

问答方法以及问答模型训练方法

交叉援引

本公开要求于 2023 年 04 月 11 日提交中国专利局、申请号 202310412430.3、申请名称“问答方法以及问答模型训练方法”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

技术领域

本公开实施例涉及计算机技术领域，特别涉及一种问答方法。

背景技术

随着多媒体数据的丰富，任务型对话系统得到了广泛的关注，例如，电商领域的智能客服，任务型对话系统可以帮助用户自助得到问题的回复。随着用户对任务型对话系统的要求不断提高，对话系统需要先对回复用户的模态进行判断，即是否回复图片或文本，再完成内容的回复，但由于多模态的对话数据难以收集，导致训练数据较少，对话生成模型难以训练，使得对话生成模型的应用效果较差。

发明内容

有鉴于此，本公开实施例提供了一种问答方法。本公开一个或者多个实施例同时涉及另一种问答方法，一种问答模型训练方法，一种文本问答方法，一种问答装置，一种计算设备，一种计算机可读存储介质以及一种计算机程序，以解决现有技术中存在的技术缺陷。

根据本公开实施例的第一方面，提供了一种问答方法，包括：

获取待答复的问题信息；

将所述问题信息输入问答模型，获得所述问题信息对应的答复信息；

其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练获得，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。

根据本公开实施例的第二方面，提供了另一种问答方法，应用于云侧设备，包括：

接收端侧设备发送的问答请求，其中，所述问答请求携带有问题信息；

将所述问题信息输入问答模型，获得所述问题信息对应的答复信息；

其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练得到，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的；

将所述答复信息发送至所述端侧设备。

根据本公开实施例的第三方面，提供了一种问答模型训练方法，应用于云侧设备，包括：

获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本；

基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为利用所述预训练问答模型对所述待训练图像和目标语义信

息进行匹配得到,所述目标语义信息是基于所述图像描述文本和所述对话文本生成的;

获取问题样本,其中,所述问题样本包括待训练图像、与所述待训练图像关联的对话文本;

将所述问题样本输入所述预训练问答模型,获得所述预训练问答模型输出的预测答复信息,并基于所述预测答复信息训练所述预训练问答模型,获得问答模型;

将所述问答模型的模型参数发送至端侧设备。

根据本公开实施例的第四方面,提供了一种文本问答方法,应用于端侧设备,包括:

接收文本问题信息;

将所述文本问题信息输入问答模型,获得所述文本问题信息对应的文本答复信息,其中,所述问答模型根据上述问答模型训练方法方法获得;

将所述文本答复信息进行展示。

根据本公开实施例的第五方面,提供了一种问答装置,包括:

问题获取模块,被配置为获取待答复的问题信息;

答复获得模块,被配置为将所述问题信息输入问答模型,获得所述问题信息对应的答复信息;其中,所述问答模型基于预测答复信息训练获得,所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到,所述预训练问答模型基于预测图文匹配结果训练获得,所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到,所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。

根据本公开实施例的第六方面,提供了一种计算设备,包括:

存储器和处理器;

所述存储器被设置为存储计算机可执行指令,所述处理器被设置为执行所述计算机可执行指令,该计算机可执行指令被处理器执行时实现上述方法的步骤。

根据本公开实施例的第七方面,提供了一种计算机可读存储介质,其存储有计算机可执行指令,该指令被处理器执行时实现上述方法的步骤。

根据本公开实施例的第八方面,提供了一种计算机程序,其中,当所述计算机程序在计算机中执行时,令计算机执行上述方法的步骤。

本公开一个实施例提供的问答方法,获取待答复的问题信息;将所述问题信息输入问答模型,获得所述问题信息对应的答复信息;其中,所述问答模型基于预测答复信息训练获得,所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到,所述预训练问答模型基于预测图文匹配结果训练获得,所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到,所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。

本公开实施例中,通过将待答复的问题信息输入问答模型,获得问答模型输出的答复信息,其中,问答模型可通过训练预训练问答模型实现,根据预训练问答模型对问题样本对应的答复信息进行预测,根据预测的答复信息对问答模型进行训练,并且预训练问答模型基于预测图文匹配结果训练获得,同时该预测图文匹配结果根据预训练问答模型对待训练图像和目标语义信息匹配的到,另外,目标语义信息是基于待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成;该种方式即可实现通过少部分的对话文本,以及待训练图像能够训练问答模型更好地学习目标语义文本,进而训练问答模型根据对话文本能够输出对话内容,这样无需大量的对话文本作为训练数据,也能够使得问答模型更好地学习对话文本中的语义知识,以实现训练的问答模型的答复信息更为准确,提高模型的应用效果。

附图说明

图 1 是本公开一个实施例提供的一种问答方法的应用场景示意图；
图 2 是本公开一个实施例提供的一种问答方法的流程图；
图 3a 是本公开一个实施例提供的一种问答方法中问答模型的第一阶段训练示意图；
图 3b 是本公开一个实施例提供的一种问答方法中问答模型的第二阶段训练示意图；
图 3c 是本公开一个实施例提供的一种问答方法中问答模型的第三阶段训练示意图；
图 4 是本公开一个实施例提供的另一种问答方法的流程图；
图 5 是本公开一个实施例提供的一种问答模型训练方法的流程图；
图 6 是本公开一个实施例提供的一种文本问答方法的流程图；
图 7 是本公开一个实施例提供的一种问答装置的结构示意图；
图 8 是本公开一个实施例提供的一种计算设备的结构框图。

具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本公开。但是本公开能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本公开内涵的情况下做类似推广，因此本公开不受下面公开的具体实施的限制。

在本公开一个或多个实施例使用的术语是出于描述特定实施例的目的，而非旨在限制本公开一个或多个实施例。在本公开一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本公开一个或多个实施例使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本公开一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语用来将同一类型的信息彼此区分开。例如，在不脱离本公开一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

