

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 5 月 23 日 (23.05.2024)



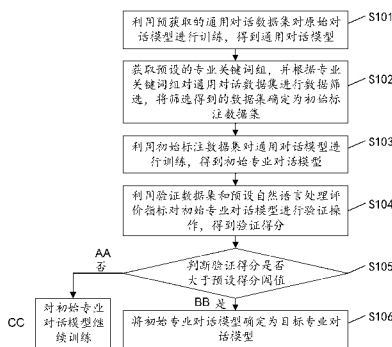
(10) 国际公布号
WO 2024/103609 A1

- (51) 国际专利分类号:
G06F 16/332 (2019.01)
- (21) 国际申请号: PCT/CN2023/086071
- (22) 国际申请日: 2023 年 4 月 4 日 (04.04.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211441290.4 2022 年 11 月 17 日 (17.11.2022) CN
- (71) 申请人: 苏州元脑智能科技有限公司 (SUZHOU METABRAIN INTELLIGENT TECHNOLOGY CO., LTD.) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。
- (72) 发明人: 刘红丽 (LIU, Hongli); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。李峰 (LI, Feng); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。

- (74) 代理人: 北京润泽恒知识产权代理有限公司 (BEIJING RUN ZEHENG INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区中关村南大街甲 18 号北京国际 C 座 6 层 606, Beijing 100081 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,

(54) Title: DIALOGUE-MODEL TRAINING METHOD AND APPARATUS, AND DIALOGUE RESPONSE METHOD AND APPARATUS

(54) 发明名称: 一种对话模型的训练方法及装置、对话响应方法及装置



- S101 Train an original dialogue model by using a pre-acquired universal dialogue data set, so as to obtain a universal dialogue model
- S102 Acquire a preset professional keyword group, perform data screening on the universal dialogue data set according to the professional keyword group, and determine a data set, which is obtained by means of screening, to be an initial labelling data set
- S103 Train the universal dialogue model by using the initial labelling data set, so as to obtain an initial professional dialogue model
- S104 Perform a validation operation on the initial professional dialogue model by using a validation data set and a preset natural language processing evaluation indicator, so as to obtain a validation score
- S105 Determine whether the validation score is greater than a preset score threshold value
- S106 Determine the initial professional dialogue model to be a target professional dialogue model
- AA No
- BB Yes
- CC Continue to train the initial professional dialogue model

图 1

(57) Abstract: Disclosed in the present application is a dialogue-model training method. The method comprises: training an original dialogue model by using a universal dialogue data set, so as to obtain a universal dialogue model; acquiring a preset professional keyword group, and performing data screening on the universal dialogue data set according to the professional keyword group; training the universal dialogue model by using an initial labelling data set, which is obtained by means of screening, so as to obtain an initial professional dialogue model; validating the initial professional dialogue model by using a validation data set and a preset natural language processing evaluation indicator, so as to obtain a validation score; determining whether the validation score is greater than a preset score threshold value; and if the validation score is greater than the preset score threshold value, determining the initial professional dialogue model as a target professional dialogue model. The present application enables a trained target professional dialogue model to have universality and specialty at the same time, thus improving the usage experience of users.



HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布：
— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本申请公开了一种对话模型的训练方法，包括：利用通用对话数据集对原始对话模型进行训练，得到通用对话模型；获取预设的专业关键词组，根据专业关键词组对通用对话数据集进行数据筛选；利用筛选到的初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证，得到验证得分；判断验证得分是否大于预设得分阈值；若验证得分大于预设得分阈值，则将初始专业对话模型确定为目标专业对话模型。本申请使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

一种对话模型的训练方法及装置、对话响应方法及装置

相关申请的交叉引用

本申请要求于 2022 年 11 月 17 日提交中国专利局，申请号为 202211441290.4，申请名称为“一种对话模型的训练方法及装置、对话响应方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能技术领域，特别是涉及一种对话模型的训练方法及装置、对话响应方法及装置、电子设备及存储介质。

背景技术

人机对话作为自然语言处理（Natural Language Processing, NLP）的一项基础应用，一直备受学术界和工业界的重视。随着人工智能技术的发展，基于生成式的对话模型越来越受欢迎，它们专门针对对话数据进行训练，在开放域对话获得了非常好的表现。但是从头开始训练对话大模型，需要大量的多类型对话数据作为训练语料，这需要的成本比较高，训练耗时长。

在专业的人机对话系统中常常也会存在不同的聊天需求，包括：闲聊、常识问答以及专业问答等。如医疗机器人在与病人聊天过程中不仅要回答医疗专业知识，生活中常识问题也会涉及到，还要会闲聊排解病人的情绪。目前专业的对话模型大部分采用检索方式，其主要原理是语义匹配，即在知识库中找到用户提问的问题的答案。虽然技术已比较成熟，但太依赖于语料，知识片面、回复单一生硬，缺乏通用性和多样性，用户体验感差。

发明内容

本申请的目的是提供一种对话模型的训练方法及装置、对话响应方法及装置、电子设备及存储介质，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升用户使用体验。

为解决上述技术问题，本申请提供如下技术方案：

一种对话模型的训练方法，包括：

利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；

获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；

利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；

利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到

验证得分；

判断验证得分是否大于预设得分阈值；

若验证得分大于预设得分阈值，则将初始专业对话模型确定为目标专业对话模型。

本申请一些实施例中，当确定验证得分小于等于预设得分阈值时，还包括：

利用初始专业对话模型针对预设未标记池中各样本数据生成对应的响应数据；

分别计算各响应数据对应的自动评估得分；

对各自动评估得分进行大小排序，从得分较小的一端选取预设数量个自动评估得分；

输出对选取得到的各自动评估得分对应的响应数据进行标注的标注提示信息；

根据标注结果对初始标注数据集进行更新，得到更新后标注数据集；

基于更新后标注数据集对初始专业对话模型进行训练，得到更新后专业对话模型；

利用验证数据集对更新后专业对话模型进行验证操作，得到验证得分，并重复执行判断验证得分是否大于预设得分阈值的步骤。

本申请一些实施例中，在得到更新后标注数据集之后，还包括：

根据更新后标注数据集对预设未标记池进行更新操作。

本申请一些实施例中，利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，包括：

通过以下公式结合验证数据集、BLEU 指标、ROUGE 指标、PPL 指标、DISTINCT 指标对初始专业对话模型进行验证操作：

$$Score_{val} = \left(Score_{BLEU} + Score_{ROUGE} + \frac{1}{Score_{PPL}} + Score_{distinct} \right) / 4 ;$$

其中， $Score_{BLEU}$ 为初始专业对话模型在 BLEU 指标上的得分， $Score_{ROUGE}$ 为初始专业对话模型在 ROUGE 指标上的得分， $Score_{PPL}$ 为初始专业对话模型在 PPL 指标上的得分，采用 PPL 指标得分的倒数形式， $Score_{distinct}$ 为初始专业对话模型在 DISTINCT 指标上的得分， $Score_{val}$ 为验证得分。

本申请一些实施例中，还包括初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程，初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程包括：

通过以下公式计算初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ ：

$$Score_{BLEU} = BP \times e^{\left(\sum_{n=1}^N w_n \times \log P_n \right)}$$

其中,
$$BP = \begin{cases} 1 & lc > lr \\ e^{(1-lr/lc)} & lc \leq lr \end{cases}$$
, lc 为机器译文的长度, lr 为最短的参考翻

译句子的长度, P_n 为 n-gram 的精确率, W_n 为 n-gram 的权重, 对于任意 n 都有 $W_n = 1/N$, BP 为惩罚因子。

本申请一些实施例中, 还包括初始专业对话模型在 ROUGE 指标上的得分

$Score_{ROUGE}$ 的计算过程, 初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ 的计算过程包括:

通过以下公式计算初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$:

$$Score_{ROUGE} = \frac{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count(gram_N)};$$

其中, $\{\text{参考译文}\}$ 表示参考译文集合, $gram_N$ 表示 N 个字的组合, $Count(gram_N)$ 表示计算译文中 N-gram 的数目, 公式的分母是统计在所有的参考译文中 N-gram 的个数, 分子是统计所有参考译文与机器译文共有的 N-gram 个数。

本申请一些实施例中, 还包括初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$ 的计算过程, 初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$ 的计算过程:

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})};$$

其中, $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率, N 代表句子长度。

本申请一些实施例中, 还包括初始专业对话模型在 DISTINCT 指标上的得分

$Score_{distinct}$ 的计算过程, 初始专业对话模型在 DISTINCT 指标上的得分

$Score_{distinct}$ 的计算过程包括:

通过以下公式计算初始专业对话模型在 DISTINCT 指标上的得分 $Score_{distinct}$:

$$Score_{distinct} = \frac{Count(unique\ ngram)}{Count(word)}$$

其中, $Count(unique\ ngram)$ 表示回复中不重复的 ngram 数量, $Count(word)$ 表示回复中 ngram 词语的总数量。

本申请一些实施例中, 在利用预获取的通用对话数据集对原始对话模型进行训练之前, 还包括:

