



(12) 发明专利申请

(10) 申请公布号 CN 113627198 A

(43) 申请公布日 2021. 11. 09

(21) 申请号 202110965077.2

(22) 申请日 2021.08.20

(71) 申请人 作业帮教育科技(北京)有限公司
地址 100085 北京市海淀区上地信息产业
基地开拓路1号4层4002

(72) 发明人 邵睿 宋旻 蒋宏飞 吕少科

(74) 专利代理机构 北京清诚知识产权代理有限
公司 11691

代理人 乔东峰

(51) Int. Cl.

G06F 40/35 (2020.01)

G06K 9/62 (2006.01)

G06F 16/9035 (2019.01)

G06Q 10/06 (2012.01)

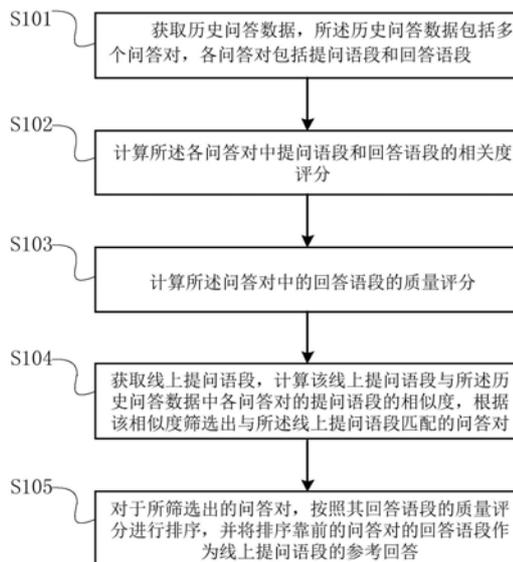
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种基于历史问答数据的参考回答生成方法、系统、装置和计算机设备

(57) 摘要

本发明属于教育领域,提供一种获取历史问答数据的参考回答生成方法、系统、装置和计算机设备,用于在线上学生与老师问答。本发明包括:获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段,计算所述各问答对中提问语段和回答语段的相关度评分,计算所述问答对中的回答语段的质量评分,获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。本发明的方法能够在用户提出线上问题时基于历史问答数据给予快速准确的答复、提高用户使用体验感。



1. 一种基于历史问答数据的参考回答生成方法,其特征在于,所述方法包括如下步骤:
获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段;

计算所述各问答对中提问语段和回答语段的相关度评分;

计算所述问答对中的回答语段的质量评分;

获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对;

对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

2. 根据权利要求1所述的基于历史问答数据的参考回答生成方法,其特征在于,所述计算所述各问答对中提问语段和回答语段的相关度评分,包括:

设定相关度评分阈值,当回答语段相关度评分在评分阈值以上,保存所述回答语段;

当回答语段相关度评分小于评分阈值,舍弃所述回答语段。

3. 根据权利要求2所述的基于历史数据的回复内容评分与推荐方法,其特征在于,设定相关度评分阈值,对所述提问语段与回答语段相关度匹配并标注评分,所述评分在0-1区间,其中,

当提问语段与回答语段完全相关时,评分为1,完全不相关则评分为0,并根据其评分对相同提问语段所对应的多个回答语段从高至低排序。

4. 根据权利要求1所述的基于历史问答数据的参考回答生成方法,其特征在于,获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,包括:

对提问语段进行分词,提取所述提问语段关键词;

建立问答对检索库,根据问答对中提问语段关键词设定倒排索引;

匹配所述提问语段与线上提问语段关键词相似度;

提取所述线上提问语段与提问语段相似度高的问答对。

5. 根据权利要求4所述的基于历史问答数据的参考回答生成方法,其特征在于,根据问答对中提问语段关键词设定倒排索引,包括:

对问答对提问语段中的单词或记录进行分词,将所述单词或记录作为关键词,将所述问答对作为记录;

当线上提问语段与提问语段相似度匹配,提取被匹配问答对中的回答语段。

6. 根据权利要求5所述的基于历史问答数据的参考回答生成方法,其特征在于,所述将线上提问语段与提问语段相似度匹配,提取被匹配问答对,根据所述问答对的回答语段质量评分排序,当线上提问语段关键词与匹配到的提问语段关键词近似度较高时,展示与所述提问语段所对应的回答语段。

7. 根据权利要求6所述的基于历史问答数据的参考回答生成方法,其特征在于,所述展示与提问语段所对应的回答语段包括:

展示线上提问语段与其相似度匹配的提问语段;

展示与所述提问语段匹配的问答对;

展示相同问答对所对应的多个回答语段并根据质量评分从高至低排序,作为线上提问

语段的参考回答选择反馈给用户。

8. 一种线上问答系统,其特征在於,所述问答系统包括:

客户端和相对端;

至少一个服务器,用于线上问答时客户端和相对端之间的交互,并且所述客户端或所述至少一个服务器使用权利要求1所述的基于历史问答数据的参考回答生成方法接收问答数据;所述至少一个服务器将生成的问答数据传输至所述相对端,并在所述客户端和/或相对端中显示。

9. 一种基于历史问答数据的参考回答装置,其特征在於,所述参考回答装置包括:

获取模块,用于获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段;

第一计算模块,用于计算所述各问答对中提问语段和回答语段的相关度评分;

