说　明　书

一种多任务大模型的训练方法及装置

技术领域

本说明书一个或多个实施例涉及人工智能领域，尤其涉及一种多任务大模型的训练方法及装置。

背景技术

随着计算机硬件算力的提升以及大规模数据集的可用性的提高，拥有庞大规模参数、能力更强的大模型走进了人们的视野。大模型一般在大规模数据集上进行预训练（pre-training）之后，再根据下游的具体任务使用相应的小规模数据集进行微调（fine-tuning），以适应具体的任务。

如果想让大模型同时适用于多种任务类型的任务，则需要在多个任务类型的数据集上分别或同时对大模型进行微调。然而，不同任务类型的任务信号可能会相互干扰，产生“跷跷板效应”（seesaw phenomenon），即在强化大模型针对于某种任务类型的任务的能力时，大模型针对于其它任务类型的能力会降低。同时，由于下游任务的数据集与预训练时的数据集之间存在异质性，直接进行微调难以达到很好的效果，甚至可能会产生负向迁移（negative transfer），即微调后的大模型的效果还不如微调之前的大模型。因此，需要一种更好的方法，在多任务场景下对大模型进行微调，使得微调后的大模型能够同时适用于多种任务类型的任务。

发明内容

本说明书一个或多个实施例描述了一种多任务大模型的训练方法及装置，可以增强训练后的大模型的泛化效果，以同时适用于多种任务场景。

第一方面，提供了一种多任务大模型的训练方法，所述大模型包括已训练的目标网络层，以及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数；所述方法包括：

获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；

将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务

网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；

基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；

根据所述预测结果对应的损失，更新所述旁路任务网络。

在一种可能的实施方式中，当所述第一样本为图像数据时，所述任务类型至少包括：图像分类、目标检测、图像分割、图像描述；

当所述第一样本为文本数据时，所述任务类型至少包括：文本分类、命名实体识别、文本摘要、文本问答、文本情感识别。

在一种可能的实施方式中，所述第一样本包括图像数据以及文本数据；所述多个预设任务类型是与图像和文本之间的关联相关的任务。

在一种可能的实施方式中，所述大模型包括图像编码器，桥接网络，自然语言处理网络；所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间；所述目标网络层属于所述桥接网络，所述第一嵌入向量对应于所述图像数据。

在一种可能的实施方式中，所述大模型包括图像编码器，桥接网络，自然语言处理网络；所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间；所述目标网络层包括第一目标层和第二目标层，所述旁路任务网络包括作为第一目标层旁路的第一任务网络，和作为第二目标层旁路的第二任务网络；所述第一目标层属于所述桥接网络，所述第二目标层属于所述自然语言处理网络。

在一种可能的实施方式中，所述权重参数包括，第一参数矩阵和第二参数集，所述第一参数矩阵的行列数目分别对应预设任务类型的数目和通用适配器的数目，其中，任意位置的第一权重参数指示对应的通用适配器在对应任务类型下的权重；所述第二参数集至少包括，所述多个专用适配器分别对应的第二权重参数。

在一种可能的实施方式中，所述专用适配器包括多个子适配器，任一专用适配器对应的所述第二权重参数包括多个子适配器对应的多个子权重参数。

在一种可能的实施方式中，根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量，包括：

从所述第一参数矩阵中获取所述第一任务类型下各个通用适配器的第一权重参数；以

及从所述第二参数集中获取所述第一专用适配器对应的第二权重参数；

将各个通用适配器处理所述第一嵌入向量的各个第一结果，和所述第一专用适配器处理所述第一嵌入向量的第二结果进行加权求和，得到第二嵌入向量，其中，各个第一结果的权重因子根据所述第一权重参数确定，所述第二结果的权重因子根据所述第二权重参数确定。

在一种可能的实施方式中，所述各个第一结果的权重因子根据以下方法确定：

将所述第一参数矩阵中所述第一任务类型下各个通用适配器的第一权重参数分别输入到 Gumbel Sigmoid 函数中，将输出结果输入到 softmax 层中，得到所述各个第一结果的权重因子。

