问答大模型的后训练方法、装置、电子设备及存储介质

相关申请的交叉引用

本公开要求于 2025 年 3 月 12 日提交中国专利局的申请号为 2025102875932、名称为“问答大模型的后训练方法、装置、电子设备及存储介质”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

本公开涉及人工智能技术领域，具体而言，涉及一种问答大模型的后训练方法、装置、电子设备及存储介质。

背景技术

现有技术中，常常通过后训练的方式使得模型回答更为准确。例如，典型的方式是RLHF（Reinforcement Learning from Human Feedback，基于人类反馈的强化学习）方法和 DPO（Direct Preference Optimization，直接偏好优化）方法，但是，RLHF 方法和 DPO 方法的本质均是让模型能够按照人类偏好回答问题，即使得模型的回答内容贴近人类的偏好，但是这两种方法均未实质增强模型的推理能力。

因此，如何增加模型的推理能力以使模型回答更为准确是亟需解决的问题。

发明内容

本公开实施例至少提供了一种问答大模型的后训练方法以及相关设备，相关设备可以包括问答大模型的后训练装置、电子设备、计算机可读存储介质和计算机程序产品，使得问答大模型能够学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

为实现上述目的中的至少一个目的，本申请采用的技术方案如下：

第一方面：本申请提供一种问答大模型的后训练方法，包括：

获取若干推理数据，所述推理数据是对原问题及其原答案进行逻辑拆分得到的；

基于全部所述推理数据，构造推理数据集；

利用所述推理数据集对所述问答大模型进行后训练处理；所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理。

一种可选的实施方式中，所述推理数据集包括第一子集、第二子集、第三子集和第四子集；所述基于全部所述推理数据，构造推理数据集的步骤，包括：

针对每份所述推理数据，基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本；所述第一子集包括每份所述推理数据对应的全部第一样本；

针对每份所述推理数据，基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本，所述第二子集包括每份所述推理数据对应的全部第二样本；

针对每份所述推理数据，基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本，所述第三子集包括每份所述推理数据对应的全部第三样本；

针对每份所述推理数据，基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本，所述第四子集包括每份所述推理数据对应的全部第四样本。

一种可选的实施方式中，所述第一子集配置成引导所述问答大模型学习对所述输入问题进行分步解答，所述第二子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解，所述第三子集配置成引导所述问答大模型学习基于分步推理的方式补全缺失的中间过程，所述第四子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解后进行分步解答。

一种可选的实施方式中，所述利用所述推理数据集对问答大模型进行后训练处理，包括：

分别利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集，对所述问答大模型进行后训练处理。

一种可选的实施方式中，所述利用所述推理数据集对问答大模型进行后训练处理，包括：

利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集的先后顺序，对所述问答大模型进行后训练处理；所述先后顺序包括：依次排列顺序或随机排列顺序。

一种可选的实施方式中，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本的步骤，包括：

将第一引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第一样本，并设置第一个第一样本的标签为所述推理数据中的第一个子答案；所述第一引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中最后一个子问题的答案；

将所述第一引导文本、所述推理数据中的原问题、第 n 个子问题以及第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第一样本，并设置第 n 个第一样本的标签为所述推理数据中的第 n 个子答案，n=2,3,…,N。

一种可选的实施方式中，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本的步骤，包括：

将第二引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第二样本，并设置第一个第二样本的标签为所述推理数据中的第二个子答案；所述第二引导文本配置成指示所述问答大模型基于输入内容从所述输入内容的原问题中拆解出所述输入内容中最后一个子问题的下一个子问题；

将所述第二引导文本、所述推理数据中的原问题、第 m 个子问题及其之前的每个子问题进行拼接得到第 m 个第二样本，并设置第 m 个第二样本的标签为所述推理数据中的第 m+1 个子问题，m=2,3,…,N-1。

一种可选的实施方式中，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本的步骤，包括：

将第三引导文本、所述推理数据中除了第一个子问题和第一个子答案以外的每个其他子问题和其他子答案进行拼接得到第一个第三样本，并设置第一个第三样本的标签为所述推理数据中的第一个子问题和第一个子答案；所述第三引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中缺失的中间子问题及其中间子答案；

将所述第三引导文本、所述推理数据中除了第 m 个子问题和第 m 个子答案以外的每个其他子问题和其他子答案进行拼接得到第 m 个第三样本，并设置第 m 个第三样本的标签为所述推理数据中的第 m 个子问题和第 m 个子答案，m=2,3,…,N-1。