第二计算模块,用于计算所述问答对中的回答语段的质量评分;

匹配模块,用于获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对;

生成模块,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

10. 一种计算机程序产品,包括计算机程序/指令,其特征在於,所述计算机程序/指令当被处理器执行时,实现权利要求1-7中任一项所述的基于历史问答数据的参考回答生成方法。

一种基于历史问答数据的参考回答生成方法、系统、装置和计算机设备

技术领域

[0001] 本发明属于教育领域,特别适用于线上教育领域,更具体的是涉及一种基于历史问答数据的参考回答生成方法、装置和计算机设备。

背景技术

[0002] 近年来,随着线上课程日益频繁的参与率,学生与老师更多课前或课后互动交流也成为线上课程必要的沟通环节,学生的数量远远要超过在线老师的数量,这也就导致了部分学生无法得到及时、准确的问题答复,降低了课程体验感间接影响学习积极性,若因此增加老师用于与学生交流,则会增加不必要的人力浪费,并且需要前期对老师进行相关培训,而在老师回答问题的准确性上也无法把关监控。

[0003] 所以,通过解决用户在提出重复问题时,如何可以在不增加人力的基础上提高用户与老师沟通时快速响应的同时保证答复的准确性以提高用户使用体验感是目前亟需解决的问题。

[0004] 因此,有必要提供基于历史数据的回复内容评分与推荐方法,以解决上述问题。

发明内容

[0005] (一)要解决的技术问题

[0006] 本发明旨在解决线上学生与老师课程互动时,问答不能得到及时的回复、答复针对性低、重复问题无法优选答复以及用户体验感差等技术问题。

[0007] (二)技术方案

[0008] 为解决上述技术问题,本发明的一方面提出的基于历史问答数据的参考回答生成方法,用于线上教育领域,所述方法包括如下步骤:获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段,计算所述各问答对中提问语段和回答语段的相关度评分,计算所述问答对中的回答语段的质量评分,获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

[0009] 可选地,所述计算所述各问答对中提问语段和回答语段的相关度评分,包括:设定相关度评分阈值,当回答语段相关度评分在评分阈值以上,保存所述回答语段,当回答语段相关度评分小于评分阈值,舍弃所述回答语段。

[0010] 根据本发明的优选实施方式,设定相关度评分阈值,对所述提问语段与回答语段相关度匹配并标注评分,所述评分在0-1区间,其中,当提问语段与回答语段完全相关时,评分为1,完全不相关则评分为0,并根据其评分对相同提问语段所对应的多个回答语段从高至低排序。

[0011] 根据本发明的优选实施方式,获取线上提问语段,计算该线上提问语段与所述历

史问答数据中各问答对的提问语段的相似度,包括:对提问语段进行分词,提取所述提问语段关键词,建立问答对检索库,根据问答对中提问语段关键词设定倒排索引,匹配所述提问语段与线上提问语段关键词相似度,提取所述线上提问语段与提问语段相似度高的问答对。

[0012] 根据本发明的优选实施方式,根据问答对中提问语段关键词设定倒排索引,包括:对问答对提问语段中的单词或记录进行分词,将所述单词或记录作为关键词,将所述问答对作为记录,当线上提问语段与提问语段相似度匹配,提取被匹配问答对中的回答语段。

[0013] 根据本发明的优选实施方式,所述将线上提问语段与提问语段相似度匹配,提取被匹配问答对,根据所述问答对的回答语段评分排序,当线上提问语段关键词与匹配到的提问语段关键词近似度较高时,展示与所述提问语段所对应的回答语段。

[0014] 根据本发明的可选实施方式,所述展示与提问语段所对应的回答语段包括:展示线上提问语段与其相似度匹配的提问语段,展示与所述提问语段匹配的问答对,展示相同问答对所对应的多个回答语段并根据质量评分从高至低排序,作为线上提问语段的参考回答选择反馈给用户。

[0015] 本发明第二方面提出一种线上问答系统,所述问答系统包括:客户端和相对端,至少一个服务器,用于线上问答时客户端和相对端之间的交互,并且所述客户端或所述至少一个服务器使用权利要求1所述的基于历史问答数据的参考回答生成方法接收问答数据,所述至少一个服务器将生成的问答数据传输至所述相对端,并在所述客户端和/或相对端中显示。

[0016] 本发明第三方面提出一种基于历史问答数据的参考回答装置,所述参考回答装置包括:获取模块,用于获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段,第一计算模块,用于计算所述各问答对中提问语段和回答语段的相关度评分,第二计算模块,用于计算所述问答对中的回答语段的质量评分,匹配模块,用于获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对,生成模块,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

[0017] 本发明第四方面提出一种计算机程序产品,包括计算机程序/指令,其特征在于,所述计算机程序/指令当被处理器执行时,实现本发明所述的基于历史问答数据的参考回答生成方法。

[0018] (三)有益效果

[0019] 与现有技术相比,本发明通过周期性的获取用户在本平台的沟通记录作为历史数据,当用户再次提出问题时,平台对历史数据的内容相关度筛选,可以快速匹配近似度高且答复质量高的优选答复反馈给客户。在这个过程中,有效的避免了相似问题,反复回答造成的人力浪费,且通过筛选优质答复确保答复的准确性,提高用户体验。