此外，需要说明的是，本公开一个或多个实施例所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

首先，对本公开一个或多个实施例涉及的名词术语进行解释。

多模态：在人工智能领域中，往往指感知信息，如图像、文本、语音等。

随着多媒体数据的丰富，互联网上的对话沟通不再局限于纯文本的模式。因此引入多模态对话将成为未来智能客服不可阻挡的趋势。多模态对话要求模型首先对于回复用户的模态进行判断，即是否回复图片亦或是文本；其次应具备检索出符合上下文的回复，分别为图像检索和文本检索；同时应该具备跟踪用户状态的能力，以追踪用户需求；最后应该具备文本生成的能力，以生成流畅自然的回复。然而，在数据驱动的今天，多模态对话数据由于隐私等系列问题并不好收集，导致其数量相对较少，模型难以训练，而多模态的非对话数据则具备较大规模。因此，本公开实施例提出了一种基于增量式的组合专家多模态对话预训练模型，利用多模态非对话数据建模型文本专家和图像专家的识别和匹配能力，通过对上述专家能力的继承，引入多模态对话数据从而建模对话专家，以实现阶梯式的模型能力提升。

在本公开中，提供了一种问答方法，本公开同时涉及另一种问答方法，一种问答模型训练方法，一种文本问答方法，一种问答装置，一种计算设备，以及一种计算机可读存储介质，在下面的实施例中逐一进行详细说明。

参见图 1，图 1 示出了根据本公开一个实施例提供的一种问答方法的应用场景示意图。

需要说明的是，本实施例提供的问答方法可应用于智能客服的问答回复场景等，具体的应用场景可根据不同的项目需求适应不同的需求，本实施例对此不做限定；为了便于理解，本实施例中以用户使用智能客服的应用场景为例进行说明。

实际应用中，用户向智能客服发送需要答复的问题信息，其中，问题信息可以理解为不同应用场景中的信息，包括但不限于文本信息、图像信息等，由于该智能客服使用了本实施例中提供的问答方法中的问答模型，在该问答模型接收到用户发送的问题信息之后，可输出该问题信息对应的答复信息，其中，答复信息可以理解为针对问题信息所反馈的答复内容，包括但不限于文本信息、图像信息等；进而，用户可接收到该智能客服反馈的答复内容，需要说明的是，在该问答模型投入使用之前，需要对该问答模型进行预先训练，具体的，为了提高问答模型的训练效果，本实施例中可采取分阶段训练模型，通过组合不同的专家网络实现，引入多模态对话数据对问答模型进行建模，实现阶梯式模型能力提升。

参见图 2，图 2 示出了根据本公开一个实施例提供的一种问答方法的流程图，具体包括以下步骤。

步骤 202：获取待答复的问题信息。

实际应用中，执行主体可获取待答复的问题信息，其中，问题信息可以理解为用户需等待答复的问题内容，包括但不限于文本内容、图像内容、文本和图像内容等。

步骤 204：将所述问题信息输入问答模型，获得所述问题信息对应的答复信息。

其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练获得，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。

实际应用中，执行主体可将问题信息输入至问答模型中，以获得该问题信息对应的答复信息，其中，答复信息可以理解为针对问题信息确定的相应答复内容，包括文本内容和/或图像内容等。需要说明的是，本实施例提供的问答模型可通过训练预训练问答模型实现，并根据预训练问答模型根据问题样本所输出的预测答复信息进行训练；进一步地，预训练问答模型还可根据待训练图像和目标语义信息训练获得，其中，目标语义信息可以理解为待训练图像对应的图像描述文本、与该待训练图像相关联的对话文本生成。

为了便于理解，下述实施例对问答模型的训练过程进行介绍，该问答模型的训练过程可包括三个阶段的训练，即先通过图像和图像标注文本训练第一阶段的初始问答模型，再通过图像、图像标注文本和对话文本训练第二阶段的预训练问答模型，最后再在根据图像和对话文本训练第三阶段的问答模型；需要说明的是，初始问答模型、预训练问答模型和问答模型均指代同一问答模型，可理解为在不同的训练阶段，问答模型所拥有的模型输出能力和效果不同；本实施例提供的训练过程为先训练预训练问答模型，再训练问答模型，即第二阶段和第三阶段的训练过程；具体的，所述将所述问题信息输入问答模型，获得所述问题信息对应的答复信息之前，还包括：

获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本；

基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为利用所述预训练问答模型对所述待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述图像描述文本和所述对话文本生成的；

获取问题样本，其中，所述问题样本包括待训练图像、与所述待训练图像关联的

对话文本；

将所述问题样本输入所述预训练问答模型，获得所述预训练问答模型输出的预测答复信息，并基于所述预测答复信息训练所述预训练问答模型，获得问答模型。

其中，预训练样本可以理解为训练预训练问答模型的样本数据，包括待训练图像、待训练图像对应的图像描述文本、与该待训练图像关联的对话文本；待训练图像可理解为任意类型的训练图像，对图像类别不做限定；待训练图像对应的图像描述文本可以理解为对待训练图像中的图像内容进行描述的文本信息，比如描述图像中的物体内容和状态等；与该待训练图像关联的对话文本可以理解为讨论与该待训练图像有关的对话信息，该对话信息可以理解为不同主体之间的对话数据，但对于对话内容的形式和对话主体本实施例中也不做任何限定。