分别对通用对话数据集中的问答数据和闲聊数据进行过滤。

本申请一些实施例中, 利用预获取的通用对话数据集对原始对话模型进行训练, 得到通用对话模型, 包括:

将通用对话数据集输入至原始对话模型进行模型迭代训练;

获取当前迭代数和本轮迭代训练得到的损失标准差;

根据当前迭代数和损失标准差确定是否达到模型训练截止条件;

若达到模型训练截止条件, 则将本轮迭代训练得到的对话模型确定为通用对话模型。

本申请一些实施例中, 根据当前迭代数和损失标准差确定是否达到模型训练截止条件, 包括:

判断当前迭代数是否大于第一预设值且损失标准差小于第二预设值。

本申请一些实施例中, 当确定当前迭代数大于第一预设值且损失标准差大于等于第二预设值时, 还包括:

判断当前迭代数是否大于第三预设值; 其中, 第三预设值大于第一预设值;

若是, 则执行将本轮迭代训练得到的对话模型确定为通用对话模型的步骤;

若否, 则将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练, 并重复执行获取当前迭代数和本轮迭代训练得到的损失标准差的步骤。

本申请一些实施例中, 根据专业关键词组对通用对话数据集进行数据筛选, 包括:

利用 DFA 算法根据专业关键词组对通用对话数据集进行数据筛选。

一种对话响应方法, 应用于包含如前训练得到的目标专业对话模型的对话系统, 包括:

接收待响应的目标提问语音;

利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音;

对目标响应语音进行输出操作。

本申请一些实施例中, 还包括:

当目标专业对话模型对目标提问语音响应失败时, 基于预设检索算法从数据库中搜索相关答案;

对相关答案进行语音输出。

一种对话模型的训练装置, 包括:

通用对话模型获得模块，用于利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；

初始标注数据集确定模块，用于获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；

初始专业对话模型获得模块，用于利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；

验证得分获得模块，用于利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分；

判断模块，用于判断验证得分是否大于预设得分阈值；

目标专业对话模型确定模块，用于当验证得分大于预设得分阈值时，将初始专业对话模型确定为目标专业对话模型。

一种对话响应装置，包括：

提问语音接收模块，用于接收待响应的目标提问语音；

响应语音生成模块，用于利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音；

响应语音输出模块，用于对目标响应语音进行输出操作。

一种电子设备，包括：

存储器，用于存储计算机程序；

处理器，用于执行计算机程序时实现如前对话模型的训练方法或对话响应方法的步骤。

一种非易失性可读存储介质，非易失性可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现如前对话模型的训练方法或对话响应方法的步骤。

本申请所提供的对话模型的训练方法，利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分；判断验证得分是否大于预设得分阈值；若验证得分大于预设得分阈值，则将初始专业对话模型确定为目标专业对话模型。

由上述技术方案可知，通过预先基于通用对话模型训练得到应用于特定对话场景的目标专业对话模型，大大减少了对数据量和计算力的需求，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

相应的，本申请还提供了与上述对话模型的训练方法相对应的对话模型的训练装置、对话响应方法及装置、电子设备及存储介质，具有上述技术效果，在此不再赘述。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本申请实施例中对话模型的训练方法的一种实施流程图；

图 2 为本申请实施例中对话模型的训练方法的另一种实施流程图；

图 3 为本申请实施例中一种对话响应方法的实施流程图；

图 4 为本申请实施例中一种对话模型的训练装置的结构框图；

图 5 为本申请实施例中一种对话响应装置的结构框图；

图 6 为本申请实施例中一种电子设备的结构框图；

图 7 为本申请实施例提供的一种电子设备的具体结构示意图。

具体实施方式

为了使本技术领域的人员更好地理解本申请方案，下面结合附图和具体实施方式对本申请作进一步的详细说明。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。

参见图 1，图 1 为本申请实施例中对话模型的训练方法的一种实施流程图，该方法可以包括以下步骤：

S101：利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型。

预先在公开数据集中收集通用对话数据集，通用对话数据集可以分为问答和闲聊两大类。问答数据可以涉及常识、实事、母婴、医疗、法律、保险、航空、心理、中医以及疫情等多领域。闲聊数据可以包括微博讨论、电视剧对白、贴吧讨论、豆瓣评论、电商对话等多个数据集，可以涉及历史、电影、天气、娱乐、体育等日常生活的各种话题讨论。

构建通用对话数据集的具体示例如下：

词条解释类 prompt 格式为标题：“title”，文章：“text”。原始语料示例{"id": "0", "url": "https://xxx", "title": "经济学", "text": "经济学是一门对产品和服务的生产、分配以及消费进行研究的社会科学....."}，按 prompt 格式组成后：标题：“经济学”，文章：“经济学是一门对产品和服务的生产、分配以及消费进行研究的社会科学.....”。

问答类 prompt 格式：问：“title+desc” 答：“answer”。原始语料示例{"qid": 0, "title": "AlphaGo 只会下围棋吗，阿法狗能写小说吗", "desc": "现在会不会有智能机器人能从事文学创作，
如果有，能写出什么水平的作品", "answer": "AlphaGo 只会下围棋，因为它的设计目的，架构，技术方案以及训练数据，都是围绕下围棋这个核心进行的....."}，按 prompt 格式组成：问：“AlphaGo 只会下围棋吗，阿法狗能写小说吗，现在会不会有智能机器人能从事文学创作，如果有，能写出什么水平的作品” 答：“AlphaGo 只会下围棋，因为它的设计目的，架构，技术方案以及训练数据，都是围绕下围棋这个核心进行的.....”。

阅读理解类 prompt 格式：context 问：“question” 答：“answer”。原始语料示例{"id": "0", "context": "胆石症的治疗应区别不同情况分别处理，无症状胆囊结石可不作治疗，但应定期观察并注意良好的饮食习惯.....", "question": "什么类型的胆囊结石可不作治疗", "answer": "无症状胆囊结石"}，按 prompt 格式组成后：胆石症的治疗应区别不同情况分别处理，无症状胆囊结石可不作治疗，但应定期观察并注意良好的饮食习惯.....问：“什么类型的胆囊结石可不作治疗” 答：“无症状胆囊结石”。

单轮或多轮对话类 prompt 格式：对话：“dialog1” “dialog2” “dialog3”。按 prompt 格式组成后：对话：“怎么不直播了，我都看不到你人” “不播了” “可惜我这么喜欢

你”……

利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型。

S102：获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集。

专业对话数据集一般由专家标记，虽然数据需求量远小于通用对话数据集，但是只靠专家标注的话也非常耗时耗力，因此预先设置专业关键词组。在利用通用对话数据集对原始对话模型进行训练，得到通用对话模型之后，获取预设的专业关键词组，并根据专业关键词组

对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集，记为 L_0 。通过设置专业关键词组从通用对话数据集中筛选得到初始标注数据集，相较于单纯的人工标注方法，较大地提升了专业对话数据集的生成效率。

本申请一些实施例中，根据专业关键词组对通用对话数据集进行数据筛选，可以包括以下步骤：

利用 DFA 算法根据专业关键词组对通用对话数据集进行数据筛选。

当从通用对话数据集中筛选专业对话数据集时，利用 DFA 算法根据专业关键词组对通用对话数据集进行数据筛选。从而充分利用 DFA 算法可以实现高效的关键词匹配的同时，也可以实现敏感词过滤的优势。

本申请实施例采用 DFA 算法实现关键词匹配，从通用对话数据集中筛选出专业对话数据的流程可以包括以下步骤：

(1) 专家提供专业关键词组；

(2) 将专业关键词组通过建立嵌套字典的方式构建专业词链表（以一个特定字符‘\x00’作为结束）；

(3) 遍历通用对话数据集中的每组对话，将其作为输入遍历专业词链表，如果遇到特定字符\x00，说明该组对话包含专业关键词，筛选出来。

专业对话数据虽然可以通过如上关键词匹配筛选出一部分，但是通常通用话数据集中涉及的专业对话有限，特别是一些偏门的专业，所以还是需要专家标注。专家标注的数据如涉及隐私，需添加脱敏处理（隐藏对话中姓名、手机号、邮箱等私密信息）。与构建通用对话数据集一样，将专业对话数据集按表 1 的 prompt 格式组成。

构建服务器专业对话数据集的具体示例如下：

如服务器智能客服属于多轮对话，对话内容：“您好，请问有什么可以帮您。”“status 灯红就是跟电源有关是嘛，这个不影响服务器正常运行吧”“status 是一个总的灯，机器有问题就会亮，建议把 4 路电都插上。”“现场没条件插 4 路电源，有没有办法不让 status 灯亮”“有，用指令，把电源策略刷成双电。”

S103：利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型。

在得到初始标注数据集之后，利用初始标注数据集 L_0 对通用对话模型进行训练，得到初始专业对话模型，记为 G_0 。

S104: 利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分。

在训练得到初始专业对话模型 G_0 之后，利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分，记为 $Score_{vai}$ 。通过验证得分预估初始专业对话模型对语音提问的响应性能。

S105: 判断验证得分是否大于预设得分阈值，若验证得分大于预设得分阈值，则执行步骤 **S106**，若验证得分不大于预设得分阈值，则对初始专业对话模型继续训练。

