

文本处理方法、装置、电子设备和存储介质

相关申请

本申请要求 2023 年 09 月 18 日申请的，申请号为 2023112010037，名称为“文本处理方法、装置、电子设备和存储介质”的中国专利申请的优先权，在此将其全文引入作为参考。

技术领域

本申请涉及人工智能技术领域，具体涉及一种文本处理方法、装置、电子设备和存储介质。

背景技术

近年来，随着人工智能技术的快速发展，基于人工智能技术的智能语音助手产品已经广泛地应用到了人们的日常生活中，通过智能语音助手产品，人们能够仅使用语音就实现对智能设备的控制。其中，口语理解是智能语音助手产品的核心算法能力。

在实际应用中，智能语音助手产品在利用口语理解来梳理用户提出的问题，通常会面临多种口语理解场景，例如多意图口语理解场景、单意图口语理解场景等。其中，多意图口语理解场景，是指从口语的一句话中识别出多个意图的场景，而单意图口语理解场景则是从口语的一句话中识别出单个意图的场景。

然而，在相关技术中，智能语音助手产品的系统中通常会设置多个单独的模型来分别支持上述多种口语理解问题场景，例如，多个模型中的每个模型可以专门处理一种口语理解问题场景下产生的文本，从而导致智能语音助手产品的系统模型复杂，且系统中模型的训练以及持续迭代成本较高。

发明内容

本申请实施例提供一种文本处理方法、装置、电子设备和存储介质。

一方面，提供一种文本处理方法，由电子设备执行，包括：

获取表达口语请求的口语请求文本；

通过预测模型，利用所述口语请求文本，预测所述口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本；及

通过所述预测模型根据所述子请求的数量 N 以及 N 个子请求文本，预测所述 N 个子请求文本各自对应的目标指令信息，所述目标指令信息包括意图、领域以及槽位，所述 N 个子请求文本各自对应的目标指令信息，用于确定对于所述口语请求的响应。

另一方面，本申请实施例还提供一种文本处理装置，包括：

获取单元，用于获取表达口语请求的口语请求文本；

第一预测单元，用于通过预测模型，利用所述口语请求文本，预测所述口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本，其中， N 为正整数；及

第二预测单元，用于通过所述预测模型根据所述子请求的数量 N 以及 N 个子请求文本，预测所述 N 个子请求文本各自对应的目标指令信息，所述目标指令信息包括意图、领域以及槽位，所述 N 个子请求文本各自对应的目标指令信息，用于确定对于所述口语请求的响应。

再一方面，本申请实施例还提供一种电子设备，包括存储器存储有多条指令；所述处理器从所述存储器中加载指令，以执行本申请实施例所提供的任一种文本处理方法中的步骤。

再一方面，本申请实施例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多条指令，所述指令适于处理器进行加载，以执行本申请实施例所提供的任一种文本处理方法中的步骤。

再一方面，本申请实施例还提供一种计算机程序产品，包括计算机指令，计算机指令被处理器执行时，实现本申请实施例所提供的任一种文本处理方法中的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例或传统技术中的技术方案，下面将对实施例或传统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申请的实施例，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，还可以根据公开的附图获得其他的附图。

图 1 是本申请实施例提供的相关技术中多场景口语理解流程示意图；

图 2 是本申请实施例提供的文本处理方法的应用场景示意图；

图 3 是本申请实施例提供了一种文本处理方法的流程示意图；

图 4 是本申请实施例提供的构造训练数据集的流程示意图；

图 5 是本申请实施例提供的对样本文本的更新流程示意图；

图 6 是本申请实施例提供的另一种文本处理方法的流程示意图；

图 7 是本申请实施例提供的多轮口语理解场景下对样本文本的合并流程示意图；

图 8 是本申请实施例提供的单意图场景下对样本文本的拼接流程示意图；

图 9 是本申请实施例提供的生成式大规模语言模型的输入和输出示意图；

图 10 是本申请实施例提供的负样本训练的损失函数示意图；

图 11 是本申请实施例提供的预测模型的训练过程和推理过程示意图；

图 12 是本申请实施例提供的文本处理装置的结构示意图；

图 13 是本申请实施例提供的电子设备的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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，都属于本申请保护的范围。

由于口语理解是智能语音助手产品的核心算法能力，其在智能家居、智能音箱、智能车机控制等诸多领域起着关键性作用，是提升用户体验的重要环节。因此，口语理解技术如今已被广泛应用在各种具有语音助手的智能设备中。

然而，口语理解在实际运用中也存在很多的难点，比如庞大的标签系统、复杂多变的用户口语表达、多轮的用户连续请求、否定的用户指令等。针对以上问题，在相关技术中，主要是采用设置多个子系统来独立支持。由于设置了多个模型，导致了其训练和迭代成本较高。另外，系统迁移能力也比较弱，例如切换语言（如切换外语、方言等）时，需要重新训练部署多个模型。

作为一种示例，如图 1 所示，以相关技术中采用多子系统的模式来解决单意图口语理解（SLU，Spoken Language Understanding 的缩写）场景、一语多意图口语理解场景以及多轮次口语理解场景中的用户口语请求（以下可简称口语请求）为例，其中，整体的口语理解系统可以包括多个子系统，多个子系统分别为系统 1：一语多意图口语理解模块、系统 2：单意图口语理解模块、以及系统 3：多轮口语理解模块。

其中，系统 1 中包括一语多意图判别模型和多意图口语理解模型，一语多意图判别模型用于判别接收到的用户口语请求是否具有多个意图，多意图口语理解模型用于根据多意图的用户口语请求输出相应的口语理解结果。其中，单意图口语理解模型可以实现 3 种单意图口语理解方法，分别是基于规则的模板口语理解，预料匹配口语理解，还有基于 BERT 深度模型的口语理解识别结果。模板口语理解可以为用户口语请求匹配模板对应的口语理解结果。预料匹配口语理解可以为用户口语匹配语料库中的口语理解结果。

举例说明，若用户口语请求为：附近最近的交通银行在哪，一语多意图判别模型用于判别该用户口语请求是否具有多个意图，根据该用户口语请求，多意图口语理解模型可用

于输出相应的口语理解结果，包括领域、意图和槽位。其中，领域为导航，意图为地点查询，槽位包括排序类别和地点类别，排序类别为附近和/或最近，地点类别为交通银行。

系统2中包括单意图口语理解模型，单意图口语理解模型用于根据单意图的用户口语请求输出相应的口语理解结果。

系统3中包括多轮意图判别模型和多轮改写模型。多轮意图判别模型用于判别接收到的用户口语请求是否是多轮次意图请求，例如当前轮次的口语请求是否依赖于之前轮次的口语请求。多轮改写模型用于将多轮意图的用户口语请求改写为单意图的用户口语请求，例如将当前轮次的口语请求改写成和在当前轮次的之前轮次的口语请求无关的完整文本。

可见，在相关技术中采用多子系统的模式，系统模型复杂，系统的部署以及持续迭代成本高，系统其他语言迁移能力弱。此外，口语理解识别能力较差，针对冷启动、小样本场景口语理解识别能力较弱。另外，多轮口语理解能力弱，只能处理2轮的请求，实际运用场景较局限。

针对上述问题，本申请实施例提供一种文本处理方法、装置、电子设备和存储介质。

其中，该文本处理装置具体可以集成在电子设备中，该电子设备可以为终端、服务器等设备。其中，终端可以包括但不限于手机、电脑、智能语音交互设备、智能家电、车载终端、飞行器等；服务器可以是单一服务器，也可以是由多个服务器组成的服务器集群。

在一些实施例中，该文本处理装置还可以集成在多个电子设备中，比如，文本处理装置可以集成在多个服务器中，由多个服务器协作来实现本申请的文本处理方法。在一些实施例中，终端也可以用作服务器，实现服务器部分或全部的功能。

作为一种示例，参考图2，图2示出了本申请实施例提供的文本处理方法的一种应用场景示意图。

如图2所示，该应用场景可以包括电子设备10和智能设备20，该文本处理方法可以应用于该电子设备10中，该电子设备10包括但不限于手机、平板电脑、智能蓝牙设备、笔记本电脑、个人电脑、车载终端、智能可穿戴设备等等。其中，该电子设备10可以与智能设备20通信连接。

其中，该智能设备20可以为用户当前需要通过语音进行控制的对象，该智能设备20可以包括但不限于智能音箱、智能家居设备、智能汽车等等。可选地，智能家居设备包括但不限于：智能晾衣架、智能冰箱、智能电视、智能照明设备等等。其中，智能设备20的数量可以一个或多个，对此不做限定。

其中，该电子设备10可以获取表达口语请求的口语请求文本，并通过预测模型预测口语请求的子请求的数量N以及N个子请求文本，其中，N为正整数，然后通过预测模型根据子请求的数量N以及N个子请求文本，预测N个子请求文本各自对应的目标指令信息，目标指令信息包括意图、领域以及槽位，N个子请求文本各自对应的目标指令信息，用于确定对于口语请求的响应。在实际应用中，当电子设备10接收用户针对智能设备20的目标语音时，可以将目标语音识别为口语请求文本，并将口语请求文本输入到预测模型中，以得到口语请求文本对应的目标指令信息，再将与目标指令信息对应的控制指令发送至智能设备20，以控制智能设备20执行相应的动作。

在一些实施方式中，用户当前需要通过语音进行控制的对象也可以是电子设备10本身，例如，电子设备10为手机，用户可以通过目标语音向手机进行提问，当手机接收到目标语音后，可以将目标语音识别为口语请求文本，并将口语请求文本输入到预测模型中，以得到口语请求文本对应的目标指令信息，此后，该手机可以查询到与目标指令信息对应的控制指令，基于该控制指令回答用户的提问。

以下对本申请所涉及的一些概念进行说明。

人工智能（Artificial Intelligence，AI）是一种利用数字计算机来模拟人类感知环境、获取知识并使用知识的技术，该技术可以使机器具有类似于人类的感知、推理与决策的功能。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据

处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习、自动驾驶、智慧交通等几大方向。

语音技术(Speech Technology)的关键技术有自动语音识别技术和语音合成技术以及声纹识别技术。让计算机能听、能看、能说、能感觉，是未来人机交互的发展方向，其中语音成为未来最被看好的人机交互方式之一。

自然语言处理(Nature Language processing, NLP)是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言，所以它与语言学的研究有着密切的联系。自然语言处理技术通常包括文本处理、语义理解、机器翻译、机器人问答、知识图谱等技术。