在一种可能的实施方式中，所述适配器至少包含以下之一：LoRA 适配器、AdaLoRA 适配器、$(IA)^3$ 适配器。

在一种可能的实施方式中，所述大模型为基于 Transformer 架构的模型；所述目标网络层为以下之一：Query 层、Key 层、Value 层、Output 层、MLP 层。

第二方面，提供了一种多任务大模型的训练装置，所述大模型包括已训练的目标网络层，以及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数；所述装置包括：

获取单元，配置为，获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；

嵌入向量处理单元，配置为，将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；

预测结果确定单元，配置为，基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；

旁路网络更新单元，配置为，根据所述预测结果对应的损失，更新所述旁路任务网络。

第三方面，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，令计算机执行第一方面的方法。

第四方面，提供了一种计算设备，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，实现第一方面的方法。

本说明书实施例提出的一种多任务大模型的训练方法及装置，使用旁路任务网络对大模型中需要微调的网络层进行微调，可以在不改变大模型的原始参数的基础上微调大模型，降低了微调大模型所需的运算规模。同时，将旁路任务网络层中的适配器显式地划分为适用于所有任务的通用适配器、以及适用于特定任务的专用适配器，然后分别训练这些适配器的参数以及对多个适配器的输出结果加权求和时的权重。通过上述方案，可以降低跷跷板效应和负向迁移的影响，增强了训练后的大模型的泛化效果，以同时适用于多种任务场景。

附图说明

为了更清楚地说明本说明书披露的多个实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书披露的多个实施例，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它的附图。

图 1 示出根据一个实施例的多任务大模型的训练方法的架构示意图；

图 2 示出根据一个实施例的多任务大模型的训练方法的流程图；

图 3 示出根据一个实施例的大模型架构示意图；

图 4 示出根据一个实施例的对自然语言网络中的网络层进行微调的示意图；

图 5 示出根据一个实施例的对桥接网络中的网络层进行微调的示意图；

图 6 示出根据一个实施例的同时对桥接网络和自然语言网络中的网络层进行微调的示意图；

图 7 示出根据一个实施例的多任务大模型的训练装置的示意性框图。

具体实施方式

下面结合附图，对本说明书提供的方案进行描述。

如前所述，在多任务场景下直接对大模型进行微调会产生跷跷板效应并可能出现负向迁移的结果，使得微调后的大模型的效果不如微调前的大模型。所以需要一种方案能够使得微调后的大模型能够同时适用于多种任务类型的任务。

目前，在人工智能领域存在着多种类型的机器学习任务。例如，对于文本内容，任务可以是对文本内容进行分类、命名实体识别、基于文本生成摘要、文本问答、文本情感识别等等；对于图像内容，任务可以是图像分类、目标检测、图像内容分割、图像描述等等。

同时，针对于同一个输入内容，也可以进行多种任务。例如，对于一张包含猫的图像，可以同时进行目标检测、图像内容分割以及图像描述：检测图像中猫的位置，将图像中包含猫的部分分割出来，以及生成图像描述：“一只橘色的猫正在睡觉”。

根据发明人的研究和观察，不同种类的任务之间并非完全独立的，多种任务类型之间存在着潜在的可以互相借鉴和迁移的知识，与此同时，每种任务也有区别于其他任务的、属于自身独有的内容。因此，可以根据上述特点，有效地将不同任务类型之间的共有部分和特有部分利用起来。

基于以上分析，在本说明书的实施例中，提出了一种适用于多任务的旁路微调方式，其取代于直接调整预训练好的网络层，在该网络层的旁路设置可训练的旁路任务网络，并将该旁路任务网络显式地划分为任务通用网络和任务专用网络。图 1 示出根据一个实施例的多任务大模型的训练方法的架构示意图。在图 1 的示例中，右侧为需要微调的大模型，其拥有多个网络层，其中包含待微调的目标网络层。在该目标网络层的旁路，设置旁路任务网络。从而，在前向传播的过程中，旁路任务网络与目标网络层同时接收来自大模型上一层网络的输入向量，分别对输入向量进行处理后，将各自的输出结果进行求和，得到本层输出的结果向量，并将输出向量作为大模型下一层网络的输入。在微调的过程中，保持目标网络层原本的权重参数不变（这一过程也可以被称为“冻结”），只对旁路任务网络中的权重参数进行调整，以降低微调时的计算规模。