预先设置得分阈值，在利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作之后，判断验证得分是否大于预设得分阈值，若验证得分大于预设得分阈值，则说明模型已经训练好，执行步骤 **S106**，若验证得分不大于预设得分阈值，则说明需要对初始专业对话模型继续训练。

S106: 将初始专业对话模型确定为目标专业对话模型。

当确定验证得分大于预设得分阈值时，说明模型已经训练好，将初始专业对话模型确定为目标专业对话模型。还可以输出目标专业对话模型和当前所有专家标注数据集。通过根据预设得分阈值判别专业对话模型是否已训练完成，保证了训练得到的目标专业对话模型能够对提问语音有较好的答案生成能力。

由上述技术方案可知，通过预先基于通用对话模型训练得到应用于特定对话场景的目标专业对话模型，大大减少了对数据量和计算力的需求，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

参见图 2，图 2 为本申请实施例中对话模型的训练方法的另一种实施流程图，该方法可以包括以下步骤：

S201: 利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型。

本申请一些实施例中，在步骤 **S201** 之前，该对话模型的训练方法还可以包括以下步骤：分别对通用对话数据集中的问答数据和闲聊数据进行过滤。

在获取到通用对话数据集之后，分别对通用对话数据集中的问答数据和闲聊数据进行过滤。例如，由于问答数据集整体噪音较小，可以只进行简单过滤，包括去除包含敏感词的对话、去除不足 $m(m \leq 10)$ 个字的对话、去除问题与答案相同的对话、去除语料中无意义字符等。由于闲聊数据集整体噪音较大，需进行严格过滤。采用的过滤方式包括去除包含敏感词的对话、去除不足 $m(m \leq 10)$ 个字的对话、去除只有一句的对话、去除不包含中文字符的对话、删除广告对话、删除重复对话、去除语料中无意义字符等。通过利用经过滤后的通用对话数据集对原始对话模型进行训练，避免了无用数据的干扰，降低了模型训练复杂度，提升了模型训练效率，提升了训练得到的模型的准确性。

为了使训练效果更好，还可以将数据集按不同类别分别按一定的 prompt（提示）格式组成，如下：

表 1

类别	Prompt 格式
词条解释类	标题: “title”文章: “text”
问答类	问: “title+desc”答: “answer”
阅读理解类	Context (上下文) 问: “question”答: “answer”
单轮对话类	对话: “dialog1”“dialog2”
多轮对话类	对话: “dialog1”“dialog2”“dialog3”.....

通过固定的 prompt 格式, 减少后续处理工作。

本申请一些实施例中, 步骤 S201 可以包括以下步骤:

步骤一: 将通用对话数据集输入至原始对话模型进行模型迭代训练;

步骤二: 获取当前迭代数和本轮迭代训练得到的损失标准差;

步骤三: 根据当前迭代数和损失标准差确定是否达到模型训练截止条件, 若达到模型训练截止条件, 则执行步骤四, 若未达到模型训练截止条件, 则执行步骤五;

步骤四: 将本轮迭代训练得到的对话模型确定为通用对话模型;

步骤五: 将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练, 并返回执行步骤二。

为方便描述, 可以将上述五个步骤结合起来进行说明。

利用通用对话数据集对原始对话模型进行训练得到通用对话模型的过程可以包括将通用对话数据集输入至原始对话模型进行模型迭代训练, 获取当前迭代数和本轮迭代训练得到的损失标准差, 根据当前迭代数和损失标准差确定是否达到模型训练截止条件, 若达到模型训练截止条件, 则说明当前训练得到的模型已经能够对通用提问给出较好的语音响应, 将本轮迭代训练得到的对话模型确定为通用对话模型, 若未达到模型训练截止条件, 则说明当前训练得到的模型还不能对通用提问给出较好的语音响应, 将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练, 并再次获取当前迭代数和本轮迭代训练得到的损失标准差, 通过多次训练迭代, 使得模型得到不断优化。

需要说明的是, 模型训练截止条件可以根据实际情况进行设定和调整, 本申请实施例对此不做限定, 如可以设置为迭代次数上限, 还可以设置为损失阈值。

本申请一些实施例中, 根据当前迭代数和损失标准差确定是否达到模型训练截止条件, 可以包括以下步骤:

步骤一: 将通用对话数据集输入至原始对话模型进行模型迭代训练;

步骤二: 获取当前迭代数和本轮迭代训练得到的损失标准差;

步骤三: 判断当前迭代数是否大于第一预设值且损失标准差小于第二预设值, 若当前迭代数大于第一预设值且损失标准差小于第二预设值, 则执行步骤四, 若不满足当前迭代数大于第一预设值且损失标准差小于第二预设值, 则当确定当前迭代数大于第一预设值且损失标准差大于等于第二预设值时, 执行步骤五;

步骤四: 将本轮迭代训练得到的对话模型确定为通用对话模型;

步骤五: 判断当前迭代数是否大于第三预设值, 若当前迭代数大于第三预设值, 则返回执行步骤四, 若当前迭代数不大于第三预设值, 则执行步骤六;

其中, 第三预设值大于第一预设值;

步骤六：将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练，并返回执行步骤二。

为方便描述，可以将上述六个步骤结合起来进行说明。

预先设置模型训练中的超参数，超参数可以包括对迭代数 $iters$ 进行预训练得到的预训练最小迭代数 N_1 （即第一预设值）、损失标准差 $\Delta loss$ 的标准差阈值 T （即第二预设值），损失标准差 $\Delta loss$ 表示最新十个迭代 loss 的标准差。在获取当前迭代数和本轮迭代训练得到的损失标准差之后，通过判断当前迭代数是否大于第一预设值且损失标准差小于第二预设值，即 $iters > N_1 \text{ and } \Delta loss < T$ ，从而确定是否已经达到模型训练截止条件。通过结合当前迭代数和损失标准差进行模型训练阶段判定，保证了已对满足训练截止条件的模型进行了一定次数的迭代，提升了模型性能。

预先设置的模型训练中的超参数还可以包括对迭代数 $iters$ 进行预训练得到的预训练最大迭代数 N_2 （即第三预设值），第三预设值大于第一预设值，即 $N_2 > N_1$ 。当确定当前迭代数大于第一预设值且损失标准差大于等于第二预设值时，判断当前迭代数是否大于第三预设值，若当前迭代数大于第三预设值，则说明认为损失值下降缓慢，模型已训练至接近全局最优，将本轮迭代训练得到的对话模型确定为通用对话模型，若当前迭代数不大于第三预设值，则说明需要对模型继续训练，将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练，并再次获取当前迭代数和本轮迭代训练得到的损失标准差，基于本轮迭代的数据判断是否达到模型训练截止条件，如此往复直到达到预设的模型训练截止条件，从而得到能够对通用提问语音进行良好响应的通用对话模型。

S202：获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集。

S203：利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型。

S204：利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分。

本申请一些实施例中，利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，可以包括以下步骤：

通过以下公式结合验证数据集、BLEU 指标、ROUGE 指标、PPL 指标、DISTINCT 指标对初始专业对话模型进行验证操作：

$$Score_{val} = \left(Score_{BLEU} + Score_{ROUGE} + \frac{1}{Score_{PPL}} + Score_{distinct} \right) / 4 ;$$

其中， $Score_{BLEU}$ 为初始专业对话模型在 BLEU 指标上的得分， $Score_{ROUGE}$ 为

初始专业对话模型在 ROUGE 指标上的得分， $Score_{PPL}$ 为初始专业对话模型在 PPL 指标

上的得分, 采用 PPL 指标得分的倒数形式, $Score_{distinct}$ 为初始专业对话模型在 DISTINCT 指标上的得分, $Score_{vai}$ 为验证得分。

在对初始专业对话模型进行验证时, 可以结合验证数据集、BLEU 指标、ROUGE 指标、PPL 指标、DISTINCT 指标对初始专业对话模型进行验证操作。如可以通过以下公式进行计算验证得分:

$$Score_{vai} = \left(Score_{BLEU} + Score_{ROUGE} + \frac{1}{Score_{PPL}} + Score_{distinct} \right) / 4 ;$$

其中, $Score_{BLEU}$ 为初始专业对话模型在 BLEU 指标上的得分, $Score_{ROUGE}$ 为初始专业对话模型在 ROUGE 指标上的得分, $Score_{PPL}$ 为初始专业对话模型在 PPL 指标上的得分, 采用 PPL 指标得分的倒数形式, $\frac{1}{Score_{PPL}}$ 越小说明模型生成效果越差, $Score_{distinct}$ 为初始专业对话模型在 DISTINCT 指标上的得分。

通过采用 BLEU、ROUGE、PPL、DISTINCT 四个指标综合评判模型在验证数据集上的表现。在保证模型生成的通顺性和多样性的同时, 保证生成答案的精确率和召回率。

本申请一些实施例中, 该对话模型的训练方法还可以包括初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程, 初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程可以包括可以步骤:

通过以下公式计算初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$:

$$Score_{BLEU} = BP \times e^{\left(\sum_{n=1}^N w_n \times \log P_n \right)}$$

其中, $BP = \begin{cases} 1 & lc > lr \\ e^{(1-lr/lc)} & lc \leq lr \end{cases}$, lc 为机器译文的长度, lr 为最短的参考翻译句