附图说明

[0020] 图1是本发明的基于历史问答数据的参考回答生成的流程图;

[0021] 图2是本发明的基于历史问答数据的参考回答生成的一示例流程图。

- [0022] 图3是本发明的线上问答系统的一示例的示意图。
- [0023] 图4是本发明的基于历史问答数据的参考回答生成装置的示意图；
- [0024] 图5是本发明的基于历史问答数据的参考回答生成装置的一示例的示意图。
- [0025] 图6是本发明的一个实施例的计算机设备的结构示意图。
- [0026] 图7是本发明的一个实施例的计算机程序产品的示意图。

具体实施方式

[0027] 在对于具体实施例的介绍过程中,对结构、性能、效果或者其他特征的详细描述是为了使本领域的技术人员对实施例能够充分理解。但是,并不排除本领域技术人员可以在特定情况下,以不含有上述结构、性能、效果或者其他特征的技术方案来实施本发明。

[0028] 附图中的流程图仅是一种示例性的流程演示,不代表本发明的方案中必须包括流程图中的所有内容、操作和步骤,也不代表必须按照图中所显示的顺序执行。例如,流程图中的有的操作/步骤可以分解,有的操作/步骤可以合并或部分合并,等等,在不脱离本发明的发明主旨的情况下,流程图中显示的执行顺序可以根据实际情况改变。

[0029] 附图中的框图一般表示的是功能实体,并不一定必然与物理上独立的实体相对应。即,可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理单元装置和/或微控制器装置中实现这些功能实体。

[0030] 各附图中相同的附图标记表示相同或类似的元件、组件或部分,因而下文中可能省略了对相同或类似的元件、组件或部分的重复描述。还应理解,虽然本文中可能使用第一、第二、第三等表示编号的定语来描述各种器件、元件、组件或部分,但是这些器件、元件、组件或部分不应受这些定语的限制。也就是说,这些定语仅是用来将一者与另一者区分。例如,第一器件亦可称为第二器件,但不偏离本发明实质的技术方案。此外,术语“和/或”、“及/或”是指包括所列出项目中的任一个或多个的所有组合。

[0031] 为了进一步提升线上针对问答的能够快速的反馈答复,本发明提出一种基于历史问答数据的参考回答生成方法,用于线上教育,该方法通过对获取的历史问答数据中间问答对根据相关度进行评分,针对问答对中的回答语段进行质量评分排序,在后续用户反馈线上提问语段,筛选出与所述线上提问语段匹配的问答对,对于所筛选出的问答对,按照其回答语段的质量评分将排序靠前的问答对的回答语段作为线上提问语段的参考回答。由此,本发明的方法能够根据用户提出的线上问题给予快速准确的答复并能够提高用户使用体验感,还能够进一步优化基于历史问答数据的参考回答生成方法。

[0032] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明作进一步的详细说明。

[0033] 需要说明的是,为便于理解,本文中以上述线上教育场景对本发明实施例进行了说明,但本领域技术人员可以理解的是,本发明的应用并不限于此。本发明还可以用于其他用户提问服务器优选答复再反馈的场景。

[0034] 图1是本发明的实施例1的基于历史问答数据的参考回答生成的流程图。

[0035] 如图1所示,本发明提供了一种基于历史问答数据的参考回答生成方法,所述方法包括:

[0036] 步骤S101,获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段;

[0037] 步骤S102,计算所述各问答对中提问语段和回答语段的相关度评分;

[0038] 步骤S103,计算所述问答对中的回答语段的质量评分;

[0039] 步骤S104,获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对;

[0040] 步骤S105,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

[0041] 在本示例中,在线上教育的应用场景下,包括客户端和相对端,运行在这二者或其中之一基于历史问答数据的参考回答系统,该回答方法用于线上教育通过客户端与相对端之间例如提出疑问和反馈答复的用户交互,在本发明中,客户端是指任何能够与其他电子设备进行数据传输的电子设备,包括台式电脑、笔记本电脑、平板电脑、手机等设备。所述的客户端和相对端通常提供有人机交互接口,包括例如显示器或显示屏、鼠标、键盘等,显示器或显示屏通常具有可视化交互界面。诸如手机等移动智能客户端,人机交互接口通常是触摸屏和按键。无论如何,用户可以通过人机接口与客户端和相对端进行交互。

[0042] 首先,步骤S101,获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段。

[0043] 在本示例中,客户端用户为学生,对应端用户为老师,在线上问答过程中,周期性的获取学生与老师日常沟通记录作为历史问答数据,每个历史问答数据包括多个问答对,每个问答对是由单独的提问语段和回答语段组成。

[0044] 作为一种优选的实施方式,通过创建不同的时间间隔将历史问答数据切分为单独的问答对,所述时间间隔可以根据不同要求具体设定,举例说明例如工作人员可以根据学生与老师的平均问答时间作为切分当前历史问答数据时间间隔,比方说当学生提出问题后老师的平均回答时间为15秒内,那么这个时间间隔可以设定在15秒,来确保能够切分到完整的问答对。为了更准确的切分历史问答数据中的问答对,也可以设定切分历史问答数据的方式为当学生提问,老师完成对该提问的答复后立即切分,使每个问答对都只包括单独的提问语段和回答语段。