随着人工智能技术研究和进步，人工智能技术在多个领域展开研究和应用，例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服、车联网、自动驾驶、智慧交通等，相信随着技术的发展，人工智能技术将在更多的领域得到应用，并发挥越来越重要的价值。

口语理解 (Spoken Language Understanding, 口语理解)：理解人类的口头语言中的指令请求信息，包括指令的领域、意图和槽位。之后用于控制机器（如智能家居、车机控制、智能音箱等）来执行相应的指令。

BERT，全称为 Bidirectional Encoder Representations from Transformers，是一种基于 Transformer 架构的预训练语言模型，它通过大规模的无监督训练从大量文本数据中学习语言知识，然后可以用于各种自然语言处理任务。

生成式大规模语言模型 (Large Language Model, LLM)：指使用大量文本数据训练的深度学习生成模型，可以生成自然语言文本或理解语言文本的含义。大语言模型可以处理多种自然语言任务，如文本分类、问答、对话等，是通向人工智能的一条重要途径。常见的大语言模型有 ChatGPT, Llama, Bloom 等。

思维链生成 (Chain-of-Thought, CoT)：通过训练模型学习根据人类的思维方式，将复杂的问题拆解成多个子步骤，一步一步推理最后得出正确的结果的方法。能够促进模型面对复杂问题的推理能力。

其中，口语请求可以为用户发送给机器设备的口头语言请求，例如语音指令或口头提问，通过语音指令或口头提问，用户可以与智能助手或语音交互系统进行交互，并根据用户的需求实现各种功能。口语请求可以包括但不限于：单意图请求、一语多意图请求以及多轮次意图请求等。

单意图请求：发送给机器设备的一轮次的口头语言请求，且该口头语言请求中只包含一个领域下的一个意图，例如，口头语言请求：“我要打开空调”，可见，该口头语言请求包括“设备控制”领域下的一个意图“打开空调”。

一语多意图请求：发送给机器设备的一轮次的口头语言请求，且请求中只包含 2 个或 2 个以上的领域和意图组合，具体可以包括 2 个或 2 个以上的领域下各自的意图，或者一个领域下 2 个或 2 个以上的意图。例如，口头语言请求：“我要打开空调并导航去公司”，可见，该口头语言请求包括“设备控制”领域下的一个意图“打开空调”以及“导航”领域下的一个意图“导航去公司”。

多轮次意图请求：发送给机器设备的多个轮次的口头语言请求，当前轮次的口语请求和在当前轮次之前的其他口语请求有关联，例如，多轮次意图请求包括其它轮次的口语请求：“今天天气怎么样”，当前轮次的请求：“明天呢？”。

其中，可以理解的是，在本申请的具体实施方式中，涉及到用户口语请求、训练数据集、样本文本、目标语音、口语请求文本等相关数据，当本申请以下实施例运用到具体产

品或技术中时，需要获得许可或者同意，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。

在本实施例中，提供了一种涉及人工智能技术的文本处理方法，该方法可以应用于图2中的电子设备，如图3所示，该文本处理方法的具体流程可以如下：

301、获取表达口语请求的口语请求文本。

其中，口语请求文本，是表达口语请求的文本。其中，口语请求表示通过口语表达的请求，该口语请求可以通过语音形式表达。该口语请求可以是针对可控对象的，用于请求可控对象执行一些动作。其中，口语请求可以包括一个或多个子请求，在此不做限定。其中，可控对象是可以被控制的对象，可以包括车载终端、移动电话、远程服务器或者传感器等，例如，可控对象可以是车载终端，口语请求可以是请求车载终端显示天气信息，也可以是请求车载终端播放导航信息，还可以是请求车载终端控制车内温度等。

可选地，口语请求文本可以是对该口语请求进行转换而得到的文本。其中，语音场景可以是指口语理解场景，在口语理解场景中，电子设备可以接收到用户的口语请求，并通过自然语言处理技术进行解析和理解，从而进行相应操作或提供准确的回答。在本实施例中，语音场景可以包括但不限于单意图请求场景、多意图请求场景以及多轮请求场景。

其中，口语请求文本为单意图请求场景下口语请求的文本、多意图请求场景下口语请求的文本或多轮请求场景中一轮口语请求的文本。

其中，单意图请求场景（以下也可称单意图场景）的文本包括单意图请求内容，单意图请求内容表征包含一个意图的请求内容。

作为一种示例，单意图请求内容为“我要打开空调”，其中，该请求内容中只包括一个意图“打开空调”。

其中，多意图请求场景（以下也可称多意图场景）的文本包括多意图请求内容，多意图请求内容表征包含多个意图的请求内容。

其中，多意图请求可以参考上述一语多意图请求。作为一种示例，多意图请求内容可以为“我要打开空调并导航去公司”，其中，该请求内容中包括意图“打开空调”以及意图“我要去公司”。

其中，多轮请求场景（以下也可称多轮次场景）的文本包括当前轮的文本和当前轮的之前轮次的文本。

可以理解的是，多轮请求可以参考上述多轮次意图请求。作为一种示例，例如，用户发出第一轮口语请求：“我想预订一张明天下午从A地点到B地点的机票。”；电子设备：“好的，请问您需要经济舱还是商务舱？”；用户发出第二轮（如当前轮次）口语请求：“经济舱”的场景。其中，第一轮口语请求对应的文本为当前轮的之前轮次的文本。第二轮口语请求对应的文本为当前轮的文本。其中，单意图场景是指用户发出的口语请求为上述单意图请求的场景。多意图场景是指用户发出的口语请求为上述一语多意图请求的场景。

302、通过预测模型，利用口语请求文本，预测口语请求的子请求的数量N以及N个子请求一一对应的N个子请求文本，N为正整数。

其中，口语请求可以包括一个或多个子请求，该子请求可以用于请求上述可控对象执行一些相应的动作。

其中，预测模型可以是预先通过思维链生成训练得到的，由于思维链生成是通过训练模型学习根据人类的思维方式，将复杂的问题拆解成多个子步骤，一步一步推理最后得出正确的结果的方法，所以在本实施例中，可以通过思维链生成训练，使得预测模型学习如何预测口语请求文本中的子请求的数量N，并将口语请求文本拆解成N个子请求文本。

303、通过预测模型根据子请求的数量N以及N个子请求文本，预测N个子请求文本各自对应的目标指令信息，目标指令信息包括意图、领域以及槽位，N个子请求文本各自对应的目标指令信息，用于确定对于口语请求的响应。

其中，预测模型可以是基于生成式大规模语言模型训练得到，因此结合预设指示信息

进行训练，可以是训练得到的预测模型根据子请求的数量 N 以及 N 个子请求文本，预测 N 个子请求文本各自对应的目标指令信息。其中，指示信息可以指示生成式大规模语言模型输出包括意图、领域以及槽位的目标指令信息。可以理解的是，预测模型可以针对 N 个子请求文本中的每个子请求文本都预测出一个目标指令信息。

在一些实施方式中，在步骤 302 之前，该文本处理方法还可以包括预测模型的训练方法，该预测模型的训练方法可以包括如下步骤：

S1、获取多种场景下的训练数据集，训练数据集包括每一场景对应的至少一个样本文本，多种场景包括单意图请求场景、多意图请求场景或多轮请求场景中的至少两个。

其中，样本文本可以是用户的口语请求经过语音识别转换而得到的文本，因此，口语请求对应的语音场景也就是该口语请求转换得到的样本文本对应的语音场景。

在一些实施方式中，获取多种语音场景下的训练数据集的具体实施方式可以包括：获取多个口语请求样本，并标注每个口语请求样本的类型，然后，将每个口语请求样本通过语音识别转换为相应的样本文本，此后，再根据样本文本对应的口语请求样本的类型，将样本文本与语音场景建立对应关系。例如，当样本文本对应的口语请求的类型为一语多意图请求时，可以将该样本文本与多意图请求场景建立对应关系。又例如，当样本文本对应的口语请求的类型为单意图请求时，可以将该样本文本与单意图请求场景建立对应关系。再例如，当样本文本对应的口语请求的类型为多轮次意图请求时，可以将该样本文本与多轮请求场景建立对应关系，从而得到多种语音场景下的训练数据集。其中，多个口语请求样本可以是电子设备记录的历史对话记录中提取出的，也可以是电子设备根据预设的口语请求模板自动生成的，在此不做限定。

在一些实施方式中，在步骤 S1 中，获取多种场景下的训练数据集的具体实施方式可以包括：

S11、获取首轮对话模板集合和至少一个次轮对话模板集合，首轮对话模板集合中的每个首轮对话模板和次轮对话模板集合中的每个次轮对话模板均包括空白槽位。

其中，首轮对话模板集合中包括至少一个首轮对话模板，首轮对话模板可以是预先标注的包括空白槽位的问句模板。其中，空白槽位可以看作是在对话中留下的待填写或待补充的位置。这些位置需要根据上下文或问题提供的提示来填写合适的信息，如图 4 所示，首轮对话模板可以为“[location][datetime]天气怎么样”，其含义为某时某地的天气怎么样，其中，[location]和[datetime]为空白槽位，[location]表示位置、地址，[datetime]表示日期和/或时间。

其中，每个次轮对话模板集合中包括至少一个次轮对话模板，次轮对话模板可以包括指代模板、省略模板等，其中，省略模板用于在对话中省略重复信息。当某个信息已经在上下文中明确提及过时，可以通过省略模板来简化对话，只提供新的或变化的信息。其中，指代模板用于在对话中使用特定词语或短语来替代之前提到的信息，以减少重复和提高交流效率。沿用上述示例，例如，次轮对话模板可以为“[location]呢”，结合上述示例中的首轮对话模板“[location][datetime]天气怎么样”来看，该次轮对话模板是在首轮对话模板的基础上省略了“天气怎么样”而得到省略模板。其中，请再次参阅图 4，空白槽位除了[location]和[datetime]以外，还可以包括[activity]、[description]等等，[activity]表示活动等，[description]表示描述对象，示例性地，例如次轮对话模板集合中包括“适合[activity]吗”，其含义是适合某个活动吗，又例如次轮对话模板集合中包括“[datetime]会有[description]吗”，其含义是某时会有某个描述对象吗。其中，描述对象可以为雪、雨等。

在一些实施方式中，首轮对话模板集合和至少一个次轮对话模板集合可以预先标注好后存储在预设模板数据库中，在电子设备需要通过预设模板数据库中的对话模板来构造多轮次场景下的训练数据时，可以从预设模板数据库中调取出需要的首轮对话模板集合和至少一个次轮对话模板集合。其中，该预设模板数据库可以设置在电子设备的本地，也可以设置在与电子设备通信连接的云端设备中，对此不做限定。