子的长度, P_n 为 n-gram 的精确率, W_n 为 n-gram 的权重, 对于任意 n 都有 $W_n = 1/N$, BP 为惩罚因子。

BLEU 的核心思想是比较候选译文和参考译文里的 n-gram 的重合程度, 重合程度越高

就认为译文质量越高。实践中，通常是取 $N=1\sim4$ ，然后进行加权平均。

$$Score_{BLEU} = BP \times e^{\left(\sum_{n=1}^N W_n \times \log P_n\right)} ;$$

其中， $BP = \begin{cases} 1 & lc > lr \\ e^{(1-lr/lc)} & lc \leq lr \end{cases}$ ， lc 为机器译文的长度， lr 为最短的参考翻译

句子的长度， P_n 为 n -gram 的精确率， W_n 为 n -gram 的权重，一般设为均匀权重，即对

于任意 n 都有 $W_n = 1/N$ 。BP 为惩罚因子，如果译文的长度小于最短的参考译文，则 BP 小于 1。BLEU 的 1-gram 精确率表示译文忠于原文的程度，而其他 n -gram 表示翻译的流畅程度。

本申请一些实施例中，该对话模型的训练方法还可以包括初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ 的计算过程，初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ 的计算过程可以包括：

通过以下公式计算初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ ：

$$Score_{ROUGE} = \frac{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count(gram_N)} ;$$

其中， $\{\text{参考译文}\}$ 表示参考译文集合， $gram_N$ 表示 N 个字的组合， $Count(gram_N)$ 表示计算译文中 N -gram 的数目，公式的分母是统计在所有的参考译文中 N -gram 的个数，分子是统计所有参考译文与机器译文共有的 N -gram 个数。

ROUGE-N 专注于召回率而非精度。查看有多少个参考译句中的 n 元词组出现在了输出之中。“ N ”指的是 N -gram，其计算方式与 BLEU 类似，只是 BLEU 基于精确率，而 ROUGE 基于召回率。ROUGE-N 主要统计 N -gram 上的召回率，对于 N -gram，可以计算得到 ROUGE-N 分数，计算公式如下：

$$Score_{ROUGE} = \frac{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count(gram_N)} ;$$

其中， $\{\text{参考译文}\}$ 表示参考译文集合，在实际应用中参考译文可能是多个。

$gram_N$ 表示 N 个字的组合, $Count(gram_N)$ 表示计算译文中 N-gram 的数目。
公式的分母是统计在所有的参考译文中 N-gram 的个数, 而分子是统计所有参考译文与机器译文共有的 N-gram 个数。

本申请一些实施例中, 该对话模型的训练方法还可以包括初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$ 的计算过程, 初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$ 的计算过程:

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})};$$

其中, $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率, N 代表句子长度。

PPL 指的是语言模型中的 perplexity, 困惑度(Perplexity)是衡量一句话是否通顺的指标。定义为:

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})};$$

其中, $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率, N 代表句子长度。PPL 值越小, 说明模型生成的回复越自然、语句越通顺。通过 PPL 来评价回复质量, 可以避免模型生成的回复有乱序、前后颠倒的情形。

本申请一些实施例中, 该方法还可以包括初始专业对话模型在 DISTINCT 指标上的得分

$Score_{distinct}$ 的计算过程, 初始专业对话模型在 DISTINCT 指标上的得分

$Score_{distinct}$ 的计算过程包括:

通过以下公式计算初始专业对话模型在 DISTINCT 指标上的得分 $Score_{distinct}$:

$$Score_{distinct} = \frac{Count(unique\ ngram)}{Count(word)};$$

其中, $Count(unique\ ngram)$ 表示回复中不重复的 ngram 数量, $Count(word)$ 表示回复中 ngram 词语的总数量。

Distinct 评价指标判断机器回复的多样性，Distinct 指标判断是否出现大量的通用性、重复性回复。Distinct 的定义如下：

$$Score_{\text{distinct}} = \frac{\text{Count}(\text{unique ngram})}{\text{Count}(\text{word})};$$

其中， $\text{Count}(\text{unique ngram})$ 表示回复中不重复的 ngram 数量，

$\text{Count}(\text{word})$ 表示回复中 ngram 词语的总数量。

越大表示生成回复的多样性越高。

S205：判断验证得分是否大于预设得分阈值，若验证得分大于预设得分阈值，则执行步骤 S106，若验证得分不大于预设得分阈值，则执行步骤 S207。

S206：将初始专业对话模型确定为目标专业对话模型。

S207：利用初始专业对话模型针对预设未标记池中各样本数据生成对应的响应数据。

当确定验证得分小于等于预设得分阈值时，说明需要对模型进行继续训练，利用初始专

业对话模型 G_0 针对预设未标记池中各样本数据生成对应的响应数据。

S208：分别计算各响应数据对应的自动评估得分。

在利用初始专业对话模型针对预设未标记池中各样本数据生成对应的响应数据之后，分别计算各响应数据对应的自动评估得分。如可以根据 PPL 指标和 Distinct 指标计算各响应数据对应的自动评估得分，计算公式如下：

$$Score_U = \left(\frac{1}{Score_{PPL}} + Score_{\text{distinct}} \right) / 2;$$

从而得到每个响应数据分别对应的自动评估得分。

S209：对各自动评估得分进行大小排序，从得分较小的一端选取预设数量个自动评估得分。

在分别计算得到各响应数据对应的自动评估得分之后，对各自动评估得分进行大小排序，从得分较小的一端选取预设数量个自动评估得分，如选择最低的 N 个 $Score_U$ 得分。

S210：输出对选取得到的各自动评估得分对应的响应数据进行标注的标注提示信息。

在从得分较小的一端选取预设数量个自动评估得分之后，输出对选取得到的各自动评估得分对应的响应数据进行标注的标注提示信息，从而提示对最低的 N 个 $Score_U$ 得分对应的响应数据进行专家标注。

S211：根据标注结果对初始标注数据集进行更新，得到更新后标注数据集。

在输出对选取得到的各自动评估得分对应的响应数据进行标注的标注提示信息之后，获取标注结果，根据标注结果对初始标注数据集进行更新，得到更新后标注数据集，从而实现当前专业对话模型生成响应数据效果不好的数据的有效标注。

本申请一些实施例中，在步骤 S211 之后，该对话模型的训练方法还可以包括以下步骤：根据更新后标注数据集对预设未标记池进行更新操作。

在得到更新后标注数据集之后，根据更新后标注数据集对预设未标记池进行更新操作，从而实现对预设未标记池中未标记样本数据的及时更新。

S212：基于更新后标注数据集对初始专业对话模型进行训练，得到更新后专业对话模型。

在根据标注结果对初始标注数据集进行更新，得到更新后标注数据集之后，基于更新后标注数据集对初始专业对话模型进行训练，得到更新后专业对话模型。

本申请实施例通过采用主动学习的方式，尽可能减少了专家标注样本量，降低了对模型性能的影响。不断从预设未标记池中选择对模型性能提升最大的“困难样本”，提升了模型性能。

S213：利用验证数据集对更新后专业对话模型进行验证操作，得到验证得分，并返回执行步骤 S205。

在基于更新后标注数据集对初始专业对话模型进行训练，得到更新后专业对话模型之后，利用验证数据集对更新后专业对话模型进行验证操作，得到验证得分，并返回执行判断验证得分是否大于预设得分阈值的步骤，如此往复直到计算得到的验证得分大于预设得分阈值，从而得到能够对接收到的提问语音进行良好响应的目标专业对话模型。

参见图 3，图 3 为本申请实施例中一种对话响应方法的实施流程图，应用于包含如前训练得到的目标专业对话模型的对话系统，该方法可以包括以下步骤：

S301：接收待响应的目标提问语音。

当用户需要进行情景对话时，向对话响应控制中心输出目标提问语音，对话响应控制中心接收待响应的目标提问语音。

对话响应控制中心可以为部署有对话模型的处理器。

目标提问语音可以为闲聊、常识问答以及专业问答等。

S302：利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音。

预先训练通用对话模型，如可以基于大模型在通用对话数据集上进行模型训练，获得通用对话模型，其中大模型可以是基于 Transformer 结构，适用于生成任务，诸如 GPT（Generative Pre-Training）模型、BERT（Bidirectional Encoder Representation from Transformers）模型等。再基于通用对话模型训练得到目标专业对话模型。在接收到待响应的目标提问语音之后，利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音。

通过在大模型的基础上进行再训练，大大减少了对数据量和计算力的需求，并且采用两阶段训练模型方式，使得训练出来的目标专业对话模型同时拥有通用性和专业性。

S303：对目标响应语音进行输出操作。

在利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音之后，对目标响应语音进行输出操作，从而实现对目标提问语音的响应。

由于模型训练过程相较于模型应用过程需要较多的资源，因此可以预先为模型训练过程分配较多的资源，为模型应用过程分配相对较少的资源。如可以预先划分 8 块及以上 80G 大小的 GPU（Graphics Processing Unit，图像处理器）用于进行模型训练，划分 1 块及以上

