



(12) 发明专利

(10) 授权公告号 CN 108446286 B

(45) 授权公告日 2023. 04. 25

(21) 申请号 201710084397.0

G06F 40/284 (2020.01)

(22) 申请日 2017.02.16

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 105701253 A, 2016.06.22

申请公布号 CN 108446286 A

审查员 吴单单

(43) 申请公布日 2018.08.24

(73) 专利权人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四  
层847号邮箱

(72) 发明人 张增明 姜飞俊 陈永健 汪洋

胡于响 沈慧 王成龙 杨洋

(74) 专利代理机构 北京三友知识产权代理有限

公司 11127

专利代理师 李辉

(51) Int. Cl.

G06F 16/332 (2019.01)

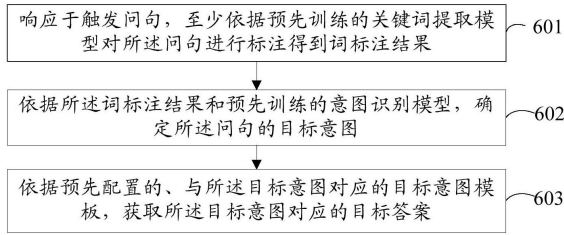
权利要求书5页 说明书41页 附图6页

(54) 发明名称

一种自然语言问句答案的生成方法、装置及  
服务器

(57) 摘要

本申请提供了一种自然语言问句答案的生成方法、装置及服务器,其中,问句答案的生成方法包括:响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型;依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图;依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。采用本申请实施例,基于意图模板来训练得到能够更为准确的识别用户意图的意图识别模型,以及采用关键词提取模型来提取用户的问句的关键词来为意图模板中的意图参数赋值,从而可以执行意图模板中的逻辑来自动生成问句的答案。



1. 一种自然语言问句答案的生成方法,其特征在于,该方法包括:

响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型;

依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图;

依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案,

其中,所述响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,包括:

对所述问句进行分词,得到所述问句的分词集合;

基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;以及,基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果;所述领域词典用于保存各领域词及其对应的词类型之间的对应关系;

依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

2. 根据权利要求1所述的方法,其特征在于,所述依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果,包括:

针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都标注为关键词,如果是,则将该分词确定为关键词,并依据所述第一标记结果或第二标记结果确定所述关键词的词类型;

如果第一标记结果标注为关键词而第二标记结果标注为非关键词,则将该分词确定为非关键词。

3. 根据权利要求1所述的方法,其特征在于,所述领域词典通过以下方式生成:

获取领域内各实体的各属性;

将所述各属性对应的属性值作为关键词,并将实体标签和属性名作为关键词的词类型,以及,将所述关键词的关联词汇也确定为关键词,所述关联词汇包括:同义词、近义词、口语表达词和别名。

4. 根据权利要求1所述的方法,其特征在于,所述依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图,包括:

将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度;

将所述多个候选意图中可信度最高的候选意图确定为当前意图;

依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整;所述意图模板包括:确定一意图所需要的全部意图参数;

如果完整,则将所述当前意图确定为目标意图;

如果不完整,则按照可信度从高到低的顺序选择下一个候选意图作为当前意图,并执行所述依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整的步骤,直至所述多个候选意图都选择完毕。

5. 根据权利要求4所述的方法,其特征在于,所述依据所述当前意图对应的当前意图模

板,判断所述词标注结果中各关键词是否完整,包括:

获取所述当前意图模板中配置的全部意图参数;

采用所述词标注结果中的各关键词的词类型,为所述全部意图参数进行赋值,并判断所述全部意图参数是否能赋值成功。

6.根据权利要求1所述的方法,其特征在于,所述依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图,包括:

至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及其对应的可信度;

将所述多个候选意图中可信度最高的候选意图确定为当前意图;

判断所述当前意图的可信度是否大于预设可信度阈值,如果大于,则将所述当前意图确定为所述目标意图;

如果不大于,则选取预设个数的前若干个候选意图作为待选择意图,并将所述待选择意图发送至终端以供用户选择,以及,将所述终端返回的、用户选择的意图确定为目标意图;

依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键词是否完整,如果不完整,则将不完整的关键词对应的意图参数确定为缺少的意图参数;依据所述缺少的意图参数和所述目标意图模板,生成所述目标意图相关的参数问题;将所述参数问题返回给客户端以便显示给用户;所述意图模板包括:所述意图参数和参数问题之间的对应关系;

根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数。

7.根据权利要求6所述的方法,其特征在于,还包括:

将用户输入的对话内容保存至会话数据库中,所述对话内容包括:用户选择的意图和/或用户输入的参数答案。

8.根据权利要求7所述的方法,其特征在于,还包括:

对所述会话数据库中的对话内容进行更新,

或者,在到达预设的时间周期的时候,将用户输入的对话内容进行删除。

9.根据权利要求7所述的方法,其特征在于,所述至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及其对应的可信度,包括:

从会话数据库中获取所述对话内容;

将所述词标注结果和所述对话内容输入预先训练的意图识别模型的,得到多个候选意图及其对应的可信度。

10.根据权利要求6所述的方法,其特征在于,所述依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键词是否完整,包括:

获取所述目标意图模板中配置的全部意图参数;

采用所述词标注结果中的各关键词的词类型,为所述目标意图模板中的全部意图参数进行赋值,并判断所述全部意图参数是否能赋值成功。

11.根据权利要求1所述的方法,其特征在于,所述意图模板通过以下方式生成:

依据领域内各本体所表示出的各实体之间的关系,确定各领域内所有问句的意图;所述意图用于表述问句的询问目的,所述意图包括:询问实体的属性,询问实体之间的所有关

系,通过实体的一个或多个属性询问该实体的其他属性或所有属性,和/或,通过实体的一个或多个属性询问该实体的关系;

为确定的各意图配置对应的意图模板,所述意图模板包括:模板标识,模板描述,意图参数及意图参数对应的参数问题,意图对应的例句组,以及,各例句组的组合规则;所述模板标识用于唯一表示一个意图模板,所述模板描述用于描述所述意图模板的目的;所述意图参数为确定一个意图所对应的参数,所述参数问题为:对应的意图参数缺失时需要向用户发送的问题,所述例句组中的例句包括实体的变量名,用于表示所述意图模板能解决的问题,所述各例句组的组合规则用于表示:不同例句组中的例句的组合方式。

12. 根据权利要求11所述的方法,其特征在于,所述意图识别模型和关键词提取模型通过以下方式训练:

依据所述意图模板中配置的例句组和例句组合规则,生成训练语料;

依据所述训练语料采用机器学习算法,训练得到意图分类模型和关键词提取模型。

13. 根据权利要求12所述的方法,其特征在于,所述依据所述意图模板中配置的例句组和例句组合规则,生成训练语料,包括:

获取实体的变量值列表,所述变量值列表包括:实体的变量名、实体的变量值及两者的对应关系;

依据意图模板中配置的例句组和例句组的组合规则,生成包括实体的变量名的变量例句;

将所述变量值列表中各实体的变量值,分别对应替换所述变量例句中的变量名,得到真实例句;

对所述真实例句进行标记,得到训练语料;所述训练语料包括:所述真实例句对应的模板标识和所述真实例句中包括的各关键词及其类型。

14. 根据权利要求1所述的方法,其特征在于,依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案,包括:

依据目标意图模板中配置的意图参数,为所述意图参数赋值,得到查询参数;

依据所述查询参数获取所述目标答案的源数据;

对所述源数据按照预设处理方式进行处理得到处理后的源数据;

依据所述处理后的源数据生成所述目标答案。

15. 一种自然语言问句答案的生成装置,其特征在于,该生成装置集成于服务器上,该生成装置包括:

关键词标注单元,用于响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型;

意图确定单元,用于依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图;

答案获取单元,用于依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案,

其中,所述关键词标注单元包括:

分词子单元,用于对所述问句进行分词,得到所述问句的分词集合;

第一标记子单元,用于基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;

第二标记子单元,用于基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果;所述领域词典用于保存各领域词及其对应的词类型之间的对应关系;

第一确定子单元,用于依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

16.根据权利要求15所述的装置,其特征在于,所述第一确定子单元包括:

第一判断子单元,用于针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都标注为关键词;

第二确定子单元,用于在所述第一判断子单元的结果为是的情况下,将该分词确定为关键词,并依据所述第一标记结果或第二标记结果确定所述关键词的词类型;

第三确定子单元,用于在所述第一判断子单元的结果为第一标记结果标注为关键词而第二标记结果标注为非关键词的情况下,将该分词确定为非关键词。

17.根据权利要求15所述的装置,其特征在于,所述意图确定单元包括:

第一模型处理子单元,用于将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度;

第四确定子单元,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图;

第二判断子单元,用于依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整;所述意图模板包括:确定一意图所需要的全部意图参数;

第五确定子单元,用于在所述第二判断子单元的结果为是的情况下,将所述当前意图确定为目标意图;

选择子单元,用于在所述第二判断子单元的结果为否的情况下,按照可信度从高到低的顺序选择下一个候选意图作为当前意图;

触发子单元,用于在选择子单元选择下一个候选意图之后,触发所述第二判断子单元,直至所述多个候选意图都选择完毕。

18.根据权利要求15所述的装置,其特征在于,所述意图确定单元包括:

第二模型处理子单元,用于至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及其对应的可信度;

第六确定子单元,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图;

第三判断子单元,用于判断所述当前意图的可信度是否大于预设可信度阈值;

第七确定子单元,用于在所述第三判断子单元的结果为大于的情况下,将所述当前意图确定为所述目标意图;

选取子单元,用于在所述第三判断子单元的结果为否的情况下,选取预设个数的前若干个候选意图作为待选择意图;

发送子单元,用于将所述待选择意图发送至终端以供用户选择;

第八确定子单元,用于将所述终端返回的、用户选择的意图确定为目标意图;

第四判断子单元,用于依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键词是否完整;

第九确定子单元,用于在所述第四判断子单元的结果为否的情况下,将不完整的关键词对应的意图参数确定为缺少的意图参数;

问题生成子单元,用于依据所述缺少的意图参数和所述目标意图模板,生成所述目标意图相关的参数问题;

问题返回子单元,用于将所述参数问题返回给客户端以便显示给用户;所述意图模板包括:所述意图参数和参数问题之间的对应关系;

参数获取子单元,用于根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数。

19. 根据权利要求15所述的装置,其特征在于,所述答案获取单元包括:

赋值子单元,用于依据目标意图模板中配置的意图参数,为所述意图参数赋值,得到查询参数;

源数据获取子单元,用于依据所述查询参数获取所述目标答案的源数据;

处理子单元,用于对所述源数据按照预设处理方式进行处理得到处理后的源数据;

生成子单元,用于依据所述处理后的源数据生成所述目标答案。

20. 一种服务器,其特征在于,所述服务器包括:权利要求15~19任意一项所述的装置。

## 一种自然语言问句答案的生成方法、装置及服务器

### 技术领域

[0001] 本申请涉及智能问答技术领域,特别涉及一种基于意图识别模型和关键词提取模型的、自然语言问句答案的生成方法,装置及服务器。

### 背景技术

[0002] 在互联网越来越发展的今天,很多网站或者网店等都采用智能问答系统来回答各种问题的咨询。智能问答系统是以一问一答的形式,精确的定位网站或网店等用户所需要的提问知识,通过与用户进行交互,为用户提供个性化的信息服务的系统。

[0003] 在现有技术中,通过采用预先设定规则的方式来实现智能问答。例如,枚举所有可能的问句以及问句的回答,然后对每种问句抽象出一个规则。具体在生成当前问句的回答时,依次检查当前问句和每个规则是否匹配,如果匹配,则执行该规则指定的回答方式获取答案。

### 发明内容

[0004] 但是发明人在研究过程中发现,现有技术中,对规则的配置比较麻烦,将规则配错的可能性很大,而且各个规则之间可能互相矛盾,添加新规则的时候还需要不和其他的已有规则冲突。因此,基于规则的智能问答的维护成本较大。具体执行时,因为问句要和每个规则的对应关系都尝试匹配,如果规则的数量较多,就会使得匹配的效率也较低。并且,匹配不到规则的时候,就无法回答用户提出的问题,因此灵活性较低,即如果希望能够回答某个问句,就必须增加一条和该问句对应的规则才行。

[0005] 基于此,本申请提供了一种基于预先训练的意图识别模型和关键词提取模型的问句答案的自动生成方法,用以采用预先建立本体的方式枚举领域内的所有意图,即一个问句的询问目的,再依据意图来配置相对应的意图模板,即配置好什么样的问句能够被该意图模板回答以及如何回答,进一步的,基于意图模板来训练得到能够更为准确的识别用户意图的意图识别模型,以及采用关键词提取模型来提取用户的问句的关键词来为意图模板中的意图参数赋值,从而可以执行意图模板中的逻辑来自动生成问句的答案。

[0006] 本申请还提供了一种问句答案的自动生成装置和服务器例如问答引擎等,用以保证上述方法在实际中的实现及应用。

[0007] 为了解决上述问题,本申请公开了一种问句答案的生成方法,该方法包括:

[0008] 响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型;

[0009] 依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图;

[0010] 依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0011] 其中,所述响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进

行标注得到词标注结果,包括:

[0012] 对所述问句进行分词,得到所述问句的分词集合;

[0013] 基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;以及,基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果;所述领域词典用于保存各领域词及其对应的词类型之间的对应关系;

[0014] 依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

[0015] 其中,所述依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果,包括:

[0016] 针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都标注为关键词,如果是,则将该分词确定为关键词,并依据所述第一标记结果或第二标记结果确定所述关键词的词类型;

[0017] 如果第一标记结果标注为关键词而第二标记结果标注为非关键词,则将该分词确定为非关键词。

[0018] 其中,所述领域词典通过以下方式生成:

[0019] 获取领域内各实体的各属性;

[0020] 将所述各属性对应的属性值作为关键词,并将实体标签和属性名作为关键词的词类型,以及,将所述关键词的关联词汇也确定为关键词,所述关联词汇包括:同义词、近义词、口语表达词和别名。

[0021] 其中,所述依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图,包括:

[0022] 将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度;

[0023] 将所述多个候选意图中可信度最高的候选意图确定为当前意图;

[0024] 依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整;所述意图模板包括:确定一意图所需要的全部意图参数;

[0025] 如果完整,则将所述当前意图确定为目标意图;

[0026] 如果不完整,则按照可信度从高到低的顺序选择下一个候选意图作为当前意图,并执行所述依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整的步骤,直至所述多个候选意图都选择完毕。

[0027] 其中,所述依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整,包括:

[0028] 获取所述当前意图模板中配置的全部意图参数;

[0029] 采用所述词标注结果中的各关键词的词类型,为所述全部意图参数进行赋值,并判断所述全部意图参数是否能赋值成功。

[0030] 其中,所述依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图,包括:

[0031] 至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选

意图及其对应的可信度；

[0032] 将所述多个候选意图中可信度最高的候选意图确定为当前意图；

[0033] 判断所述当前意图的可信度是否大于预设可信度阈值，如果大于，则将所述当前意图确定为所述目标意图；

[0034] 如果不大于，则选取预设个数的前若干个候选意图作为待选择意图，并将所述待选择意图发送至终端以供用户选择，以及，将所述终端返回的、用户选择的意图确定为目标意图；

[0035] 依据所述目标意图对应的目标意图模板，判断所述词标注结果中各关键词是否完整，如果不完整，则将不完整的关键词对应的意图参数确定为缺少的意图参数；依据所述缺少的意图参数和所述目标意图模板，生成所述目标意图相关的参数问题；将所述参数问题返回给客户端以便显示给用户；所述意图模板包括：所述意图参数和参数问题之间的对应关系；

[0036] 根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数。

[0037] 其中，所述方法还包括：

[0038] 将用户输入的对话内容保存至会话数据库中，所述对话内容包括：用户选择的意图和/或用户输入的参数答案。

[0039] 其中，所述方法还包括：

[0040] 对所述会话数据库中的对话内容进行更新，

[0041] 或者，在到达预设的时间周期的时候，将用户输入的对话内容进行删除。

[0042] 其中，所述至少将所述词标注结果作为所述预先训练的意图识别模型的输入，得到多个候选意图及其对应的可信度，包括：

[0043] 从会话数据库中获取所述对话内容；

[0044] 将所述词标注结果和所述对话内容输入预先训练的意图识别模型的，得到多个候选意图及其对应的可信度。

[0045] 其中，所述依据所述目标意图对应的目标意图模板，判断所述词标注结果中各关键词是否完整，包括：

[0046] 获取所述目标意图模板中配置的全部意图参数；

[0047] 采用所述词标注结果中的各关键词的词类型，为所述目标意图模板中的全部意图参数进行赋值，并判断所述全部意图参数是否能赋值成功。

[0048] 其中，所述意图模板通过以下方式生成：

[0049] 依据领域内各本体所表示出的各实体之间的关系，确定各领域内所有问句的意图；所述意图用于表述问句的询问目的，所述意图包括：询问实体的属性，询问实体之间的所有关系，通过实体的一个或多个属性询问该实体的其他属性或所有属性，和/或，通过实体的一个或多个属性询问该实体的关系；

[0050] 为确定的各意图配置对应的意图模板，所述意图模板包括：模板标识，模板描述，意图参数及意图参数对应的参数问题，意图对应的例句组，以及，各例句组的组合规则；所述模板标识用于唯一表示一个意图模板，所述模板描述用于描述所述意图模板的目的；所述意图参数为确定一个意图所对应的参数，所述参数问题为：对应的意图参数缺失时需要向用户发送的问题，所述例句组中的例句包括实体的变量名，用于表示所述意图模板能解

决的问题,所述各例句组的组合规则用于表示:不同例句组中的例句的组合方式。

[0051] 其中,所述意图识别模型和关键词提取模型通过以下方式训练:

[0052] 依据所述意图模板中配置的例句组和例句组合规则,生成训练语料;

[0053] 依据所述训练语料采用机器学习算法,训练得到意图分类模型和关键词提取模型。

[0054] 其中,所述依据所述意图模板中配置的例句组和例句组合规则,生成训练语料,包括:

[0055] 获取实体的变量值列表,所述变量值列表包括:实体的变量名、实体的变量值及两者的对应关系;

[0056] 依据意图模板中配置的例句组和例句组的组合规则,生成包括实体的变量名的变量例句;

[0057] 将所述变量值列表中各实体的变量值,分别对应替换所述变量例句中的变量名,得到真实例句;

[0058] 对所述真实例句进行标记,得到训练语料;所述训练语料包括:所述真实例句对应的模板标识和所述真实例句中包括的各关键词及其类型。

[0059] 其中,依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案,包括:

[0060] 依据目标意图模板中配置的意图参数,为所述意图参数赋值,得到查询参数;

[0061] 依据所述查询参数获取所述目标答案的源数据;

[0062] 对所述源数据按照预设处理方式进行处理得到处理后的源数据;

[0063] 依据所述处理后的源数据生成所述目标答案。

[0064] 本申请还公开了一种自然语言问句答案的生成装置,该生成装置集成于服务器上,该生成装置包括:

[0065] 关键词标注单元,用于响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型;

[0066] 意图确定单元,用于依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图;

[0067] 答案获取单元,用于依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0068] 其中,所述关键词标注单元包括:

[0069] 分词子单元,用于对所述问句进行分词,得到所述问句的分词集合;

[0070] 第一标记子单元,用于基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;

[0071] 第二标记子单元,用于基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果;所述领域词典用于保存各领域词及其对应的词类型之间的对应关系;

[0072] 第一确定子单元,用于依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

[0073] 其中,所述第一确定子单元包括:

[0074] 第一判断子单元,用于针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都标注为关键词;

[0075] 第二确定子单元,用于在所述第一判断子单元的结果为是的情况下,将该分词确定为关键词,并依据所述第一标记结果或第二标记结果确定所述关键词的词类型;

[0076] 第三确定子单元,用于在所述第一判断子单元的结果为第一标记结果标注为关键词而第二标记结果标注为非关键词的情况下,将该分词确定为非关键词。

[0077] 其中,所述意图确定单元包括:

[0078] 第一模型处理子单元,用于将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度;

[0079] 第四确定子单元,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图;

[0080] 第二判断子单元,用于依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整;所述意图模板包括:确定一意图所需要的全部意图参数;

[0081] 第五确定子单元,用于在所述第二判断子单元的结果为是的情况下,将所述当前意图确定为目标意图;

[0082] 选择子单元,用于在所述第二判断子单元的结果为否的情况下,按照可信度从高到低的顺序选择下一个候选意图作为当前意图;

[0083] 触发子单元,用于在选择子单元选择下一个候选意图之后,触发所述的第二判断子单元,直至所述多个候选意图都选择完毕。

[0084] 其中,所述意图确定单元包括:

[0085] 第二模型处理子单元,用于至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及其对应的可信度;

[0086] 第六确定子单元,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图;

[0087] 第三判断子单元,用于判断所述当前意图的可信度是否大于预设可信度阈值;

[0088] 第七确定子单元,用于在所述第三判断子单元的结果为大于的情况下,将所述当前意图确定为所述目标意图;

[0089] 选取子单元,用于在所述第三判断子单元的结果为否的情况下,选取预设个数的前若干个候选意图作为待选择意图;

[0090] 发送子单元,用于将所述待选择意图发送至终端以供用户选择;

[0091] 第八确定子单元,用于将所述终端返回的、用户选择的意图确定为目标意图;

[0092] 第四判断子单元,用于依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键词是否完整;

[0093] 第九确定子单元,用于在所述第四判断子单元的结果为否的情况下,将不完整的关键词对应的意图参数确定为缺少的意图参数;

[0094] 问题生成子单元,用于依据所述缺少的意图参数和所述目标意图模板,生成所述目标意图相关的参数问题;

[0095] 问题返回子单元,用于将所述参数问题返回给客户端以便显示给用户;所述意图

模板包括：所述意图参数和参数问题之间的对应关系；

[0096] 参数获取子单元，用于根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数。

[0097] 其中，所述答案获取单元包括：

[0098] 赋值子单元，用于依据目标意图模板中配置的意图参数，为所述意图参数赋值，得到查询参数；

[0099] 源数据获取子单元，用于依据所述查询参数获取所述目标答案的源数据；

[0100] 处理子单元，用于对所述源数据按照预设处理方式进行处理得到处理后的源数据；

[0101] 生成子单元，用于依据所述处理后的源数据生成所述目标答案。

[0102] 本申请还公开了一种服务器，其特征在于，所述服务器包括：前述的生成装置。

[0103] 与现有技术相比，本申请实施例包括以下优点：

[0104] 在本申请实施例中，因为训练的意图识别模型和关键词提取模型能够更为准确的识别意图，并且提取更为准确的关键词为执行意图逻辑做意图参数的准备，所以不仅提高了答案生成的效率和准确度，还能与用户进行对话，在用户没有表达意图的时候进一步的明确用户的真实意图，或者，在明确用户的意图的情况下进一步的获取完整的意图参数等，从而使得本申请实施例的灵活性也比现有技术中基于规则的方式更好，且能与用户进行对话，用户的体验也较高。

[0105] 当然，实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

## 附图说明

[0106] 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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性的前提下，还可以根据这些附图获得其他的附图。

[0107] 图1是本申请中配置意图模板的流程图；

[0108] 图2是本申请中本体的一个结构示意图；

[0109] 图3是本申请的本体的另一个结构示意图；

[0110] 图4是本申请中训练意图识别模型和关键词提取模型的流程图；

[0111] 图5是本申请在实际应用中的场景架构图；

[0112] 图6是本申请的问句答案的生成方法实施例1的流程图；

[0113] 图7是本申请的问句答案的生成方法实施例2的流程图；

[0114] 图8是本申请的问句答案的生成方法实施例3的流程图；

[0115] 图9是本申请的问句答案的生成装置实施例1的结构框图；

[0116] 图10是本申请的问句答案的生成装置实施例2中意图确定单元902的结构框图；

[0117] 图11是本申请的问句答案的生成装置实施例3中意图确定单元902的结构框图。

## 具体实施方式

[0118] 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。

[0119] 为了方便本领域技术人员对本申请的理解,下面首先对本申请中出现的术语进行介绍。

[0120] 本体(ontology)

[0121] 本体是知识图谱里面的概念,可以理解为对现实世界的抽象。人类经过对现实世界的分析之后,可以抽象出来很多的实体,并且,每个实体又有自己的属性,实体和实体之间存在着某种关系。那么本体就可以理解为这些实体和实体之间关系的表达和描述。

[0122] 例如,在诗歌领域,诗人和诗为两个实体,这两个实体分别有各自的属性,其中,诗人的属性可以包括:名字、朝代等;诗的属性可以包括:内容、标题、朝代、类型等。并且,诗人和诗这两个实体之间存在着写与被写的关系。

[0123] 因为本体描述了实体和实体之间的关系,所以从本体就能了解到这个领域的所有模式,从而可以枚举出关于这个领域的所有问题的意图。例如,在诗歌领域,从诗歌领域存在的本体中,可以枚举出下面这些意图的问题:问诗人的作品,问诗人的朝代,问诗人的性别,问诗的内容,问诗的朝代,问诗的类型,或者,问诗的作者。

[0124] 问答引擎

[0125] 问答引擎在用户输入问题后,自动给出答案,并且问答引擎允许用户以自然语言的形式输入问题。问答引擎根据用户输入的问题得到答案的过程为:首先,通过解析用户给出的关键词语,理解用户表达的意图;然后通过查询知识库或者通过某种算法得到答案;最后以自然语言的形式给出答案。问答引擎可以包括单轮的问答引擎和多轮的问答引擎。其中,单轮的问答引擎是一问一答,即用户问一个问题,在此问题中明确给出了意图和关键词(slot),问答引擎输出一个答案。多轮的问答引擎可以根据用户的当前问题以及过往的问答记录(上下文,Context)综合识别出用户的意图并解析slot,当遇到意图不清晰或者意图所需的各种参数没有完全满足条件时,问答引擎会掌控对话,对用户进行提问,直到用户清晰的表达了意图以及完整的slot,问答引擎才会给出答案。多轮的问答引擎通过对话的形式问问题,给用户的体验就像是在和一个真实的人交流。