S12、将首轮对话模板集合中的每个首轮对话模板分别与次轮对话模板集合中的每个次轮对话模板进行组合，得到多个对话组合。

示例性地，请再次参阅图 4，以一个首轮对话模板和两个次轮对话模板集合（次轮对话模板 1 和次轮对话模板 2）为例，其中，首轮对话模板集合包括 7 个首轮对话模板，次轮对话模板集合 1 包括 6 个次轮对话模板，次轮对话模板集合 2 包括 6 个次轮对话模板，首轮对话模板集合中的每个首轮对话模板分别与次轮对话模板集合 1 中的 6 个对轮对话模板进行组合，可以得到 42 个对话组合。此后，这 42 个对话组合还可以再分别与次轮对话模板集合 2 中的 6 次轮对话模板进行组合，从而得到 42*6 个对话组合。以此类推，当次轮对话模板集合超过 2 个时，可以通过上述方式进行组合，以便获得更多的对话组合。

在一些实施方式中，在将首轮对话模板和次轮对话模板进行组合时，可以先设定好每一轮对话模板集合的对话轮次顺序（例如第二轮次、第三轮次等），然后按照该对话轮次顺序，将每一轮对话模板集合中的次轮对话模板与首轮对话模板进行组合。例如，从首轮对话模板集合（如图 4 中的第 1 轮）中选取的首轮对话模板为“[location][datetime]天气怎么样”，从第二轮次的次轮对话模板集合（如图 4 中的第 2 轮）中选取的次轮对话模板为“[location]呢”，从第三轮次的次轮对话模板集合（如图 4 中的第 3 轮）中选取的次轮对话模板为“[datetime]呢”，将这三轮对话模板进行组合，得到的对话组合为：第一轮对话“[location][datetime]天气怎么样”，第二轮对话“[location]呢”，第三轮对话“[datetime]呢”。

S13、基于对话组合中空白槽位，从预设实体库中提取出实体信息。

其中，实体信息在自然语言处理技术中表示关键性的信息，通常可以分为两种，一是与业务基本无关的，可以算作通识的信息，比如手机号、邮箱、日期、时间、地址等；二是与业务相关的，根据实际场景定制的。

在一些实施方式中，可以预先将划分了实体类别（人名、地名、日期、机构名或其他）实体信息存储至预设实体库，然后根据空白槽位表示的含义，从实体库中选取对应实体类别的实体信息，例如，针对空白槽位[datetime]，其表示时间的含义，因此可以将实体库中提取出日期对应的实体信息（如 3 月 9 日）。

S14、将实体信息填充到对话组合中，得到多轮请求场景下的训练数据集。

沿用上述示例，以对话组合中的首轮对话模板为“[location][datetime]天气怎么样？”，针对[location]选取出的实体信息为代表某城市的“A 地点”，针对[datetime]选取出的实体信息为“3 月 9 日”为例，将实体信息填充到首轮对话模板中可以得首轮对话文本“A 地点 3 月 9 日天气怎么样”，同理，通过上述方式将相应的实体信息填充至该对话组合中的次轮对话模板中，即可得到填充后的对话组合，然后将填充后的对话组合作为多轮次场景下的训练数据集。

考虑到多轮次场景中，训练数据集的数量稀少，且标注成本较高，在本实施方式中，通过获取首轮对话模板集合和次轮对话模板集合，并基于首轮对话模板集合和次轮对话模板集合进行对话组合，再将对话组合中每个对话模板的空白模板进行填充，即可得到多轮次场景下的训练数据集，从而通过少量标注对话模板即可获得大量的多轮次场景下的训练数据集，可以增强后续对初始预测模型中多轮次场景的训练效果。

在一些实施方式中，在步骤 S14 中，将实体信息填充到对话组合中，得到多轮请求场景下的训练数据集的具体实施方式可以包括：

S141、将实体信息填充到对话组合中，得到多轮次对话样本文本。

其中，在得到多个对话组合以后，可以针对每个对话组合，对该对话组合中的空白槽位进行相应实体信息的填充，从而得到该对话组合对应的多轮次对话样本文本。示例性地，具体的空白槽位的填充方式可以参考步骤 S4 的示例，例如，得到的多轮次对话样本文本可以为：第一轮“A 地点 3 月 9 日天气怎么样”，第二轮“B 地点呢”，第三轮“3 月 10 日呢”。

S142、获取至少一个干扰文本，并将至少一个干扰文本插入多轮对话样本文本中，得到目标多轮对话样本文本。

其中，针对一个多轮对话样本文本，其干扰文本可以是与该多轮对话样本文本中每轮对话内容不相关的内容。其中，干扰文本可以包括单意图请求对应的文本和多意图请求对应的文本。

在一些实施方式中，干扰文本可以预先标注并存储至预设的文本数据库中，电子设备可以从预设的文本数据库中调取需要的干扰文本，其中，干扰文本还可以被标注出类型（如人名、地名等），干扰文本的类型可以预先和空白槽位的类型建立关联关系，具体的，一个空白槽位的类型可以预先关联一个或多个与其不相关的干扰文本的类型，因此，电子设备可以在确定空白槽位的类型后，依据空白槽位的类型从预设的文本数据库中筛选出相应的干扰文本。

然后，便可以将干扰文本插入多轮对话样本文本中。沿用上述示例，以多轮次对话样本文本：第一轮对话“A 地点 3 月 9 日天气怎么样”，第二轮对话“B 地点呢”，第三轮对话“3 月 10 日呢”为例，干扰文本可以包括干扰文本 1（如某人的电话号码是多少）、干扰文本 2（如某地的邮编是多少），可以将干扰文本 1 插入至第一轮对话和第二轮对话之间，将干扰文本 2 插入至第二轮对话和第三轮对话之间，从而得到目标多轮对话样本文本。其中，相邻的两个轮次之间可以插入一个或多个干扰文本。其中，图 4 中的 Q_n 表示第 N 轮对话，其中， n 为正整数。

S143、将目标多轮对话样本文本作为多轮请求场景下的训练数据集。

在本实施方式中，通过在多轮对话样本文本插入不相关的干扰文本，得到目标多轮对话样本文本，并将目标多轮对话样本文本作为多轮次场景下的训练数据集，从而能够在后续使用该训练数据集训练初始预测模型时，增强模型的上文理解能力。

S2、通过初始预测模型，基于预测样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，其中， M 为正整数。

其中，初始预测模型可以是指待训练的生成式大规模语言模型，该大语言模型的模型参数还未确定，因此需要通过样本文本进行训练以得到预测模型。在本实施例中，初始预测模型中可以根据不同的语音场景（如单意图场景、多意图场景、多轮次场景），建立多个训练任务，在训练时可以对多个训练任务进行联合训练。

其中，单意图请求场景下口语请求的样本文本包括单意图请求内容，单意图请求内容表征包含一个意图的请求内容。该样本子请求的数量 M 为单意图请求内容中的意图数量 1，且样本子请求的文本是根据单意图请求内容得到的。

其中，多意图请求场景下口语请求的样本文本包括多意图请求内容，多意图请求内容表征包含多个意图的请求内容。

其中，多轮请求场景的样本文本包括当前轮的文本和当前轮的之前轮次的文本。

在一些实施方式中，单意图请求场景下口语请求的样本文本包括多个单意图请求内容，在步骤 S2 中，通过初始预测模型，基于样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，可以包括：

将多个单意图请求内容进行拼接，得到一个拼接内容。

示例性地，例如当输入到初始预测模型的单意图请求场景下口语请求的样本文本包括：样本文本 1（如当前轮：请调频广播）、样本文本 2（如当前轮：把地图连续放大）、样本文本 3（如当前轮：附近最近的交通银行在哪）时，可以将单意图请求场景下口语请求的样本文本包括的多个单意图请求内容直接拼接成一个包含上述多个样本文本的拼接内容，拼接内容可以为“当前轮：样本文本 1 样本文本 2 样本文本 3（如请调频广播把地图连续放大附近最近的交通银行在哪）”。

其中，上述拼接操作也可以通过对初始预测模型进行相应的思维链生成训练来实现。

在本实施例中通过将多个单意图请求内容进行拼接，得到一个拼接内容，并基于拼接内容对初始预测模型进行训练，能够提升初始预测模型的口语理解解耦能力，并避免局部小样本领域意图训练出现过拟合现象，进而提高了生成式大规模语言模型的训练效率。

通过初始预测模型，基于拼接内容，预测样本子请求的数量 M 以及 M 个样本子请求的文本，其中，样本子请求的数量 M 为单意图请求内容的数量，且每个样本子请求的文本是根据一个单意图请求内容得到的。

沿用上述示例，例如拼接内容为“请调频广播把地图连续放大附近最近的交通银行在哪”。通过初始预测模型，基于拼接内容，预测样本子请求的数量 M 以及 M 个样本子请求的文本，可以得到样本子请求的数量 M 为 3，3 个样本子请求的文本分别为：“请调频广播”、“把地图连续放大”以及“附近最近的交通银行在哪”。

在另一些实施方式中，针对多意图请求场景下口语请求的样本文本，在步骤 S2 中，通过初始预测模型，基于样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，可以包括：

通过初始预测模型，预测多意图请求场景下口语请求的样本文本中的样本子请求的数量 M 和以及 M 个样本子请求的文本，样本子请求的数量 M 为多意图请求内容中的意图数量，且每个样本子请求的文本是根据多意图请求内容中一个意图的请求内容得到的。

示例性地，针对多意图请求场景下口语请求的样本文本，可以通过初始预测模型对多意图请求场景下口语请求的样本文本进行拆解，得到 M 个样本子请求的文本。

示例性地，通过初始预测模型可以先识别样本文本中的意图数量。以样本文本“打开 XX 应用，车辆的音量设置为 5”为例，通过对该样本文本进行意图识别，可以识别到该样本文本包含 2 个意图的请求内容，分别为“打开 XX 应用”以及“车辆的音量设置为 5”。然后可以初始预测模型可以依据这 2 个意图，将样本文本拆解成两个样本子请求的文本“打开 XX 应用”以及“车辆的音量设置为 5”。

其中，上述拆解操作也可以通过对初始预测模型进行相应的思维链生成训练来实现，使模型得到思维链生成拆解能力。

在又一些实施方式中，针对多轮请求场景的样本文本，在步骤 S2 中，通过初始预测模型，基于样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，可以包括：