80G 大小的 GPU 用于模型应用。

由上述技术方案可知，通过预先基于通用对话模型训练得到应用于特定对话场景的目标专业对话模型，大大减少了对数据量和计算力的需求，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

本申请一些实施例中，该对话响应方法还可以包括以下步骤：

步骤一：当目标专业对话模型对目标提问语音响应失败时，基于预设检索算法从数据库中搜索相关答案；

步骤二：对相关答案进行语音输出。

为方便描述，可以将上述两个步骤结合起来进行说明。

本申请实施例预先设置兜底方案，利用专业数据集构建专业数据库，当目标专业对话模型对目标提问语音响应失败时，即当目标专业对话模型输出为空时，基于预设检索算法从数据库中搜索相关答案，对相关答案进行语音输出。从而优化专业对话模型应用流程，进一步保证用户提问语音不会落空，提升用户体验。

相应于上面的方法实施例，本申请还提供了一种对话模型的训练装置，下文描述的对话模型的训练装置与上文描述的对话模型的训练方法可相互对应参照。

参见图 4，图 4 为本申请实施例中一种对话模型的训练装置的结构框图，该对话模型的训练装置可以包括：

通用对话模型获得模块 41，用于利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；

初始标注数据集确定模块 42，用于获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；

初始专业对话模型获得模块 43，用于利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；

验证得分获得模块 44，用于利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分；

判断模块 45，用于判断验证得分是否大于预设得分阈值；

目标专业对话模型确定模块 46，用于当验证得分大于预设得分阈值时，将初始专业对话模型确定为目标专业对话模型。

由上述技术方案可知，通过预先基于通用对话模型训练得到应用于特定对话场景的目标专业对话模型，大大减少了对数据量和计算力的需求，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

本申请一些实施例中，该对话模型的训练装置还可以包括：

响应数据生成模块，用于当确定验证得分小于等于预设得分阈值时，利用初始专业对话模型针对预设未标记池中各样本数据生成对应的响应数据；

自动评估得分计算模块，用于分别计算各响应数据对应的自动评估得分；

自动评估得分选取模块，用于对各自动评估得分进行大小排序，从得分较小的一端选取预设数量个自动评估得分；

标注提示信息输出模块，用于输出对选取得到的各自动评估得分对应的响应数据进行标注的标注提示信息；

标注数据集更新模块，用于根据标注结果对初始标注数据集进行更新，得到更新后标注数据集；

专业对话模型更新模块，用于基于更新后标注数据集对初始专业对话模型进行训练，得到更新后专业对话模型；

重复执行模块，用于利用验证数据集对更新后专业对话模型进行验证操作，得到验证得分，并重复执行判断验证得分是否大于预设得分阈值的步骤。

本申请一些实施例中，该对话模型的训练装置还可以包括：

未标记池更新模块，用于在得到更新后标注数据集之后，根据更新后标注数据集对预设未标记池进行更新操作。

本申请一些实施例中，验证得分获得模块 44，具体用于通过以下公式结合验证数据集、BLEU 指标、ROUGE 指标、PPL 指标、DISTINCT 指标对初始专业对话模型进行验证操作：

$$Score_{val} = \left(Score_{BLEU} + Score_{ROUGE} + \frac{1}{Score_{PPL}} + Score_{distinct} \right) / 4 ;$$

其中， $Score_{BLEU}$ 为初始专业对话模型在 BLEU 指标上的得分， $Score_{ROUGE}$ 为初始专业对话模型在 ROUGE 指标上的得分， $Score_{PPL}$ 为初始专业对话模型在 PPL 指标上的得分，采用 PPL 指标得分的倒数形式， $Score_{distinct}$ 为初始专业对话模型在 DISTINCT 指标上的得分， $Score_{val}$ 为验证得分。

本申请一些实施例中，该对话模型的训练装置还可以包括：

BLEU 指标上的得分计算模块，用于通过以下公式计算初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ ：

$$Score_{BLEU} = BP \times e^{\left(\sum_{n=1}^N W_n \times \log P_n \right)} ;$$

其中， $BP = \begin{cases} 1 & lc > lr \\ e^{1-lr/lc} & lc \leq lr \end{cases}$ ， lc 为机器译文的长度， lr 为最短的参考翻译句子的长度， P_n 为 n-gram 的精确率， W_n 为 n-gram 的权重，对于任意 n 都有 $W_n = 1/N$ ，BP 为惩罚因子。

本申请一些实施例中，该对话模型的训练装置还可以包括：

ROUGE 指标上的得分计算模块，用于通过以下公式计算初始专业对话模型在 ROUGE

指标上的得分 $Score_{ROUGE}$:

$$Score_{ROUGE} = \frac{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count(gram_N)} ;$$

其中, $\{\text{参考译文}\}$ 表示参考译文集合, $gram_N$ 表示 N 个字的组合, $Count(gram_N)$ 表示计算译文中 N -gram 的数目, 公式的分母是统计在所有的参考译文中 N -gram 的个数, 分子是统计所有参考译文与机器译文共有的 N -gram 个数。

本申请一些实施例中, 该对话模型的训练装置还可以包括:

PPL 指标上的得分计算模块, 用于通过以下公式计算初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$:

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})} ;$$

其中, $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率, N 代表句子长度。

本申请一些实施例中, 该对话模型的训练装置还可以包括:

DISTINCT 指标上的得分计算模块, 用于通过以下公式计算初始专业对话模型在 DISTINCT 指标上的得分 $Score_{distinct}$:

$$Score_{distinct} = \frac{Count(unique\ ngram)}{Count(word)} ;$$

其中, $Count(unique\ ngram)$ 表示回复中不重复的 $ngram$ 数量, $Count(word)$ 表示回复中 $ngram$ 词语的总数量。

本申请一些实施例中, 该对话模型的训练装置还可以包括:

数据过滤模块, 用于在利用预获取的通用对话数据集对原始对话模型进行训练之前, 分别对通用对话数据集中的问答数据和闲聊数据进行过滤。

本申请一些实施例中, 通用对话模型获得模块 41 包括:

迭代训练子模块, 用于将通用对话数据集输入至原始对话模型进行模型迭代训练;

损失标准差获取子模块，用于获取当前迭代数和本轮迭代训练得到的损失标准差；

训练截止判断子模块，用于根据当前迭代数和损失标准差确定是否达到模型训练截止条件；

通用对话模型确定子模块，用于当根据当前迭代数和损失标准差确定达到模型训练截止条件时，将本轮迭代训练得到的对话模型确定为通用对话模型。

本申请一些实施例中，训练截止判断子模块具体为判断当前迭代数是否大于第一预设值且损失标准差小于第二预设值的模块。

本申请一些实施例中，该对话模型的训练装置还可以包括：

迭代数统计子模块，用于当确定当前迭代数大于第一预设值且损失标准差大于等于第二预设值时，判断当前迭代数是否大于第三预设值；其中，第三预设值大于第一预设值；

通用对话模型确定子模块，还用于当当前迭代数大于第三预设值时，将本轮迭代训练得到的对话模型确定为通用对话模型；

迭代训练子模块，还用于当当前迭代数小于等于第三预设值时，将通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练，并重复执行获取当前迭代数和本轮迭代训练得到的损失标准差的步骤。

本申请一些实施例中，初始标注数据集确定模块 42 具体为利用 DFA 算法根据专业关键词组对通用对话数据集进行数据筛选的模块。

相应于上面的方法实施例，本申请还提供了一种对话响应装置，下文描述的对话响应装置与上文描述的对话响应方法可相互对应参照。

参见图 5，图 5 为本申请实施例中一种对话响应装置的结构框图，该对话响应装置可以包括：

提问语音接收模块 51，用于接收待响应的目标提问语音；

响应语音生成模块 52，用于利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音；

响应语音输出模块 53，用于对目标响应语音进行输出操作。

由上述技术方案可知，通过预先基于通用对话模型训练得到应用于特定对话场景的目标专业对话模型，大大减少了对数据量和计算力的需求，使得训练出来的目标专业对话模型同时拥有通用性和专业性，提升了用户使用体验。

本申请一些实施例中，该对话响应装置还可以包括：

答案搜索模块，用于当目标专业对话模型对目标提问语音响应失败时，基于预设检索算法从数据库中搜索相关答案；

语音输出模块，用于对相关答案进行语音输出。

相应于上面的方法实施例，参见图 6，图 6 为本申请所提供的电子设备的示意图，该设备可以包括：

存储器 332，用于存储计算机程序；

处理器 322，用于执行计算机程序时实现上述方法实施例的对话模型的训练方法或对话响应方法的步骤。

具体的，请参考图7，图7为本申请实施例提供的一种电子设备的具体结构示意图，该电

电子设备可因配置或性能不同而产生比较大的差异，可以包括处理器（central processing units, CPU）322（例如，一个或一个以上处理器）和存储器332，存储器332存储有一个或一个以上的计算机程序342或数据344。其中，存储器332可以是短暂存储或持久存储。存储在存储器332的程序可以包括一个或一个以上模块（图示没标出），每个模块可以包括对数据处理设备中的一系列指令操作。更进一步地，处理器322可以设置为与存储器332通信，在电子设备301上执行存储器332中的一系列指令操作。