[0126] 意图模板

[0127] 意图是在用户当前提出的问题中,用户想表达的目的。例如,用户问“周杰伦唱过哪些歌曲”的意图是找某个歌手唱的歌;例如,用户问“床前明月光的下一句是什么”的意图是寻找一句诗的下一句。在本申请实施例中,使用意图模板来表示一个意图。一个意图模板可以包括意图的名称、满足意图执行的各种参数(slot),以及意图的执行逻辑,在本申请实施例中,意图的执行逻辑以groovy动态脚本的形式配置在意图模板中。

[0128] 具体回答用户某个带有意图的问题时,首先要提取用户问题中的slot,给意图模板中的参数赋值,然后执行groovy脚本配置的执行逻辑,最终得到groovy脚本返回的答案。

[0129] 关键词slot

[0130] slot是意图执行的参数,例如,“周杰伦唱过哪些歌曲”这个问题,“周杰伦”就是一个slot。“床前明月光的下一句是什么”这个问题,“床前明月光”就是一个slot。如果把意图看成一个句子,那么slot就是问题中的关键词;如果把意图看成一个函数,那么slot就是参

数。

[0131] slot提取就是提取用户问题中的slot。如果要正确的回答一个问题,slot提取是十分重要的。如果slot提取错误,直接导致意图的参数错误,进而造成意图执行的结果错误。因此,要正确的回答一个问题,slot提取是十分重要的。

[0132] groovy动态脚本

[0133] groovy是一种脚本语言,完全兼容java语法,并且可在java中动态执行。基于这种特性,groovy语言可以作为意图模板中的执行逻辑的承载者,即执行逻辑是可以被配置的,所以执行逻辑可以在不需要重新编译或者重启系统的情况下被改写,这极大的方便了系统的维护,也增强了系统的可扩展性。

[0134] 知识库

[0135] 知识库即知识图谱,是对实体、属性以及实体之间的关系进行建模和存储的系统。知识图谱是以图的方式存储知识,这点区别于关系型数据库。

[0136] 推理(reasoning)

[0137] 本申请实施例中包括两种推理,一种是对知识的推理,另一种是对用户意图的推理。本申请实施例主要侧重于对用户意图的推理,因为对知识的推理是知识图谱系统的范畴。例如,知识库中记载了A是B的父亲,B是C的父亲,当查询C的爷爷是谁,就需要使用“爷爷=父亲的父亲”推理规则来推理。而对用户意图的推理是根据过往的上下文和当前的用户问题,以及自动生成的对用户的提问和用户的回答,推理出用户的意图。例如,下面表1所示的关于旅游领域的人机对话,其中,Q是用户提问,A是引擎回答:

[0138] 表1

[0139]

Q: 埃菲尔铁塔

A: 您想知道关于埃菲尔铁塔这个景点的什么呢

Q: 它有多高?

A: 埃菲尔铁塔高 300 米, 天线高 24 米, 总高 324 米。

Q: 怎么去哪里呢

A: 您想从哪里出发?

Q: 凯旋门

A: 从凯旋门去埃菲尔铁塔, 为您找到以下路线: ..... (具体结果略)

Q: 如果从巴黎圣母院呢?

A: 从巴黎圣母院去埃菲尔铁塔, 为您找到以下路线: ..... (具体结果略)

[0140] 从上述表1的人机对话可以看出,机器会试图询问用户的意图,并根据用户的回答以及上下文推理出用户当前的意图。当用户只说埃菲尔铁塔时,机器并不知道用户的具体意图,于是机器提出了问题,然后用户表达了埃菲尔铁塔有多高的具体意图。此时,机器明白了用户的意图是查询一个实体的高度,并且实体是埃菲尔铁塔,于是机器去知识库里面获取埃菲尔铁塔的高度属性,然后返回给用户。

[0141] 接着用户提出“怎么去那里呢”,表达了一个明确规划去某地路线的意图。但是规

划去某地路线的意图有两个参数,一个是起点,一个是终点。机器从上下文中可以得到终点是埃菲尔铁塔,但是不知道起点,因此机器无法执行这个意图,于是机器提出关于起点的问题,最后机器得到用户所说起点是凯旋门。

[0142] 一般情况下,当用户说出一个实体的名称但是没有表达任何意图时,机器会像第一个Q那样,询问用户的意图。当机器已经有了足够的上下文时,机器可以将用户所说的凯旋门理解为,规划去某地路线这个意图的一个起点参数。此时,机器得到规划去某地路线意图的起点是凯旋门,终点是埃菲尔铁塔后,执行这个意图得出规划后的路线。当用户提出“如果从巴黎圣母院呢”的问题时,机器通过上下文可以得出用户提出问题的目的是更改起点,于是,机器给出从巴黎圣母院到埃菲尔铁塔的规划路线的执行结果。

[0143] 为了方便本领域技术人员对本申请中的意图模板有更深刻的理解,参考图1所示,为本申请中配置意图模板的流程图,配置意图模板的流程可以包括步骤:

[0144] 步骤101:建立领域内的本体。

[0145] 当对一个新领域设计智能问答引擎的时候,需要建立该领域内的本体。具体的,通过分析该领域,抽象出该领域包含什么内容,以及这些内容之间的关系,那么抽象出的内容就是实体(entity),内容之间的关系就是实体之间的关系(relation),最终建立该领域的本体(ontology)。

[0146] 例如,当对诗歌领域设计智能问答引擎时,通过数据调研发现这个领域里面有诗人和诗,那么诗人和诗就是实体。并且,诗人和诗都有自己的一些属性,例如,诗人有姓名、性别、朝代等属性,诗有名称、内容等属性。此外,诗人和诗之间存在着写与被写的关系,建立如图2所示的本体。

[0147] 例如,当对旅游领域进行分析时,为了方便将问题简化,只考虑旅游领域里面有城市和景点两个实体。其中,城市有名称、国家等属性,景点有名称、类型等属性。城市和景点之间存在包含和被包含的关系,以及,景点之间存在如何到达的关系,可以建立如图3所示的本体。

[0148] 步骤102:依据领域内的本体生成领域词典。

[0149] 根据已经确定出的包括实体和实体间关系的本体,生成一个包含实体属性值和实体的属性名的领域词典。具体的,可以根据以下步骤A1~步骤A2生成领域词典:

[0150] 步骤A1:获取领域内各实体的各属性。

[0151] 根据领域内确定出的本体,获取本体中各实体所包含的各属性。例如,对旅游领域确定出的本体,可以获取到本体所包含的实体有城市和景点等,并且可以获取到,城市具有名称、所属的国家等属性,景点具有名称、类型、营业时间等属性。

[0152] 步骤A2:将所述各属性对应的属性值作为关键词,并将实体标签和属性名作为关键词的词类型,以及,将所述关键词的同义词、近义词和别名确定为关键词。

[0153] 在获取到城市的名称、所属的国家等属性后,将获取的各属性对应的属性值作为关键词,将实体标签和属性名作为关键词的词类型。例如,城市的名称有“北京”和“上海”,所属的国家为“中国”,那么将“北京”、“上海”和“中国”作为关键词;将“城市名称”和“所属国家”作为词类型。进一步的,还可以将确定出的关键词的同义词、近义词和别名确定为关键词。例如,“中国的首都”是关键词“北京”的别名,所以,将“中国的首都”同样作为关键词。

[0154] 此外,还可以通过人工添加和补充的方式来来确定关键词和关键词的词类型。

[0155] 此时,将获取的各实体的各属性对应的属性值,以及各实体的各属性名作为词类型,就可以构成本领域的领域词典。由于通过查阅领域词典就可以确定一个句子中的关键词,因此,在后续的流程中,利用领域词典,就可以解析句子并提取句子中的关键词(slot)。特别的,领域词典对于冷门领域的作用是非常重要的。

[0156] 例如,有一份如下表2所示的词典:

[0157] 表2

|        |                         |
|--------|-------------------------|
|        | 床前明月光 #### Poem_content |
| [0158] | 李白 #### Poem_author     |
|        | 疑是地上霜 #### Poem_content |

[0159] 其中,“###”是分隔符,分隔符前面是关键词,后面是关键词对应的词类型。特别情况下,领域词典中只有关键词列表,即只有关键词,没有关键词对应的词类型,这样的领域词典只能识别出句子中的关键词,不能确定关键词的类型。

[0160] 当对如下的句子进行处理的时候:{窗前明月光是李白写的吗?},就可以提取到如下所示的关键词:{\$ {Poem\_content,窗前明月光} 是\$ {Poem\_author,李白} 写的吗?}。

[0161] 步骤103:依据领域内各本体所表示出的各实体之间的关系,确定各领域内所有问句的意图。

[0162] 其中,本实施例中的意图用于表述问句的询问目的,意图可以包括以下四点:

[0163] 第一点、询问实体的属性。

[0164] 例如,对上述的城市实体,可以有如下表3所示的意图:

[0165] 表3

|        |                         |
|--------|-------------------------|
|        | xxx 城市的 ID 是什么? (ID 属性) |
| [0166] | xxx 城市的名称是什么? (名称属性)    |
|        | xxx 城市属于哪个国家? (国家属性)    |

[0167] 例如,对上述的景点实体,可以有如下表4所示的意图:

[0168] 表4

|        |                        |
|--------|------------------------|
|        | xxx 景点的 ID 是什么 (ID 属性) |
| [0169] | xxx 景点叫什么? (名称属性)      |
|        | xxx 景点是什么类型的? (类型属性)   |
|        | xxx 景点的营业时间? (营业时间属性)  |

[0170] 第二点、询问实体之间的所有关系。

[0171] 例如,上述的旅游领域本体,一共有三种关系,可以有如下所示的三种意图:

[0172] {城市有哪些景点?(城市到景点的关系) 景点属于哪个城市?(景点到城市的关系) 从这个景点怎么到另一个景点?(景点到景点的关系)}

[0173] 第三点、通过实体的一个或多个属性询问该实体的其他属性或所有属性。

[0174] 具体的,通过一个实体的某个或多个属性定位到实体,然后查询实体的其他属性。

[0175] 例如上述的城市实体,可以有如下意图:{名称是xxx的城市是哪个国家的?(通过名称定位到城市,查询国家属性) 名称是xxx的国家包含哪些城市?(通过国家定位到城市,查询名称属性)}。

[0176] 第四点、通过实体的一个或多个属性询问该实体的关系。

[0177] 具体的,第四点和上述第三点所描述的方法相同的地方是先定位到实体;不同的地方是上述第三点是查询实体的属性,而第四点是查询实体的关系。例如,对于上述城市实体,可以有如下表5所示的意图:

[0178] 表5

[0179]

名称是 xxx 的城市包含哪些景点? (通过名称定位到城市, 查询其和景点的关系)

国家是 xxx 的城市包含哪些景点? (通过国家定位到城市, 查询其和景点的关系, 这句实际上的查询结果是国家包含的景点)

[0180] 通过上述第一点到第四点介绍的四种方法,可以通过本体枚举出所有可能的意图,这四种方法构成了意图枚举算法。这种枚举出来的意图有些是在真实场景下没有实际用处的。例如,问ID属性,因此意图枚举的结果可能需要一定的人工验证。实际上第一点是第三点的一个特例,即是通过名称属性定位到实体,然后查询实体的其他属性。

[0181] 步骤104:为确定的各意图配置对应的意图模板。

[0182] 具体的,意图模板负责对某一类意图的解答,例如,下面所示的问句:{杭州包含什么景点?西安有哪些景点?南京有哪些好玩的地方?},其中的每一个问句都是一个意图,即问一个城市有哪些景点,所有这些问题都可以用一个意图模板来解决。再例如,{从埃菲尔铁塔怎么去卢浮宫?规划一条从巴黎圣母院到埃菲尔铁塔的路线带我从巴黎圣母院去卢浮宫},所示的规划到某地的路线的问句,也都可以用一个意图模板来解决。可以认为,意图模板是意图的一个配置文件,用来表示与该配置文件对应的意图应该如何生成答案。

[0183] 具体的,意图模板可以包括:模板标识,模板描述,意图参数及意图参数对应的参数问题,意图对应的例句组,以及,各例句组的组合规则。

[0184] 其中,所述模板标识用于唯一表示一个意图模板,所述模板描述用于描述所述意图模板的目的;所述意图参数为确定一个意图所对应的参数;所述参数问题为:对应的意图参数缺失时需要向用户发送的问题;所述例句组中的例句包括实体的变量名,用于表示所述意图模板能解决的问题;所述各例句组的组合规则用于表示:不同例句组中的例句的组合方式。

[0185] 具体的,上述第一个意图模板的模板描述可以为“找某个城市的景点”;第二个意图模板的模板描述可以为“规划从某地到某地的路线”。第一个意图模板的意图参数是城市名称;第二个意图模板的意图参数有两个,一个是起点(startPoi),一个是终点(endPoi)。如果用户在对话的过程中没有给出意图参数,需要使用默认的参数问题来询问用户,下面表6是完整的意图参数的配置:

[0186] 表6

[0187]

startPoi  
q:您想从哪里出发呢  
q:起点是哪里呢

[0188]

endPoi:  
q:您想到哪里去呢  
q:终点是哪里呢

[0189] 在上述意图参数后面的以“q:”开头的句子就是参数问题,一个意图参数可以对应多个参数问题,具体用的时候随机取一个参数问题即可。

[0190] 具体的,意图模板对应的例句组指示了什么样的问题适合用这个模板来解决。其中,例句组由模板的设计者自己填写,一般通过分析领域经常问到的问题总结得到。例如,上述第一个寻找某个城市的景点的意图模板,其例句组可以如下表7所示:

[0191] 表7

[0192]

$\{city\_name\}$ 包含什么景点?  
 $\{city\_name\}$ 有哪些景点?  
 $\{city\_name\}$ 有哪些好玩的地方?