旁路任务网络中包含一个或多个适配器（adapter），多个适配器之间彼此相互独立。适配器内部具有可以调整的参数，其接收输入的嵌入向量，处理后并输出相应的嵌入向量，适配器例如可以被实现为一个神经网络。旁路任务网络中的适配器被显式地划分为，适用于所有任务类型的通用适配器，以及适用于单一任务的专用适配器。其中，当任务类型的总数为 T 时，可以设置 A 个通用适配器，以及 T 个专用适配器。当针对嵌入向量执行任务类型为 t 的任务时，使用全部 A 个通用适配器，以及与任务类型 t 对应的专用适配器 t。具体的，将输入样本对应的嵌入向量（在图 1 中以小正方形表示）分别输入到 A 个通用适配器以及专用适配器 t 中分别进行处理，并将这 A+1 个适配器的输出进行加权求和后，作为旁路任务网络针对于该嵌入向量的结果向量进行输出。

在训练（微调）时，将基于上述流程得到的大模型的预测结果与输入样本的标签之间的损失反向传播，调整旁路任务网络中的各个适配器内部的参数，以及加权求和时各个适配器对应的权重。在推理时，则直接基于上述流程，得到关于输入样本的推理结果。

可以理解的是，由于目标网络层可以是大模型中的任意一层，所以本说明书实施例描

述的旁路任务网络可以用于对大模型的任意一层网络层进行微调。更进一步地，还可以选择大模型的多个目标网络层，在多个目标网络层的旁边分别设置各自的旁路任务网络，以同时对大模型的多个目标网络层进行微调，实现更好的微调效果。

以下结合具体的实施例，描述上述多任务大模型的训练方法的具体实施步骤。图 2 示出根据一个实施例的多任务大模型的训练方法的流程图，所述大模型包括已训练的目标网络层，以及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数；所述方法的执行主体可以为任何具有计算、处理能力的平台或服务器或设备集群等。如图 2 所示，所述方法至少包括：步骤 202，获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；步骤 204，将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；步骤 206，基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；步骤 208，根据所述预测结果对应的损失，更新所述旁路任务网络。

首先，先对上述的通用适配器、专用适配器以及权重参数进行描述。假设任务类型的总数为T，则可以设置A个通用适配器，记为$\{\phi_1, \phi_2, \ldots \phi_A\}$，以及$T$个专用适配器，记为$\{\phi^1, \phi^2, \ldots \phi^T\}$，各个专用适配器对应于各个任务类型。任一适配器$\phi$可以接收输入的嵌入向量$x$，并输出处理后的向量$\phi(x)$。

可以使用多种类型的适配器作为本说明书实施例中所使用的适配器。例如，可以使用 LoRA 适配器（Low-Rank Adaptation）、AdaLoRA 适配器（Adaptive Low-Rank Adaptation）、$(IA)^3$ 适配器（论文 Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning 中提出的适配器）等等，这里不做限定。

权重参数可以包括第一参数矩阵W_A和第二参数集，其中，第一参数矩阵W_A的行列数目分别对应预设任务类型的数目T和通用适配器的数目A，即为一个$T * A$的矩阵，其任意位置的第一权重参数w_i^t指示对应的通用适配器t在对应任务类型i下的权重。

在一个实施例中，第一参数矩阵W_A可以记为如公式（1）所示的形式：

$$W_A = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_A^1 \\ w_1^2 & w_2^2 & \dots & w_A^2 \\ \dots & \dots & \dots & \dots \\ w_1^T & w_2^T & \dots & w_A^T \end{bmatrix} \quad (1)$$