电子设备301还可以包括一个或一个以上电源326，一个或一个以上有线或无线网络接口350，一个或一个以上输入输出接口358，和/或，一个或一个以上操作系统341。

上文所描述的对话响应方法中的步骤可以由电子设备的结构实现。

相应于上面的方法实施例，本申请还提供一种非易失性可读存储介质，非易失性可读存储介质上存储有计算机程序，计算机程序被处理器执行时可实现如下步骤：

利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；获取预设的专业关键词组，并根据专业关键词组对通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；利用初始标注数据集对通用对话模型进行训练，得到初始专业对话模型；利用验证数据集和预设自然语言处理评价指标对初始专业对话模型进行验证操作，得到验证得分；判断验证得分是否大于预设得分阈值；若验证得分大于预设得分阈值，则将初始专业对话模型确定为目标专业对话模型；

或，

接收待响应的目标提问语音；利用基于对通用对话模型训练得到的目标专业对话模型生成目标提问语音对应的目标响应语音；对目标响应语音进行输出操作。

该非易失性可读存储介质可以包括：U 盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

对于本申请提供的存储介质的介绍请参照上述方法实施例，本申请在此不做赘述。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。对于实施例公开的装置、电子设备及存储介质而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

本文中应用了具体个例对本申请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申请的技术方案及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这些改进和修饰也落入本申请权利要求的保护范围内。

权 利 要 求 书

- 1、一种对话模型的训练方法，其特征在于，包括：
 - 利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；
 - 获取预设的专业关键词组，并根据所述专业关键词组对所述通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；
 - 利用所述初始标注数据集对所述通用对话模型进行训练，得到初始专业对话模型；
 - 利用验证数据集和预设自然语言处理评价指标对所述初始专业对话模型进行验证操作，得到验证得分；
 - 判断所述验证得分是否大于预设得分阈值；
 - 若所述验证得分大于所述预设得分阈值，则将所述初始专业对话模型确定为目标专业对话模型。
- 2、根据权利要求 1 所述的对话模型的训练方法，其特征在于，当确定所述验证得分小于等于所述预设得分阈值时，还包括：
 - 利用所述初始专业对话模型针对预设未标记池中各样本数据生成对应的响应数据；
 - 分别计算各所述响应数据对应的自动评估得分；
 - 对各所述自动评估得分进行大小排序，从得分较小的一端选取预设数量个自动评估得分；
 - 输出对选取得到的各所述自动评估得分对应的响应数据进行标注的标注提示信息；
 - 根据标注结果对初始标注数据集进行更新，得到更新后标注数据集；
 - 基于所述更新后标注数据集对所述初始专业对话模型进行训练，得到更新后专业对话模型；
 - 利用所述验证数据集对所述更新后专业对话模型进行验证操作，得到验证得分，并重复执行所述判断所述验证得分是否大于预设得分阈值的步骤。
- 3、根据权利要求 2 所述的对话模型的训练方法，其特征在于，在得到更新后标注数据集之后，还包括：
 - 根据所述更新后标注数据集对所述预设未标记池进行更新操作。
- 4、根据权利要求 1 所述的对话模型的训练方法，其特征在于，利用验证数据集和预设自然语言处理评价指标对所述初始专业对话模型进行验证操作，包括：
 - 通过以下公式结合所述验证数据集、BLEU 指标、ROUGE 指标、PPL 指标、DISTINCT 指标对所述初始专业对话模型进行验证操作：

$$Score_{val} = \left(Score_{BLEU} + Score_{ROUGE} + \frac{1}{Score_{PPL}} + Score_{distinct} \right) / 4 ;$$

其中， $Score_{BLEU}$ 为所述初始专业对话模型在 BLEU 指标上的得分，

$Score_{ROUGE}$ 为所述初始专业对话模型在 ROUGE 指标上的得分， $Score_{PPL}$ 为所述初始专业对话模型在 PPL 指标上的得分，采用 PPL 指标得分的倒数形式，

$Score_{distinct}$ 为所述初始专业对话模型在 DISTINCT 指标上的得分, $Score_{val}$ 为验证得分。

5、根据权利要求 4 所述的对话模型的训练方法, 其特征在于, 还包括所述初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程, 所述初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$ 的计算过程包括:

通过以下公式计算所述初始专业对话模型在 BLEU 指标上的得分 $Score_{BLEU}$:

$$Score_{BLEU} = BP \times e^{\left(\sum_{n=1}^N w_n \times \log P_n\right)}$$

其中, $BP = \begin{cases} 1 & lc > lr \\ e^{(1-lr/lc)} & lc \leq lr \end{cases}$, lc 为机器译文的长度, lr 为最短的参

考翻译句子的长度, P_n 为 n-gram 的精确率, w_n 为 n-gram 的权重, 对于任意 n 都有 $w_n = 1/N$, BP 为惩罚因子。

6、根据权利要求 4 所述的对话模型的训练方法, 其特征在于, 还包括所述初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ 的计算过程, 所述初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$ 的计算过程包括:

通过以下公式计算所述初始专业对话模型在 ROUGE 指标上的得分 $Score_{ROUGE}$:

$$Score_{ROUGE} = \frac{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{\text{参考译文}\}} \sum_{gram_N \in S} Count(gram_N)}$$

其中, $\{\text{参考译文}\}$ 表示参考译文集合, $gram_N$ 表示 N 个字的组合,

$Count(gram_N)$ 表示计算译文中 N-gram 的数目, 公式的分母是统计在所有的参考译文中 N-gram 的个数, 分子是统计所有参考译文与机器译文共有的 N-gram 个数。

7、根据权利要求 4 所述的对话模型的训练方法, 其特征在于, 还包括所述初始专业对话模型在 PPL 指标上的得分 $Score_{PPL}$ 的计算过程, 所述初始专业对话模型在 PPL

指标上的得分 $Score_{PPL}$ 的计算过程：

$$Score_{PPL} = 2^{-\frac{1}{N} \sum_{i=1}^N \log_2 P(x_i | x_1, x_2, \dots, x_{i-1})} ;$$

其中， $P(x_i | x_1, x_2, \dots, x_{i-1})$ 表示根据上文词语预测第 i 个词的概率，N 代表句子长度。

8、根据权利要求 4 所述的对话模型的训练方法，其特征在于，还包括所述初始专业

对话模型在 DISTINCT 指标上的得分 $Score_{distinct}$ 的计算过程，所述初始专业对话模

型在 DISTINCT 指标上的得分 $Score_{distinct}$ 的计算过程包括：

通过以下公式计算所述初始专业对话模型在 DISTINCT 指标上的得分

$Score_{distinct}$ ：

$$Score_{distinct} = \frac{Count(unique\ ngram)}{Count(word)} ;$$

其中， $Count(unique\ ngram)$ 表示回复中不重复的 ngram 数量，

$Count(word)$ 表示回复中 ngram 词语的总数量。

9、根据权利要求 1 所述的对话模型的训练方法，其特征在于，在利用预获取的通用对话数据集对原始对话模型进行训练之前，还包括：

分别对所述通用对话数据集中的问答数据和闲聊数据进行过滤。

10、根据权利要求 1 所述的对话模型的训练方法，其特征在于，利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型，包括：

将所述通用对话数据集输入至所述原始对话模型进行模型迭代训练；

获取当前迭代数和本轮迭代训练得到的损失标准差；

根据所述当前迭代数和所述损失标准差确定是否达到模型训练截止条件；

若达到模型训练截止条件，则将本轮迭代训练得到的对话模型确定为所述通用对话模型。

11、根据权利要求 10 所述的对话模型的训练方法，其特征在于，根据所述当前迭代数和所述损失标准差确定是否达到模型训练截止条件，包括：

判断所述当前迭代数是否大于第一预设值且所述损失标准差小于第二预设值。

12、根据权利要求 11 所述的对话模型的训练方法，其特征在于，当确定所述当前迭

代数大于所述第一预设值且所述损失标准差大于等于所述第二预设值时，还包括：

判断所述当前迭代数是否大于第三预设值；其中，所述第三预设值大于所述第一预设值；

若所述当前迭代数大于第三预设值，则执行所述将本轮迭代训练得到的对话模型确定为所述通用对话模型的步骤；

若所述当前迭代数不大于第三预设值，则将所述通用对话数据集输入至本轮迭代训练得到的对话模型进行模型迭代训练，并重复执行所述获取当前迭代数和本轮迭代训练得到的损失标准差的步骤。