一种可选的实施方式中，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本的步骤，包括：

将第四引导文本以及所述推理数据中的原问题进行拼接得到第一个第四样本，并设置第一个第四样本的标签为所述推理数据中的每个子问题及其子答案；所述第四引导文本配置成指示所述问答大模型基于输入内容分步推理出后续的子问题和子答案；

将所述第四引导文本、所述推理数据中的原问题、第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第四样本，并设置第 n 个第四样本的标签为所述推理数据中的第 n-1 个子问题之后的每个子问题及其子答案，n=2,3,…,N。

一种可选的实施方式中，所述后训练处理的过程中包括：监督微调方式或指令微调方式。

第二方面：本申请还提供一种问答大模型的后训练装置，包括：

数据获取模块，配置成获取若干推理数据，所述推理数据是对原问题及其原答案进行逻辑拆分得到的；

构造模块，配置成基于全部所述推理数据，构造推理数据集；

后训练模块，配置成利用所述推理数据集对所述问答大模型进行后训练处理；所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理。

第三方面：本申请还提供一种电子设备，包括：存储器和处理器，所述存储器存储有软件程序，当所述电子设备运行时所述处理器执行所述软件程序以实现如第一方面所述的问答大模型的后训练方法。

第四方面：本申请还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现第一方面所述的问答大模型的后训练方法。

与现有技术相比，本公开提供了一种问答大模型的后训练方法、装置、电子设备及存储介质，首先获取若干推理数据，再基于全部推理数据，构造推理数据集；最后利用推理数据集对问答大模型进行后训练处理；由于每份推理数据是对原问题及其原答案进行逻辑拆分得到的，故推理数据集可以配置成引导问答大模型学习对输入问题进行分步推理。本公开能够使得问答大模型能够学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

为使本公开的上述目的、特征和优点能更明显易懂，下文特举较佳实施例，并配合所附附图，作详细说明如下。

附图说明

为了更清楚地说明本公开实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，此处的附图被并入说明书中并构成本说明书中的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。应当理解，以下附图仅示出了本公开的某些实施例，因此不应被看作是对范围的限定，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相关的附图。

图 1 为本公开实施例提供的一种问答大模型的后训练方法的流程示意图之一。

图 2 为本公开实施例提供的一种问答大模型的后训练方法的流程示意图之二。

图 3 为本公开实施例提供的一种问答大模型的后训练装置的结构示意图。

图 4 为本公开实施例提供的一种电子设备的结构示意图。

具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本公开一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本公开实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本公开的实施例的详细描述并非旨在限制要求保护的本公开的范围，而是仅仅表示本公开的选定实施例。基于本公开的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本公开保护的范围。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步定义和解释。

此外，若出现术语“第一”、“第二”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

需要说明的是，在不冲突的情况下，本公开的实施例中的特征可以相互结合。

现有技术中，为了让模型回答的内容更为准确，存着以下两种方式：

第一种是通过后训练增强模型的推理能力，以使得模型的回答更为准确。典型的方式是 RLHF 方法和 DPO 方法。其中，RLHF 方法的实现原理为：首先需要给指令微调后的模

型喂入数据，得到数据的答案后，人为标注偏好数据，列如标注数据（x, ylose, ywin）；再利用标注的偏好数据训练一个奖励模型 modelaward；为了训练后的模型的分布与原来模型的分布相差不是太大，还需要一个参考模型 modelrefer；因此训练的过程中需要用到三个模型。而 DPO 方法是直接在偏好数据集上微调模型，不需要奖励模型和参考模型，是 RLHF 方法的一种简化方式。

第二种是通过增加推理时间，使得模型回答更为准确。最常见的就是让模型链接到外部知识库或搜索引擎，使得模型以获取到的外部知识作为参考来回答问题。而链接到外部知识库，需要用 BGE（Bidirectional Generative Embedding，双向生成嵌入）模块来提前对外部知识库的内容进行分割和向量化，以便模型能够高效地检索和利用这些知识。

对于上述第一种方式，其本质上只是让模型根据人类的偏好回答问题，并未实质上增强模型推理能力，并且标准的 RLHF 方法在训练时同时需要三个模型，对硬件有更高的要求。