A1、将当前轮的文本和当前轮的之前轮次的文本进行合并，得到合并文本，合并文本用于表征当前轮的文本和当前轮的之前轮次的文本。

其中，当前轮可以是多轮次中的最后一轮次，例如多轮次包括第一轮、第二轮以及第三轮，那么可以将第三轮确定为当前轮。

示例性地，以第一轮：“AAA 活了多久”，其中，AAA 为某历史人物，第二轮：“50 岁我确定”，当前轮：“为什么”为例，其中，当前轮的文本为“为什么”，当前轮的之前轮次的文本为“AAA 活了多久”和“50 岁我确定”。

作为一种实施方式，步骤 A1 中，步骤将当前轮的文本和当前轮的之前轮次的文本进行合并，得到合并文本，可以包括：

从当前轮的之前轮次的文本中提取出与当前轮的文本关联的关键信息，关键信息包括关键词和关键句中的至少一种。

沿用上述示例，根据上述三轮文本整体所表示的语义，可以从当前轮的之前轮次的文本中的第一轮的文本“AAA 活了多久”中提取出关键句“AAA 活了”，从第二轮的文本“50 岁我确定”中提取出关键词“50 岁”“确定”。

将关键信息与当前轮的文本进行合并，得到合并文本。

沿用上述示例，可以将关键词“50 岁”“确定”、关键句“AAA 活了”以及当前样本文本“为什么”进行合并，得到合并文本：“为什么确定 AAA 活了 50 岁”。

考虑到在多轮对话文本中，当前轮次对应的当前样本文本通常最为重要，其它轮次对应的备选样本文本可能只有部分与当前样本文本相关，在本实施方式中，通过从当前轮的之前轮次的文本中提取出与当前轮的文本关联的关键信息，将关键信息与当前轮的文本进行合并，得到合并文本，从而可以提升合并的效率和准确性。

A2、通过初始预测模型，基于合并文本，预测样本子请求的数量 M 以及 M 个样本子

请求的文本，样本子请求的数量 M 为当前轮的文本的数量 1，且样本子请求的文本是结合当前轮的之前轮次的文本对当前轮的文本进行处理得到的。

沿用上述示例，以合并文本“为什么确定 AAA 活了 50 岁”为例，将多轮次的文本进行合并后，得到的合并文本，只包含了一个最终的请求，因此，通过初始预测模型，基于合并文本，预测样本子请求的数量 M 为 1。

可见，通过初始预测模型对多轮请求场景的样本文本进行合并，得到合并文本后，再通过初始预测模型，基于合并文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本，可以有效提升预测效率。

其中，上述合并操作也可以通过对初始预测模型进行相应的思维链生成训练来实现，使模型得到思维链生成合并能力。

在一些实施方式中，在步骤 S2 中，通过预测模型，利用口语请求文本，预测口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本的具体实施方式可以包括：

基于预设提示信息和样本文本，生成输入信息，预设提示信息包含得到预测指令信息对应的任务描述信息。

示例性地，例如预设提示信息可以为“下面是一条描述一个任务的指令，带有一条提供更多信息的输入。写一个能够完成请求的回答”。

将输入信息输入至初始预测模型，以通过初始预测模型，基于样本文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本。

在另一些实施方式中，样本文本为口语文本，在步骤 S2 中，通过预测模型，利用口语请求文本，预测口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本的具体实施方式可以包括：

基于口语文本、预设提示信息以及预设指示信息，生成输入信息，预设指示信息用于指示初始预测模型输出指令信息，指令信息包括意图、领域以及槽位。

示例性地，例如预设指示信息可以为“假设你是一个还能语音助手，你需要对输入的多轮用户请求进行语义解析，输出当前轮请求对应的领域、意图和槽位”。

其中，口语文本可以是对口语请求进行转换或翻译得到的文本，示例性地，口语请求可以为“附近最近的交通银行在哪”。

S3、通过初始预测模型根据样本子请求的数量 M 以及 M 个样本子请求的文本，预测样本文本在对应的场景下的指令信息，以得到预测指令信息。

在一些实施方式中，在步骤 S3 中，通过初始预测模型根据样本子请求的数量 M 以及 M 个样本子请求的文本，预测样本文本在对应的场景下的指令信息，以得到预测指令信息的具体实施方式可以包括：

S31、将 M 个样本子请求的文本分别与预设实体库进行匹配，得到每一样本子请求对应的实体信息。

其中，可以通过实体最大匹配（Maximum Matching of Entities）技术来从预设实体库中得到与样本文本对应的实体信息，其中，实体最大匹配是一种用于文本处理和自然语言处理的技术。它主要用于在给定的文本中，识别和提取出现的实体（如人名、地名、组织机构等）。实体最大匹配的基本思想是将待处理文本按照一定的词库或规则进行分词，然后从左到右逐步匹配最长的可能实体。通过这种方式，能够较好地识别出连续的词语组合成为实体的情况。例如，对于句子：“我喜欢 C 公司的产品。”，使用实体最大匹配的方法，可以辨别出“C 公司”作为一个组织机构实体。

作为一种示例，如图 5 所示，例如以当前轮样本文本为“play 名称 X”为例，如果预设实体库中存储了关于名称 X 的实体信息包括“音乐：名称 X”、“地址：名称 X”和“电子游戏：名称 X”，那么可以将“音乐：名称 X”、“地址：名称 X”和“电子游戏：名称 X”都作为与该样本文本对应的实体信息。

S32、针对每一样本子请求，基于样本子请求对应的实体信息对样本子请求进行更新，得到更新后的样本子请求。

沿用上述示例，可以基于“音乐：名称 X”、“地址：名称 X”和“电子游戏：名称 X”生成对该样本文本“play 名称 X”的补充说明，并将该补充说明添加到该样本文本中，得到更新后的样本文本（如“play 名称 X，其中，名称 X 可能是音乐、地址或电子游戏”）。

其中，上述更新操作也可以通过对初始预测模型进行相应的思维链生成训练来实现。

其中，图 5 中 \n 表示换行，答案可以为标注更新后的样本文本，该标注更新后的样本文本与更新后的样本文本可以用于对初始预测模型进行相应的思维链生成训练，以训练模型的思维链生成更新能力。

S33、通过初始预测模型根据更新后的样本子请求，预测样本文本在对应的场景下的指令信息，以得到预测指令信息。

在本实施方式中，通过将样本子请求与预设实体库进行匹配，得到与样本子请求对应的实体信息，并基于实体信息对样本子请求进行更新，得到更新后的样本子请求，再基于更新后的样本子请求对初始预测模型进行预测指令信息的预测，得到的预测结果用于模型训练时，可以有效提升模型的预测准确性。

S4、基于预测指令信息对初始预测模型进行训练，以得到预测模型。

在一些实施方式中，在步骤 S4 中，基于预测指令信息对初始预测模型进行训练，以得到预测模型的具体实施方式可以包括：

S41、基于预测指令信息，确定多种场景中每一场景对应的初始损失值。

在一些实施方式中，样本文本包括负样本文本和正样本文本，负样本文本包括标注错误领域的样本文本，正样本文本包括样本文本中除负样本文本以外的样本文本；在步骤 S41 中，基于预测指令信息，确定多种场景中每一场景对应的初始损失值（以下也可称初始损失）的具体实施方式可以包括：

S411、在预测指令信息中筛选正样本文本对应的预测指令信息，得到正样本指令信息。

示例性地，负样本文本可以是标注了“不属于领域 x”的样本，在实际应用中，为了简化标注，可以将“不属于领域 x”的标签设置为“0”，那么在多个样本文本中可以将不携带“0”的样本可以作为正样本文本。由于预测指令信息是初始预测模型针对所有样本文本的输出，所以该预测指令信息是包括正样本文本对应的预测指令信息和负样本文本对应的预测指令信息的，对于正样本文本对应的预测指令信息，可以从该预测指令信息中筛选出并作为正样本指令信息，其中，正样本指令信息还可以包括 P_{pos} ， P_{pos} 表示初始预测模型输出该正样本指令信息的概率。

S412、在预测指令信息中筛选负样本文本对应的预测指令信息，得到负样本指令信息，并对负样本指令信息进行调整，得到目标负样本指令信息。

对于负样本文本对应的预测指令信息，可以从该预测指令信息中筛选出并作为负样本指令信息，其中，负样本指令信息还可以包括 P_{neg} ， P_{neg} 表示初始预测模型输出该负样本指令信息的概率，也可以表示标签“0”的预测概率。

其中，对负样本指令信息进行调整，得到目标负样本指令信息，具体可以是将负样本指令信息中“不属于领域 x”的概率调整为 $1-P_{neg}$ 。

S413、基于正样本指令信息和负样本指令信息，确定每一场景对应的初始损失值。

其中，步骤 S413 的具体实施方式可以包括：

S4131、获取正样本文本对应的正样本标注指令信息和负样本文本对应的负样本标注指令信息。

其中，正样本标注指令信息用于为初始预测模型基于正样本文本输出的正样本指令信息做参考，为正样本指令信息的目标值，因此可以通过比对正样本标注指令信息和正样本指令信息，来确定初始预测模型的训练情况。同理，负样本标注指令信息用于为初始预测

模型基于负样本文本输出的负样本指令信息做参考，为负样本指令信息的目标值，因此可以通过比对负样本标注指令信息和负样本指令信息，来确定初始预测模型的训练情况。其中，每个正样本文本对应的正样本标注指令信息和每个负样本文本对应的负样本标注指令信息，可以预先被存储在上述训练数据集中，因此可以从上述训练数据集获取。

S4132、基于正样本标注指令信息和正样本指令信息，确定每一场景对应的正样本损失。

其中，正样本损失计算的目的是评估初始预测模型预测的准确性或误差程度。由于正样本标注指令信息为正样本指令信息的目标值，因此可以通过比对正样本标注指令信息和正样本指令信息，来确定正样本标注指令信息与正样本指令信息来之间的误差值，并根据误差值来确定正样本损失。具体可以采用常用的损失函数计算，如均方误差（Mean Squared Error）、交叉熵（Cross Entropy）、绝对误差（Absolute Error）等。