可以使用基于(0, 1)之间的均匀采样（uniform sampling）对第一参数矩阵W_A中的各个权重参数值进行初始化。

第二参数集至少包括，所述多个专用适配器分别对应的第二权重参数w^j。在一个实施例中，第二参数集可以实现为第二参数矩阵W_B，第二参数矩阵W_B为一对角方阵，其尺寸与预设任务类型的数目T相同,即为一个$T*T$的矩阵,其任意对角位置的第二权重参数w^j指示对应的专用适配器j的权重。

在一个实施例中，第二参数矩阵W_B可以记为如公式（2）所示的形式：

$$W_B = \begin{bmatrix} w^1 & 0 & \dots & 0 \\ 0 & w^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w^T \end{bmatrix} \quad (2)$$

在初始化时，可以将第二参数集中的各个权重参数值都初始化为 1。

由于公式（2）所示的第二参数矩阵为对角矩阵，其中仅对角线上的元素不是 0，在一个实施例中，可以将第二参数集表示为由这 T 个对角元素构成的数组。

基于此，在一个实施例中，将第一参数矩阵W_A和第二参数矩阵W_B进行横向拼接，可以得到形式为矩阵的权重参数W，如公式（3）所示：

$$W = [W_A|W_B] = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_A^1 & w^1 & 0 & \dots & 0 \\ w_1^2 & w_2^2 & \dots & w_A^2 & 0 & w^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ w_1^T & w_2^T & \dots & w_A^T & 0 & 0 & \dots & w^T \end{bmatrix} \quad (3)$$

下面基于上述记号，描述前述各个步骤的具体执行过程。

首先，在步骤 202，获取第一样本对应的第一嵌入向量x^t，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型t，所述第一任务类型t属于所述多个预设任务类型。

第一嵌入向量x^t可以是将第一样本输入到大模型之后，目标网络层在大模型中的上一层的输出向量。预设任务类型的总数为T，第一任务类型t为T个任务类型之一。

在一个实施例中，所述大模型在所述目标网络层之前具有第一网络；所述第一嵌入向量x^t基于以下方法确定：将所述第一样本输入到所述第一网络中，得到所述第一嵌入向量x^t。

在一个实施例中，当所述第一样本为图像数据时，所述任务类型至少包括：图像分类、

目标检测、图像分割、图像描述；当所述第一样本为文本数据时，所述任务类型至少包括：文本分类、命名实体识别、文本摘要、文本问答、文本情感识别。

然后，在步骤 204，将所述第一嵌入向量x^t分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器$\{\phi_1,\phi_2,...\phi_A\}$，以及所述多个专用适配器中对应于所述第一任务类型$t$的第一专用适配器$\phi^t$分别处理所述第一嵌入向量$x^t$，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量z_2。

在一个实施例中，步骤 204 中所述根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量z_2，具体包括：从所述第一参数矩阵W_A中获取所述第一任务类型t下各个通用适配器$\{\phi_1,\phi_2,...\phi_A\}$的第一权重参数$[w_1^t \quad w_2^t \quad ... \quad w_A^t]$；以及从所述第二参数集中获取所述第一专用适配器$\phi^t$对应的第二权重参数$w^t$；将各个通用适配器处理所述第一嵌入向量$x^t$的各个第一结果，和所述第一专用适配器$\phi^t$处理所述第一嵌入向量$x^t$的第二结果进行加权求和，得到第二嵌入向量$z_2$，其中，各个第一结果的权重因子根据所述第一权重参数确定，所述第二结果的权重因子根据所述第二权重参数确定。