对于上述第二种方式，其只是提供了更多的只是供模型参考，同样并未实质上增加模型的推理能力。并且推理过程需要接入 bge 模块，增加了额外的推理时间。

因此，如何通过增加模型的推理能力，来使得模型回答更为准确是亟需解决的问题。

基于上述技术问题的发现，发明人经过创造性劳动提出下述技术方案以解决或者改善上述问题。需要注意的是，以上现有技术中的方案所存在的缺陷，均是发明人在经过实践并仔细研究后得出的结果，因此，上述问题的发现过程以及下文中本申请实施例针对上述问题所提出的解决方案，都应该是发明人在发明创造过程中对本申请做出的贡献，而不应当理解为本领域技术人员所公知的技术内容。

本公开实施例提供一种问答大模型的后训练方法，能够获得对原问题及其原答案进行逻辑拆分得到的推理数据，利用若干份这样的推理数据构造推理数据集，从而利用该推理数据集对问答大模型进行后训练，使得问答大模型在后训练过程中逐渐学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

请参考图 1，图 1 为本公开实施例提供的一种问答大模型的后训练方法的流程示意图，该方法的执行主体可以是可以是但不限于智能手机、个人笔记本、个人计算机、服务器等计算设备。该问答大模型的后训练方法包括以下步骤：

S101、获取若干推理数据。

在本实施例中，每份推理数据可以是对一个原问题及其原答案进行逻辑拆分得到的。

S102、基于全部所述推理数据，构造推理数据集。

其中，推理数据集为配置成训练和评估问答大模型的逻辑推理、多步问题解决能力的结构化数据集合。

S103、利用所述推理数据集对问答大模型进行后训练处理。

其中，问答大模型，是一种基于大规模预训练的人工智能模型，用于理解和回答用户提出的自然语言，如 GPT 等。

在本实施例中，所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理，因此，在后训练阶段，问答大模型可以逐渐学会对输入问题进行对问题进行分步逻辑推理。

本公开实施例通过的问答大模型的后训练方法，首先获取若干推理数据，再基于全部推理数据，构造推理数据集；最后利用推理数据集对问答大模型进行后训练处理；由于每份推理数据是对原问题及其原答案进行逻辑拆分得到的，故推理数据集可以配置成引导问答大模型学习对输入问题进行分步推理。本公开能够使得问答大模型能够学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

可选的，可以收集若干个原问题及其原答案，然后通过人工拆分或者利用额外的逻辑拆分模型（如递进推理式）的方式，分别对每个原问题进行逻辑拆分为多个子问题及其子答案。其中，若是采用逻辑拆分模型，需要人为地对子问题及其子答案进行校验，保证不

出现逻辑错误。

因此，每份所述推理数据可以包括一个原问题、基于该原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案，其中，推理数据中第 N 个子答案即为原问题的原答案。

可选的示例中，假设原问题为：1+3×9+9÷2 的结果是多少，原答案为 32.5，进行拆分可以得到以下 4 个子问题及其子答案：

第一个子问题 Q1：3×9 的结果是多少，第一个子答案 A1：27；

第二个子问题 Q2：9÷2 的结果是多少，第二个子答案 A2：4.5；

第三个子问题 Q3：1+27 的结果是多少，第三个子答案 A3：28；

第四个子问题 Q4：28+4.5 的结果是多少，第四个子答案 A4：32.5。

该举例仅为一种简单的逻辑推理问题的拆分示例，在此不作限定。

可选的实现方式中，若是将问答大模型学会进行逻辑推理的过程进行拆分，那么希望问答大模型可以由浅入深地学会：递进式进行问题拆解、递进式进行子问题推理、完整地进行分步逻辑推理。并且，在一些场景下，需要问答大模型解答的逻辑问题可能是：抛出一个完整问题以及完整问题的部分子问题，需要推导出中间缺失问题及其答案。由此，本公开中构建的推理数据集可以包括第一子集、第二子集、第三子集和第四子集这四个子集，这四个子集分别采用四种不同的构造方式。

请参见图 2，即上述步骤 S102 中“基于全部所述推理数据，构造推理数据集”的过程可以包括以下子步骤 S1021~S1024。

S1021、针对每份所述推理数据，基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本。

在本实施例中，所述第一子集包括每份所述推理数据对应的全部第一样本。即，每份推理数据在第一子集中对应至少一个第一样本。

其中，第一子集是通过第一种构造方式获得的，第一种构造方式采用了递进推理式构造规则，具体的，对于每份推理样本，步骤 S1021 中“基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本”的过程可以包括以下子步骤 S10211~S10212：

S10211、将第一引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第一样本，并设置第一个第一样本的标签为所述推理数据中的第一个子答案；

S10212、将所述第一引导文本、所述推理数据中的原问题、第 n 个子问题以及第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第一样本，并设置第 n 个第一样本的标签为所述推理数据中的第 n 个子答案，$n = 2,3,\cdots,N$。