实际应用中，可根据预训练样本对应的预测图文匹配结果训练预训练问答模型，其中，预测图文匹配结果可基于预训练问答模型中的信息匹配层（FFN）对待训练图像和目标语义信息进行匹配获得，需要说明的是，目标语义信息可以理解为根据图像描述文本与对话文本之间进行向量匹配获得，即可理解为根据图像描述文本，使得预训练问答模型能够更好地理解对话文本中的语义信息。进一步地，在预训练问答模型训练结束之后，还可获取问题样本，该问题样本可以理解为应用于问答模型的训练样本，包括待训练图像、与该待训练图像相关联的对话文本，再将问题样本输入预训练问答模型，可获得该预训练问答模型中的信息生成模块输出的预测答复信息，并根据预测答复信息继续训练预训练问答模型，最后，可获得训练效果较好的问答模型。

本实施例对问答模型的训练方式，采取先训练预训练问答模型的方式，再在下一个阶段继续训练该预训练问答模型，以使得训练获得的问答模型拥有阶梯式的学习能力。

进一步地，在训练问答模型的第三阶段中，可基于预训练问答模型输出的预测答复信息进行训练，本实施例可通过自回归的方式进行模型训练；具体的，所述基于所述预测答复信息训练所述预训练问答模型，获得问答模型，包括：

基于所述预测答复信息计算自回归损失值；

基于所述自回归损失值训练所述预训练问答模型，获得所述问答模型。

实际应用中，执行主体可通过预训练问答模型输出的预测答复信息，计算自回归损失值，并根据自回归损失值训练预训练问答模型，进而获得问答模型；本实施例对自回归损失值的确定可参考下述公式 1：

$$\mathcal{L}_{\text{stage}}^{\text{III}} = -\sum_{n=1}^N \log p_{\text{rgm}}(C_n | V, C_{<n}) \quad \text{公式 1}$$

其中， $\mathcal{L}_{\text{stage}}^{\text{III}}$ 表示第三阶段的损失值， C 表示对话历史， V 表示图像， p^* 表示模型预测结果。

更进一步地，对于训练预训练问答模型的阶段，即第二阶段，需要让模型学习到图像和对话文本之间的匹配能力，进而，需要训练预训练问答模型中图文对齐层（FFN）；具体的，所述基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，包括：

针对所述预训练样本中的待训练图像，利用所述预训练问答模型对所述待训练图像进行图像处理，获得图像向量；

针对所述预训练样本中的与所述待训练图像相关联的对话文本，利用所述预训练问答模型对所述对话文本进行文本处理，获得对话向量；

基于所述图像向量和所述对话向量，生成预测图文匹配结果；

基于所述预测图文匹配结果确定图文对齐损失值，并基于所述图文对齐损失值训练所述预训练问答模型。

实际应用中，在训练预训练问答模型阶段，可将预训练样本中的待训练图像输入

至预训练问答模型中的图像处理层 (FFN)，获得图像向量；再将预训练样本中的与该待训练图像相关联的对话文本，输入至预训练问答模型中的对话处理层 (FFN)，以获得对话向量；在对图像和文本分别进行向量处理后，可根据图像向量和对话向量进行匹配，生成预测图文匹配结果，并根据预测图文匹配结果确定图文对齐层 (FFN) 的损失值，根据该损失值训练预训练问答模型。

该种方式，通过利用图像和对话的多模态数据，让模型学习到图像和对话文本之间的匹配信息，使得预训练问答模型具有图像和对话匹配的能力。

进一步地，在训练预训练问答模型的阶段，为了提高预训练问答模型的训练效果，除了根据图文对齐损失值进行训练，还可参考待训练图像对应的损失值、对话文本对应的损失值；具体的，所述基于所述图文对齐损失值训练所述预训练问答模型，包括：

基于所述图像向量计算对话标注对齐损失值，基于所述对话向量计算图像对话损失值；

基于所述对话标注对齐损失值、所述图像对话损失值和所述图文对齐损失值训练所述预训练问答模型。

实际应用中，执行主体还可根据图像处理模块对待训练图像进行编码处理获得的图像向量，计算相应的对话标注对齐损失值，进一步地，可根据对话处理模块对对话文本进行编码处理获得的对话向量，计算相应的图像对话损失值；最后，根据对话标注对齐损失值、图像对话损失值、图文对齐损失值对预训练问答模型继续进行训练；本实施例中可参考下述公式 2 作为示例性的训练目标：

$$\mathcal{L}_{\text{stage}}^{\text{II}} = \mathcal{L}_{\text{icm}} + \mathcal{L}_{\text{ica}} + \mathcal{L}_{\text{mim}} \quad \text{公式 2}$$

其中， $\mathcal{L}_{\text{mim}}$ 、 $\mathcal{L}_{\text{ica}}$ 、 $\mathcal{L}_{\text{icm}}$ 具体可参考下述公式 3-公式 5：

$$\mathcal{L}_{\text{mim}} = \mathbb{E}_{(\mathbf{V}, \hat{\mathbf{T}}) \sim \{\mathbf{D}_d\}} \text{CE}(\mathbf{y}_{\text{mask}}, \mathbf{p}_{\text{mask}}(\mathbf{V}, \hat{\mathbf{T}})) \quad \text{公式 3}$$

$$\mathcal{L}_{\text{icm}} = \mathbb{E}_{(\mathbf{V}, \mathbf{T}, \mathbf{C}) \sim \mathbf{D}_d} \text{CE}(\mathbf{y}_{\text{icm}}, \mathbf{p}_{\text{icm}}(\mathbf{V}, \mathbf{T}, \mathbf{C})) \quad \text{公式 4}$$

$$\mathcal{L}_{\text{ica}} = \|\mathbf{H}_{\text{L-F}}^* - \mathbf{H}_{\text{L-F}}\|_2^2 \quad \text{公式 5}$$

其中， $\mathcal{L}_{\text{stage}}^{\text{II}}$ 表示第二阶段的损失值， $\mathbf{D}_d$ 表示对话多模态数据， $\hat{\mathbf{T}}$ 表示掩码后的字符， $\mathbf{y}^*$ 表示正确值， $\mathbf{H}_{\text{L-F}}^*$ 表示模型 L-F 层的中间表示。