在一种实施方式中，直接使用权重参数作为加权求和时各个适配器的输出结果的权重因子。此时，计算第二嵌入向量z_2的方法可以如公式（4）所示：

$$z_2 = \sum_{i=1}^{A} w_i^t \, \phi_i(x^t) + w^t \phi^t(x^t) \quad (4)$$

在另一实施方式中，对原始的权重参数进行进一步处理，作为加权求和时的权重因子。在一个更加具体的实施例中，所述各个第一结果的权重因子根据以下方法确定：

将第一参数矩阵W_A中第一任务类型t下各个通用适配器$\{\phi_1,\phi_2,...\phi_A\}$的第一权重参数$[w_1^t \quad w_2^t \quad ... \quad w_A^t]$分别输入到 Gumbel Sigmoid 函数中，将输出结果输入到 softmax 层中，得到所述各个第一结果的权重因子$[\widehat{w_1^t} \quad \widehat{w_2^t} \quad ... \quad \widehat{w_A^t}]$。在本实施例中，计算第二嵌入向量$z_2$的方法可以如公式（5）所示：

$$z_2 = \sum_{i=1}^{A} \widehat{w_i^t} \, \phi_i(x^t) + w^t \phi^t(x^t) \quad (5)$$

在一些可能的实施方式中，所述任一专用适配器包括多个子适配器，任一专用适配器对应的所述第二权重参数包括多个子适配器对应的多个子权重参数。将多个子适配器的输出结果基于各自的子权重参数进行加权求和，作为针对于该任务类型的专用适配器的输出。

在得到旁路任务网络输出的第二嵌入向量z_2之后，接下来，在步骤 206，基于所述第

二嵌入向量z_2以及所述目标处理输出的第三嵌入向量z_3确定输出向量z_{out}，根据所述输出向量z_{out}确定预测结果。

在一个实施例中，所述大模型在所述目标网络层之后具有第二网络。此时，步骤 206 具体包括：基于所述第二嵌入向量z_2以及所述第三嵌入向量z_3的融合结果，例如求和结果，确定所述输出向量z_{out}，即$z_{out} = z_2 + z_3$。然后将所述输出向量z_{out}输入到所述第二网络中，得到所述预测结果。

综合大模型以及旁路任务网络，根据输入样本确定预测结果的流程可以如图 3 所示。图 3 示出根据一个实施例的大模型架构示意图。在图 3 中，大模型按照数据流动方向的顺序包含第一网络、目标网络层和第二网络。将任务类型为t的输入样本输入到第一网络中，得到第一嵌入向量x^t。将第一嵌入向量x^t分别输入到目标网络层和旁路任务网络中，将基于前述方法得到的输出向量z_{out}输入到第二网络中，得到针对于输入样本的预测结果。

最后，在步骤 208，根据所述预测结果对应的损失，更新所述旁路任务网络。

在一些可能的实施方式中，所述大模型为基于 Transformer 架构的模型；所述目标网络层为以下之一：Query 层、Key 层、Value 层、Output 层、MLP（Multilayer Perceptron）层。

上述步骤 202 至步骤 208 描述了使用旁路任务网络对大模型中需要微调的目标网络层进行微调的过程。使用不同任务类型的多个输入样本，分别基于步骤 202 至步骤 208 描述的微调方法对旁路任务网络进行多轮训练，可以得到同时适用于多种任务类型的训练完成的大模型。

以下将结合输入样本的不同类型，描述多个具体的实施例。

在一个实施例中，所述第一样本包括文本数据，所述多个预设任务类型是与文本相关的任务。此时，所述大模型包含用于处理文本内容的自然语言处理网络，所述目标网络层属于所述自然语言处理网络，所述第一嵌入向量对应于所述文本数据。所述旁路任务网络用于对自然语言处理网络中的网络层进行微调。本实施例的模型架构可以如图 4 所示。

在另一个实施例中，大模型为适用于图像和文本处理的多模态大模型。所述第一样本包括图像数据以及文本数据，所述多个预设任务类型是与图像和文本之间的关联相关的任务，例如，以图生文，以文生图，文图匹配，等等。此时，所述大模型包括图像编码器，桥接网络，自然语言处理网络，所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间。在一个具体的例子中，所述目标网络层可以属于所述桥接网络，所述第一嵌入向量对应于所述图像数据。所述旁路任务