[0193] 由于上述寻找城市的景点的意图适合用这种类型的语句来询问,所以,在问题分类中这种类型的问题会被分类到这个意图模板中。

[0194] 上述例句组中给出了具体的例句句式,重点是对句式的总结,构成了所有可能问题模式的集合。由于在实际的应用场景中,实体是一个变量,所以将变量填充到例句之后,可以生成多个真实例句。例如, $\{city\_name\}$ 可以是西安、伦敦、杭州、上海等,将这些实体名称填充到表7所示例句的 $\{city\_name\}$ 中,可以生成如下所示的真实语句:{杭州包含什么景点?杭州有哪些景点?杭州有哪些好玩的地方?西安包含什么景点?西安有哪些景点?西安有哪些好玩的地方?……}。

[0195] 上面列举的例句组中没有组的标记,默认是一组例句,因此不存在例句间的组装逻辑,这种采用一问一答形式来对话的意图模板属于单轮对话意图模板。单轮对话意图模板适用于一问一答的应用场景,并且采用单轮对话意图模板中的例句生成的真实例句不包含上下文的信息。

[0196] 在意图模板中除了单轮对话意图模板外,还存在多轮对话意图模板。多轮对话意图模板中的问题有多个,且包含上下文(Context)的多句问题,具体的多句问题可以如下表8所示:

[0197] 表8

[0198]

Q:埃菲尔铁塔

A:请问您要干什么呢?

Q:帮我规划去那里的路线

A:好的, 请问您要从哪里出发呢?

Q:从巴黎圣母院出发

A:好的, 从巴黎圣母院到埃菲尔铁塔可以这么走: .....

[0199] 从上述的多句问题可以看出,服务器向用户询问了多句,才得到了用户的意图,并且得到了较多的意图参数。因此,服务器在回答多轮对话的问题时需要考虑上下文,并且当没有清楚的得到用户的意图,或者没有得到较多的以意图参数时,需要进一步的对用户询问。进一步的,在回答多轮对话的问题时可以充分考虑到上下文中用户表达的意图参数或意图。

[0200] 需要说明的是,多轮对话意图模板在配置例句的时候需要明确设置组的概念,例如下面表9所示的例句配置:一共有六组,每组以“#”符号开始。如果组号后面有“intent”关键字,则表示这个组是有明确意图的给出,否则,这个组可能只是表达某个意图参数。

[0201] 表9

[0202]

# 1 intent

怎么从\${startPoi}去\${endPoi}

从\${startPoi}去\${endPoi}, 有什么路线

规划从\${startPoi}去\${endPoi}的路线

# 2

\${endPoi}

# 3 intent

怎么到那里

怎么去那里呢

到那里怎么走啊

[0203]

# 4  
{startPoi}  
从\${startPoi}  
从\${startPoi}出发  
起点是\${startPoi}

# 5  
那到\${endPoi}呢  
那去\${endPoi}的路线呢  
那去\${endPoi}怎么走呢

# 6  
那从\${startPoi}出发呢  
那从\${startPoi}呢

[0204] 从上述多轮对话意图模板配置的例句中,可以看出,第一组既给出了意图又给出了所需的所有参数,是一个完整的问题的表达。如果只有第一组,多轮对话意图模板就转化为单轮对话意图模板,可见,单轮对话意图模板是多轮对话意图模板的特殊形式。

[0205] 如果将上述多轮对话意图模板中的各例句组按照一定的组合规则来组合,具体的组合方式如下表10所示:

[0206] 表10

[0207]

- 1) 1
- 2) 1,5
- 3) 1,6
- 4) 2,3
- 5) 2,3,4
- 6) 3,4
- 7) 2,3,5
- 8) 2,3,4,5

[0208]

|             |
|-------------|
| 9) 3,4,5    |
| 10) 2,3,6   |
| 11) 2,3,4,6 |
| 12) 3,4,6   |

[0209] 上述的组合规则表中的每一行代表一个组合规则,其中数字代表组号。例如组合规则10“2,3,6”,所表示的意思是从第二组中随机选取一句话,再从第三组中随机选取一句话,再从第六组中随机选取一句话,最后将这三句话顺序连接起来,构成了一组多句问题例句。

[0210] 按照同样的方式,表10中的每一种组合规则都可以得到一种多句问题的例句。例如,对于组合规则1:“1”,就是表9中第一组例句的单独组合,对例句中包括的实体的变量名进行填充之后,可以得到如下所示的三个不同的多句问题的真实例句:{怎么从巴黎圣母院去埃菲尔铁塔?从巴黎圣母院去埃菲尔铁塔,有什么路线?规划从巴黎圣母院去埃菲尔铁塔的路线?},其中,这里三个多句问题的真实例句,都表示了一个含义。

[0211] 再例如,根据组合规则12:“3,4,6”,可以得到如下表11所示的多句问题的真实例句,其中每组真实例句包含三句:

[0212] 表11

[0213]

|                                                                                     |
|-------------------------------------------------------------------------------------|
| 多句例句 1:<br>怎么到那里?<br>从巴黎圣母院出发<br>那从卢浮宫出发呢?<br>多句例句 2:<br>到那里怎么走啊<br>卢浮宫<br>那从埃菲尔铁塔呢 |
|-------------------------------------------------------------------------------------|

[0214] 可以看出,多句问题所对应的多句例句有可能是不完整的,例如缺少某个参数,也有可能是完整的,例如,组合规则5所组合出来的多句例句。但是,每个组合规则生成的多句问题的多句例句,总会包含一个表达意图的例句。例如,表9所示的例句组1和3这两个例句组都是表达意图的例句组,所以表10中的组合规则里,总会采用例句组1或者例句组3里的句子来进行组合。这是为了模拟真实场景中,在用户没有提供足够的参数和意图时,用户的问题不会被分类到任何一个意图模板。因此,为了获取用户意图,在所有的例句组合规则都至少包含一个表达意图的句子。

[0215] 此外,除了在意图模板中配置上述模板标识,模板描述,意图参数及意图参数对应的参数问题,意图对应的例句组,以及,各例句组的组合规则,还可以配置采用groovy脚本所表示的执行逻辑,执行该groovy脚本就可以生成对用户提问的答案。具体的,执行groovy

脚本生成用户提问的答案的步骤可以包括：获取数据、处理数据和生成答案。

[0216] 其中，获取数据指的是要获取生成答案的源数据。解决任何问题都需要有相应的数据。例如，回答一个城市有哪些景点，需要获取这个城市的所有景点的信息，这些景点信息就是生成答案的源数据。回答一个景点的开关门时间需要有这个景点的实体数据。

[0217] 处理数据指的是对生成答案的源数据进行处理。一般情况下，获取到了源数据后一般无法直接展示给用户，需要对源数据进行处理，包括控制检查、格式转化和算法加工等，才能得到可以直接展示给用户的数据。

[0218] 生成答案是将处理后的源数据进行口语化，并展示给用户。经过了对源数据的处理后得到的数据，如果直接展示给用户显得系统有些生硬。因此，对处理后的源数据进行口语化，在口语化的过程中可以套用一些回复模板，或者采用字符串拼接的方式。

[0219] 需要说明的是，意图模板是一种配置而不是代码。当需要改变意图模板中的执行逻辑时，可以通过改变意图模板中执行逻辑的配置来实现。当需要增加对某个新意图问题的回答时，只需增加相应的意图模板中执行逻辑的配置即可。为了增加意图模板的灵活性，本申请实施例使用groovy脚本来实现上述获取数据、处理数据和生成答案三个步骤。当一个用户问题被分配到具体的意图模板时，只需动态执行相应意图模板中的groovy脚本即可生成答案。

[0220] 为了更清楚的介绍意图模板，意图模板的典型结构如下表12所示：

[0221] 表12

|        |                        |                                |
|--------|------------------------|--------------------------------|
| [0222] | 项目                     | 介绍                             |
|        | ID                     | 模板的标识（ID）                      |
|        | Desc                   | 模板的描述                          |
|        | Args                   | 模板的意图参数，以及模板的意图参数的参数问题。        |
|        | Argument_script        | 为模板的意图参数赋值的 groovy 脚本。         |
|        | Data_retrieve_script   | 获取数据的 groovy 脚本。               |
|        | Data_process_script    | 数据处理的 groovy 脚本。               |
|        | Answer_generate_script | 答案生成的 groovy 脚本。               |
|        | Corpus                 | 符合此模板的例句组。如果是单轮对话模板，则只有一组。     |
|        | Assemble_rule          | 例句组的组合规则。（如果是单轮对话意图模板，此部分可以省略） |

[0223] 更详细的，意图为“规划从某地到某地的路线”的多轮对话意图模板的示例，如下表13所示：

[0224] 表13

[0225]

| 项 目  | 内 容                                                            |
|------|----------------------------------------------------------------|
| ID   | plan_route_to_poi                                              |
| Desc | 规划从某地去某地的路线                                                    |
| Args | startPoi<br>q:您想从哪里出发呢<br>q:起点是哪里呢<br><br>endPoi:<br>q:您想到哪里去呢 |

[0226]

|                        |                                                                                                                                                                                                                                                                                                        |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                        | q:终点是哪里呢                                                                                                                                                                                                                                                                                               |
| Argument_script        | <pre> Map&lt;String,Object&gt;      argmap      =      new HashMap&lt;String,Object&gt;(); argmap.put("startPoi", startPoi.get(0)); argmap.put("endPoi", endPoi.get(0)); return argmap; </pre>                                                                                                         |
| Data_retrieve_script   | <pre> Object  route  =  planner.plan("\${startPoi}", "\${endPoi}"); return route; </pre>                                                                                                                                                                                                               |
| Data_process_script    | <pre> Map&lt;String,Object&gt;      rtn      =      new HashMap&lt;String,Object&gt;(); Object route = __data_retrieve_output__; if(route != null){     rtn.put("success", true);     rtn.put("route", route); }else{     rtn.put("success", false); }; return rtn; </pre>                             |
| Answer_generate_script | <pre> Map&lt;String,Object&gt;      map      = __data_process_output__; Map&lt;String,Object&gt;      rtn      =      new HashMap&lt;String,Object&gt;(); if((Boolean) map.get("success")){     rtn.put("__answer__", " 从 \${startPoi} 出发 到 \${endPoi}, 为您找到以下路线: " + map.get("route")); }else{ </pre> |

[0227]

|        |                                                                                                                                                                                                                                                                                   |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | <pre>rtn.put("__answer__", "对不起，找不到从 \${startPoi}到\${endPoi}的路线。") }; return rtn;</pre>                                                                                                                                                                                           |
| Corpus | <pre># 1 intent 怎么从\${startPoi}去\${endPoi} 从\${startPoi}去\${endPoi}，有什么路线 规划从\${startPoi}去\${endPoi}的路线  # 2 \${endPoi}  # 3 intent 怎么到那里 怎么去那里呢 到那里怎么走啊  # 4 {startPoi} 从\${startPoi} 从\${startPoi}出发 起点是\${startPoi}  # 5 那到\${endPoi}呢 那去\${endPoi}的路线呢 那去\${endPoi}怎么走呢</pre> |

|        |               |                                                                                                  |
|--------|---------------|--------------------------------------------------------------------------------------------------|
| [0228] |               | # 6<br>那从\${startPoi}出发呢<br>那从\${startPoi}呢                                                      |
|        | Assemble_rule | 1<br>1,5<br>1,6<br>2,3<br>2,3,4<br>3,4<br>2,3,5<br>2,3,4,5<br>3,4,5<br>2,3,6<br>2,3,4,6<br>3,4,6 |

[0229] 在表13中,“ID”为模板标识,用于唯一标识一个模板,“Desc”为模板描述,用于采用自然语言描述一个模板的目的,“Args”为意图参数,为该意图模板所需要的所有参数,“Argument\_script”为意图参数赋值的脚本,“Data\_retrieve\_script”为数据获取脚本,“Data\_process\_script”为数据处理脚本,“Answer\_generate\_script”为生成答案脚本,“Corpus”为例句组,“Assemble\_rule”为例句组的组合规则。