13、根据权利要求 1 所述的对话模型的训练方法，其特征在于，根据所述专业关键词组对所述通用对话数据集进行数据筛选，包括：

利用 DFA 算法根据所述专业关键词组对所述通用对话数据集进行数据筛选。

14、根据权利要求 13 所述的对话模型的训练方法，其特征在于，利用 DFA 算法根据所述专业关键词组对所述通用对话数据集进行数据筛选，包括：

获取专业关键词组；

将所述专业关键词组通过建立嵌套字典的方式构建专业词链表；

遍历所述通用对话数据集中的每组对话，将每组对话作为输入遍历所述专业词链表，筛选出包含专业关键词的对话。

15、一种对话响应方法，其特征在于，应用于包含如权利要求 1 至 14 任一项训练得到的目标专业对话模型的对话系统，包括：

接收待响应的目标提问语音；

利用基于对通用对话模型训练得到的目标专业对话模型生成所述目标提问语音对应的目标响应语音；

对所述目标响应语音进行输出操作。

16、根据权利要求 15 所述的对话响应方法，其特征在于，还包括：

当所述目标专业对话模型对所述目标提问语音响应失败时，基于预设检索算法从数据库中搜索相关答案；

对所述相关答案进行语音输出。

17、一种对话模型的训练装置，其特征在于，包括：

通用对话模型获得模块，用于利用预获取的通用对话数据集对原始对话模型进行训练，得到通用对话模型；

初始标注数据集确定模块，用于获取预设的专业关键词组，并根据所述专业关键词组对所述通用对话数据集进行数据筛选，将筛选得到的数据集确定为初始标注数据集；

初始专业对话模型获得模块，用于利用所述初始标注数据集对所述通用对话模型进行训练，得到初始专业对话模型；

验证得分获得模块，用于利用验证数据集和预设自然语言处理评价指标对所述初始专业对话模型进行验证操作，得到验证得分；

判断模块，用于判断所述验证得分是否大于预设得分阈值；

目标专业对话模型确定模块，用于当所述验证得分大于预设得分阈值时，将所述初始专业对话模型确定为目标专业对话模型。

18、一种对话响应装置，其特征在于，包括：

提问语音接收模块，用于接收待响应的目标提问语音；

响应语音生成模块，用于利用基于对通用对话模型训练得到的目标专业对话模型生成所述目标提问语音对应的目标响应语音；

响应语音输出模块，用于对所述目标响应语音进行输出操作。

19、一种电子设备，其特征在于，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序时实现如权利要求 1 至 14 任一项所述对话模型的训练方法或权利要求 15 至 16 任一项所述对话响应方法的步骤。

20、一种非易失性可读存储介质，其特征在于，所述非易失性可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求 1 至 14 任一项所述对话模型的训练方法或权利要求 15 至 16 任一项所述对话响应方法的步骤。