在本实施例中，预设的所述第一引导文本可以配置成指示所述问答大模型基于输入内容推理出所述输入内容中最后一个子问题的答案。其中，第一引导文本的实际文本内容在此不作限制。

S1022、针对每份所述推理数据，基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本。

在本实施例中，所述第二子集包括每份所述推理数据对应的全部第二样本。即，每份推理数据在第二子集中对应至少一个第二样本。

其中，第二子集是通过第二种构造方式获得的，第二种构造方式采用了递进拆解式构造规则，具体的，对于每份推理样本，步骤 S1022 中“基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本”的过程可以包括以下子步骤 S10221~S10222：

S10221、将第二引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第二样本，并设置第一个第二样本的标签为所述推理数据中的第二个子答案；

S10222、将所述第二引导文本、所述推理数据中的原问题、第 m 个子问题及其之前的每个子问题进行拼接得到第 m 个第二样本，并设置第 m 个第二样本的标签为所述推理数据中的第 m+1 个子问题，$m = 2,3,\cdots,N-1$。

在本实施例中，预设的所述第二引导文本可以配置成指示所述问答大模型基于输入内容从所述输入内容的原问题中拆解出所述输入内容中最后一个子问题的下一个子问题。其中，第二引导文本的实际文本内容在此不作限制。

S1023、针对每份所述推理数据，基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本。

在本实施例中，所述第三子集包括每份所述推理数据对应的全部第三样本。即，每份推理数据在第三子集中对应至少一个第三样本。

其中，第三子集是通过第三种构造方式获得的，第三种构造方式采用了中间截取是构造规则，具体的，对于每份推理样本，步骤 S1023 中“基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本”的过程可以包括以下子步骤 S10231~S10232：

S10231、将第三引导文本、所述推理数据中除了第一个子问题和第一个子答案以外的每个其他子问题和其他子答案进行拼接得到第一个第三样本，并设置第一个第三样本的标签为所述推理数据中的第一个子问题和第一个子答案；

S10232、将所述第三引导文本、所述推理数据中除了第 m 个子问题和第 m 个子答案以外的每个其他子问题和其他子答案进行拼接得到第 m 个第三样本，并设置第 m 个第三样本的标签为所述推理数据中的第 m 个子问题和第 m 个子答案，$m = 2,3,\cdots,N-1$。

在本实施例中，所述第三引导文本可以配置成指示所述问答大模型基于输入内容推理出所述输入内容中缺失的中间子问题及其中间子答案。其中，第三引导文本的实际文本内容在此不作限制。

S1024、针对每份所述推理数据，基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本。

在本实施例中，所述第四子集包括每份所述推理数据对应的全部第四样本。即，每份推理数据在第四子集中对应至少一个第四样本。

其中，第四子集是通过第四种构造方式获得的，第四种构造方式采用了分步推理式构造规则。具体的，对于每份推理样本，步骤 S1024 中“基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本”的过程可以包括以下子步骤 S10241~S10242：

S10241、将第四引导文本以及所述推理数据中的原问题进行拼接得到第一个第四样本，并设置第一个第四样本的标签为所述推理数据中的每个子问题及其子答案；

S10242、将所述第四引导文本、所述推理数据中的原问题、第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第四样本，并设置第 n 个第四样本的标签为所述推理数据中的第 n-1 个子问题之后的每个子问题及其子答案，$n = 2,3,\cdots,N$。

在本实施例中，预设的所述第四引导文本可以配置成指示所述问答大模型基于输入内容分步推理出后续的子问题和子答案。其中，第四引导文本的实际文本内容在此不作限制。

可选的示例中，假设一份推理数据中包括原问题、5 个子问题（Q1~Q5）及和 5 个子问题一一对应的 5 个子答案(A1~A5)。那么分别利用四种构造方式对该推理数据进行处理，得到的样本如下：

（1）利用第一种构造方式，可以得到如下表 1 所示的 5 个第一样本，其中，prompt1 即代表第一引导文本：

表 1

序号	第一样本	标签
1	prompt1+原问题+Q1	A_1
2	prompt1+原问题+（Q1+A1）+Q2	A_2
3	prompt1+原问题+（Q1+A1）+（Q2+A2）+Q3	A_3
4	prompt1+原问题+（Q1+A1）+（Q2+A2）+(Q3+A3)+Q4	A