S4133、根据负样本标注指令信息和负样本指令信息，确定每一场景对应的负样本损失。

其中，在步骤 S4133 中“根据负样本标注指令信息和负样本指令信息，确定每一场景对应的负样本损失”的具体实施方式可以包括：

基于负样本标注信息和负样本指令信息，确定负样本对应的目标样本文本的初始负样本损失。

示例性地，例如针对多个语音场景中的每个语音场景，该语音场景对应的多个样本文本，再针对多个样本文本中的每个样本文本，可以基于该样本文本对应的负样本标注信息和负样本指令信息，确定出该样本文本的初始负样本损失。例如，单意图场景对应的训练数据集包括样本文本 1、样本文本 2、样本文本 3，基于样本文本 1 对应的负样本标注信息和负样本指令信息可以得到初始负样本损失 1，基于样本文本 2 对应的负样本标注信息和负样本指令信息可以得到初始负样本损失 2，基于样本文本 3 对应的负样本标注信息和负样本指令信息可以得到初始负样本损失 3。又例如，多意图场景对应的训练数据集包括样本文本 4、样本文本 4、样本文本 5，基于样本文本 4 对应的负样本标注信息和负样本指令信息可以得到初始负样本损失 4，基于样本文本 5 对应的负样本标注信息和负样本指令信息可以得到初始负样本损失 5。

在初始负样本损失中筛选出同一场景下至少一个初始负样本损失，得到每一场景对应的初始负样本损失集合。

沿用上述示例，可以将初始负样本损失 1、初始负样本损失 2、初始负样本损失 3，确定为单意图场景对应的初始负样本损失集合。将初始负样本损失 4、初始负样本损失 5，确定为多意图场景对应的初始负样本损失集合。

将初始负样本损失集合中的初始负样本损失进行融合，得到每一场景对应的负样本损失。

沿用上述示例，针对于单意图场景，可以将初始负样本损失 1、初始负样本损失 2、初始负样本损失 3 进行融合，从而得到单意图场景对应的负样本损失。针对于多意图场景，可以将初始负样本损失 4、初始负样本损失 5 进行融合，从而得到多意图场景对应的负样本损失。其中，针对初始负样本损失的融合操作，具体可以是权重相加，例如，预先为初始负样本损失 1、初始负样本损失 2、初始负样本损失 3 分别配置权重 w_1 、 w_2 以及 w_3 ，那么单意图场景对应的负样本损失为：初始负样本损失 1* w_1 +初始负样本损失 2* w_2 +初始负样本损失 3* w_3 。其中， w_1 、 w_2 、 w_3 可以为 1:1:1，也可以为其他比例关系，具体可以更具实际需求设置，在此不做限定。

其中，在步骤 S4132 中，基于正样本标注指令信息和正样本指令信息，确定每一语音场景对应的正样本损失的具体实施方式也可以参考步骤 S4133，故不在此赘述。

S4134、将正样本损失和负样本损失进行融合，得到每一场景对应的初始损失值。

示例性地，可以通过如下公式，对正样本损失和负样本损失进行融合，得到每一语音

场景对应的初始损失值：

$$Loss = loss_{pos} + \lambda loss_{neg} = -\log P_{pos} - \lambda \log(1 - P_{neg})$$

中， $Loss$ 为一个语音场景对应的

的初始损失值， $loss_{pos}$ 为该语音场景对应的正样本损失， $loss_{neg}$ 为该语音场景对应的负样本损失， λ 为反向梯度权重， λ 具体可以为0.1，也可以根据实际需要设置为其它值比如0.2，可测试不同反向梯度权重下训练得到模型的表现，从而选择合适的反向梯度权重。

S42、将初始损失值进行融合，得到目标损失值。

其中，步骤S42的具体实施方式可以包括：

获取每一语音场景对应的融合权重，并基于融合权重，对初始损失进行加权；将加权后的初始损失进行融合，得到目标损失值（以下也可称目标损失）。

其中，在通过步骤S4134得到每一语音场景对应的初始损失之后，可以从预设权重数据库中获取每一语音场景对应的融合权重，其中，预设权重数据库包括每一语音场景的预先设定的融合权重。

作为一种示例，例如单意图场景对应的融合权重为 $k1$ ，多意图场景对应的融合权重为 $k2$ ，单意图场景对应的初始损失为 $Loss1$ ，单意图场景对应的初始损失为 $Loss2$ 。那么可以确定目标损失值为 $(k1 * Loss1 + k2 * Loss2)$ 。

在一些实施方式中，将加权后的初始损失进行融合，得到目标损失值的具体实施方式可以是对每一语音场景对应初始损失进行加权求和，然后对加权求和的结果求解均值，如目标损失值为 $(k1 * Loss1 + k2 * Loss2)/2$ 。

其中，在多轮请求场景的样本文本进行合并，得到合并文本后，该文本处理方法还包括：

从训练数据集中提取出与该多轮请求场景的样本文本对应的标注合并样本文本。

其中，针对一个多轮请求场景的样本文本，该多轮请求场景的样本文本对应的标注合并样本文本可以为该多轮请求场景的样本文本对应的合并样本文本的目标值。

基于标注合并样本文本和合并后的样本文本，确定合并损失。

其中，在确定标注合并样本文本和合并后的样本文本之后，可以参考步骤S4132中计算损失的方式来确定合并损失。

相应地，在S42中，将初始损失值进行融合，得到目标损失值的具体实施方式可以包括：

将初始损失和合并损失进行融合，得到目标损失值。

沿用上述示例，例如合并损失为 $Loss3$ ，合并损失预先对应的融合权重为 $L3$ ，则可以得到目标损失值为 $(k1 * Loss1 + k2 * Loss2 + k3 * Loss3)$ 或 $(k1 * Loss1 + k2 * Loss2 + k3 * Loss3)/3$ 。在本实施方式中，通过将初始损失和合并损失进行融合，得到目标损失值，可以同时训练初始预测模型的思维链生成合并能力。

可以理解的是，上述实施例中提到的初始预测模型的思维链生成拼接能力、思维链生成拆解能力等，也可以通过上述训练思维链生成合并能力的方式来进行训练。

S43、根据目标损失值，对初始预测模型进行收敛，以得到预测模型。

其中，可以采用梯度下降法来优化初始预测模型，例如，可以根据目标损失值确定初始预测模型的梯度（Gradient），然后基于梯度来对该初始预测模型的模型参数进行调整，从而初始预测模型进行收敛。其中，梯度是一个向量，其方向指向函数在某一点上具有最大增长率的方向，其大小表示该最大增长率。

在一些实施方式中，在对初始预测模型进行训练的过程中，每一次基于梯度来对该初始预测模型的模型参数进行调整，可以看作成一次迭代，当迭代次数达到预设次数时，可以得到预测模型。其中，预设次数可以是根据对初始预测模型的历史训练记录确定。

在一些实施例中，在步骤303之后，该文本处理方法还可以包括：

通过目标指令信息对可控对象进行控制。

其中，通过目标指令信息对可控对象进行控制可以包括：

从预设的控制指令库中获取与目标指令信息对应的控制指令；基于控制指令，对可控对象进行控制。

其中，控制指令库中预先存储了多个控制指令，且多个控制指令预先与指令信息建立了映射关系，因此可以从预设的控制指令库中获取与目标指令信息对应的控制指令，并基于控制指令，对可控对象进行控制。例如得到目标指令信息包括领域：导航，意图：查询路线路况；槽位：从家到公司，电子设备可以根据该目标指令信息，得到控制指令“输出从起始地（家）到目的地(公司)的路线”，并将该控制指令发送给可控对象，以指示可控对象输出从起始地（家）到目的地(公司)的路线，其中，可控对象可以包括车载终端，移动电话等，输出的方式可以包括但不限于：图像展示、语音播报等。

可见，在本实施例中，通过在获取表达口语请求的口语请求文本后，可以通过预测模型预测口语请求的子请求的数量 N 以及 N 个子请求文本，其中， N 为正整数；然后，通过预测模型根据子请求的数量 N 以及 N 个子请求文本，预测 N 个子请求文本各自对应的目标指令信息，目标指令信息包括意图、领域以及槽位， N 个子请求文本各自对应的目标指令信息，用于确定对于口语请求的响应。由于不同口语理解问题场景下的文本所包含的子请求的数量往往不同，所以在本申请实施例中，通过预测模型预测口语请求的子请求的数量 N 以及 N 个子请求文本，可以将对不同口语理解问题场景下的文本的处理，转换为对文本中不同数量的子请求的处理，再通过通过预测模型根据子请求的数量 N 以及 N 个子请求文本，预测 N 个子请求文本各自对应的目标指令信息，从而实现只通过一个预测模型模型就解决对不同口语理解问题场景下文本的处理问题，避免了针对多个口语理解问题场景需要分别设置多个模型来支持时，需要训练多个模型以及对多个模型进行持续迭代的操作，大大降低了模型的复杂程度、模型的训练成本和模型的迭代成本，提高了文本处理效率。

根据上述实施例所描述的方法，以下将作进一步详细说明。

在本实施例中，将以文本处理为例，对本申请实施例的方法进行详细说明。

如图 6 所示，一种文本处理方法具体流程如下：

601、电子设备获取多种语音场景下的训练数据集，训练数据集包括每一语音场景对应的至少一个样本文本。

其中，步骤 601 可以包括：获取首轮对话模板集合和至少一个次轮对话模板集合，首轮对话模板集合中的每个首轮对话模板和次轮对话模板集合中的每个次轮对话模板均包括空白槽位。

将首轮对话模板集合中的每个首轮对话模板分别与次轮对话模板集合中的每个次轮对话模板进行组合，得到多个对话组合。

基于对话组合中空白槽位，从预设实体库中提取出实体信息。

将实体信息填充到对话组合中，得到多轮次场景下的训练数据集。

其中，将实体信息填充到对话组合中，得到多轮次场景下的训练数据集，包括：

将实体信息填充到对话组合中，得到多轮次对话样本文本；获取至少一个干扰文本，并将至少一个干扰文本插入多轮对话样本文本中，得到目标多轮对话样本文本；将目标多轮对话样本文本作为多轮次场景下的训练数据集。

示例性地，请再次参阅图 4，在实际应用中，步骤 601 的实施流程可以包括：先人工标注少量的首轮模板+多轮领域模板(包含指代/省略的模板以及改写后完整的模板)，其中，改写后完整的模板可以用于训练初始预测模型的相应思维链生成能力，即自动生成多轮次场景下的训练数据集的能力。然后进行对话模板采样：第一步采样首轮对话模板集合，之后循环采样次轮对话模板集合，构造出多轮模板数据（即上述实施例中对话组合）；第二步采用实体信息对多轮模板数据进行槽位的填充；第三步在填充后的多轮数据中插入不相关的单/多意图请求对应的文本，以增强模型上写文理解能力。

602、电子设备基于语音场景的类型，对样本文本进行改写，得到目标样本文本。

可以理解的是，基于语音场景的类型，对样本文本进行改写，得到目标样本文本，可以对应于上述实施例中的步骤 S2、通过初始预测模型，基于样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，其中， M 为正整数。其中，目标样本文本可以看作是 M 个样本子请求的文本。