1/4

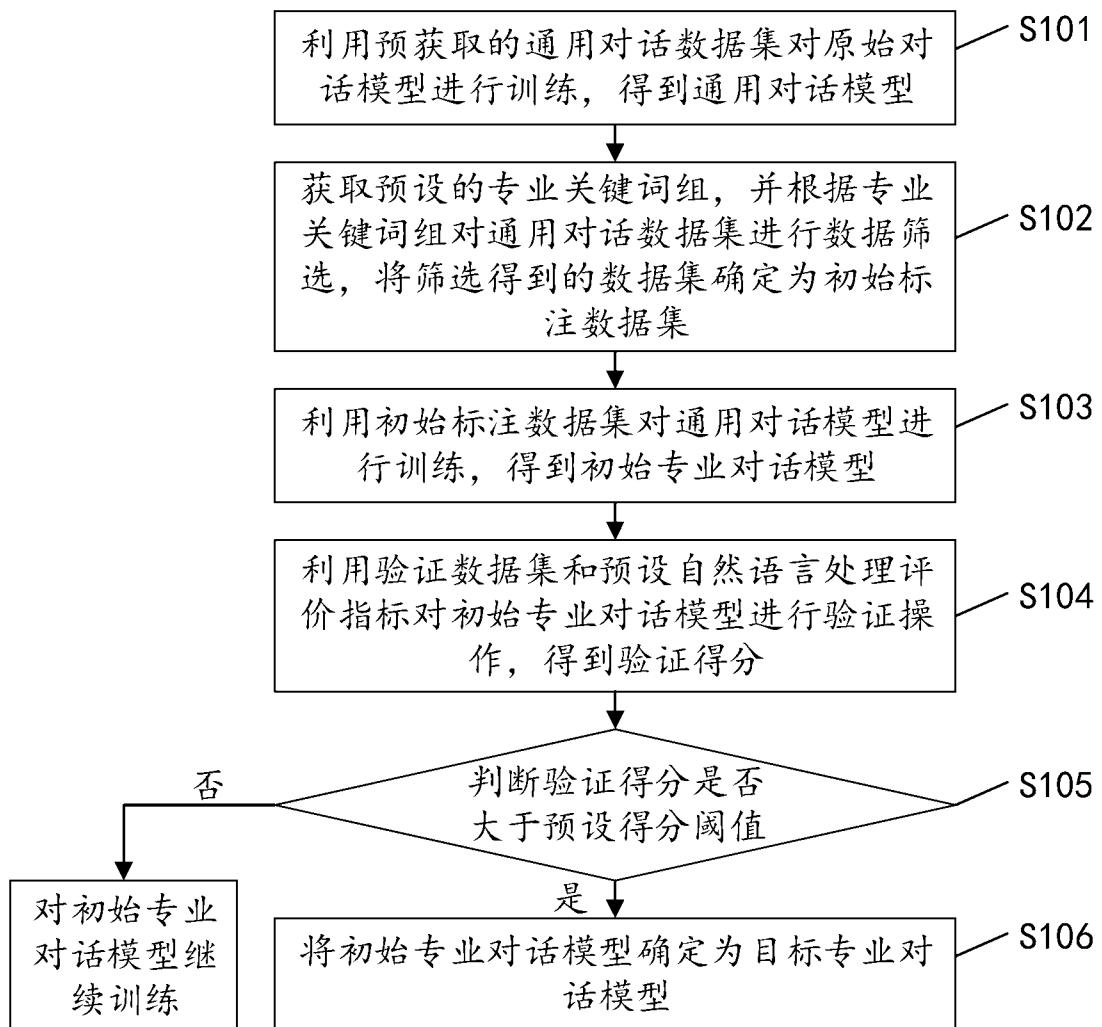


图 1

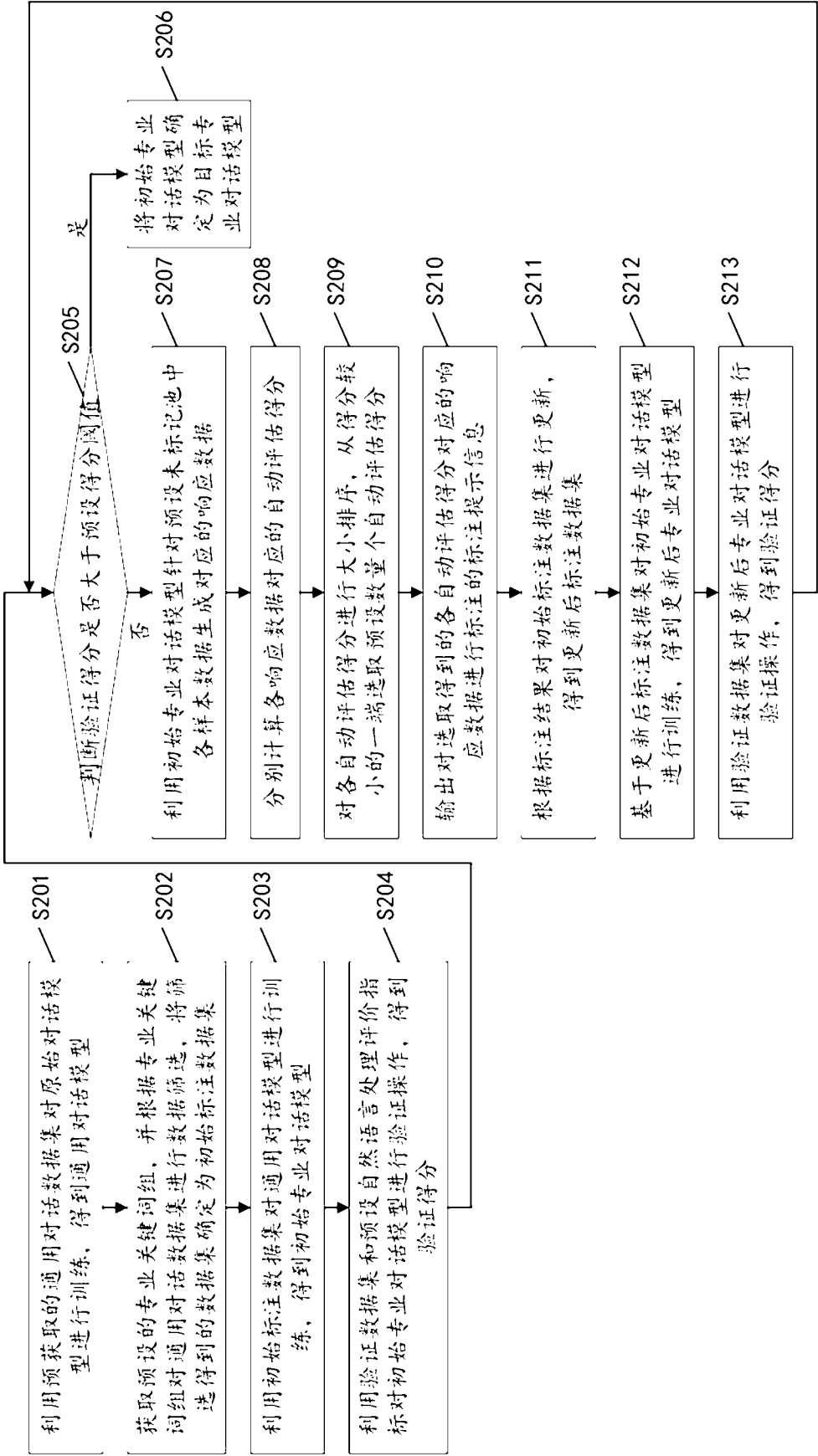


图 2

3/4

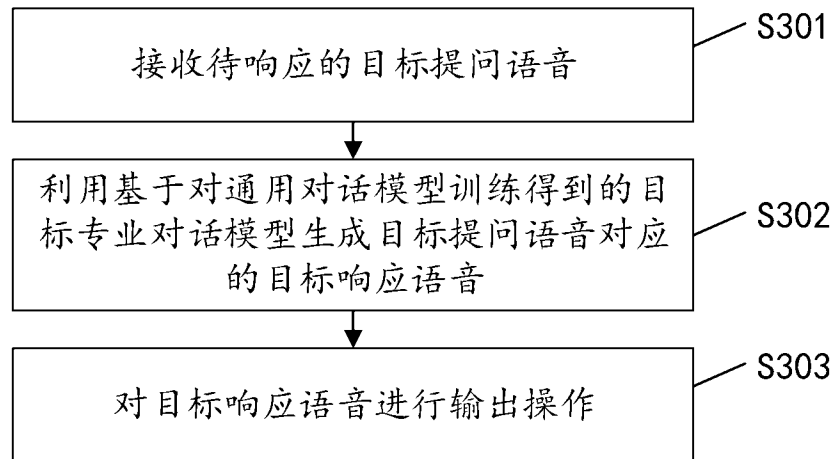


图 3

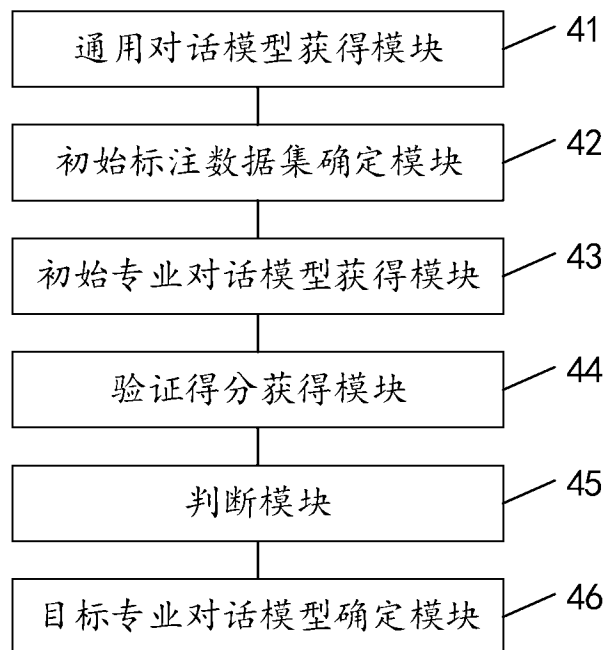


图 4

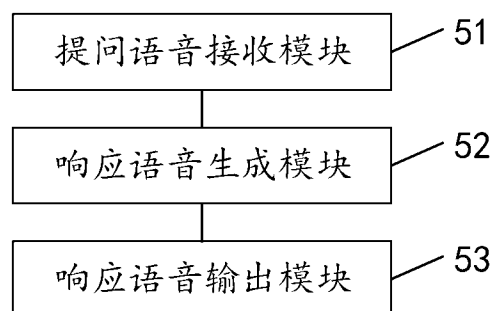


图 5

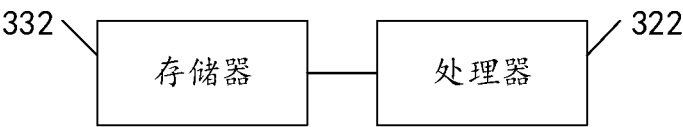


图 6

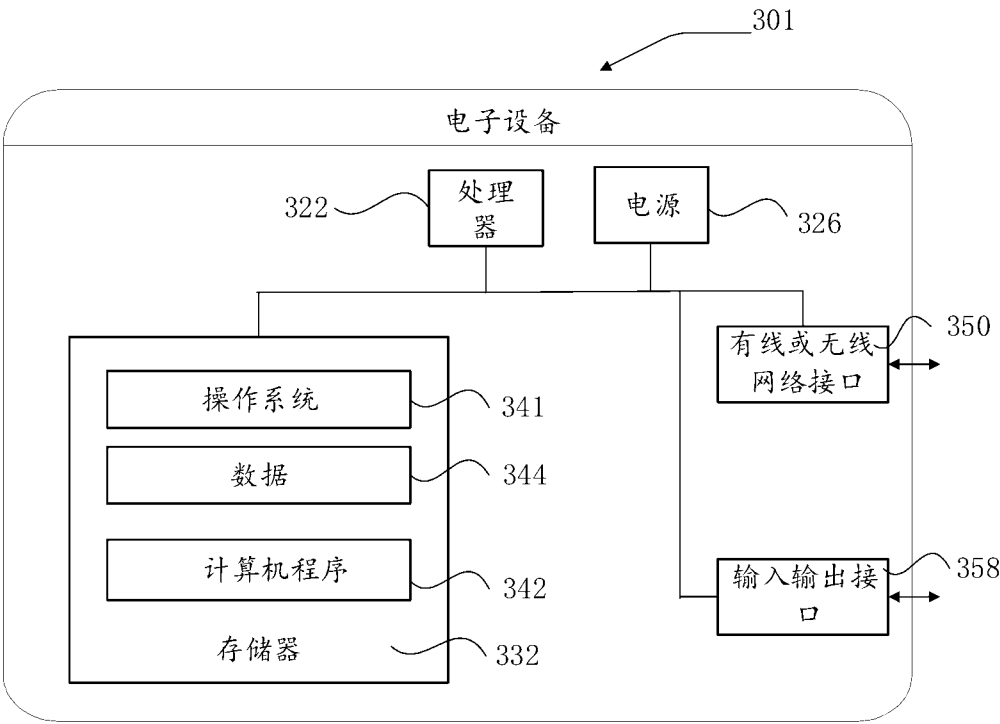


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/086071

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/332(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 模型, 对话, 训练, 目标, 数据集, 数据筛选, 标注, 关键词, 专业, 初始, 指标, 评价, 目标, 验证, 得分, set, evaluating, aiming, score, training, model, dialog, target, original, collecting, conversation, loss

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115495568 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 20 December 2022 (2022-12-20) claims 1-19	1-20
Y	CN 108897797 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 27 November 2018 (2018-11-27) description, paragraphs 69-92, and figure 3	1-20
Y	CN 110188331 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 30 August 2019 (2019-08-30) description, paragraphs 60-67	1-20
A	CN 114968788 A (ZHEJIANG UNIVERSITY) 30 August 2022 (2022-08-30) entire document	1-20
A	US 2021303555 A1 (ADOBE INC.) 30 September 2021 (2021-09-30) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “D” document cited by the applicant in the international application
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

19 May 2023

Date of mailing of the international search report

25 May 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/086071

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	115495568	A	20 December 2022	None	
CN	108897797	A	27 November 2018	None	
CN	110188331	A	30 August 2019	None	
CN	114968788	A	30 August 2022	None	
US	2021303555	A1	30 September 2021	None	

A. 主题的分类 G06F 16/332(2019.01) i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC: G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 模型, 对话, 训练, 目标, 数据集, 数据筛选, 标注, 关键词, 专业, 初始, 指标, 评价, 目标, 验证, 得分, set, evaluating, aiming, score, training, model, dialog, target, original, collecting, conversation, loss		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 115495568 A (苏州浪潮智能科技有限公司) 2022年12月20日 (2022 - 12 - 20) 权利要求1-19	1-20
Y	CN 108897797 A (腾讯科技(深圳)有限公司) 2018年11月27日 (2018 - 11 - 27) 说明书第69-92段, 附图3	1-20
Y	CN 110188331 A (腾讯科技(深圳)有限公司) 2019年8月30日 (2019 - 08 - 30) 说明书第60-67段	1-20
A	CN 114968788 A (浙江大学) 2022年8月30日 (2022 - 08 - 30) 全文	1-20
A	US 2021303555 A1 (ADOBE INC.) 2021年9月30日 (2021 - 09 - 30) 全文	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2023年5月19日		国际检索报告邮寄日期 2023年5月25日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		授权官员 王晶 电话号码 (+86) 010-53961318

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/086071

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	115495568	A	2022年12月20日	无	
CN	108897797	A	2018年11月27日	无	
CN	110188331	A	2019年8月30日	无	
CN	114968788	A	2022年8月30日	无	
US	2021303555	A1	2021年9月30日	无	