[0045] 需要说明的是,上述仅作为示例进行说明,不能理解成对本发明的限制。在其他示例中,还可以通过工作人员对历史问答数据进行人工切分等等。

[0046] 在步骤S102中,计算所述各问答对中提问语段和回答语段的相关度评分。

[0047] 切分后的历史问答数据包括提问语段以及回答语段,在这个过程中,通过计算提问语段和回答语段的相关度评分来筛选是否保留当前问答对。

[0048] 具体地,通过利用内容相关度模型对每个问答对进行评分,并设定相关度评分阈值,当回答语段相关度评分在评分阈值以上,保存所述回答语段,当回答语段相关度评分小于评分阈值,舍弃所述回答语段。所述设定相关度评分阈值,是对所述提问语段与回答语段相关度匹配并标注评分,所述评分区间在0-1,当提问语段与回答语段完全相关时,评分为1,完全不相关则评分为0,并根据其评分对相同提问语段所对应的多个回答语段从高至低排序。在实施例中,所述相关度评分可以根据老师进行人工评分也可以通过设定具体评分标准,例如根据提问语段和回答语段中出现的关键词次数进行评定,根据提问语段中出现

的关键词结合回答语段中出现的关键词出现频率作为评分标准,关键词出现频率越高则评分越高。所述相关度评分阈值可以设定为0.6,保存所述问答对中提问语段与回答语段相关度评分在0.6以上的问答对,相关度评分小于0.6的问答对直接舍弃。所述相关度评分阈值可根据不同场景等原因做调整,这里不具体限定。

[0049] 其中,所述内容相关度训练方法包括:周期性的获取学生与老师日常沟通记录作为历史问答数据,对所述历史问答数据切分生成的问答对中的提问语段和问答语段进行相关度评分标注产生0-1分值的评分,并分别对学生的提问语段和老师的回答语段进行向量化,放入bert模型中进行fine-tuning。

[0050] 在步骤S103中,计算所述问答对中的回答语段的质量评分。

[0051] 在本示例中,经过对各问答对中提问语段和回答语段的相关度评分筛选,舍弃了低于所设定相关度评分阈值的问答对,对保留下的问答对中的回答语段进行质量评分,值得一提的是,所述质量评分规则可根据应用场景的不同、内容长度、内容复杂度、段落数、语气词、用户后期满意度打分等来设定,例如某评分规则为该回答语段运用多个“请”来指导用户操作作为加分项,内容长度达到指定字数作为加分项,回复段落在2段以上作为加分项或在后续满意度评分上被给予5颗星评分作为加分项,则所述回答语段根据具体加分项来计算质量评分,并根据其质量评分对相同提问语段所对应的多个回答语段的质量分从高至低排序并保存。

[0052] 进一步地,对保留下来的回答语段所对应的提问语段进行分词,所述分词为记录提问语段中的关键词,例如学生的提问语段为:儒家的思想主张是?那么该提问语段中的关键词及为儒家和/或思想。为了更好地存储不同关键词,针对历史问答数据的多个问答对中通过相关度评分筛选的提问语段和回答语段中的关键词在一服务器中建立检索库并建立倒排索引。

[0053] 接下来,在步骤S104中,获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对。

[0054] 具体地,当学生通过客户端对老师提出线上提问,在获取线上提问语段后,服务器会计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,其中,服务器会再次对线上提问语段进行分词,提取所述线上提问语段的关键词,服务器通过匹配所述提问语段与线上提问语段关键词相似度,从检索库中提取所述线上提问语段与提问语段相似度高的问答对。

[0055] 优选地,所述根据问答对中提问语段关键词设定倒排索引,是对历史问答数据问答对提问语段中的单词或记录进行分词,将所述单词或记录作为关键词,将所述问答对作为记录,当线上提问语段与提问语段相似度匹配,提取被匹配问答对中的回答语段。

[0056] 与现有技术相比,所述计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度是通过改良后的bm25算法进行完成,bm25算法是目前比较实用的一种用来评价搜索词和文档之间相关性的算法,但由于学生提出的提问语段比较短,所以本发明通过修改bm25算法中的参数b,主要是通过减小b的值,作为改良算法来削弱长度对最终分值的惩罚程度。用于更好的实现本发明所提到的对线上提问语段与提问语段的相似度进行匹配。

[0057] 步骤S105,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。

[0058] 具体地,当服务器通过近似度匹配到线上提问语段与提问语段,服务器会根据所述问答对回答语段质量评分进行排序,并展示与所述提问语段所对应的回答语段。

[0059] 在实施例中,所述展示与提问语段所对应的回答语段包括:

[0060] 在老师的对应端展示线上提问语段与其相似度匹配的提问语段,展示与所述提问语段匹配的问答对,展示相同问答对所对应的多个回答语段并根据质量评分从高至低排序。也就是说,老师的对应端上会显示包括:学生提出的问题、与学生问题匹配的历史问题以及与历史问题对应的历史回复,其中,所述回复通过质量评分从高至低排序。

[0061] 作为一种优选的实施方式,老师可以通过服务器反馈给对应端的回答语段进行浏览并选择,老师可以根据线上提问语段来优选回答语段反馈给学生,并在学生的客户端显示老师反馈回答语段。