其中，语音场景包括单轮次场景和多轮次场景，基于语音场景的类型，对样本文本进行改写，得到目标样本文本，包括：

当样本文本对应的语音场景为多轮次场景时，将样本文本转换为单轮次场景对应的备选样本文本，并将备选样本文本作为单轮次场景对应的样本文本；当样本文本对应的语音场景为单轮次场景时，基于样本文本的意图类型，对样本文本进行改写，得到目标样本文本。

其中，样本文本包括多轮次的对话样本文本，将样本文本转换为单轮次场景对应的备选样本文本，包括：

在多轮次中确定出当前轮次，并在对话样本文本中提取出当前轮次对应的当前样本文本和其他轮次对应的备选样本文本，其他轮次包括多轮次中除当前轮次以外的轮次；将当前样本文本和备选样本文本进行合并，得到当前轮次对应的合并后的样本文本；将合并后的样本文本作为单轮次场景对应的样本文本。

其中，将当前样本文本和备选样本文本进行合并，得到当前轮次对应的合并后的样本文本，包括：从备选样本文本中提取出与当前样本文本关联的关键信息，关键信息包括关键词和关键句中的至少一种；将关键信息与当前样本文本进行合并，得到当前轮次对应的合并后的样本文本。其中，备选样本文本可以相当于上述实施例中的在当前轮的文本之前轮次的文本。

示例性地，在实际应用，如图 7 所示，由于在多轮口语理解场景中，多轮的标注数据稀少，且标注成本高，为提升生成式大规模语言模型（初始预测模型）在多轮用户口语请求理解上的能力，本实施例可以利用已经公开的通用多轮对话改写数据来训练生成式大规模语言模型的对多轮口语理解的样本文本的思维链生成合并能力。例如，在训练时，具体的输入格式可以为：“###输入：第 1 轮：{第 1 轮对话}\n... \n第 i 轮次{第 i 轮对话}\n 当前轮：{第 $i+1$ 轮对话}\n###答案：”。具体的输出格式可以为：“当前轮包含 1 个请求\n 1.{改写后的对话\n 领域：</s>}”。例如，输入的第 1 轮：AAA 活多久，第 2 轮：50 岁我确定，当前轮：为什么。通过初始预测模型合并，可以得到“为什么确定 AAA 活了 50 岁”。其中， i 为正整数。 \n 表示换行。

其中，基于样本文本的意图类型，对样本文本进行改写，得到目标样本文本，包括：

获取样本文本的意图标识信息，并基于意图标识信息，确定样本文本的意图类型；当样本文本的意图类型为单意图时，将单意图对应的至少一个样本文本进行拼接，得到目标样本文本。

示例性地，如图 8 所示，为了提高生成式大规模语言模型的训练效率，并避免局部小样本领域意图训练出现过拟合现象。本实施例提出多样本拼接的思维链生成训练策略。具体的，在训练的过程中，可以随机采样 1-3 个单意图口语理解训练样本，按照输入格式为：“###输入：当前轮：{口语请求 1}{口语请求 2}{口语请求 3}\n ###答案：”形式组织。按照输出为：“当前轮包含 $\{n=1,2,3\}$ 个请求\n 1.{口语请求 1}领域：{xxx}\n 意图：{xxx}\n 槽位：{xxx}\n 2.{口语请求 2}领域：{xxx}\n 意图：{xxx}\n 槽位：{xxx}\n 3.{口语请求 3}领域：{xxx}\n 意图：{xxx}\n 槽位：{xxx}”的形式构造。其中，Sample#1、Sample#2、Sample#3，分别表示对应口语请求 1 的样本文本 1、对应口语请求 2 的样本文本 2 以及对应口语请求 3 的样本文本 3。

当样本文本的意图类型为多意图时，对样本文本进行拆解，得到多个文本子样本，并将文本子样本作为目标样本文本。

其中，基于样本文本的意图类型，对样本文本进行改写，得到目标样本文本，包括：获取样本文本的意图标识信息，并基于意图标识信息，确定样本文本的意图类型；当样本文本的意图类型为单意图时，将单意图对应的至少一个样本文本进行拼接，得到目标样本文本；当样本文本的意图类型为多意图时，对样本文本进行拆解，得到多个文本子样本，并将文本子样本作为目标样本文本。

其中，对样本文本进行拆解，得到多个文本子样本，包括：

对样本文本进行文本识别，并基于文本识别结果，确定样本文本中的领域数量和意图数量；基于领域数量和意图数量，对样本文本进行拆解，得到多个文本子样本。

603、电子设备采用初始预测模型预测目标样本文本对应的指令信息，以得到每一目标样本文本的预测指令信息。

可以理解的是，初始预测模型预测目标样本文本对应的指令信息，以得到每一目标样本文本的预测指令信息，可以对应于上述实施例中步骤 S3、通过初始预测模型根据样本子请求的数量 M 以及 M 个样本子请求的文本，预测样本文本在对应的场景下的指令信息，以得到预测指令信息。其中，预测指令信息可以看作样本文本在对应的场景下的指令信息。

示例性地，请参阅图 9，对于生成式大规模语言模型训练的输入，本实施例可以将用户口语请求 Input 和提示（Prompt）以及指示（Instruction）进行拼接：Prompt + Instruction + Input。具体的单/多意图口语理解场景训练时，口语请求 Input 按照“###输入:当前轮:{用户口语请求}\n ###答案:”的形式构造。

对于多轮次口语理解训练时，口语请求 Input 按照“###输入: 第 1 轮: {xxx}, ...,第 i 轮: {xxx}, 当前轮: {用户口语请求}\n ###答案”的形式进行构造。

其中，单意图自然语言理解（NLU, Natural Language Understanding）生成上述单意图场景对应的预测指令信息，多意图自然语言理解生成上述多意图场景对应的预测指令信息，多轮自然语言理解生成上述多轮次场景对应的预测指令信息。

可以理解是，本实施例中的用户口语请求，指的是用户口语请求对应的文本。

604、电子设备基于预测指令信息，确定每一语音场景对应的初始损失，并将初始损失进行融合，得到目标损失值。

示例性地，上述训练数据集中可以包含大量的负样本数据（即上述负样本文本），该负样本数据只有“不属于领域 x ”的标签，没有“属于领域 y ”的标签，这部分数据可以明显提升模型的精确率，减少误召回情况。在多类 2 分类判别式训练中，负样本数据可直接将领域 x 的标签设置为 0 即可。在生成式训练中，本实施例提出负向梯度策略来进行学习。具体的，可以将负样本数据输出标签按照单意图场景对应的正样本文本输出标签进行组织，即“当前轮包含 1 个请求\n 1. {用户口语请求}\n 领域: {负样本领域 x } </s>”（其中“</s>”为句子结束符）。之后在计算损失 $Loss$ 时，将生成序列中“负样本领域 x ”处的预测概率 P_{neg} 变换

为 $1 - P_{neg}$ ，即训练模型最小化“负样本领域 x ”的预测概率： $-\log(1 - P_{neg})$ ，实现模型不预测出“负样本领域 x ”的能力。同时结束符“</s>”停止更新梯度，避免影响正样本训练中，继续预测出意图、槽位结果。其中，负样本生成序列剩余部分按照正样本更是进行更新 $-\log(P_{pos})$ 。

如图 10 所示，负样本训练总的损失函数表示为：

$$Loss = loss_{pos} + \lambda loss_{neg} = -\log P_{pos} - \lambda \log(1 - P_{neg})$$

中， $Loss$ 为一个语音场景对应的

的初始损失， $loss_{pos}$ 为该语音场景对应的正样本损失， $loss_{neg}$ 为该语音场景对应的负样本损失， λ 为反向梯度权重， λ 具体可以为 0.1，也可以根据实际需要设置为其它值比如 0.2，

可测试不同反向梯度权重下训练得到模型的表现，从而选择合适的反向梯度权重。

其中，样本文本包括负样本文本和正样本文本，负样本文本包括标注错误领域的样本文本，正样本文本包括样本文本中除负样本文本以外的样本文本；基于预测指令信息，确定每一语音场景对应的初始损失，包括：在预测指令信息中筛选正样本文本对应的预测指令信息，得到正样本指令信息；在预测指令信息中筛选负样本文本对应的预测指令信息，得到负样本指令信息，并对负样本指令信息进行调整，得到目标负样本指令信息；基于正样本指令信息和负样本指令信息，确定每一语音场景对应的初始损失。

其中，基于正样本指令信息和负样本指令信息，确定每一语音场景对应的初始损失，包括：

获取正样本文本对应的正样本标注指令信息和负样本文本对应的负样本标注指令信息；基于正样本标注指令信息和正样本指令信息，确定每一语音场景对应的正样本损失；根据负样本标注指令信息和负样本指令信息，确定每一语音场景对应的负样本损失；将正样本损失和负样本损失进行融合，得到每一语音场景对应的初始损失。

其中，根据负样本标注指令信息和负样本指令信息，确定每一语音场景对应的负样本损失，包括：

基于负样本标注信息和负样本指令信息，确定负样本对应的目标样本文本的初始负样本损失；在初始负样本损失中筛选出同一语音场景下至少一个初始负样本损失，得到每一语音场景对应的初始负样本损失集合；将初始负样本损失集合中的初始负样本损失进行融合，得到每一语音场景对应的负样本损失。

其中，在将当前样本文本和备选样本文本进行合并，得到当前轮次对应的合并后的样本文本之后，还包括：从训练数据集中提取出与多轮次对话样本文本对应的标注合并样本文本；基于标注合并样本文本和合并后的样本文本，确定合并损失。相应地，将初始损失进行融合，得到目标损失值，包括：将初始损失和合并损失进行融合，得到目标损失值。