更进一步地，预训练问答模型训练之前，还包括训练初始问答模型阶段，即第一阶段，能够使得初始问答模型具有对图像和文本匹配的能力；具体的，所述获取预训练样本之前，还包括：

获取初始训练样本，其中，所述初始训练样本包括待训练图像、所述待训练图像对应的图像标注文本；

针对所述待训练图像，利用输入初始问答模型对所述待训练图像进行图像处理，获得图像向量；

针对所述待训练图像对应的图像标注文本，利用所述初始问答模型对所述图像标注文本进行文本处理，获得文本标注向量；

基于所述图像向量和所述文本标注向量，确定预测图像描述文本；

基于所述预测图像描述文本训练所述初始问答模型。

实际应用中，获取初始训练样本，基于初始训练样本对初始问答模型进行训练，具体的，可将初始训练样本中的待训练图像输入至初始问答模型中的图像处理层 (FFN)，获得图像向量，再将初始训练样本中的图像标注文本输入至初始问答模型中的图像标注层 (FFN)，以获得文本标注向量，进而，根据图像向量和文本标注向量，确定预测图像描述文本，使得模型学习对图像和文本之间的匹配能力，再根据预测图像描述文本训练初始问答模型；需要说明的是，图像标注文本可以理解为人工对

待训练图像的图像标注信息,也可以理解为利用基础图像标注模型实现,本实施例对此不做限定。

进一步地,本实施例还提供了根据预测图像描述文本确定损失值,以训练初始问答模型;具体的,所述基于所述预测图像描述文本训练所述初始问答模型,包括:

基于所述图像向量和所述文本标注向量,计算图文标注损失值;

基于所述预测图像描述文本,计算图文对齐损失值;

根据所述图文标注损失值和所述图文对齐损失值,训练所述初始问答模型。

实际应用中,初始问答模型中的图像处理模块和图像标注模块分别对图像和文本进行编码处理后,还可分别计算图文标注损失值和图文对齐损失值,进而,根据图文标注损失值和图文对齐损失值,训练初始问答模型;本实施例中可参考下述训练目标:

$$\mathcal{L}_{\text{stage}}^I = \mathcal{L}_{\text{itm}} + \mathcal{L}_{\text{itm}} \quad \text{公式 6}$$

$$\mathcal{L}_{\text{itm}} = \mathbb{E}_{\{V, T\} \sim D_n} \text{CE}(\mathbf{y}_{\text{mask}}, \mathbf{g}_{\text{mask}}(V, T)) \quad \text{公式 7}$$

$$\mathcal{L}_{\text{itm}} = \mathbb{E}_{\{V, T\} \sim D_n} \text{CE}(\mathbf{y}_{\text{itm}}, \mathbf{p}_{\text{itm}}(V, T)) \quad \text{公式 8}$$

其中, $\mathcal{L}_{\text{stage}}^I$ 表示第一阶段的损失值, D_n 表示非对话多模态数据(图像描述文本), CE表示交叉熵计算。

基于此,在训练问答模型的第一阶段中,通过计算图文标注损失值,和图文对齐损失值,训练初始问答模型,便于后续在初始问答模型的基础上继续对问答模型的对话文本和图像之间的多模态数据的匹配能力进行训练,以提高问答模型对多模态数据的处理能力。

此外,执行主体经过三个阶段获得了问答模型之后,可将问答模型输出的针对问题信息对应的答复信息发送至用户,以获取到用户针对答复信息的反馈结果,实现对问答模型的微调过程;具体的,所述获得所述问题信息对应的答复信息之后,还包括:

将所述答复信息发送至用户;

接收所述用户针对所述答复信息的反馈结果;

基于所述反馈结果调整所述问答模型中的模型参数。

实际应用中,执行主体可将答复信息发送至用户,用户在接收到答复信息之后,可根据答复信息的具体内容给予相应的反馈或者评价,并将该反馈结果发送至执行主体,执行主体可根据接收到的反馈结果对问答模型中的模型参数进行微调,以便于获得应用效果更好的问答模型。

综上,本实施例提供的问答方法,通过利用问答模型,完成对问题信息的答复,其中,问答模型的训练过程还采用三阶段的训练过程,使得问答模型不断地学习图像文本匹配的能力,图像对话匹配的能力和答复回复生成的能力,具体的训练过程采用了一种增量式组合专家网络的方式,对问答模型中的各个处理模块进行训练,引入了多模态对话数据的建模对话的能力,实现阶梯式的模型能力提升。

下述结合附图 3a、图 3b 和图 3c,分别示出了本公开一个实施例提供的一种问答方法中问答模型的各个阶段的训练示意图。

需要说明的是,每个阶段的训练中,问答模型中均包括 L 层结构,且每层可包括五个不同的专家网络,分别是图像标注、对话上下文、图像、图文对齐和文本生成这五个网络层,另外每层还有一个共享的多头自注意机制层。

图 3a 示出了本公开一个实施例提供的一种问答方法中问答模型的第一阶段训练示意图。

实际应用中,图 3a 中的一阶段预训练过程,分别利用对图像和图像标注文本进行编码,输入至后续的 L-F 层中,完成图像标注文本和图像之间的匹配,进而在 F 层中利用图文对齐网络,实现一阶段预训练问答模型。

图 3b 示出了本公开一个实施例提供的一种问答方法中问答模型的第二阶段训练

示意图。

实际应用中，图 3b 中的二阶段预训练过程，模型的输入为图像、图像标注文本、对话文本，先输入至 L-F 层中，完成了图像标注文本和对话文本之间的匹配能力，获得了目标语义文本后，再输入模型的 F 层中，与图像进行匹配，完成图文对齐的训练过程。