[0230] 需要说明的是,在表13的字段“data\_retrieve\_script”中,直接使用了一个planner变量的plan函数来获取路线。而脚本里面并没有说明planner变量是如何初始化的。实际上planner变量是一个路线规划引擎,plan函数是路线规划的入口函数。意图模板在执行脚本的时候,在初始化groovy执行器之前,是需要初始化planner变量的,这个是在实际的引擎里面完成的,初始化完成之后,将planner这个变量设置到groovy执行引擎的环境变量里面,然后在groovy脚本里面就可以直接使用了,所以上述脚本是不会报错的。

[0231] 采用groovy脚本可以使意图模板的问答引擎可以回答任何的问题,因为执行数据获取脚本可以获取到任何的数据,无论是直接查询数据库,查找知识图谱,调用搜索引擎,或者调用其他的相关系统,等等,都可以实现。对于在其他非groovy脚本,也可以将复杂的数据处理操作(可能需要复杂的算法来进行数据处理),首先在问答引擎里面实现,然后将访问入口放置在groovy的环境变量里面,这样就可以直接调用这个环境变量来进行数据处理。所以本申请实施例中采用groovy脚本,可以大量减少意图模板中的groovy脚本的量,却又可以使这些groovy脚本执行从简单到复杂的任意逻辑操作。

[0232] 通过上述对单轮对话意图模板和多轮对话意图模板的介绍,可以得出这二者都有

自己特定的使用场景,各有特点。单轮对话意图模板简单,例句简单,维护也简单,适用于一问一答这种简单的场景(即图7所示的实施例),并且单轮对话意图模板是多轮对话意图模板的一种特殊形式。多轮对话意图模板稍微复杂,但是可以适用于更多的场景(即图8所示的实施例),并且,多轮对话意图模板可以以对话的形式与用户沟通来挖掘用户的潜在意图,进而根据意图来回答问题,因此更加智能化。

[0233] 在介绍完意图模板如何配置之后,意图模板中包含用户可能提出的每个问题所对应的例句组,并且意图模板中配置了可以完成用户提出的每个问题的意图的执行逻辑的groovy脚本,接着需要利用配置好的意图来训练本申请实施例中的意图识别模型和关键词提取模型。其中,意图识别模型可以用于针对一个问句,识别出该问句的意图,例如,一个问句是“请问从巴黎圣母院到卢浮宫怎么走”,将该问句的词标注结果输入意图识别模型可以得到该问句的多个候选意图及所述多个候选意图的可信度,可信度可以用于表示对于的候选意图的可能性大小,可信度越高,对应的候选意图是用户询问的真实意图的可能性就越大。其中,候选意图可以采用模板标识ID来表示。

[0234] 在本申请实施例中,意图识别模型是自然语言分类器,而自然语言分类器是机器学习模型,无论采用哪种机器学习算法,都需要同样的训练语料,即带有标记的自然语言文本。其中,参考表14所示,为带有标记的自然语言文本的例子示意表。

[0235] 表14

|        |                |                      |
|--------|----------------|----------------------|
| [0236] | 自然语言语句         | 标记(label)            |
|        | 杭州有哪些景点        | get_city_scenery     |
|        | 西安有什么好玩的地方     | get_city_scenery     |
|        | 巴黎圣母院什么时候关门呢?  | get_scenery_opentime |
|        | 怎么从卢浮宫去埃菲尔铁塔呢? | plan_route_to_poi    |

[0237] 利用上述表14所示的大量的训练语料,就可以采用机器学习算法训练意图识别模型。例如,支持向量机(SVM,Support Vector Machine)是一个有监督的机器学习模型,通常用来进行模式识别、分类、以及回归分析等。其中,生成训练语料的过程,就是执行配置的意图模板中的例句组合规则,并将例句中实体的变量名替换为变量值生成真实例句的过程。

[0238] 而关键词提取模型是一种序列标注模型,序列标注就是对句子中的各个序列(即是分词之后的各分词)标注类型。因此,训练关键词提取模型的时候,就可以依据各真实例句(例如表14中的真实例句)的词标注结果进行训练,其中词标注结果可以包括:各个分词的词性标注,各个分词是否为关键词的标注,以及各个关键词的词类型的标注,等等。

[0239] 例如,对于表14中的真实例句“怎么从卢浮宫去埃菲尔铁塔呢?”,其词标注结果可以为:“怎么\_r从\_p卢浮宫\_ns去\_v埃菲尔铁塔\_ns###plan\_route\_to\_poi###0 0 B-startPoi 0 B-endPoi”,在标注结果中,第一个分隔符“###”前的内容“怎么\_r从\_p卢浮宫\_ns去\_v埃菲尔铁塔\_ns”为词性标注结果,第一个分隔符和第二个分隔符之间的内容“plan\_route\_to\_poi”为该真实例句对应的意图模板ID,而第二个分隔符之后的内容“0 0 B-startPoi 0 B-endPoi”为关键词标注结果,表示分词“怎么”和“从”不是关键词,而“卢浮宫”是关键词,且词类型为“起点”,分词“去”也不是关键词,而“埃菲尔铁塔”是关键词,且词类型为“终点”。依据大量真实例句的词标注结果即可训练得到关键词提取模型,该关键词提取模型有对线上问句进行关键词提取。

[0240] 因此,在介绍完意图模板的配置过程之后,参考图4,示出了本申请训练意图识别模型和关键词提取模型的流程图,本实施例可以包括以下步骤401~步骤405:

[0241] 步骤401:获取实体的变量值列表,所述变量值列表包括:实体的变量名、实体的变量值及两者的对应关系。

[0242] 在本实施例的流程中,步骤401~步骤404是生成训练语料的过程。

[0243] 在本实施例中,意图模板中虽然配置了例句,但是这些例句包括了实体的变量名,例如“City\_name”,并且只是问法的集合,是对所有可能的问句的模式总结。这种例句不是真实的句子,因为例句中有变量名。而如果将这些变量名都替换为真实的变量值,例如将“City\_name”替换为“北京”,则可以生成真实例句。因此生成真实记录的前提,是一个实体的变量名的所有可能的变量值的集合。

[0244] 例如,对于意图模板中配置的例句“\${City\_name}是哪个国家的?”,其中的变量名就是“City\_name”,其可能的变量值的集合有:{杭州,苏州,北京,上海,伦敦,巴黎,等等}。因此,将该变量值的集合中的变量值替换为例句中的变量名,就能生成如下所示的真实例句的集合:{杭州是哪个国家的?苏州是哪个国家的?北京是哪个国家的?上海是哪个国家的?伦敦是哪个国家的?巴黎是哪个国家的?等等}

[0245] 而在本在步骤中,针对一个领域,为了得到实体的变量值的列表,可以有如下所示的方法:第一,从知识库中获取实体的类型和属性名的属性的值。例如,上述的变量名“City\_name”,实体的类型是“City”,属性名是“name”,所以可以查询知识库获取实体“City”的属性名“name”的属性值,就可以得到相当多的实体的变量值。第二,从外部获取。例如,从外部的网页爬取等。第三,人工维护。依靠用户的经验自己添加相应的变量值。第四,随机产生,不过第四种方式适合某些不可枚举的变量值,例如身高的变量名为“\${height}”、体重的变量名为“\${weight}”等。这种变量名的变量值可以通过随机的方式产生,“身高”可以生成100到230之间的随机整数,就代表可能的变量值(单位为厘米)。

[0246] 一般情况下,如果希望生成的真实例句不会重复,即有相同的变量值被填充到例句的变量名中,所以变量值的个数一般要比生成的真实例句的个数要多。当然,如果变量值太多的话,也会增加获取实体的变量值列表的工作量,也会影响系统性能,所以也不是越多越好。本领域技术人员可以在获取实体的变量值列表时自主选择变量值的个数。

[0247] 步骤402:依据意图模板中配置的例句组和例句组的组合规则,生成包括实体的变量名的变量例句。

[0248] 在本步骤中,依据已经配置好的意图模板中的例句组和例句组的组合规则,来生成包括变量名的变量例句。其中,意图模板中的例句是非结构化的,可以通过解析意图模板中已经编辑好的例句,并将例句存储在一定的数据结构中,例如存在内存里面,方便后续的使用。这里的存储例句的数据结构可以有很多种选择,主要看是否方便输出给后续的处理程序来使用。

[0249] 这里需要注意的是,因为意图模板中,针对于直接根据用户询问的问句生成答案的单轮形式的情况,例句组可以只有一组,即所有例句都编号为“1”,或者,添加其他的默认组编号均可。而对于需要向用户返回问题并根据用户的回答才能确定用户的意图的多轮形式的情况,例句组就有多个,可以分别采用“1、2、3……”等为各例句组编号。

[0250] 而例句组的组合规则,也可以通过读取配置好的意图模板中而得到。同样的,存储

例句组的组合规则的数据结构也可以有很多种选择,选择的依据是看是否方便后续的处理程序来使用。例句组的组合规则用于表示需要将哪几个例句组的例句组合在一起,例如例句组的组合规则为“1,2”,则说明需要从第一个例句组和第二个例句组中分别选择一个例句来生成变量例句。而对于单轮形式的意图目标,其中的例句组的组合规则可以直接为“1”,或者其他的默认规则。

[0251] 因为例句组和例句组的组合规则都存储在特定的数据结构中,接着就可以读取例句组及其组合规则来生成一定数量的真实例句。假设,例句组及其组合规则已经解析成如表15所示的数据结构(json格式)。

[0252] 表15

|        | 例句                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 拼装规则                                                                                                                                                                                   |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [0253] | <pre> {     "1":{         "questions":[             "怎么从\${startPoi} 去\${endPoi}",             "从 \${startPoi} 去 \${endPoi}, 有什么路线",             "规划从\${startPoi} 去\${endPoi}的路线"         ],         "type":"intent"     },     "2":{         "questions":[             "\${endPoi}"         ],         "type":"slot"     },     "3":{         "questions":[             "怎么到那里?",             "怎么去那里呢?",             "到那里怎么走啊?"         ],         "type":"intent"     } } </pre> | <pre> [     ["1"],     ["2","3"] ] </pre> <p>---<br/>注：这里有两条拼装规则，第一条为：["1"], 意思是全部从第一组里面选。</p> <p>第二条为 ["2", "3"], 意思是分别从第二组和第三组的例句里面分别选择一条，然后按照第二组在前，第三组在后的顺序，将这两条例句组装在一起，行成多轮例句。</p> |
| [0254] | <pre> } </pre>                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                        |

[0255] 在实际应用中,为了避免真实例句的数量过多也会造成系统性能的降低,可以预

先设定一个组合规则最多生成多少条真实例句,例如,最多生成100条真实例句。

[0256] 在生成真实例句的过程中,可以从组合规则中指定的每个例句组里面随机选择一条例句,并带上例句组的类型。如果例句组的类型是“intent”,则将该真实例句标记为模板标识ID,如果例句组的类型是关键词slot,则在模板ID后面添加标记“-slot”,说明该真实例句不是直接表明意图,而是提供意图参数。接着按照组合规则指定的例句组,就可以将包括变量名的例句拼装起来,形成多轮变量例句。

[0257] 参考表16所示,为按照组合规则1和2,得到的变量例句的例子。

[0258] 表16

|        |            |                                                                         |
|--------|------------|-------------------------------------------------------------------------|
| [0259] | 组合<br>规则 1 | 怎么从 \${startPoi} 去 \${endPoi} ###<br>plan_route_to_poi                  |
|        |            | 从 \${startPoi} 去 \${endPoi} , 有什么路线 ###<br>plan_route_to_poi            |
|        |            | 从 \${startPoi} 去 \${endPoi} , 有什么路线 ###<br>plan_route_to_poi            |
|        | 组合<br>规则 2 | \${endPoi} ### plan_route_to_poi-slot<br>怎么到那里? ### plan_route_to_poi   |
|        |            | \${endPoi} ### plan_route_to_poi-slot<br>怎么去哪里呢? ### plan_route_to_poi  |
|        |            | \${endPoi} ### plan_route_to_poi-slot<br>到那里怎么走啊? ### plan_route_to_poi |

[0260] 步骤403:将所述变量值列表中各实体的变量值,分别对应替换所述变量例句中的变量名,得到真实例句。

[0261] 接着将步骤401中获取到的变量值列表中各实体的变量值,来对应替换表16所示的变量例句中的变量,就可以得到真实例句。例如,从变量值列表中随机获取一个变量值,替换变量例句中相应的变量名。因此,上述表16中的生成的变量例句,经过步骤403可以得到表17所示的真实例句的例子。在表17中,“###”符号是分隔符,该分隔符前的内容是真实例句,该分隔符后的内容该真实例句对应的模板标识ID。