网络用于对桥接网络中的网络层进行微调。本实施例的模型架构可以如图 5 所示。

在又一个实施例中，所述第一样本包括图像数据以及文本数据，所述多个预设任务类型是与图像和文本之间的关联相关的任务。所述大模型包括图像编码器，桥接网络，自然语言处理网络，所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间。所述目标网络层包括第一目标层和第二目标层，所述旁路任务网络包括作为第一目标层旁路的第一任务网络，和作为第二目标层旁路的第二任务网络。所述第一目标层属于所述桥接网络，所述第二目标层属于所述自然语言处理网络。所述两个旁路任务网络分别用于对桥接网络和自然语言处理网络中的网络层进行微调。本实施例的模型架构可以如图 6 所示。

在又一个实施例中，可以在多模态大模型的不同网络部分，设置多个旁路任务网络。以具有图 5 结构的大模型为例，可以在桥接网络的一个或多个网络层的旁路，设置其对应的旁路任务网络，此外还在自然语言处理网络的一个或多个网络层的旁路，也设置对应的旁路任务网络。具体而言，该大模型包括图像编码器，桥接网络，自然语言处理网络，所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间。目标网络层可以包括第一目标层和第二目标层，所述第一目标层属于所述桥接网络，所述第二目标层属于所述自然语言处理网络。相应的，旁路任务网络包括设置于第一目标层旁路的第一任务网络，和设置于第二目标层旁路的第二任务网络。所述两个旁路任务网络分别用于对桥接网络和自然语言处理网络中的网络层进行微调。本实施例的模型架构可以如图 6 所示。

在还一个实施例中，也可以针对多模态大模型，仅在自然语言处理网络部分设置旁路任务网络。此时，与旁路任务网络有关的模型架构类似于或退化为图 4 的示例，这里不复赘述。

综合以上，本说明书实施例提出的多任务大模型的训练方法，通过明确区分任务通用适配器和任务专用适配器，即使在计算资源有限的情况下也能在大语言模型上实现高效的多任务微调。专用适配器和通用适配器的分离有效地缓解了多任务学习中常见的跷跷板问题和负向迁移问题，从而提升了整体性能。此外，本说明书提出的旁路任务网络还能根据所选的适配器学习到明确的任务层次结构，从而深入了解任务与有效完成任务所需适配器之间的关系（即各个适配器对应的权重大小），提供了令人信服的可解释性。

根据再一方面的实施例，还提供一种多任务大模型的训练装置。图 7 示出根据一个实施例的多任务大模型的训练装置的示意性框图，所述大模型包括已训练的目标网络层，以

及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数，该装置可以部署在任何具有计算、处理能力的设备、平台或设备集群中。如图 7 所示，所述装置 700 包括：

获取单元 701，配置为，获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；

嵌入向量处理单元 702，配置为，将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；

预测结果确定单元 703，配置为，基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；

旁路网络更新单元 704，配置为，根据所述预测结果对应的损失，更新所述旁路任务网络。

根据另一方面的实施例，还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，令计算机执行上述任一实施例所描述的方法。

根据再一方面的实施例，还提供一种计算设备，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，实现上述任一实施例所描述的方法。

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。

上述对本说明书特定实施例进行了描述。其它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求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。

需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或

者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，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、物品或者装置不仅包括那些要素，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、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、物品或者装置中还存在另外的相同要素。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权　利　要　求　书

1. 一种多任务大模型的训练方法，所述大模型包括已训练的目标网络层，以及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数；所述方法包括：

获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；

将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；

基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；

根据所述预测结果对应的损失，更新所述旁路任务网络。

2. 根据权利要求 1 所述的方法，其中，当所述第一样本为图像数据时，所述任务类型至少包括：图像分类、目标检测、图像分割、图像描述；

当所述第一样本为文本数据时，所述任务类型至少包括：文本分类、命名实体识别、文本摘要、文本问答、文本情感识别。

3. 根据权利要求 1 所述的方法，其中，所述第一样本包括图像数据以及文本数据；所述多个预设任务类型是与图像和文本之间的关联相关的任务。

4. 根据权利要求 3 所述的方法，其中，所述大模型包括图像编码器，桥接网络，自然语言处理网络；所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间；所述目标网络层属于所述桥接网络，所述第一嵌入向量对应于所述图像数据。