605、电子设备根据目标损失值，对初始预测模型进行收敛，得到预测模型。

606、电子设备获取针对可控对象的目标语音对应的目标文本，采用预测模型预测目标文本在对应语音场景下的指令信息，得到目标指令信息。

607、电子设备从预设的控制指令库中获取与目标指令信息对应的控制指令。

608、电子设备基于控制指令，对可控对象进行控制。

示例性地，请参阅图 11，图 11 示出了本实施例的预测模型的训练过程和推理过程，在训练过程中，首先对用户口语请求（即上述样本文本）进行提示信息和指示信息的拼接，然后将拼接后的结果，通过统一思维链生成的多场景生成式大规模语言模型（即上述初始预测模型）进行预测，通过单/多意图口语理解训练任务可以得到单意图口语理解结果（即上述单意图场景对应的预测指令信息）和多意图口语理解结果（即上述多意图场景对应的预测指令信息），通过多轮次口语理解训练任务，可以得到多轮次口语理解结果（即上述多轮次场景对应的预测指令信息），此后，将上述三中结果结合预设单意图口语理解标签、多意图口语理解标签和多轮次口语理解标签，可以对初始预测模型进行损失计算，并以基于损失更新模型参数，直到获得预测模型。在推理过程中，首先对需要预测的用户口语请求输入至训练好的统一思维链生成的多场景生成式大规模语言模型（即上述预测模型）中进行预测，预测模型可以先确定输入数据对应的语音场景，然后基于输入数据，输出与该语音场景对应的口语理解结果。

可见，本实施例结合生成式大语言模型 LLM 和思维链生成训练思维链生成策略将单意图口语理解、多意图口语理解、多轮次口语理解三个场景进行统一任务建模，实现不同场景之间的信息交互，口语理解能力互相补充增强，大幅度提升冷启动场景和多轮次场景的口语理解效果。另外，将多个口语理解场景进行了单模型统一的整合、简化了部署复杂度和语言迁移的难度；同时利用生成式大规模语言模型的泛化能力大幅度提升冷启动、低资源口语理解场景的效果；还能大幅度提升多轮口语理解场景的泛化能力和效果。并且在

多轮次口语理解场景中，本方案能够处理更多的上下文问题，大幅度提升用户在和机器设备的多轮交互体验。

为了更好地实施以上方法，本申请实施例还提供一种文本处理装置，该文本处理装置具体可以集成在电子设备中，该电子设备可以为终端、服务器等设备。其中，终端可以为手机、平板电脑、智能蓝牙设备、笔记本电脑、个人电脑等设备；服务器可以是单一服务器，也可以是由多个服务器组成的服务器集群。

比如，在本实施例中，将以文本处理装置具体集成在文本处理为例，对本申请实施例的方法进行详细说明。

例如，如图 12 所示，该文本处理装置可以包括：

获取单元 1201，用于获取表达口语请求的口语请求文本；

第一预测单元 1202，用于通过预测模型，利用口语请求文本，预测口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本，其中， N 为正整数；

第二预测单元 1203，用于通过该预测模型根据该子请求的数量 N 以及 N 个子请求文本，预测该 N 个子请求文本各自对应的目标指令信息，该目标指令信息包括意图、领域以及槽位，该 N 个子请求文本各自对应的目标指令信息，用于确定对于该口语请求的响应。

在一些实施例中，该口语请求文本为单意图请求场景下口语请求的文本、多意图请求场景下口语请求的文本或多轮请求场景中的一轮口语请求的文本。

在一些实施例中，该文本处理装置还包括：

数据集获取单元 1201，用于获取多种场景下的训练数据集，该训练数据集包括每一场景对应的至少一个样本文本，该多种场景包括单意图请求场景、多意图请求场景或多轮请求场景中的至少两个；

第三预测单元，用于通过初始预测模型，基于样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，其中， M 为正整数；

第四预测单元，用于通过该初始预测模型根据该样本子请求的数量 M 以及 M 个样本子请求的文本，预测该样本文本在对应的场景下的指令信息，以得到预测指令信息；

训练单元，用于基于该预测指令信息对该初始预测模型进行训练，以得到该预测模型。

在一些实施例中，该单意图请求场景下口语请求的样本文本包括单意图请求内容，该单意图请求内容表征包含一个意图的请求内容；

该样本子请求的数量 M 为该单意图请求内容中的意图数量 1，且该样本子请求的文本是根据该单意图请求内容得到的。

在一些实施例中，该单意图请求场景下口语请求的样本文本包括多个该单意图请求内容，第三预测单元，包括：

拼接子单元，用于将该多个该单意图请求内容进行拼接，得到一个拼接内容；

第一预测子单元，用于通过初始预测模型，基于拼接内容，预测样本子请求的数量 M 以及 M 个样本子请求的文本，该样本子请求的数量 M 为该单意图请求内容的数量，且每个样本子请求的文本是根据一个该单意图请求内容得到的。

在一些实施例中，该多意图请求场景下口语请求的样本文本包括多意图请求内容，该多意图请求内容表征包含多个意图的请求内容。

在一些实施例中，第三预测单元，包括：

第二预测子单元，用于通过初始预测模型，预测该多意图请求场景下口语请求的样本文本中的样本子请求的数量 M 和该以及 M 个样本子请求的文本，该样本子请求的数量 M 为该多意图请求内容中的意图数量，且每个样本子请求的文本是根据该多意图请求内容中一个意图的请求内容得到的。

在一些实施例中，该多轮请求场景的样本文本包括当前轮的文本和该当前轮的之前轮次的文本。

在一些实施例中，第三预测单元，包括：

合并子单元，用于将该当前轮的文本和该当前轮的之前轮次的文本进行合并，得到合并文本，该合并文本用于表征该当前轮的文本和该当前轮的之前轮次的文本；

第三预测子单元，用于通过初始预测模型，基于该合并文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本，该样本子请求的数量 M 为该当前轮的文本的数量 1 ，且该样本子请求的文本是结合该当前轮的之前轮次的文本对该当前轮的文本进行处理得到的。

在一些实施例中，合并子单元，具体用于从该当前轮的之前轮次的文本中提取出与该当前轮的文本关联的关键信息，该关键信息包括关键词和关键句中的至少一种；

将该关键信息与该当前轮的文本进行合并，得到该合并文本。

在一些实施例中，第三预测单元，还包括：

生成子单元，用于基于预设提示信息和该样本文本，生成输入信息，该预设提示信息包含得到预测指令信息对应的任务描述信息；

输入子单元，用于将该输入信息输入至该初始预测模型，以通过初始预测模型，基于该样本文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本。

在一些实施例中，该样本文本为口语文本，生成子单元，具体用于：

基于该口语文本、预设提示信息以及预设指示信息，生成输入信息，该预设指示信息用于指示该初始预测模型输出指令信息，该指令信息包括意图、领域以及槽位。

在一些实施例中，训练单元，包括：

损失确定子单元，用于基于该预测指令信息，确定该多种场景中每一场景对应的初始损失值；

融合子单元，用于将该初始损失值进行融合，得到目标损失值；

收敛子单元，用于根据该目标损失值，对该初始预测模型进行收敛，以得到该预测模型。

在一些实施例中，该样本文本包括负样本文本和正样本文本，该负样本文本包括标注错误领域的样本文本，该正样本文本包括该样本文本中除该负样本文本以外的样本文本；损失确定子单元，具体用于：

在该预测指令信息中筛选该正样本文本对应的预测指令信息，得到正样本指令信息；

在该预测指令信息中筛选该负样本文本对应的预测指令信息，得到负样本指令信息，并对该负样本指令信息进行调整，得到目标负样本指令信息；

基于该正样本指令信息和负样本指令信息，确定每一场景对应的初始损失值。

在一些实施例中，数据集获取单元 1201，包括：

模板获取子单元，用于获取首轮对话模板集合和至少一个次轮对话模板集合，该首轮对话模板集合中的每个首轮对话模板和该次轮对话模板集合中的每个次轮对话模板均包括空白槽位；

组合子单元，用于将该首轮对话模板集合中的每个首轮对话模板分别与该次轮对话模板集合中的每个次轮对话模板进行组合，得到多个对话组合；

提取子单元，用于基于该对话组合中空白槽位，从预设实体库中提取出实体信息；

填充子单元，用于将该实体信息填充到该对话组合中，得到该多轮请求场景下的训练数据集。

在一些实施例中，填充子单元，具体用于：

将该实体信息填充到该对话组合中，得到多轮次对话样本文本；

获取至少一个干扰文本，并将该至少一个干扰文本插入该多轮对话样本文本中，得到目标多轮对话样本文本；

将该目标多轮对话样本文本作为该多轮请求场景下的训练数据集。

在一些实施例中，第三预测单元，具体用于：

将该 M 个样本子请求的文本分别与预设实体库进行匹配，得到每一样本子请求对应的

实体信息；

针对每一样本子请求，基于该样本子请求对应的实体信息对该样本子请求进行更新，得到更新后的样本子请求；

通过该初始预测模型根据该更新后的样本子请求，预测该样本文本在对应的场景下的指令信息，以得到预测指令信息。

具体实施时，以上各个单元可以作为独立的实体来实现，也可以进行任意组合，作为同一或若干个实体来实现，以上各个单元的具体实施可参见前面的方法实施例，在此不再赘述。

本申请实施例还提供一种电子设备，该电子设备可以为终端、服务器等设备。其中，终端可以为手机、平板电脑、智能蓝牙设备、笔记本电脑、个人电脑，等等；服务器可以是单一服务器，也可以是由多个服务器组成的服务器集群，等等。

在一些实施例中，该文本处理装置还可以集成在多个电子设备中，比如，文本处理装置可以集成在多个服务器中，由多个服务器来实现本申请的文本处理方法。

在本实施例中，将以本实施例的电子设备为例进行详细描述，比如，如图 13 所示，其示出了本申请实施例所涉及的电子设备的结构示意图，具体来讲：

该电子设备可以包括一个或者一个以上处理核心的处理器 1301、一个或一个以上计算机可读存储介质的存储器 1302、电源 1303、输入模块 1304 以及通信模块 1305 等部件。本领域技术人员可以理解，图 13 中示出的电子设备结构并不构成对电子设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。其中：

处理器 1301 是该电子设备的控制中心，利用各种接口和线路连接整个电子设备的各个部分，通过运行或执行存储在存储器 1302 内的软件程序和/或模块，以及调用存储在存储器 1302 内的数据，执行电子设备的各种功能和处理数据，从而对电子设备进行整体检测。在一些实施例中，处理器 1301 可包括一个或多个处理核心；在一些实施例中，处理器 1301 可集成应用处理器和调制解调处理器，其中，应用处理器主要处理操作系统、用户界面和应用程序等，调制解调处理器主要处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器 1301 中。

存储器 1302 可用于存储软件程序以及模块，处理器 1301 通过运行存储在存储器 1302 的软件程序以及模块，从而执行各种功能应用以及数据处理。存储器 1302 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据电子设备的使用所创建的数据等。此外，存储器 1302 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地，存储器 1302 还可以包括存储器控制器，以提供处理器 1301 对存储器 1302 的访问。

电子设备还包括给各个部件供电的电源 1303，在一些实施例中，电源 1303 可以通过电源管理系统与处理器 1301 逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源 1303 还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