[0262] 表17

|        |      |                                                                                |
|--------|------|--------------------------------------------------------------------------------|
| [0263] | 拼装规则 | 怎么从\${startPoi, 巴黎圣母院}去\${endPoi, 卢浮宫} ### plan_route_to_poi                   |
|        | 1    | 从\${startPoi, 巴黎圣母院}去\${endPoi, 埃菲尔铁塔}, 有什么路线 ### plan_route_to_poi            |
|        |      | 从\${startPoi, 埃菲尔铁塔}去\${endPoi, 巴黎圣母院}, 有什么路线 ### plan_route_to_poi            |
|        | 拼装规则 | \${endPoi, 卢浮宫} ### plan_route_to_poi-slot<br>怎么到那里? ### plan_route_to_poi     |
|        | 2    | \${endPoi, 巴黎圣母院} ### plan_route_to_poi-slot<br>怎么去哪里呢? ### plan_route_to_poi  |
|        |      | \${endPoi, 埃菲尔铁塔} ### plan_route_to_poi-slot<br>到那里怎么走啊? ### plan_route_to_poi |

[0264] 步骤404:对所述真实例句进行标记,得到训练语料;所述训练语料包括:所述真实例句对应的模板标识和所述真实例句中包括的各关键词及其类型。

[0265] 本步骤需要对表17所示的真实例句进行标记处理,从而得到训练语料,可以用来训练本申请实施例中的意图识别模型和关键词提取模型。其中,本步骤中得到的训练语料可以直接作为训练意图识别模型和关键词提取模型需要的、特定格式的输入。其中,表18示出了训练语料的一种格式示例。

[0266] 表18

|        |                                |
|--------|--------------------------------|
| [0267] | BOS                            |
|        | 分词和词性标注结果 ### 句子标记 ### slot 标记 |
| [0268] | 分词和词性标注结果 ### 句子标记 ### slot 标记 |
|        | .....                          |
|        | EOS                            |

[0269] 其中“BOS(begin of session)”是一组训练语料的开始标记,而“EOS(end of session)”是一组训练语料的结束标记。“分词和词性标注结果”可以是对一个真实例句进行分词和词性标记的结果,句子标记即是该真实例句的模板标识ID,“关键词slot标记”是对该真实例句中的每个分词进行是否为关键词slot以及关键词slot类型的标记。在“slot标记”中,可以采用“0”来表示这个分词不是slot,可以采用“B”表示这个分词是slot且是该slot的开始,可以采用“I”表示这个分词是一个slot的继续,因此,从前面最相邻的“B”到连

续出现的“I”对应的分词,就构成了一个完整的slot。“B”和“I”后的标记内容,是该slot的词类型。

[0270] 因此,针对表18中生成的真实例句,就可以通过标记处理,得到如表19所示的词标注结果,该词标注结果可以直接用于机器学习算法的训练输入。

[0271] 表19

[0272]

```
BOS
  怎么 _r 从 _p 巴黎圣母院 _ns 去 _v 卢浮宫 _ns ###
plan_route_to_poi ### O O B-startPoi O B-endPoi
EOS
BOS
  从 _p 巴黎圣母院 _ns 去 _v 埃菲尔 _nz 铁塔 _n , _w 有 _v 什
么 _r 路线 _n ### plan_route_to_poi ### O B-startPoi O B-endPoi
I-endPoi O O O O
EOS
BOS
  从 _p 埃菲尔 _nz 铁塔 _n 去 _v 巴黎圣母院 _ns , _w 有 _v 什
么 _r 路线 _n ### plan_route_to_poi ### O B-startPoi I-startPoi O
B-endPoi O O O O
```

[0273]

```

EOS
BOS
卢浮宫_ns ### plan_route_to_poi-slot ### B-endPoi
怎么_r 到_v 那里_r ? _w ### plan_route_to_poi ### O O O O
EOS
BOS
巴黎圣母院_ns ### plan_route_to_poi-slot ### B-endPoi
怎么_r 去_v 哪里_r 呢_u ? _w ### plan_route_to_poi ### O O
O O O
EOS
BOS
埃菲尔_nz 铁塔_n ### plan_route_to_poi-slot ### B-endPoi
I-endPoi
到_v 那里_r 怎么_r 走_v 啊_u ? _w ### plan_route_to_poi ###
O O O O O O
EOS

```

[0274] 在表19中,针对第一条训练语料“怎么\_r从\_p巴黎圣母院\_ns去\_v卢浮宫\_ns###plan\_route\_to\_poi###0 0 B-startPoi 0 B-endPoi”,其中的“怎么\_r从\_p巴黎圣母院\_ns去\_v卢浮宫\_ns”,即第一个分隔符“###”之前的内容,是对该真实例句的分词和词性标记结果。其中的“plan\_route\_to\_poi”,即第一个分隔符“###”和第二个分隔符“###”之间的内容,是意图模板的模板标识。其中的“0 0 B-startPoi 0 B-endPoi”,即第二个分隔符之后的内容,是关键词标记的结果,其中第一个字符“0”用于表示对应的分词“怎么”不是关键词,第二个字符“0”用于表示对应的分词“从”不是关键词,第三个字符“B”表示对应的分词“巴黎圣母院”是关键词,其词类型为起点“startPoi”,第四个字符“0”用于表示对应的分词“去”不是关键词,第五个字符“B”表示对应的分词“卢浮宫”是关键词,其词类型为终点“endPoi”。以此类推,本领域技术人员可以确定表19中其他训练语料的含义。

[0275] 以上步骤401~步骤404是依据所述意图模板中配置的例句组和例句组合规则,生成训练语料的过程。

[0276] 步骤405:依据所述训练语料采用机器学习算法,训练得到意图分类模型和关键词提取模型。

[0277] 有了步骤404生成的训练语料,就可以使用机器学习算法来训练意图分类模型和关键词提取。当然,在实际应用中,本领域技术人员可以仅训练一个模型来同时进行模板分

类和关键词slot提取,这就是联合模型(joint model)。例如,最有效的递归神经网络(RNN)。

[0278] 本领域技术人员也可以只训练意图识别模型,例如,最有效的卷积神经网络(CNN)进行文本分类,并且可以采用多通道进行输入,增强模型判断的准确率。如表19所示的训练语料一共可以提供三种不同的通道,一种是词向量通道,第二种是词性通道,第三种是slot标注通道。

[0279] 当然,本领域技术人员也可以同时训练出意图识别模型和关键词提取模型,其中,意图识别模型有进行问句的意图识别,关键词提取模型有对问句进行关键词的标记和类型标记。

[0280] 另外,单轮对话方式和多轮对话方式需要不同的模型,单轮对话方式只需要考虑对当前这一个问句的判断,仅对当前这句话进行分词和词性标注、意图识别以及slot提取即可;而多轮对话方式需要的意图识别模型和关键词提取模型都更加复杂,模型的输入不仅包括当前这句话的标记结果,也包括这句话之前的历史对话内容(即上下文),由意图识别模型和关键词提取模型参考上下文来确定当前这句话的意图并进行slot提取。在本申请实施例中,无论是单轮对话方式还是多轮对话方式,都可以采用相应的深度学习算法来解决,因此本申请实施例既可以实现单轮对话方式的意图识别,也可以同时实现单轮对话方式的意图识别和slot提取,也可以实现多轮对话方式的意图识别及其slot提取。

[0281] 在本申请实施例中,以同时训练得到意图识别模型和关键词提取模型为例进行说明。在训练得到意图识别模型和关键词提取模型之后,可以采用意图识别模型来确定一个问句的意图,以及采用关键词提取模型对该问句进行关键词提取,得到各分词是否为关键词以及词类型的标记结果。因此,本申请实施例可以应用于问题答案的自动生成领域,例如,问答引擎。参考图5,是本申请的问句答案的生成方法在实际应用中的场景架构图。在图5中,用户通过客户端101向服务器102发送一个问句,来询问自己想要的答案,例如,杭州的景点都包括哪些,服务器102则可以采用预先训练的关键词提取模型提取出杭州和景点两个关键词,并且词类型为“城市名称”和“旅游地点”,接着可以依据预先训练的意图识别模型得到该问句是要问杭州的景点,接着就可以通过确定的意图来生成该问句的答案,并通过客户端101展示给用户。其中,服务器102可以采用问答引擎等实现。

[0282] 基于图5所示的场景架构图,参考图6所示,示出了本申请一种问句答案的生成方法实施例1的流程图,本实施例可以应用于图5所示的服务器102上,例如,问答引擎等,本实施例可以包括以下步骤:

[0283] 步骤601:响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型。

[0284] 在本步骤中,用户通过客户端向服务器例如问答引擎发送了一个问题,则服务器可以对该问句进行分析,不仅依据关键词提取模型确定出该问句包括的关键词及关键词的词类型,还能依据意图识别模型确定出该问句的询问的意图。进行关键词提取的时候,可以先对问句进行分词和词性标注,例如,问句中的分词是名词还是动词等,再依据预先训练的关键词提取模型对该问句进行标注得到词标注结果,词标注结果可以包括:问句的各分词是否为关键词以及各关键词的词类型。

[0285] 具体的,步骤601可以包括步骤B1~步骤B3:

[0286] 步骤B1:对所述问句进行分词,得到所述问句的分词集合。

[0287] 首先,对问句进行分词处理并进行词性标注,得到具有词性标注的分词集合。

[0288] 步骤B2:基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;以及,基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果。

[0289] 根据该问句所属的领域对应的领域词典,对分词集合中的各分词进行关键词标记得到第一标记结果。其中,领域词典用于保存各领域词及其对应的词类型之间的对应关系,或者,有些领域词没有标注词类型,因此,基于领域词典对各分词进行标记得到的第一标记结果中,可能同时包括关键词和对应的词类型,也可能仅包括关键词。由于领域词典中可能出现一个领域词包含另一个领域词的现象,如果这两个领域词的词类型不同,则在标记的时候就会出现无法确定采用哪一个词类型的问题。在本实施例中可以这里采用最长优先标记算法来进行标记,即:当存在词类型冲突的时候,优先标记长度最长的领域词。

[0290] 例如,领域词典中有如下的词:杭州,西湖,博物馆,和西湖博物馆,而问句为“杭州西湖的边上有一座西湖博物馆”,则基于领域词典对该问句进行标记的第一标记结果为:“杭州\_B-City-name/西湖\_B/的边上有一座\_0/西湖博物馆\_B-scenic spots name”。参考领域词典中的领域词,“西湖博物馆”既可以标记为“西湖博物馆\_B-scenic spots name”,也可以标记为:“西湖\_B博物馆\_B”,但是按照最长优先标记算法,“西湖博物馆”的长度比“西湖”和“博物馆”都长,所以优先标记为“西湖博物馆\_B-scenic spots name”。可见,关键词“杭州”、“西湖”和“西湖博物馆”都标记为关键词,并且“杭州”和“西湖博物馆”同时也被标记出了词类型,而“西湖”因为在领域词典未给出词类型,所以也没有被标记出词类型。

[0291] 以上可知,基于领域词典进行标注是一种最直接的方式来识别问句中的关键词slot,只考虑各个分词本身是一种局部优先的方法。而在实际应用中,用户可能会出现输入错误等,导致基于领域词典的第一标记结果为空,而且领域词典中的领域词可能会比较杂,因此会将问句中的不重要的分词也标记出来,这就会对最终生成的答案造成比较大的影响。在这种情况下,需要采用预先训练的关键词提取模型同时对问句进行标记。

[0292] 基于关键词提取模型对问句进行关键词提取及词类型的标记,会同时考虑问句的结构,从而确定各分词是否为关键词以及确定各关键词的词类型,是一种全局性的方法。例如,假设问句为“从埃菲尔铁塔到巴黎圣母院怎么走?”,采用预先训练的关键词提取模型已经学习到了“从 $\{startPoi\}$ 到 $\{endPoi\}$ 怎么走?”的句式,因此可以做出如下的第二标记结果:“从\_0埃菲尔铁塔\_B-startPoi到\_0巴黎\_B圣母院\_I-endPoi怎么\_0走\_0?\_0”

[0293] 步骤B3:依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

[0294] 在本步骤中,针对步骤B2中得到的第一标记结果和第二标记结果,则可以确定问句的最终词标注结果。例如,可以针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都将其标注为关键词,如果是,则将该分词确定为关键词,并依据第一标记结果或第二标记结果中标记出的词类型来确定该关键词的词类型。而如果第一标记结果标注为关键词而第二标记结果标注为非关键词,则将该分词确定为非关键词。即,如果基于领域词典的第一标注结果认为一个分词是关键词,而基于关键词提取模型的第二标注结果认为

该分词不是关键词,则以第二标注结果为准,不再将该分词确定为关键词。

[0295] 例如,用户发送的问句为“帮我规划一条从巴黎圣母院到凯旋门的路线?”,根据该问句可以生成词标记结果,该词标记结果可以是词标记矩阵的形式,其中包括各分词和分词是否为关键词以及词类型的标记结果。一种词标记矩阵的示例可以参考表20所示。