5. 根据权利要求 3 所述的方法，其中，所述大模型包括图像编码器，桥接网络，自然语言处理网络；所述桥接网络连接在所述图像编码器与自然语言处理网络之间，用于将图像编码器的编码结果转换到文本表征空间；所述目标网络层包括第一目标层和第二目标层，所述旁路任务网络包括作为第一目标层旁路的第一任务网络，和作为第二目标层旁路的第二任务网络；所述第一目标层属于所述桥接网络，所述第二目标层属于所述自然语言处理网络。

6. 根据权利要求 1 所述的方法，其中，所述权重参数包括，第一参数矩阵和第二参数

集，所述第一参数矩阵的行列数目分别对应预设任务类型的数目和通用适配器的数目，其中，任意位置的第一权重参数指示对应的通用适配器在对应任务类型下的权重；所述第二参数集至少包括，所述多个专用适配器分别对应的第二权重参数。

7. 根据权利要求 6 所述的方法，其中，所述专用适配器包括多个子适配器，任一专用适配器对应的所述第二权重参数包括多个子适配器对应的多个子权重参数。

8. 根据权利要求 6 所述的方法，其中，根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量，包括：

从所述第一参数矩阵中获取所述第一任务类型下各个通用适配器的第一权重参数；以及从所述第二参数集中获取所述第一专用适配器对应的第二权重参数；

将各个通用适配器处理所述第一嵌入向量的各个第一结果，和所述第一专用适配器处理所述第一嵌入向量的第二结果进行加权求和，得到第二嵌入向量，其中，各个第一结果的权重因子根据所述第一权重参数确定，所述第二结果的权重因子根据所述第二权重参数确定。

9. 根据权利要求 8 所述的方法，其中，所述各个第一结果的权重因子根据以下方法确定：

将所述第一参数矩阵中所述第一任务类型下各个通用适配器的第一权重参数分别输入到 Gumbel Sigmoid 函数中，将输出结果输入到 softmax 层中，得到所述各个第一结果的权重因子。

10. 根据权利要求 1 所述的方法，其中，所述适配器至少包含以下之一：LoRA 适配器、AdaLoRA 适配器、$(IA)^3$ 适配器。

11. 根据权利要求 1 所述的方法，其中，所述大模型为基于 Transformer 架构的模型；所述目标网络层为以下之一：Query 层、Key 层、Value 层、Output 层、MLP 层。

12. 一种多任务大模型的训练装置，所述大模型包括已训练的目标网络层，以及待训练的旁路任务网络，所述旁路任务网络包括若干通用适配器以及与多个预设任务类型分别对应的多个专用适配器，并具有各适配器对应的权重参数；所述装置包括：

获取单元，配置为，获取第一样本对应的第一嵌入向量，其中，所述第一样本包括图像数据和文本数据中的至少一项，并具有第一任务类型，所述第一任务类型属于所述多个预设任务类型；

嵌入向量处理单元，配置为，将所述第一嵌入向量分别输入所述目标网络层进行目标处理，以及输入所述旁路任务网络进行旁路处理，所述旁路处理包括，利用所述若干通用

适配器，以及所述多个专用适配器中对应于所述第一任务类型的第一专用适配器分别处理所述第一嵌入向量，并根据所述权重参数，将各适配器处理结果加权求和，得到第二嵌入向量；

预测结果确定单元，配置为，基于所述第二嵌入向量以及所述目标处理输出的第三嵌入向量确定输出向量，根据所述输出向量确定预测结果；

旁路网络更新单元，配置为，根据所述预测结果对应的损失，更新所述旁路任务网络。

13. 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，令计算机执行权利要求 1-11 中任一项所述的方法。

14. 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 1-11 中任一项所述的方法。