图 3c 示出了本公开一个实施例提供的一种问答方法中问答模型的第三阶段训练示意图。

实际应用中，图 3c 中的三阶段预训练过程，模型的输入为对话文本和图像，由于在前两个阶段的训练中，问答模型已经拥有了对话和图像匹配的能力，进而，在三阶段的预训练中，利用 L-F 层中的对话上下文、图像这两个网络进行对话和图像之间的匹配理解，进而在模型的 F 层中，即可训练文本生成网络，以训练该问答模型的答复信息生成能力。

需要说明的是，问答模型在应用于多模态对话任务时，可包括多类执行任务：1、多模态意图识别，具体上预测当前轮是否需要图像的分享；2、多模态检索，根据对话历史检索出合适的图像、文本进行回复；3、多模态对话追踪，追踪用户关键的对话需求；4、多模态文本生成，根据对话历史生成适当回复。

基于此，本实施例提供的问答模型训练过程，将模型中每层划分为多个专家网络，在完成具体任务时将不同的专家网络进行组合，实现任务处理，由于小模块的划分，使得模型的训练上变得更为可控，其解释性也得到增强，同时，在具体任务中没有被使用的专家网络参数不会被激活，这样可以大大加速了模型的推理速度，提高参数利用率；另外，对各个专家网络赋予不同的语义知识，使得每一种模态，每一种数据类型在模型的流动都有迹可循，增强了模型的可结实性和泛化能力。

参见图 4，图 4 示出了根据本公开一个实施例提供的另一种问答方法的流程图，具体包括以下步骤。

需要说明的是，本实施例提供的问答方法应用于云侧设备，即在接收到端侧设备的问题信息之后，可根据在云侧设备的问答模型，实现对问题信息的答复。

步骤 402：接收端侧设备发送的问答请求，其中，所述问答请求携带有问题信息。

实际应用中，云侧设备可接收到端侧设备发送的问答请求，并可获取到该问答请求中携带的问题信息；需要说明的是，问题信息的具体内容和形式本实施例不做限定。

步骤 404：将所述问题信息输入问答模型，获得所述问题信息对应的答复信息。

其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练得到，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的；

实际应用中，云侧设备可将问题信息输入至问答模型中，并获得问答模型输出的答复信息，其中，云侧设备中对于问答模型的训练过程，可参考上述实施例中描述的训练过程，即三阶段的问答模型训练，在此不做过多赘述。

步骤 406：将所述答复信息发送至所述端侧设备。

进一步地，云侧设备在获得了答复信息之后，可将答复信息反馈至端侧设备。

基于此，本实施例提供应用于云侧设备的问答方法，能够实现对端侧设备发送的问题信息进行问题回复，同时，问答模型在云侧设备中训练获得，可以节省端侧设备的计算资源和存储空间，能够为用户提供的更好的使用效果。

参见图 5，图 5 示出了根据本公开一个实施例提供的一种问答模型训练方法的流程图，具体包括以下步骤。

需要说明的是，本实施例提供的问答模型训练方法可应用于端侧设备，也可应用

于云侧设备，本实施例对此不做限定。

步骤 502：获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本。

步骤 504：基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为利用所述预训练问答模型对所述待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述图像描述文本和所述对话文本生成的。

步骤 506：获取问题样本，其中，所述问题样本包括待训练图像、与所述待训练图像关联的对话文本。

步骤 508：将所述问题样本输入所述预训练问答模型，获得所述预训练问答模型输出的预测答复信息，并基于所述预测答复信息训练所述预训练问答模型，获得问答模型。

步骤 510：将所述问答模型的模型参数发送至端侧设备。

其中，预训练样本可以理解为训练预训练问答模型的样本数据，包括待训练图像、待训练图像对应的图像描述文本、与该待训练图像关联的对话文本；待训练图像可理解为任意类型的训练图像，对图像类别不做限定；待训练图像对应的图像描述文本可以理解为对待训练图像中的图像内容进行描述的文本信息，比如描述图像中的物体内容和状态等；与该待训练图像关联的对话文本可以理解为讨论与该待训练图像有关的对话信息，该对话信息可以理解为不同主体之间的对话数据，但对于对话内容的形式和对话主体本实施例中也不做任何限定。

实际应用中，可根据预训练样本对应的预测图文匹配结果训练预训练问答模型，其中，预测图文匹配结果可基于预训练问答模型中的信息匹配层（FFN）对待训练图像和目标语义信息进行匹配获得，需要说明的是，目标语义信息可以理解为根据图像描述文本与对话文本之间进行向量匹配获得，即可理解为根据图像描述文本，使得预训练问答模型能够更好地理解对话文本中的语义信息。进一步地，在预训练问答模型训练结束之后，还可获取问题样本，该问题样本可以理解为应用于问答模型的训练样本，包括待训练图像、与该待训练图像相关联的对话文本，再将问题样本输入预训练问答模型，可获得该预训练问答模型中的信息生成层（FFN）输出的预测答复信息，并根据预测答复信息继续训练预训练问答模型，最后，可获得训练效果较好的问答模型；再将问答模型中的模型参数发送至端侧设备，便于端侧设备根据模型参数即可构建使用效果较好的问答模型，提高端侧设备中问答模型的使用效果。

基于此，通过训练能够实现对话文本和图像之间匹配的预训练问答模型，再进一步地，训练该预训练问答模型能够实现对话文本和图像的图文对齐，进而对答复信息的生成与回复，该问答模型通过利用多模态数据，训练对话生成能力，以提高问答模型的答复效果。

参见图 6，图 6 示出了根据本公开一个实施例提供的一种文本问答方法的流程图，具体包括以下步骤。

需要说明的是，为了便于理解，本实施例提供的文本问答方法可应用于智能客服的应用场景，对用户提出的文本问题进行文本答复，其中，该智能客服的自动回复文本是利用问答模型的输出回复。