[0062] 参照图3,图3是本发明的线上直播系统的一示例的示意图。

[0063] 根据本发明的第二方面,本发明还提供了一种线上直播系统,所述线上问答系统包括:客户端和相对端,至少一个服务器,用于线上问答时客户端和相对端之间的交互。

[0064] 所述客户端或所述至少一个服务器使用权利要求1所述的基于历史问答数据的参考回答生成方法接收问答数据;所述至少一个服务器将生成的问答数据传输至所述相对端,并在所述客户端和/或相对端中显示。

[0065] 优选地,线上直播系统可以包括客户端设备301、302、303,对应端设备311、312、313,网络304和服务端305。网络304用以在客户端设备301、302、303和服务端305与对应端设备311、312、313之间提供通信链路的介质。网络304可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0066] 用户可以使用客户端设备301、302、303或对应端设备311、312、313通过网络304与服务端305交互,以接收或发送消息等。客户端设备301、302、303和对应端设备311、312、313上可以安装有各种通讯客户端应用,例如线上课程类应用、网页浏览器应用、金融网络平台应用、即时通信工具、邮箱客户端、社交平台软件等。

[0067] 客户端设备301、302、303和对应端设备311、312、313可以是具有显示屏并且支持线上课程的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。客户端设备301、302、303可例如获取用户的历史问答数据,所述历史问答数据包括多个问答对;客户端设备301、302、303可例如在所述用户在客户端获取问答对时,根据服务器要求创建不同的时间间隔将切分为单独的问答对,客户端设备301、302、303可例如获取所述用户在客户端递交的线上提问语段并获取服务端305反馈的回答语段,以及显示服务端305反馈的回答语段。

[0068] 对应端设备311、312、313可例如展示线上提问语段与其相似度匹配的提问语段;对应端设备311、312、313可例如展示与所述提问语段匹配的问答对;对应端设备311、312、313可例如展示相同问答对所对应的多个回答语段并根据质量评分从高至低排序,以及选择线上提问语段的参考回答选择反馈给用户。服务端305可以是提供各种服务的服务端,例如计算所述客户端设备301、302、303与对应端设备311、312、313反馈的各问答对中提问语段和回答语段的相关度评分;计算所述问答对中的回答语段的质量评分;从客户端设备

301、302、303获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对。服务端305可以对接收到的用户服务申请的流程进程监控,还可以根据对应端设备按照其回答语段的质量评分进行排序筛选出的回答语段反馈给客户端设备。

[0069] 例如用户利用客户端设备301(也可以是客户端设备302或303)向服务端305发送了线上提问语段,在提问语段发送之后,对应端设备311(也可以是对应端设备312或313)可以进行针对提问语段的筛选并回答,服务端305可例如反馈对应端设备311(也可以是对应端设备312或313)选择的回答语段到客户端设备301(也可以是客户端设备302或303)。

[0070] 服务端305可以是一个实体的服务端,还可例如为多个服务端组成,需要说明的是,本公开实施例所提供的参考回答生成方法可以由服务端305、客户端设备301、302、303和/或对应端设备311、312、313中执行,相应地,参考回答生成装置可以设置于服务端305、客户端设备301、302、303和/或对应端设备311、312、313中。

[0071] 与现有技术相比,本发明的系统通过使用实施例的基于历史问答数据的参考回答生成方法能够在用户提出线上问题时基于历史问答数据给予快速准确的答复、提高用户使用体验感。。

[0072] 下面描述本发明的装置实施例,该装置可以用于执行本发明的方法实施例。对于本发明装置实施例中描述的细节,应视为对于上述方法实施例的补充;对于在本发明装置实施例中未披露的细节,可以参照上述方法实施例来实现。

[0073] 根据本发明的第三方面,本发明还提供了一种基于历史问答数据的参考回答装置,图4是本发明的基于历史问答数据的参考回答生成装置的示意图,根据图4所示,所述装置包括:

[0074] 获取模块,用于获取历史问答数据,所述历史问答数据包括多个问答对,各问答对包括提问语段和回答语段。通过周期性的获取学生与老师日常沟通记录作为历史问答数据,每个历史问答数据包括多个问答对,每个问答对是由单独的提问语段和回答语段组成。

[0075] 第一计算模块,用于计算所述各问答对中提问语段和回答语段的相关度评分,通过设定相关度评分阈值,切分后的历史问答数据包括提问语段以及回答语段,在这个过程中,通过计算提问语段和回答语段的相关度评分来筛选是否保留当前问答对。当回答语段相关度评分在评分阈值以上,保存所述回答语段,当回答语段相关度评分小于评分阈值,舍弃所述回答语段,并设定相关度评分阈值,对所述提问语段与回答语段相关度匹配并标注评分,所述评分在0-1区间,其中,当提问语段与回答语段完全相关时,评分为1,完全不相关则评分为0,并根据其评分对相同提问语段所对应的多个回答语段从高至低排序。

[0076] 第二计算模块,用于计算所述问答对中的回答语段的质量评分,值得一提的是,所述质量评分规则可根据应用场景的不同、内容长度、内容复杂度、段落数、语气词、用户后期满意度打分等来设定。