[0296] 表20

| 分词     | 词性  | 基于领域词典的第一标注结果 | 关键词提取模型的第二标注结果 | 最终结果 | 词类型      |
|--------|-----|---------------|----------------|------|----------|
|        |     |               |                |      |          |
| [0297] | 帮   | V             | O              | O    | -        |
|        | 我   | R             | O              | O    | -        |
|        | 规划  | V             | O              | O    | -        |
|        | 一   | M             | O              | O    | -        |
| [0298] | 条   | Q             | O              | O    | -        |
|        | 从   | P             | O              | O    | -        |
|        | 巴黎  | NS            | B              | B    | POI_name |
|        | 圣母院 | N             | I              | I    | POI_name |
|        | 到   | V             | O              | O    | -        |
|        | 凯旋门 | NS            | B              | B    | POI_name |
|        | 的   | U             | O              | O    | -        |
|        | 路线  | N             | O              | O    | -        |
|        | ?   | W             | O              | O    | -        |
|        |     |               |                |      |          |

[0299] 其中,表20中的“-”代表没有标注结果。

[0300] 步骤602:依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图。

[0301] 在本步骤中,则将步骤601中得到的问句的词标注结果,作为预先训练的意图识别模型的输入,从而得到该问句多个候选意图,以及各候选意图对应的可信度。例如,针对表20中的词标注结果,根据各关键词及词类型的标注结果,意图识别模型输出以下三个候选意图:plan\_route\_to\_poi,get\_scenic spots\_of\_city,和,get\_name\_of\_city,分别对应的可信度为0.9、0.4和0.3。可见,候选意图其实是配置好的意图模板的模板标识ID。候选意图“plan\_route\_to\_poi”对应于用于规划路线的意图,而候选意图“get\_scenic spots\_of\_city”对应于用于获取一个城市的景点的意图,而候选意图“get\_name\_of\_city”对应于用

于获取一个城市的名字的意图,其中,候选意图“plan\_route\_to\_poi”对应的可信度0.9的分值最高,说明该候选意图“plan\_route\_to\_poi”最有可能是该问句的目标意图,最有可能反映出用户提问的真实目的。

[0302] 步骤603:依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0303] 在本步骤中,根据确定的目标意图可以分别对应到目标意图模板,再依据该目标意图模板,可以根据目标意图模板中配置的获取目标答案的方法,来获取该目标意图对应的目标答案,并返回给客户端。

[0304] 具体的,步骤603的实现过程可以包括步骤C1~步骤C4:

[0305] 步骤C1:依据目标意图模板中配置的意图参数,为所述意图参数赋值,得到查询参数。

[0306] 在本步骤中,因为一个意图模板中均配置了意图参数,因此,采用步骤601中得到的词标注结果为意图参数进行赋值从而得到查询参数。例如,对于目标意图“plan\_route\_to\_poi”,在目标意图模板中已经配置好意图参数,可以包括起点参数“startPoi”和终点参数“endPoi”。则可以根据表20示意的词标注结果,起点关键词的名字为:“巴黎圣母院”,终点关键词的名字为“凯旋门”,则分别为起点参数“startPoi”和终点参数“endPoi”赋值,得到查询参数“巴黎圣母院”和“凯旋门”。

[0307] 在实际应用中,可以在意图模板中配置好脚本,例如,为参数赋值的参数执行脚本,则本步骤即可直接运行目标意图模板中配置的参数执行脚本,从而为意图参数赋值。

[0308] 可以理解的是,在实际应用,可以采用groovy脚本实现本申请实施例中的所有执行脚本。在执行脚本之前,可以先对groovy执行引擎进行初始化,并在groovy引擎中设置相应的变量,这些变量可以直接在意图模板中配置的groovy脚本中进行使用而无需再行声明。采用这种方式可以大量简化groovy脚本的数量,并且增强groovy脚本的功能。例如,初始化一个java的类,该类用来执行一个比较复杂的操作,然后将这个类的实例设置到groovy执行引擎中,则在其他的groovy脚本里面,可直接调用这个类的方法。

[0309] 步骤C2:依据所述查询参数获取所述目标答案的源数据。

[0310] 然后,再依据赋值后的查询参数来获取模板答案的源数据。例如,可以直接执行目标意图模板中配置的数据获取脚本。数据获取脚本中定义了如何获取相关的数据,执行这个脚本,直接获取数据,并按照用户自定义的返回值类型进行返回。

[0311] 步骤C3:对所述源数据按照预设处理方式进行处理得到处理后的源数据。

[0312] 在本步骤中,可以执行目标意图模板中配置的数据处理脚本实现对源数据的处理。其中,数据处理脚本可以接收数据获取脚本的返回值作为输入,并进行数据处理,并返回用户自定义的返回值类型。

[0313] 步骤C4:依据所述处理后的源数据生成所述目标答案。

[0314] 在本步骤中,可以执行目标意图模板中的答案生成脚本实现目标答案的生成。其中,可以数据处理脚本的返回值作为输入,进行答案拼装,或者执行特定的行为action,并返回用户自定义的返回值类型,并将答案生成脚本的返回值作为目标答案,可以将该目标答案通过客户端返回给用户,用户就可以得到提问的问句对应的答案或者最终行为action的执行结果。

[0315] 可见,采用本申请实施例,通过意图模板的配置,可以生成足够多的训练语料,进而使得预先训练关键词提取模型和意图识别模型,对问句的关键期提取和意图的确定更为准确。并且,该意图模板中包括了获取答案的执行脚本,后续可以对执行脚本进行修改等操作,使得意图模板的可维护性更好。并且,根据确定的意图对应的意图模板可以直接执行脚本从而获得答案,从而提高了服务器例如问答引擎的效率。此外,意图模板中还配置了能回答出问题的问句的例句,因此,使用意图模板可以方便的回答用户提问的问题,自动给出问题的答案。

[0316] 参考图7所示,示出了本申请一种问句答案的生成方法实施例2的流程图,本实施例主要为单轮对话形式的示例,即根据用户提问的问句可以直接返回答案的应用场景,本实施例可以包括以下步骤:

[0317] 步骤701:响应于触发问句,依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型。

[0318] 在本实施例中,步骤701的实现可以参考步骤601的介绍,在此不再赘述。

[0319] 步骤702:将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度。

[0320] 在本步骤中,将词标注结果作为意图识别模型的输入,可以得到多个候选意图及对应的可信度。具体可以参考步骤602的介绍。

[0321] 步骤703:将所述多个候选意图中可信度最高的候选意图确定为当前意图。

[0322] 在单轮对话形式中,以各候选意图的可信度为准来确定目标意图,首先,将可信度最高的候选意图作为当前意图进行意图决策,即从多个候选意图中选择最符合实际情况的意图作为目标意图。

[0323] 步骤704:依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整,如果完整,则进入步骤705,如果不完整,则进入步骤706。

[0324] 对该当前意图对应的当前意图模板,可以获取到该当前意图模板中配置的所有意图参数,因为生成问句的答案需要所有意图参数的值,所以可以参考当前意图模板中配置的所有意图参数,来判断词标注结果中的各关键词是否完整,即是否能依据各关键词及对应的词类型,对当前意图模板中的所有意图参数成功进行赋值。

[0325] 步骤705:将所述当前意图确定为目标意图。

[0326] 如果当前意图模板的所有意图参数都能被赋值,则将当前意图确定为目标意图。

[0327] 步骤706:按照可信度从高到低的顺序选择下一个候选意图作为当前意图,并返回步骤704,直至所述多个候选意图都选择完毕,仍然无法确定目标意图的时候,向客户端返回无法生成问题答案。

[0328] 如果当前意图模板的所有意图参数不能都被赋值,则按照可信度从高到低的顺序选择下一个候选意图作为当前意图,并继续判断新选择的当前意图中的全部意图参数是否能全部被赋值,直至多个候选意图都选择完毕。如果所有的候选意图都没有被选中,则可以生成该问句的答案,进一步可以将无法生成答案的信息通过客户端返回给用户。

[0329] 步骤707:依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0330] 如果能够确定得到目标意图,则进一步执行该目标意图模板中配置的各脚本,从而得到该问句的目标答案。

[0331] 其中,在本实施例中,这里采用某一个意图模板的所有意图参数是否能被赋值作为确定目标意图的依据,因为意图分类模型可能存在着一定的失败率,可信度最高的候选意图可能不是真实的用户意图,因此,在得到候选意图之后再通过意图参数来约束,往往可以确定出到真实的用户意图。

[0332] 在本实施例中,通过预先建立本体可以枚举用户所有可能的意图,然后将各个意图以意图模板的方式存储,每个意图模板负责解决对应的意图的问题。意图模板可以包括意图参数、意图的解决逻辑以及意图的可能问句形式,针对用户提出的问题,通过预先训练的关键词提取模型和,将该问句归类到一个具体的意图上,然后将提取的关键词作为意图的参数,执行意图模板里面对应的逻辑,即执行意图模板中配置的脚本从而得到问题的答案。本申请实施例中训练模型的大量训练语料都是根据意图模板中的例句组和例句组的组合规则生成的,因此训练得到的关键词提取模型和意图识别模型更能准确的提取关键词并进行意图归类,从而使得本实施例中生成的答案也能更准确。

[0333] 参考图8所示,示出了本申请一种问句答案的生成方法实施例3的流程图,主要为多轮对话形式的示例,即不能根据用户提问的问句直接返回答案,需要向用户返回问题并根据用户的输入才能确定目标意图的应用场景,本实施例可以包括以下步骤:

[0334] 步骤801:响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型。

[0335] 在本实施例中,与步骤701不同的是,针对每一次用户提出的当前问句,服务器都会将当前问句的词标注结果,以及历史问句的关键词提取结果作为关键词提取模型的输入,来对当前问句进行关键词提取。

[0336] 其中,在多轮对话形式中,对于用户的历史对话内容,例如,历史问句,对历史问句的意图识别结果和关键词提取结果等,可以保存至会话数据库中,则本步骤就可以从会话数据库中获取对话内容。因为历史问句已经进行过关键词提取,且关键词提取结果都存储在历史对话内容即上下文中,因此,可以直接从会话数据库中获取到历史问句的关键词提取结果,将其和当前问句的词标注结果一并作为关键词提取模型的输入,来对当前问句进行关键词提取。可以理解的是,关键词提取结果也不是一成不变的,可能随着对话内容的改变而改变。

[0337] 步骤802:至少将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及其对应的可信度。

[0338] 本步骤与步骤702的不同之处在于,本步骤中意图识别模型的输入,除了用户提出的当前问句的词标注结果,还可以包括用户的历史问句的关键词提取结果和意图识别结果,其中,关键词提取结果和意图识别结果可以用于对当前问句辅助识别意图。因为本实施例应用的是多轮对话形式的应用场景,所以用户每一次输入的内容例如问句中也会包括必要的信息,例如意图参数等,因此,对用户后续输入的内容进行词标注之后,除了将词标注结果作为意图识别模型的输入,还会将历史问句的意图识别结果和关键词提取结果也作为意图识别模型的输入,从而得到当前问句的多个候选意图及对应的可信度。

[0339] 其中,本步骤就可以从会话数据库中获取对话内容。因为历史问句已经进行过关键词提取,且关键词提取结果都存储在历史对话内容即上下文中,因此,解析出了候选意图之后,候选意图的参数一部分来自于对当前语句的关键词提取结果,另一部分来自于对话内容的上下文中。具体的,针对用户每一次输入的当前问句,将该当前问句的词标注结果,以及历史问句的意图识别结果和关键词提取结果,都作为意图识别模型的输入,来获取当前问句的意图识别结果。可以理解的是,候选意图也不是一成不变的,可能随着对话内容的改变而改变。

[0340] 步骤803:将所述多个候选意图中可信度最高的候选意图确定为当前意图。

[0341] 在本步骤中将步骤802中得到的多个候选意图中可信度最高的候选意图确定为当前意图。

[0342] 步骤804:判断所述当前意图的可信度是否大于预设可信度阈值,如果大于,则进入步骤806;如果不大于,则进入步骤805。

[0343] 在本步骤中,判断可信度的最大值是否大于预设可信度阈值,例如,0.8,如果大于,则可以认为选择可信度最大的当前意图作为目标意图。

[0344] 步骤805:选取预设个数的前若干个候选意图作为待选择意图,并将所述待选择意图发送至终端以供用户选择。

[0345] 而如果可信度的最大值都小于预设可信度阈值,则可以按照可信度从大到小的顺序给候选意图进行排序,并选取前预设个数的多个候选意图作为待选择意图。例如,选取前两个候选意图作为待选择意图,并发送给客户端,以使用户可以从这两个待选择意图中选择与自己实际目的相符合的意图,进一步的,后续可以将用户选择的意图作为目标意图。

[0346] 具体的,服务器可以生成包括这两个待选择意图的选择问句返回给用户。选择问句的内容可以来自于意图模板的“desc”字段,例如有两个待选择意图的“desc”字段分别为表20所示的内容。

[0347] 表20

[0348]

|                   |               |
|-------------------|---------------|
| plan_route_to_poi | 规划到\${poi}的路线 |
| get_poi_desc      | 获取\${poi}的介绍  |