步骤 602：接收文本问题信息。

其中，文本问题信息可以理解为用户对智能客服提出的文本问题，该文本问题与当前智能客服所在的平台中的项目相关联，本实施例对此不做限定。

步骤 604：将所述文本问题信息输入问答模型，获得所述文本问题信息对应的文本答复信息。

实际应用中，将文本问题信息输入至问答模型后，该问答模型可输出该文本问题

信息对应的文本答复信息，即可理解为智能客服对用户提出的文本问题进行了文本答复，需要说明的是，本实施例提供的问答模型可参考上述实施例中所描述的训练方法获得，在此不做过多赘述，包括利用云侧设备训练的问答模型。

此外，本实施例中还提供了该问答模型在端侧设备进行训练获得，并在端侧设备进行应用；具体的，所述将所述文本问题信息输入问答模型，获得所述文本问题信息对应的文本答复信息之前，包括：

获取云侧设备中问答模型的模型参数；

获取文本问题样本；

基于所述模型参数构建待训练的问答模型，基于所述文本问题样本对所述待训练的问答模型进行训练，获得问答模型。

实际应用中，端侧设备可获取到云侧设备中问答模型的模型参数，其中，该模型参数的获得过程，可参考上述在云侧设备上训练问答模型的过程，在此不做赘述；进一步地，端侧设备还可获取到文本问题样本，并利用从云侧设备获取到的问答模型的模型参数构建待训练的问答模型，再根据端侧设备中的文本问题样本对待训练的问答模型进行训练，以获得问答模型；进而，在端侧设备中训练的问答模型可在本地进行应用。

需要说明的是，云侧设备训练问答模型的模型参数的过程，可参考上述实施例中对问答模型的训练的描述，即包括通过三阶段训练问答模型的过程，在此不做赘述。

步骤 606：将所述文本答复信息进行展示。

实际应用中，端侧设备在利用问答模型对文本问题信息进行答复后，可将该文本答复信息在端侧设备的展示界面中进行展示。

综上，本实施例提供的文本问答方法，通过利用问答模型对文本问题信息进行自动答复，同时该问答模型可应用云侧设备训练的问答模型，也可应用端侧设备训练的问题模型，均能够对文本问题信息实现快速且合理的答复，提高了用户对文本答复的使用效果。

与上述方法实施例相对应，本公开还提供了问答装置实施例，图 7 示出了本公开一个实施例提供的一种问答装置的结构示意图。如图 7 所示，该装置包括：

问题获取模块 702，被配置为获取待答复的问题信息；

答复获得模块 704，被配置为将所述问题信息输入问答模型，获得所述问题信息对应的答复信息；其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练获得，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。

可选地，所述装置，还包括：

问答模型训练模块，被配置为获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本；

基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为利用所述预训练问答模型对所述待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述图像描述文本和所述对话文本生成的；

获取问题样本，其中，所述问题样本包括待训练图像、与所述待训练图像关联的对话文本；

将所述问题样本输入所述预训练问答模型，获得所述预训练问答模型输出的预测答复信息，并基于所述预测答复信息训练所述预训练问答模型，获得问答模型。

可选地，所述问答模型训练模块，进一步被配置为：

基于所述预测答复信息计算自回归损失值；

基于所述自回归损失值训练所述预训练问答模型，获得所述问答模型。

可选地，所述问答模型训练模块，进一步被配置为：

针对所述预训练样本中的待训练图像，利用所述预训练问答模型对所述待训练图像进行图像处理，获得图像向量；

针对所述预训练样本中的与所述待训练图像相关联的对话文本，利用所述预训练问答模型对所述对话文本进行文本处理，获得对话向量；

基于所述图像向量和所述对话向量，生成预测图文匹配结果；

基于所述预测图文匹配结果确定图文对齐损失值，并基于所述图文对齐损失值训练所述预训练问答模型。

可选地，所述问答模型训练模块，进一步被配置为：

基于所述图像向量计算对话标注对齐损失值，基于所述对话向量计算图像对话损失值；

基于所述对话标注对齐损失值、所述图像对话损失值和所述图文对齐损失值训练所述预训练问答模型。

可选地，所述问答模型训练模块，进一步被配置为：

获取初始训练样本，其中，所述初始训练样本包括待训练图像、所述待训练图像对应的图像标注文本；

针对所述待训练图像，利用初始问答模型对所述待训练图像进行图像处理，获得图像向量；

针对所述待训练图像对应的图像标注文本，利用所述初始问答模型对所述图像标注文本进行文本处理，获得文本标注向量；

基于所述图像向量和所述文本标注向量，确定预测图像描述文本；

基于所述预测图像描述文本训练所述初始问答模型。

可选地，所述问答模型训练模块，进一步被配置为：

基于所述图像向量和所述文本标注向量，计算图文标注损失值；

基于所述预测图像描述文本，计算图文对齐损失值；

根据所述图文标注损失值和所述图文对齐损失值，训练所述初始问答模型。

可选地，所述装置，还包括：

模型调整模块，被配置为将所述答复信息发送至用户；

接收所述用户针对所述答复信息的反馈结果；

基于所述反馈结果调整所述问答模型中的模型参数。

综上，本公开实施例提供的问答装置，通过将待答复的问题信息输入问答模型，获得问答模型输出的答复信息，其中，问答模型可通过训练预训练问答模型实现，根据预训练问答模型对问题样本对应的答复信息进行预测，根据预测的答复信息对问答模型进行训练，并且预训练问答模型基于预测图文匹配结果训练获得，同时该预测图文匹配结果根据预训练问答模型对待训练图像和目标语义信息匹配的到，另外，目标语义信息是基于待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成；该种方式即可实现通过少部分的对话文本，以及待训练图像能够训练问答模型更好地学习目标语义文本，进而训练问答模型根据对话文本能够输出对话内容，这样无需大量的对话文本作为训练数据，也能够使得问答模型更好地学习对话文本中的语义知识，以实现训练的问答模型的答复信息更为准确，提高模型的应用效果。