[0077] 匹配模块,用于获取线上提问语段,计算该线上提问语段与所述历史问答数据中各问答对的提问语段的相似度,根据该相似度筛选出与所述线上提问语段匹配的问答对,对提问语段进行分词,提取所述提问语段关键词,建立问答对检索库,根据问答对中提问语段关键词设定倒排索引,匹配所述提问语段与线上提问语段关键词相似度,提取所述线上提问语段与提问语段相似度高的问答对。

[0078] 生成模块,对于所筛选出的问答对,按照其回答语段的质量评分进行排序,并将排序靠前的问答对的回答语段作为线上提问语段的参考回答。所述将线上提问语段与提问语段相似度匹配,提取被匹配问答对,根据所述问答对的回答语段质量评分排序,当线上提问语段关键词与匹配到的提问语段关键词近似度较高时,展示与所述提问语段所对应的回答语段。

[0079] 在实施例中,图5是本发明的基于历史问答数据的参考回答生成装置的一示例的示意图,如图5所示,还包括显示模块,该显示模块用于将所述书写轨迹在客户端中显示线上提问语段和被反馈的回答语段,在相对端中显示线上提问语段与其相似度匹配的提问语段,显示与所述提问语段匹配的问答对,显示相同问答对所对应的多个回答语段并根据质量评分从高至低排序。

[0080] 本领域技术人员可以理解,上述装置实施例中的各模块可以按照描述分布于装置中,也可以进行相应变化,分布于不同于上述实施例的一个或多个装置中。上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块。

[0081] 下面描述本发明的计算机设备实施例,该计算机设备可以视为对于上述本发明的方法和装置实施例的具体实体实施方式。对于本发明计算机设备实施例中描述的细节,应视为对于上述方法或装置实施例的补充;对于在本发明计算机设备实施例中未披露的细节,可以参照上述方法或装置实施例来实现。

[0082] 图6是本发明的一个实施例的计算机设备的结构示意图,该计算机设备包括处理器和存储器,所述存储器用于存储计算机可执行程序,当所述计算机程序被所述处理器执行时,所述处理器执行实施例中任一项所述的方法,包括但不限于图1的方法。

[0083] 如图6所示,计算机设备以通用计算设备的形式表现。其中处理器可以是一个,也可以是多个并且协同工作。本发明也不排除进行分布式处理,即处理器可以分散在不同的实体设备中。本发明的计算机设备并不限于单一实体,也可以是多个实体设备的总和。

[0084] 所述存储器存储有计算机可执行程序,通常是机器可读的代码。所述计算机可读程序可以被所述处理器执行,以使得计算机设备能够执行本发明的方法,或者方法中的至少部分步骤。

[0085] 所述存储器包括易失性存储器,例如随机存取存储单元(RAM)和/或高速缓存存储单元,还可以是非易失性存储器,如只读存储单元(ROM)。

[0086] 可选的,该实施例中,计算机设备还包括有I/O接口,其用于计算机设备与外部的设备进行数据交换。I/O接口可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0087] 应当理解,图6显示的计算机设备仅仅是本发明的一个示例,本发明的计算机设备中还可以包括上述示例中未示出的元件或组件。例如,有些计算机设备中还包括有显示屏等显示单元,有些计算机设备还包括人机交互元件,例如按钮、键盘等。只要该计算机设备能够执行存储器中的计算机可读程序以实现本发明方法或方法的至少部分步骤,均可认为是本发明所涵盖的计算机设备。

[0088] 图7是本发明的一个实施例的计算机程序产品的示意图。如图7所示,计算机程序产品中存储有计算机可执行程序,所述计算机可执行程序被执行时,实现本发明上述空中

书写方法。所述计算机程序产品可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。该计算机程序产品可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。该计算机程序产品上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF等等，或者上述的任意合适的组合。

[0089] 可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的程序代码，所述程序设计语言包括面向对象的设计语言—诸如Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网(LAN)或广域网(WAN)，连接到用户计算设备，或者，可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0090] 通过以上对实施方式的描述，本领域的技术人员易于理解，本发明可以由能够执行特定计算机程序的硬件来实现，例如本发明的系统，以及系统中包含的电子处理单元、服务器、客户端、手机、控制单元、处理器等。本发明也可以由执行本发明的方法的计算机软件来实现，例如由微处理器、电子控制单元，客户端、服务器端等执行的控制软件来实现。但需要说明的是，执行本发明的方法的计算机软件并不限于由一个或特定个的硬件实体中执行，其也可以是由不特定具体硬件的以分布式的方式来实现。对于计算机软件，软件产品可以存储在一个计算机可读的存储介质(可以是CD-ROM, U盘, 移动硬盘等)中，也可以分布式存储于网络上，只要其能使得计算机设备执行根据本发明的方法。

[0091] 以上所述的具体实施例，对本发明的目的、技术方案和有益效果进行了进一步详细说明，应理解的是，本发明不与任何特定计算机、虚拟装置或者计算机设备固有相关，各种通用装置也可以实现本发明。以上所述仅为本发明的具体实施例而已，并不用于限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

