同要素。

以上所述仅为本发明的实施例而已，并不用于限制本发明。对于本领域技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原理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权利要求范围之内。