上述为本实施例的一种问答装置的示意性方案。需要说明的是，该问答装置的技术方案与上述的问答方法的技术方案属于同一构思，问答装置的技术方案未详细描述的细节内容，均可以参见上述问答方法的技术方案的描述。

图8示出了根据本公开一个实施例提供的一种计算设备800的结构框图。该计算设备800的部件包括但不限于存储器810和处理器820。处理器820与存储器810通

过总线 830 相连接，数据库 850 被设置为保存数据。

计算设备 800 还包括接入设备 840，接入设备 840 使得计算设备 800 能够经由一个或多个网络 860 通信。这些网络的示例包括公用交换电话网（PSTN, Public Switched Telephone Network）、局域网（LAN, Local Area Network）、广域网（WAN, Wide Area Network）、个域网（PAN, Personal Area Network）或诸如因特网的通信网络的组合。接入设备 840 可以包括有线或无线的任何类型的网络接口（例如，网络接口卡（NIC, network interface controller））中的一个或多个，诸如 IEEE802.11 无线局域网（WLAN, Wireless Local Area Network）无线接口、全球微波互联接入（Wi-MAX, Worldwide Interoperability for Microwave Access）接口、以太网接口、通用串行总线（USB, Universal Serial Bus）接口、蜂窝网络接口、蓝牙接口、近场通信（NFC, Near Field Communication）。

在本公开的一个实施例中，计算设备 800 的上述部件以及图 8 中未示出的其他部件也可以彼此相连接，例如通过总线。应当理解，图 8 所示的计算设备结构框图是出于示例的目的，而不是对本公开范围的限制。本领域技术人员可以根据需要，增添或替换其他部件。

计算设备 800 可以是任何类型的静止或移动计算设备，包括移动计算机或移动计算设备（例如，平板计算机、个人数字助理、膝上型计算机、笔记本电脑、上网本等）、移动电话（例如，智能手机）、可佩戴的计算设备（例如，智能手表、智能眼镜等）或其他类型的移动设备，或者诸如台式计算机或个人计算机（PC, Personal Computer）的静止计算设备。计算设备 800 还可以是移动式或静止式的服务器。

其中，处理器 820 被设置为执行如下计算机可执行指令，该计算机可执行指令被处理器执行时实现上述问答方法、问答模型训练方法、文本问答方法的步骤。

上述为本实施例的一种计算设备的示意性方案。需要说明的是，该计算设备的技术方案与上述的问答方法、问答模型训练方法、文本问答方法的技术方案属于同一构思，计算设备的技术方案未详细描述的细节内容，均可以参见上述问答方法、问答模型训练方法、文本问答方法的技术方案的描述。

本公开一实施例还提供一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现上述问答方法、问答模型训练方法、文本问答方法的步骤。

上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是，该存储介质的技术方案与上述的问答方法、问答模型训练方法、文本问答方法的技术方案属于同一构思，存储介质的技术方案未详细描述的细节内容，均可以参见上述问答方法、问答模型训练方法、文本问答方法的技术方案的描述。

本公开一实施例还提供一种计算机程序，其中，当所述计算机程序在计算机中执行时，令计算机执行上述问答方法、问答模型训练方法、文本问答方法的步骤。

上述为本实施例的一种计算机程序的示意性方案。需要说明的是，该计算机程序的技术方案与上述的问答方法、问答模型训练方法、文本问答方法的技术方案属于同一构思，计算机程序的技术方案未详细描述的细节内容，均可以参见上述问答方法、问答模型训练方法、文本问答方法的技术方案的描述。

上述对本公开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。

所述计算机指令包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够

携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是, 所述计算机可读介质包含的内容可以根据专利实践的要求进行适当的增减, 例如在某些地区, 根据专利实践, 计算机可读介质不包括电载波信号和电信信号。

需要说明的是, 对于前述的各方法实施例, 为了简便描述, 故将其都表述为一系列的动作组合, 但是本领域技术人员应该知悉, 本公开实施例并不受所描述的动作顺序的限制, 因为依据本公开实施例, 某些步骤可以采用其它顺序或者同时进行。其次, 本领域技术人员也应该知悉, 公开中所描述的实施例均属于优选实施例, 所涉及的动作和模块并不一定是本公开实施例所必须的。

在上述实施例中, 对各个实施例的描述都各有侧重, 某个实施例中未详述的部分, 可以参见其它实施例的相关描述。

以上公开的本公开优选实施例只是用于帮助阐述本公开。可选实施例并没有详尽叙述所有的细节, 也不限制该发明为所述的具体实施方式。显然, 根据本公开实施例的内容, 可作很多的修改和变化。本公开选取并具体描述这些实施例, 是为了更好地解释本公开实施例的原理和实际应用, 从而使所属技术领域技术人员能很好地理解和利用本公开。本公开受权利要求书及其全部范围和等效物的限制。

工业实用性

本公开实施例所提供的训练文本生成模型的方法、文本生成方法及装置应用于训练文本生成场景中, 通过将待答复的问题信息输入问答模型, 获得问答模型输出的答复信息, 其中, 问答模型可通过训练预训练问答模型实现, 根据预训练问答模型对问题样本对应的答复信息进行预测, 根据预测的答复信息对问答模型进行训练, 并且预训练问答模型基于预测图文匹配结果训练获得, 同时该预测图文匹配结果根据预训练问答模型对待训练图像和目标语义信息匹配的到, 另外, 目标语义信息是基于待训练图像对应的图像描述文本、与待训练图像相关联的对话文本生成; 该种方式即可实现通过少部分的对话文本, 以及待训练图像能够训练问答模型更好地学习目标语义文本, 进而训练问答模型根据对话文本能够输出对话内容, 这样无需大量的对话文本作为训练数据, 也能够使得问答模型更好地学习对话文本中的语义知识, 以实现训练的问答模型的答复信息更为准确, 提高模型的应用效果。