		4
5	prompt1+原问题+（Q1+A1）+（Q2+A3）+(Q3+A3)+(Q4+A4)+Q5	A5

（2）利用第二种构造方式，可以得到如下表 2 所示的 4 个第二样本，其中，prompt2即代表第二引导文本：

表 2

序号	第二样本	标签
1	prompt2+原问题+Q1	Q2
2	prompt2+原问题+Q1+Q2	Q3
3	prompt2+原问题+Q1+Q2+Q3	Q4
4	prompt2+原问题+Q1+Q2+Q4	Q5

（3）利用第三种构造方式，可以得到如下表 3 所示的 4 个第三样本，其中，prompt3即代表第三引导文本：

表 3

序号	第三样本	标签
1	prompt3+原问题+（Q2+A2）+（Q3+A3）+（Q4+A4）+（Q5+A5）	Q1+A1
2	prompt3+原问题+（Q1+A1）+（Q3+A3）+（Q4+A4）+（Q5+A5）	Q2+A2
3	prompt3+原问题+（Q1+A1）+（Q2+A2）+Q4+A4）+（Q5+A5）	Q3+A3
4	prompt3+原问题+（Q1+A1）+（Q2+A2）+（Q3+A3）+（Q5+A5）	Q4+A4

（4）利用第四种构造方式，可以得到如下表 4 所示的 5 个第四样本，其中，prompt4即代表第四引导文本：

表 4

序号	第四样本	标签
1	prompt4+原问题	（Q1+A1）+（Q2+A2）+（Q3+A3）+（Q4+A4）+（Q5+A5）
2	prompt4+原问题+（Q1+A1）	（Q2+A2）+（Q3+A3）+（Q4+A4）+（Q5+A5）
3	prompt4+原问题+（Q1+A1）+（Q2+A2）	（Q3+A3）+（Q4+A4）+（Q5+A5）
4	prompt4+原问题+（Q1+A1）+（Q2+A2）+（Q3+A3）	（Q4+A4）+（Q5+A5）
5	prompt4+原问题+（Q1+A1）+（Q2+A2）+（Q3+A3）+（Q4+A4）	Q5+A5

上述举例仅为示例，本公开实施例对于一份推理数据中的子问题数量不作限制。

基于推理数据集中的四个子集，上述步骤 S103 的子步骤可以包括 S1031 或者 S1032：

S1031、分别利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集，对所述问答大模型进行后训练处理。

S1032、依次利用所述第二子集、所述第一子集、所述第四子集以及所述第三子集，对所述问答大模型进行后训练处理。

结合上述举例的四个表格内容可知，所述第一子集可以配置成引导所述问答大模型学习对所述输入问题进行分步解答，所述第二子集可以配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解，所述第三子集可以配置成引导所述问答大模型学习基于分步推

理的方式补全缺失的中间过程，所述第四子集可以配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解后进行分步解答。

所以，可以任意先后顺序利用 4 个子集来对问答大模型进行后训练处理。也可以依次利用第二子集、第一子集、第四子集以及第三子集，对所述问答大模型进行后训练处理，这样可以逐步让问答大模型学会：逻辑拆解、分步解答、逻辑拆解后进行分步解答以及基于分步推理的方式补全缺失的中间过程。或者，也可以利用第一子集、第二子集、第三子集以及第四子集这四个子集中的部分子集对问答大模型进行后训练处理，如依次利用第一子集、第二子集及第三子集对问答大模型进行后训练处理，以使得问答大模型学会能够学会：分步解答、逻辑拆解以及基于分步推理的方式补全缺失的中间过程。

可选的，进行后训练处理的过程中可以采用：监督微调（Supervised Fine-Tuning，SFT）、指令微调（Instruction Tuning）等方式。其中，后训练的具体过程为现有技术，在此不作赘述。

需要说明的是，上述方法实施例中各个步骤的执行顺序不以附图所示为限制，各步骤的执行顺序以实际应用情况为准。

与现有技术相比，本公开实施例具有以下有益效果：

本公开通过四种构造方式分别构造了四个子集来对问答大模型进行后训练处理，从而在本质上增强模型推理能力；

本公开中，第一子集以配置成引导所述问答大模型学习对所述输入问题进行分步解答，使得问答大模型学会基于原问题和之前拆分得到的子问题及其子答案对最新拆分的子问题进行解答；

本公共中，第二子集可以配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解，使得问答大模型学会基于原问题和以及拆分出来的子问题继续拆分出下一个子问题；