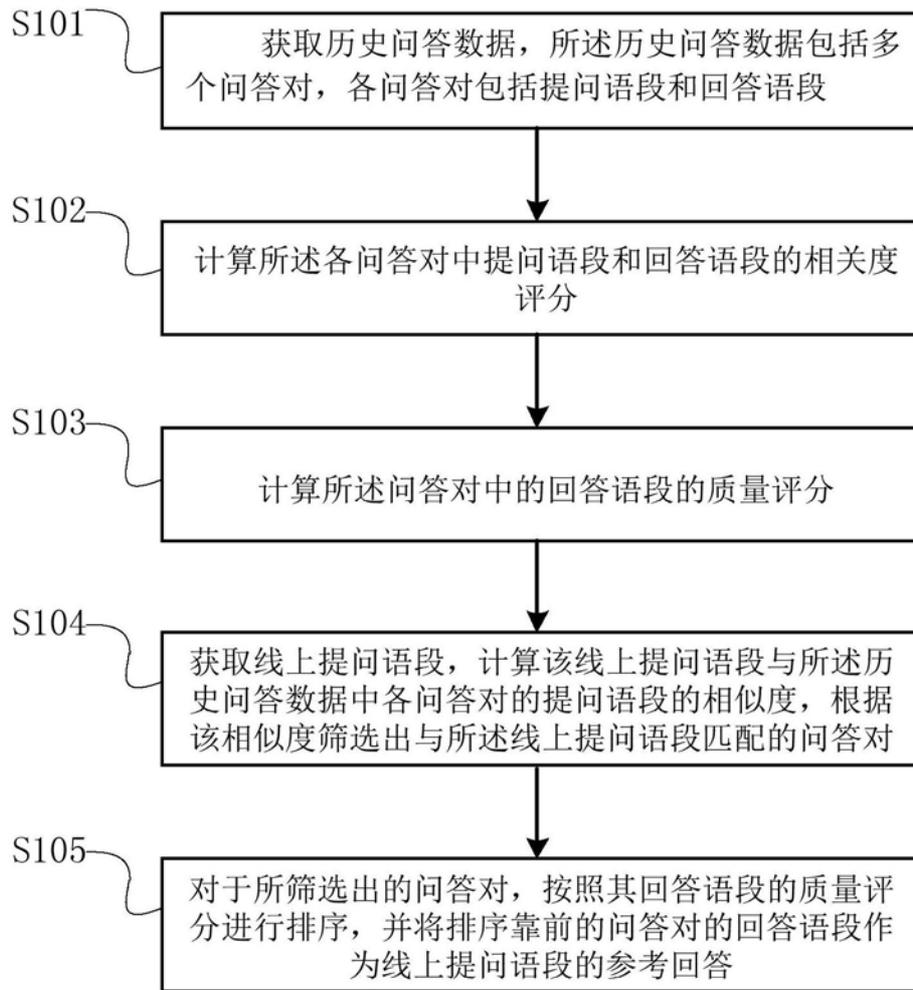


图1

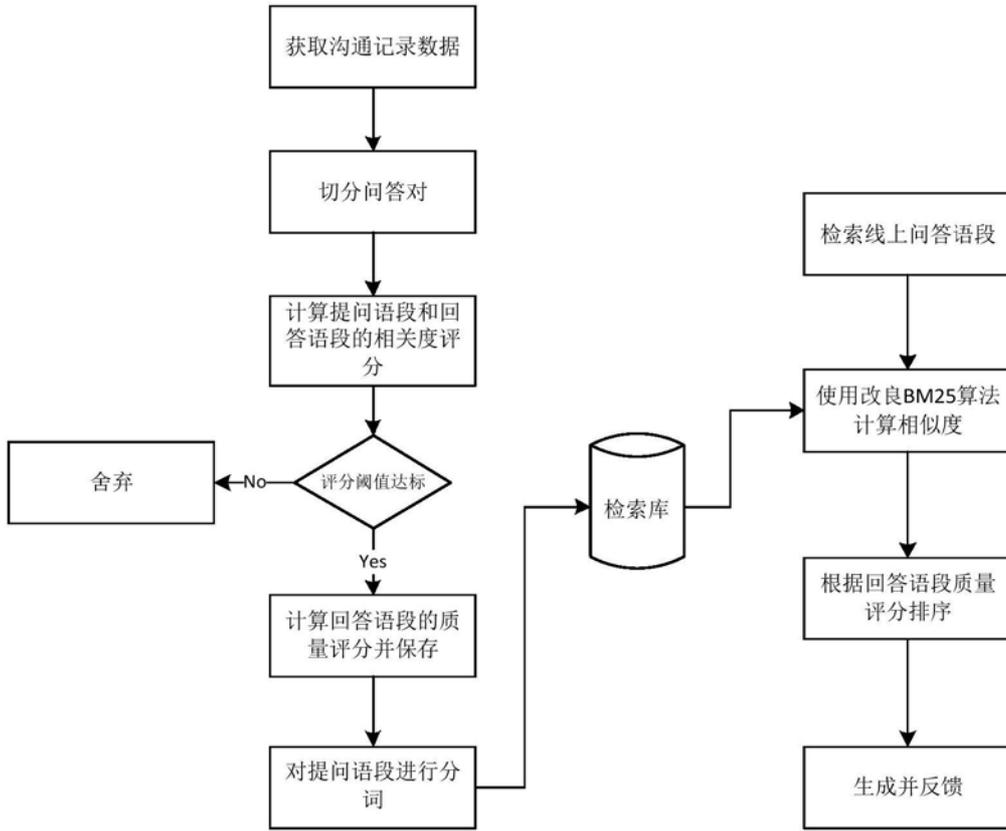


图2

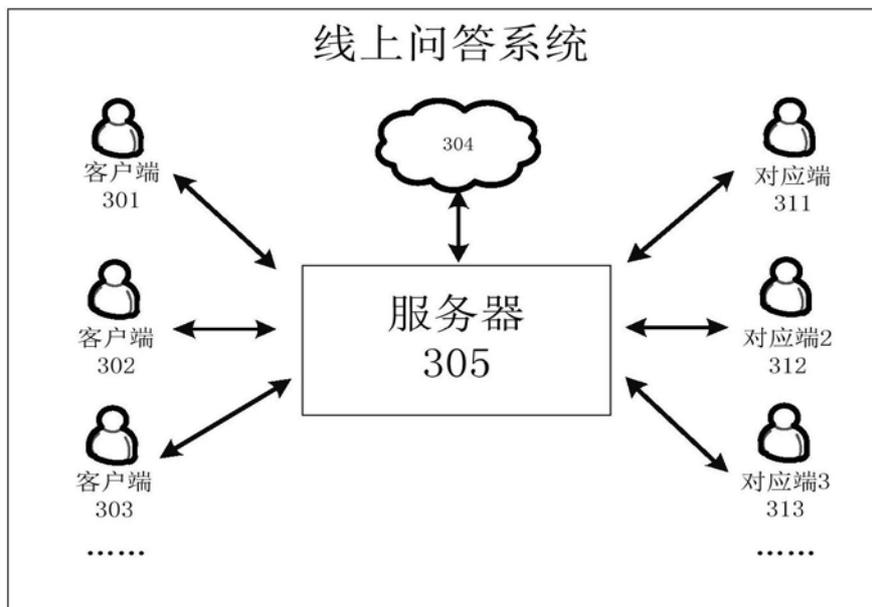


图3