权 利 要 求 书

1. 一种问答方法，包括：
 获取待答复的问题信息；
 将所述问题信息输入问答模型，获得所述问题信息对应的答复信息；
 其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练获得，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的。
2. 根据权利要求1所述的方法，所述将所述问题信息输入问答模型，获得所述问题信息对应的答复信息之前，还包括：
 获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本；
 基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为所述预训练问答模型对所述待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述图像描述文本和所述对话文本生成的；
 获取问题样本，其中，所述问题样本包括待训练图像、与所述待训练图像关联的对话文本；
 将所述问题样本输入所述预训练问答模型，获得所述预训练问答模型输出的预测答复信息，并基于所述预测答复信息训练所述预训练问答模型，获得问答模型。
3. 根据权利要求2所述的方法，所述基于所述预测答复信息训练所述预训练问答模型，获得问答模型，包括：
 基于所述预测答复信息计算自回归损失值；
 基于所述自回归损失值训练所述预训练问答模型，获得所述问答模型。
4. 根据权利要求2或3所述的方法，所述基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，包括：
 针对所述预训练样本中的待训练图像，利用所述预训练问答模型对所述待训练图像进行图像处理，获得图像向量；
 针对所述预训练样本中的与所述待训练图像相关联的对话文本，利用所述预训练问答模型对所述对话文本进行文本处理，获得对话向量；
 基于所述图像向量和所述对话向量，生成预测图文匹配结果；
 基于所述预测图文匹配结果确定图文对齐损失值，并基于所述图文对齐损失值训练所述预训练问答模型。
5. 根据权利要求4所述的方法，所述基于所述图文对齐损失值训练所述预训练问答模型，包括：
 基于所述图像向量计算对话标注对齐损失值，基于所述对话向量计算图像对话损失值；
 基于所述对话标注对齐损失值、所述图像对话损失值和所述图文对齐损失值训练所述预训练问答模型。
6. 根据权利要求2所述的方法，所述获取预训练样本之前，还包括：
 获取初始训练样本，其中，所述初始训练样本包括待训练图像、所述待训练图像对应的图像标注文本；
 针对所述待训练图像，利用初始问答模型对所述待训练图像进行图像处理，获得图像向量；

- 针对所述待训练图像对应的图像标注文本，利用所述初始问答模型对所述图像标注文本进行文本处理，获得文本标注向量；
- 基于所述图像向量和所述文本标注向量，确定预测图像描述文本；
- 基于所述预测图像描述文本训练所述初始问答模型。
7. 根据权利要求 6 所述的方法，所述基于所述预测图像描述文本训练所述初始问答模型，包括：
- 基于所述图像向量和所述文本标注向量，计算图文标注损失值；
- 基于所述预测图像描述文本，计算图文对齐损失值；
- 根据所述图文标注损失值和所述图文对齐损失值，训练所述初始问答模型。
8. 根据权利要求 1 所述的方法，所述获得所述问题信息对应的答复信息之后，还包括：
- 将所述答复信息发送至用户；
- 接收所述用户针对所述答复信息的反馈结果；
- 基于所述反馈结果调整所述问答模型中的模型参数。
9. 一种问答方法，应用于云侧设备，包括：
- 接收端侧设备发送的问答请求，其中，所述问答请求携带有问题信息；
- 将所述问题信息输入问答模型，获得所述问题信息对应的答复信息；
- 其中，所述问答模型基于预测答复信息训练获得，所述预测答复信息为利用预训练问答模型对问题样本对应的答复信息进行预测得到，所述预训练问答模型基于预测图文匹配结果训练得到，所述预测图文匹配结果为利用所述预训练问答模型对待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述待训练图像对应的图像描述文本、与所述待训练图像相关联的对话文本生成的；
- 将所述答复信息发送至所述端侧设备。
10. 一种问答模型训练方法，应用于云侧设备，包括：
- 获取预训练样本，其中，所述预训练样本包括待训练图像、所述待训练图像对应的图像描述文本、与所述待训练图像关联的对话文本；
- 基于所述预训练样本对应的预测图文匹配结果训练所述预训练问答模型，其中，所述预测图文匹配结果为利用所述预训练问答模型对所述待训练图像和目标语义信息进行匹配得到，所述目标语义信息是基于所述图像描述文本和所述对话文本生成的；
- 获取问题样本，其中，所述问题样本包括待训练图像、与所述待训练图像关联的对话文本；
- 将所述问题样本输入所述预训练问答模型，获得所述预训练问答模型输出的预测答复信息，并基于所述预测答复信息训练所述预训练问答模型，获得问答模型；
- 将所述问答模型的模型参数发送至端侧设备。
11. 一种文本问答方法，应用于端侧设备，包括：
- 接收文本问题信息；
- 将所述文本问题信息输入问答模型，获得所述文本问题信息对应的文本答复信息，其中，所述问答模型根据上述权利要求 10 所述的方法获得；
- 将所述文本答复信息进行展示。
12. 根据权利要求 11 所述方法，所述将所述文本问题信息输入问答模型，获得所述文本问题信息对应的文本答复信息之前，包括：
- 获取云侧设备中问答模型的模型参数；
- 获取文本问题样本；
- 基于所述模型参数构建待训练的问答模型，基于所述文本问题样本对所述待训练的问答模型进行训练，获得问答模型。
13. 一种计算设备，包括：
- 存储器和处理器；

所述存储器被设置为存储计算机可执行指令，所述处理器被设置为执行所述计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1 至 12 任意一项所述方法的步骤。

14. 一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1 至 12 任意一项所述方法的步骤。