本公共中，所述第三子集可以配置成引导所述问答大模型学习基于分步推理的方式补全缺失的中间过程，使得问答大模型学会分步推理来解答中间步骤缺失的推理问题；

本公开中，所述第四子集可以配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解后进行分步解答，使得问答大模型学会完整的对输入问题进行分步逻辑推理。

为了执行上述方法实施例及各个可能的实施方式中的相应步骤，下面给出一种问答大模型的后训练装置的实现方式。

请参见图 3，图 3 示出了本公开实施例提供的问答大模型的后训练装置的结构示意图。该问答大模型的后训练装置 200 包括：数据获取模块 210、构造模块 220 以及后训练模块 230。

数据获取模块 210，配置成获取若干推理数据，所述推理数据是对原问题及其原答案进行逻辑拆分得到的；

构造模块 220，配置成基于全部所述推理数据，构造推理数据集；

后训练模块 230，配置成利用所述推理数据集对问答大模型进行后训练处理；所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理。

可选的，所述推理数据集包括第一子集、第二子集、第三子集和第四子集。构造模块 220 在配置成基于全部所述推理数据，构造推理数据集的过程中，具体可以配置成：针对每份所述推理数据，基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本；所述第一子集包括每份所述推理数据对应的全部第一样本；针对每份所述推理数据，基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本，所述第二子集包括每份所述推理数据对应的全部第二样本；针对每份所述推理数据，基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本，所述第三子集包括每份所述推理数据对应的全部第三样本；针对每份所述推理数据，基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本，所述第四子集包括每份所述推理数据对应的全部第四样本。

可选的，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个

子问题以及每个子问题的子答案；

构造模块 220 在配置成所述基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本的过程中，具体可以配置成：

将第一引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第一样本，并设置第一个第一样本的标签为所述推理数据中的第一个子答案；所述第一引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中最后一个子问题的答案；

将所述第一引导文本、所述推理数据中的原问题、第 n 个子问题以及第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第一样本，并设置第 n 个第一样本的标签为所述推理数据中的第 n 个子答案，$n = 2,3,\cdots,N$。

可选的，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

构造模块 220 在配置成基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本的过程中，具体可以配置成：

将第二引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第二样本，并设置第一个第二样本的标签为所述推理数据中的第二个子答案；所述第二引导文本配置成指示所述问答大模型基于输入内容从所述输入内容的原问题中拆解出所述输入内容中最后一个子问题的下一个子问题；

将所述第二引导文本、所述推理数据中的原问题、第 m 个子问题及其之前的每个子问题进行拼接得到第 m 个第二样本，并设置第 m 个第二样本的标签为所述推理数据中的第 m+1 个子问题，$m = 2,3,\cdots,N-1$。

可选的，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

构造模块 220 在配置成所述基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本的过程中，具体可以配置成：

将第三引导文本、所述推理数据中除了第一个子问题和第一个子答案以外的每个其他子问题和其他子答案进行拼接得到第一个第三样本，并设置第一个第三样本的标签为所述推理数据中的第一个子问题和第一个子答案；所述第三引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中缺失的中间子问题及其中间子答案；将所述第三引导文本、所述推理数据中除了第 m 个子问题和第 m 个子答案以外的每个其他子问题和其他子答案进行拼接得到第 m 个第三样本，并设置第 m 个第三样本的标签为所述推理数据中的第 m 个子问题和第 m 个子答案，$m = 2,3,\cdots,N-1$。

可选的，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

构造模块 220 在配置成所述基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本的过程中，具体可以配置成：

将第四引导文本以及所述推理数据中的原问题进行拼接得到第一个第四样本，并设置第一个第四样本的标签为所述推理数据中的每个子问题及其子答案；所述第四引导文本配置成指示所述问答大模型基于输入内容分步推理出后续的子问题和子答案；

将所述第四引导文本、所述推理数据中的原问题、第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第四样本，并设置第 n 个第四样本的标签为所述推理数据中的第 n-1 个子问题之后的每个子问题及其子答案，$n = 2,3,\cdots,N$。

可选的，所述推理数据集包括第一子集、第二子集、第三子集和第四子集；

后训练模块 230 在配置成所述利用所述推理数据集对问答大模型进行后训练处理的过程中，具体可以配置成：

分别利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集，对所述问答大模型进行后训练处理；其中，所述第一子集配置成引导所述问答大模型学习对所述输

入问题进行分步解答，所述第二子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解，所述第三子集配置成引导所述问答大模型学习基于分步推理的方式补全缺失的中间过程，所述第四子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解后进行分步解答。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的问答大模型的后训练装置 200 的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