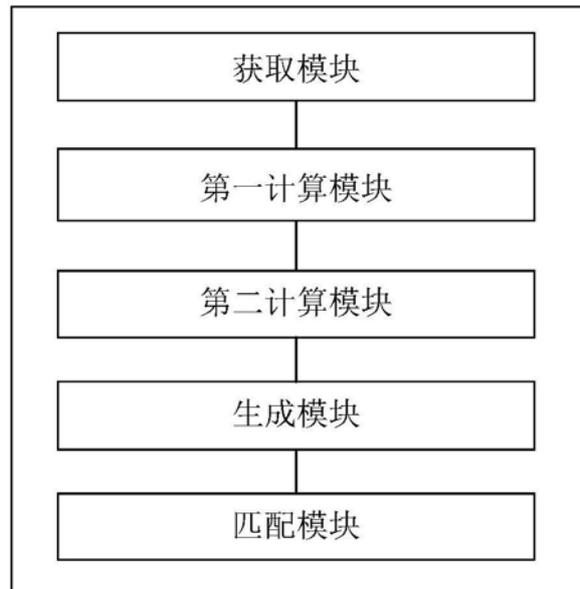


图4

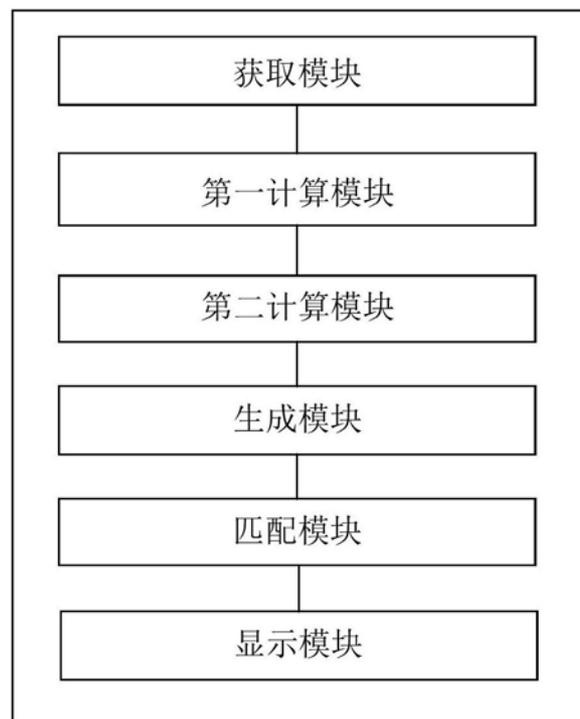


图5

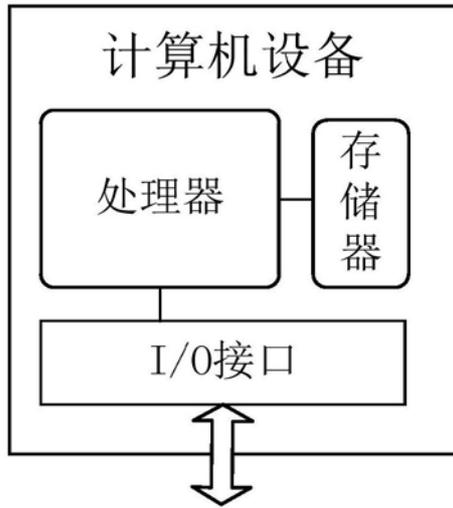


图6



图7