该电子设备还可包括输入模块 1304，该输入模块 1304 可用于接收输入的数字或字符信息，以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。

该电子设备还可包括通信模块 1305，在一些实施例中通信模块 1305 可以包括无线模块，电子设备可以通过该通信模块 1305 的无线模块进行短距离无线传输，从而为用户提供无线的宽带互联网访问。比如，该通信模块 1305 可以用于帮助用户收发电子邮件、浏览网页和访问流式媒体等。

尽管未示出，电子设备还可以包括显示单元等，在此不再赘述。具体在本实施例中，电子设备中的处理器 1301 会按照如下的指令，将一个或一个以上的应用程序的进程对应

的可执行文件加载到存储器 1302 中，并由处理器 1301 来运行存储在存储器 1302 中的应用程序，从而实现各种功能。

以上各个操作的具体实施可参见前面的实施例，在此不再赘述。

本领域普通技术人员可以理解，上述实施例的各种方法中的全部或部分步骤可以通过指令来完成，或通过指令控制相关的硬件来完成，该指令可以存储于一计算机可读存储介质中，并由处理器进行加载和执行。

为此，本申请实施例提供一种计算机可读存储介质，其中存储有多条指令，该指令能够被处理器进行加载，以执行本申请实施例所提供的任一种文本处理方法中的步骤。例如，该指令可以执行如下步骤：

其中，该存储介质可以包括：只读存储器（ROM，Read Only Memory）、随机存取记忆体（RAM，Random Access Memory）、磁盘或光盘等。

根据本申请的一个方面，提供了一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。电子设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该电子设备执行上述实施例中提供的方法。

由于该存储介质中所存储的指令，可以执行本申请实施例所提供的任一种文本处理方法中的步骤，因此，可以实现本申请实施例所提供的任一种文本处理方法所能实现的有益效果，详见前面的实施例，在此不再赘述。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求

- 1、一种文本处理方法，由电子设备执行，包括：
获取表达口语请求的口语请求文本；
通过预测模型，利用所述口语请求文本，预测所述口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本，其中， N 为正整数；及
通过所述预测模型，利用所述子请求的数量 N 以及 N 个子请求文本，预测所述 N 个子请求文本各自对应的目标指令信息，所述目标指令信息包括意图、领域以及槽位，所述 N 个子请求文本各自对应的目标指令信息，用于确定对于所述口语请求的响应。
- 2、根据权利要求 1 所述的文本处理方法，所述口语请求文本为单意图请求场景下口语请求的文本、多意图请求场景下口语请求的文本或多轮请求场景中的一轮口语请求的文本。
- 3、根据权利要求 1 或 2 所述的文本处理方法，所述方法还包括：
获取多种场景下的训练数据集，所述训练数据集包括每一场景对应的至少一个样本文本，所述多种场景包括单意图请求场景、多意图请求场景或多轮请求场景中的至少两个；
通过初始预测模型，基于所述样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，其中， M 为正整数；
通过所述初始预测模型，根据所述样本子请求的数量 M 以及 M 个样本子请求的文本，预测所述样本文本在对应的场景下的指令信息，以得到预测指令信息；
基于所述预测指令信息对所述初始预测模型进行训练，以得到所述预测模型。
- 4、根据权利要求 3 所述的文本处理方法，所述单意图请求场景下口语请求的样本文本包括单意图请求内容，所述单意图请求内容表征包含一个意图的请求内容；
所述样本子请求的数量 M 为所述单意图请求内容中的意图数量 1，且所述样本子请求的文本是根据所述单意图请求内容得到的。
- 5、根据权利要求 4 所述的文本处理方法，所述单意图请求场景下口语请求的样本文本包括多个所述单意图请求内容，所述通过初始预测模型，基于所述样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，包括：
将所述多个所述单意图请求内容进行拼接，得到一个拼接内容；
通过初始预测模型，基于所述拼接内容，预测样本子请求的数量 M 以及 M 个样本子请求的文本，所述样本子请求的数量 M 为所述单意图请求内容的数量，且每个样本子请求的文本是根据一个所述单意图请求内容得到的。
- 6、根据权利要求 3 所述的文本处理方法，所述多意图请求场景下口语请求的样本文本包括多意图请求内容，所述多意图请求内容表征包含多个意图的请求内容。
- 7、根据权利要求 6 所述的文本处理方法，所述通过初始预测模型，基于所述样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，包括：
通过初始预测模型，预测所述多意图请求场景下口语请求的样本文本中的样本子请求的数量 M 和所述以及 M 个样本子请求的文本，所述样本子请求的数量 M 为所述多意图请求内容中的意图数量，且每个样本子请求的文本是根据所述多意图请求内容中一个意图的请求内容得到的。
- 8、根据权利要求 3 所述的文本处理方法，所述多轮请求场景的样本文本包括当前轮的文本和所述当前轮的之前轮次的文本。
- 9、根据权利要求 8 所述的文本处理方法，所述通过初始预测模型，基于所述样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，包括：
将所述当前轮的文本和所述当前轮的之前轮次的文本进行合并，得到合并文本，所述合并文本用于表征所述当前轮的文本和所述当前轮的之前轮次的文本；
通过初始预测模型，基于所述合并文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本，所述样本子请求的数量 M 为所述当前轮的文本的数量 1，且所述样本子请求

的文本是结合所述当前轮的之前轮次的文本对所述当前轮的文本进行处理得到的。

10、根据权利要求 9 所述的文本处理方法，所述将所述当前轮的文本和所述当前轮的之前轮次的文本进行合并，得到合并文本，包括：

从所述当前轮的之前轮次的文本中提取出与所述当前轮的文本关联的关键信息，所述关键信息包括关键词和关键句中的至少一种；

将所述关键信息与所述当前轮的文本进行合并，得到所述合并文本。

11、根据权利要求 3 所述的文本处理方法，所述通过初始预测模型，基于所述样本文本预测样本子请求的数量 M 以及 M 个样本子请求的文本，包括：

基于预设提示信息和所述样本文本，生成输入信息，所述预设提示信息包含得到预测指令信息对应的任务描述信息；

将所述输入信息输入至所述初始预测模型，以通过初始预测模型，基于所述样本文本，预测样本子请求的数量 M 以及 M 个样本子请求的文本。

12、根据权利要求 11 所述的文本处理方法，所述样本文本为口语文本，所述基于预设提示信息和所述样本文本，生成输入信息，包括：

基于所述口语文本、预设提示信息以及预设指示信息，生成输入信息，所述预设指示信息用于指示所述初始预测模型输出指令信息，所述指令信息包括意图、领域以及槽位。

13、根据权利要求 3 至 12 任一项所述的文本处理方法，所述基于所述预测指令信息对所述初始预测模型进行训练，以得到所述预测模型，包括：

基于所述预测指令信息，确定所述多种场景中每一场景对应的初始损失值；

将所述初始损失值进行融合，得到目标损失值；

根据所述目标损失值，对所述初始预测模型进行收敛，以得到所述预测模型。

14、根据权利要求 13 所述的文本处理方法，所述样本文本包括负样本文本和正样本文本，所述负样本文本包括标注错误领域的样本文本，所述正样本文本包括所述样本文本中除所述负样本文本以外的样本文本；所述基于所述预测指令信息，确定所述多种场景中每一场景对应的初始损失值，包括：

在所述预测指令信息中筛选所述正样本文本对应的预测指令信息，得到正样本指令信息；

在所述预测指令信息中筛选所述负样本文本对应的预测指令信息，得到负样本指令信息，并对所述负样本指令信息进行调整，得到目标负样本指令信息；

基于所述正样本指令信息和负样本指令信息，确定每一场景对应的初始损失值。

15、根据权利要求 3 至 14 中任一项所述的文本处理方法，所述获取多种场景下的训练数据集，包括：

获取首轮对话模板集合和至少一个次轮对话模板集合，所述首轮对话模板集合中的每个首轮对话模板和所述次轮对话模板集合中的每个次轮对话模板均包括空白槽位；

将所述首轮对话模板集合中的每个首轮对话模板分别与所述次轮对话模板集合中的每个次轮对话模板进行组合，得到多个对话组合；

基于所述对话组合中空白槽位，从预设实体库中提取出实体信息；

将所述实体信息填充到所述对话组合中，得到所述多轮请求场景下的训练数据集。

16、根据权利要求 15 所述的文本处理方法，所述将所述实体信息填充到所述对话组合中，得到所述多轮请求场景下的训练数据集，包括：

将所述实体信息填充到所述对话组合中，得到多轮次对话样本文本；

获取至少一个干扰文本，并将所述至少一个干扰文本插入所述多轮对话样本文本中，得到目标多轮对话样本文本；

将所述目标多轮对话样本文本作为所述多轮请求场景下的训练数据集。

17、根据权利要求 3 至 16 中任一项所述的文本处理方法，所述通过所述初始预测模型根据所述样本子请求的数量 M 以及 M 个样本子请求的文本，预测所述样本文本在对应

的场景下的指令信息，以得到预测指令信息，包括：

将所述 M 个样本子请求的文本分别与预设实体库进行匹配，得到每一样本子请求对应的实体信息；

针对每一样本子请求，基于所述样本子请求对应的实体信息对所述样本子请求进行更新，得到更新后的样本子请求；

通过所述初始预测模型根据所述更新后的样本子请求，预测所述样本文本在对应的场景下的指令信息，以得到预测指令信息。

18、一种文本处理装置，包括：

获取单元，用于获取表达口语请求的口语请求文本；

第一预测单元，用于通过预测模型，利用所述口语请求文本，预测所述口语请求的子请求的数量 N 以及与 N 个子请求一一对应的 N 个子请求文本，其中， N 为正整数；及

第二预测单元，用于通过所述预测模型根据所述子请求的数量 N 以及 N 个子请求文本，预测所述 N 个子请求文本各自对应的目标指令信息，所述目标指令信息包括意图、领域以及槽位，所述 N 个子请求文本各自对应的目标指令信息，用于确定对于所述口语请求的响应。

19、一种电子设备，包括处理器和存储器，所述存储器存储有多条指令；所述处理器从所述存储器中加载指令，以执行如权利要求 1~17 任一项所述的文本处理方法中的步骤。

20、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多条指令，所述指令适于处理器进行加载，以执行权利要求 1~17 任一项所述的文本处理方法中的步骤。

21、一种计算机程序产品，包括计算机指令，所述计算机指令被处理器执行时，实现如权利要求 1~17 任一项所述的文本处理方法中的步骤。