请参见图 4，图 4 为本公开实施例提供的一种电子设备的结构示意图。该电子设备 300 包括处理器 310、存储器 320 和总线 330，处理器 310 通过总线 330 与存储器 320 连接。

存储器 320 可配置成存储软件程序，例如，如本公开实施例所提供的问答大模型的后训练装置 200 对应的软件程序。处理器 310 通过运行存储在存储器 320 内的软件程序，从而执行各种功能应用以及数据处理以实现如本公开实施例所提供的问答大模型的后训练方法。

其中，存储器 320 可以是但不限于：RAM(Random Access Memory，随机存取存储器)，ROM(Read Only Memory，只读存储器)，FLASH（闪存存储器），PROM(Programmable Read-Only Memory，可编程只读存储器)，EPROM(Erasable Programmable Read-Only Memory，可擦除只读存储器)，EEPROM(Electric Erasable Programmable Read-Only Memory，电可擦除只读存储器)等。

处理器 310 可以是一种集成电路芯片，具有信号处理能力。该处理器 310 可以是通用处理器，包括：CPU(Central Processing Unit，中央处理器)、NP(Network Processor，网络处理器)、SoC（System on Chip，系统级芯片）等；还可以是：DSP(Digital Signal Processing，数字信号处理器)、ASIC(Application Specific Integrated Circuit，专用集成电路)、FPGA(Field-Programmable Gate Array，现场可编程门阵列)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。

可以理解，图 4 所示的结构仅为示意，电子设备 300 还可以包括比图 4 中所示更多或者更少的组件，或者具有与图 4 所示不同的配置。图 4 中所示的各组件可以采用硬件、软件或其组合实现。

本公开实施例还提供一种计算机可读存储介质，该计算机可读存储介质上存储有计算机程序，该计算机程序被处理器运行时实现上述实施例揭示的问答大模型的后训练方法。该计算机可读存储介质可以是但不限于：U 盘、移动硬盘、ROM、RAM、PROM、EPROM、EEPROM、FLASH 磁碟或者光盘等各种可以存储程序代码的介质。

综上所述，本公开实施例提供了一种问答大模型的后训练方法、装置、电子设备及存储介质，首先获取若干推理数据，再基于全部推理数据，构造推理数据集；最后利用推理数据集对问答大模型进行后训练处理；由于每份推理数据是对原问题及其原答案进行逻辑拆分得到的，故推理数据集可以配置成引导问答大模型学习对输入问题进行分步推理。本公开能够使得问答大模型能够学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

以上对本说明书一些实施例所提供的一种问答大模型的后训练方法、装置以及相关电子设备、计算机可读存储介质、计算机程序产品进行了详细介绍，本说明书应用了具体个例对本技术方案的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本技术方案的方法及其核心思想；同时，对于本领域的技术人员，依据本技术方案的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本技术方案的限制。

工业实用性

采用上述方案，在训练问答大模型的过程中，通过获取若干推理数据，再基于全部推理数据，构造推理数据集；最后利用推理数据集对问答大模型进行后训练处理；由于每份

推理数据是对原问题及其原答案进行逻辑拆分得到的，故推理数据集可以配置成引导问答大模型学习对输入问题进行分步推理，能够使得问答大模型能够学会对问题进行分步推理，从而在根本上提升了模型的逻辑推理能力。

权利要求书

1.一种问答大模型的后训练方法，其特征在于，包括：

获取若干推理数据，所述推理数据是对原问题及其原答案进行逻辑拆分得到的；

基于全部所述推理数据，构造推理数据集；

利用所述推理数据集对问答大模型进行后训练处理；所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理。

2. 根据权利要求 1 所述的方法，其特征在于，所述推理数据集包括：第一子集、第二子集、第三子集和第四子集；

所述基于全部所述推理数据，构造推理数据集，包括：

针对每份所述推理数据，基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本；第一子集包括每份所述推理数据对应的全部第一样本；

针对每份所述推理数据，基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本；第二子集包括每份所述推理数据对应的全部第二样本；

针对每份所述推理数据，基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本；第三子集包括每份所述推理数据对应的全部第三样本；

针对每份所述推理数据，基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本；第四子集包括每份所述推理数据对应的全部第四样本。

3. 根据权利要求 2 所述的方法，其特征在于，所述第一子集配置成引导所述问答大模型学习对所述输入问题进行分步解答，所述第二子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解，所述第三子集配置成引导所述问答大模型学习基于分步推理的方式补全缺失的中间过程，所述第四子集配置成引导所述问答大模型学习对所述输入问题进行逻辑拆解后进行分步解答。

4. 根据权利要求 3 所述的方法，其特征在于，所述利用所述推理数据集对问答大模型进行后训练处理，包括：

分别利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集，对所述问答大模型进行后训练处理。

5. 根据权利要求 3 所述的方法，其特征在于，所述利用所述推理数据集对问答大模型进行后训练处理，包括：

利用所述第一子集、所述第二子集、所述第三子集以及所述第四子集的先后顺序，对所述问答大模型进行后训练处理；所述先后顺序包括：依次排列顺序或随机排列顺序。

6.根据权利要求 2 所述的方法，其特征在于，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用递进推理式构造规则获得至少一个第一样本的步骤，包括：

将第一引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第一样本，并设置第一个第一样本的标签为所述推理数据中的第一个子答案；所述第一引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中最后一个子问题的答案；

将所述第一引导文本、所述推理数据中的原问题、第 n 个子问题以及第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第一样本，并设置第 n 个第一样本的标签为所述推理数据中的第 n 个子答案，n=2,3,…,N。

7.根据权利要求 2 所述的方法，其特征在于，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用递进拆解式构造规则获得至少一个第二样本的步骤，包括：

将第二引导文本、所述推理数据中的原问题和第一个子问题进行拼接得到第一个第二

样本，并设置第一个第二样本的标签为所述推理数据中的第二个子答案；所述第二引导文本配置成指示所述问答大模型基于输入内容从所述输入内容的原问题中拆解出所述输入内容中最后一个子问题的下一个子问题；

将所述第二引导文本、所述推理数据中的原问题、第 m 个子问题及其之前的每个子问题进行拼接得到第 m 个第二样本，并设置第 m 个第二样本的标签为所述推理数据中的第 m+1 个子问题，m=2,3,…,N-1。

8.根据权利要求 2 所述的方法，其特征在于，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用中间截取式构造规则获得至少一个第三样本的步骤，包括：

将第三引导文本、所述推理数据中除了第一个子问题和第一个子答案以外的每个其他子问题和其他子答案进行拼接得到第一个第三样本，并设置第一个第三样本的标签为所述推理数据中的第一个子问题和第一个子答案；所述第三引导文本配置成指示所述问答大模型基于输入内容推理出所述输入内容中缺失的中间子问题及其中间子答案；

将所述第三引导文本、所述推理数据中除了第 m 个子问题和第 m 个子答案以外的每个其他子问题和其他子答案进行拼接得到第 m 个第三样本，并设置第 m 个第三样本的标签为所述推理数据中的第 m 个子问题和第 m 个子答案，m=2,3,…,N-1。

9.根据权利要求 2 所述的方法，其特征在于，所述推理数据包括原问题、基于所述原问题拆分得到的逻辑层次递进的 N 个子问题以及每个子问题的子答案；

所述基于所述推理数据，利用分步推理式构造规则获得至少一个第四样本的步骤，包括：

将第四引导文本以及所述推理数据中的原问题进行拼接得到第一个第四样本，并设置第一个第四样本的标签为所述推理数据中的每个子问题及其子答案；所述第四引导文本配置成指示所述问答大模型基于输入内容分步推理出后续的子问题和子答案；

将所述第四引导文本、所述推理数据中的原问题、第 n 个子问题之前的每个子问题及其子答案进行拼接得到第 n 个第四样本，并设置第 n 个第四样本的标签为所述推理数据中的第 n-1 个子问题之后的每个子问题及其子答案，n=2,3,…,N。

10.根据权利要求 1-9 任一项所述的方法，其特征在于，所述后训练处理的过程中包括：监督微调方式或指令微调方式。

11.一种问答大模型的后训练装置，其特征在于，包括：

数据获取模块，配置成获取若干推理数据，所述推理数据是对原问题及其原答案进行逻辑拆分得到的；

构造模块，配置成基于全部所述推理数据，构造推理数据集；

后训练模块，配置成利用推理数据集对所述问答大模型进行后训练处理，所述推理数据集配置成引导所述问答大模型学习对输入问题进行分步推理。

12.一种电子设备，其特征在于，包括：存储器和处理器，所述存储器存储有软件程序，当所述电子设备运行时所述处理器执行所述软件程序以实现如权利要求 1-10 中任一项所述的问答大模型的后训练方法。

13.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现权利要求 1-10 中任一项所述的问答大模型的后训练方法。