[0349] 则生成的选择问句可以为:“请问您是想规划到凯旋门的路线,还是获取凯旋门的介绍呢?”。

[0350] 步骤806:将所述终端返回的、用户选择的意图确定为目标意图,或者,将当前意图确定为所述目标意图。

[0351] 将选择问句返回至客户端展示给用户之后,用户会根据选择问句输入答案,例如输入“规划到凯旋门的路线”,则可以认为用户选择的意图为“规划到凯旋门的路线”,从而将该意图确定为目标意图。

[0352] 可以理解的是,如果发送一次选择问句无法确定目标意图,服务器可以再次或者多次向用户发送选择问句,直至确定出用户的目标意图,或者当所有的“desc”字段都组合完毕,仍然不能确定出用户的目标意图,则返回无法确定意图的内容给客户端,结束生成问句的答案的流程;或者,用户如果改变了其最初的意图,在这种情况下,也可以结束当前问句的答案生成的流程,转而对用户新意图的处理和执行。

[0353] 步骤807:依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键

词是否完整,如果不完整,则进入步骤808,如果完整,则进入步骤811。

[0354] 在确定目标意图之后,依据对应的目标意图模板中配置的全部意图参数,来判断词标注结果中各关键词是否完整。具体的判断过程可以参考步骤704的介绍,在此不再赘述。

[0355] 步骤808:将不完整的关键词对应的意图参数确定为缺少的意图参数;依据所述缺少的意图参数和所述目标意图模板,生成所述目标意图相关的参数问题。

[0356] 在本实施例中,如果依据词标注结果无法对目标意图模板中的意图参数进行赋值,则说明词标注结果中的关键词不完整,不完整的关键词就对应了无法赋值的那些意图参数,将这些意图参数确定为缺少的意图参数。

[0357] 步骤809:将所述参数问题返回给客户端以便显示给用户,所述意图模板包括:所述意图参数和参数问题之间的对应关系。

[0358] 进一步的,依据缺少的意图参数和目标意图模板中该缺少的意图参数对应的参数问题,来生成询问问题发送至客户端展示给用户。

[0359] 其中,参数问题的生成来自于意图模板,在意图模板中,每个意图参数在定义的时候都被设置了一些如果该意图参数缺失需要询问的问题,因此,这里从缺失的参数对应的参数问题列表中随机选择一个问题返回给用户即可。

[0360] 步骤810:根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数,进入步骤811。

[0361] 生成参数问题询问用户后,用户可能在下一轮对话中给出相应的意图参数,就可以从用户输入的参数答案中获取到缺少的意图参数,在这种情况下,目标意图模板的全部意图参数就都可以赋值了。或者,如果用户改变意图,则服务器也会停止对当前意图缺失的意图参数的提问。

[0362] 步骤811:依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0363] 在本步骤中,如果目标意图模板的全部意图参数都能被赋值,就可以接着执行该目标意图模板中预先配置的脚本,来生成该目标意图的目标答案,具体过程可以参考步骤603所示,在此不再赘述。

[0364] 需要说明的是,在多轮对话形式中,确定目标意图和目标意图的意图参数的完整性判断,实际上是一种推理,即根据用户提供的内容,推理出用户真正的意图。与单轮对话形式不同的是,本实施例并非基于意图参数优先匹配来确定目标意图,而是基于可信度优先的原则来确定目标意图。因为在单轮对话形式中用户会在一个问句中尽量全的表达意图和意图参数,而多轮对话形式时,用户可能倾向于随意的表达,即一个问句或者一句话里面可能只表达了意图,但是没有任何的参数,例如“可以帮我规划一条路线吗?”,对于这个问句,如果按照意图参数优先匹配的原则,则一条候选意图都匹配不上,因为用户没有提供任何意图参数,但是用户却明确的表达了意图,因此在多轮对话形式中采用可信度优先的方式,可以更为准确的发掘用户的意图。

[0365] 在本实施例中,为了方便挖掘用户的潜在意图,进行用户的真实意图的推理,本实施例还可以包括:

[0366] 将用户输入的对话内容保存至会话数据库中,所述对话内容包括:用户选择的意

图和/或用户输入的参数答案。

[0367] 在服务器生成问句返回给客户端并展示给用户之后,用户可能会根据问句来输入自己的答案,或者,用户会根据问句来选择自己的意图,因此,可以将用户输入的这些对话内容暂时保存至会话数据库中。或者,也可以缓存至内存中。后续可以再从会话数据库或内存中获取到用户选择的意图,或者从对话内容中查询到意图参数等。

[0368] 进一步的,在将对话内容保存至会话数据库或者内存后,还可以对所述会话数据库中的对话内容进行更新,或者,在到达预设的时间周期的时候,将用户输入的对话内容进行删除。因为在多轮对话形式中,涉及到用户和服务器例如问答引擎之间的对话,需要保存用户的历史对话内容以及涉及到的意图、意图参数等。可以理解的是,这些对话内容也不是随着时间一成不变的,随着用户和服务器的对话内容的推进,原本的意图或意图参数的值也会被新的值覆盖。同样,这些对话内容也不是永久存在的,例如可以设置一个时间阈值,当达到该时间阈值的时候就可以删除该对话内容。或者,在用户发送了新的问题产生了新的对话内容的时候,上一次的对话内容就不能影响到当前这次对话,因此,可以将上一次的这些对话内容删除。

[0369] 可见,在本实施例中,不仅会将用户最开始提问的问句作为关键词提取模型或意识识别模型的输入,还会在无法确定用户的意图或者能确定用户的意图可是意图参数不完整的情况下,向用户继续提问以便用户能够根据问题来选择自己的意图或者补充意图参数。因此,本实施例经过合理的推理就有可能能够明确用户的意图及获取到完整的意图参数,从而实现与用户对话并且根据对话内容能够自动生成用户所需的答案,不仅提高了答案生成的准确率,还能使得生成的答案更符合用户的习惯,与用户的对话交互也非常友好,用户不需要去学习一种专门的交互方式,直接通过日常生活中的对话的方式就可以与服务器例如问答引擎交流。

[0370] 对于前述的方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0371] 与上述本申请一种问句答案的生成方法实施例1所提供的方法相对应,参见图9,本申请还提供了一种问句答案的生成装置实施例,在本实施例中,该生成装置可以集成于服务器端,例如问答引擎等,该生成装置可以包括:

[0372] 关键词标注单元901,用于响应于触发问句,至少依据预先训练的关键词提取模型对所述问句进行标注得到词标注结果,所述词标注结果包括:所述问句的各分词是否为关键词以及各关键词的词类型。

[0373] 其中,所述关键词标注单元901可以包括:

[0374] 分词子单元,用于对所述问句进行分词,得到所述问句的分词集合;第一标记子单元,用于基于所述问句所属的领域对应的领域词典,对所述分词集合中的各分词进行关键词标记得到第一标记结果;第二标记子单元,用于基于预先训练的关键词提取模型,对所述分词集合中的各分词进行关键词标记得到第二标记结果;所述领域词典用于保存各领域词及其对应的词类型之间的对应关系;以及,第一确定子单元,用于依据所述第一标记结果和第二标记结果,确定所述分词集合中的各分词的词标注结果。

[0375] 其中,所述领域词典可以通过以下方式生成:获取领域内各实体的各属性;将所述各属性对应的属性值作为关键词,并将实体标签和属性名作为关键词的词类型,以及,将所述关键词的同义词、近义词和别名确定为关键词。

[0376] 其中,所述第一确定子单元可以包括:

[0377] 第一判断子单元,用于针对分词集合中的各分词,判断所述第一标记结果和第二标记结果是否都标注为关键词;第二确定子单元,用于在所述第一判断子单元的结果为是的情况下,将该分词确定为关键词,并依据所述第一标记结果或第二标记结果确定所述关键词的词类型;以及,第三确定子单元,用于在所述第一判断子单元的结果为第一标记结果标注为关键词而第二标记结果标注为非关键词的情况下,将该分词确定为非关键词。

[0378] 意图确定单元902,用于依据所述词标注结果和预先训练的意图识别模型,确定所述问句的目标意图。

[0379] 答案获取单元903,用于依据预先配置的、与所述目标意图对应的目标意图模板,获取所述目标意图对应的目标答案。

[0380] 其中,所述答案获取单元903可以包括:

[0381] 赋值子单元,用于依据目标意图模板中配置的意图参数,为所述意图参数赋值,得到查询参数;源数据获取子单元,用于依据所述查询参数获取所述目标答案的源数据;处理子单元,用于对所述源数据按照预设处理方式进行处理得到处理后的源数据;以及,生成子单元,用于依据所述处理后的源数据生成所述目标答案。

[0382] 其中,意图模板可以通过以下方式生成:

[0383] 依据领域内各本体所表示出的各实体之间的关系,确定各领域内所有问句的意图;为确定的各意图配置对应的意图模板。其中,所述意图用于表述问句的询问目的,所述意图包括:询问实体的属性,询问实体之间的所有关系,通过实体的一个或多个属性询问该实体的其他属性或所有属性,和/或,通过实体的一个或多个属性询问该实体的关系;所述意图模板包括:模板标识,模板描述,意图参数及意图参数对应的参数问题,意图对应的例句组,以及,各例句组的组合规则;所述模板标识用于唯一表示一个意图模板,所述模板描述用于描述所述意图模板的目的;所述意图参数为确定一个意图所对应的参数,所述参数问题为:对应的意图参数缺失时需要向用户发送的问题,所述例句组中的例句包括实体的变量名,用于表示所述意图模板能解决的问题,所述各例句组的组合规则用于表示:不同例句组中的例句的组合方式。

[0384] 其中,意图识别模型和关键词提取模型可以通过以下方式训练:

[0385] 依据所述意图模板中配置的例句组和例句组合规则,生成训练语料;依据所述训练语料采用机器学习算法,训练得到意图分类模型和关键词提取模型。

[0386] 其中,训练语料可以通过以下方式生成:

[0387] 获取实体的变量值列表,所述变量值列表包括:实体的变量名、实体的变量值及两者的对应关系;依据意图模板中配置的例句组和例句组的组合规则,生成包括实体的变量名的变量例句;将所述变量值列表中各实体的变量值,分别对应替换所述变量例句中的变量名,得到真实例句;对所述真实例句进行标记,得到训练语料;所述训练语料包括:所述真实例句对应的模板标识和所述真实例句中包括的各关键词及其类型。

[0388] 可见,采用本申请实施例的生成装置,预先通过意图模板的配置,可以生成足够多

的训练语料,进而使得预先训练关键词提取模型和意图识别模型,对问句的关键期提取和意图的确定更为准确。并且,该意图模板中包括了获取答案的执行脚本,后续可以对执行脚本进行修改等操作,使得意图模板的可维护性更好。并且,根据确定的意图对应的意图模板可以直接执行脚本从而获得答案,从而提高了服务器例如问答引擎的效率。此外,意图模板中还配置了能回答出问题的问句的例句,因此,使用意图模板可以方便的回答用户提问的问题,自动给出问题的答案。

[0389] 与上述本申请一种问句答案的生成方法实施例2所提供的方法相对应,参见图10,本申请还提供了一种问句答案的生成装置实施例2,在本实施例中,该生成装置可以集成于服务器端,例如问答引擎等,本实施例可以生成单轮对话形式中关于用户的问句的答案,该生成装置中的意图确定单元902可以包括:

[0390] 第一模型处理子单元1001,用于将所述词标注结果作为所述预先训练的意图识别模型的输入,得到多个候选意图及对应的可信度。

[0391] 第四确定子单元1002,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图。

[0392] 第二判断子单元1003,用于依据所述当前意图对应的当前意图模板,判断所述词标注结果中各关键词是否完整;所述意图模板包括:确定一意图所需要的全部意图参数。

[0393] 第五确定子单元1004,用于在所述第二判断子单元的结果为是的情况下,将所述当前意图确定为目标意图。

[0394] 选择子单元1005,用于在所述第二判断子单元的结果为否的情况下,按照可信度从高到低的顺序选择下一个候选意图作为当前意图。

[0395] 触发子单元1006,用于在选择子单元选择下一个候选意图之后,触发所述的第二判断子单元,直至所述多个候选意图都选择完毕。

[0396] 其中,所述第二判断子单元具体可以包括:

[0397] 参数获取子单元,用于获取所述当前意图模板中配置的全部意图参数;赋值子单元,用于采用所述词标注结果中的各关键词的词类型,为所述全部意图参数进行赋值;以及,赋值判断子单元,用于判断所述全部意图参数是否能赋值成功。

[0398] 采用本实施例的生成装置,通过预先建立本体可以枚举用户所有可能的意图,然后将各个意图以意图模板的方式存储,每个意图模板负责解决对应的意图的问题。意图模板可以包括意图参数、意图的解决逻辑以及意图的可能问句形式,针对用户提出的问题,通过预先训练的关键词提取模型和,将该问句归类到一个具体的意图上,然后将提取的关键词作为意图的参数,执行意图模板里面对应的逻辑,即执行意图模板中配置的脚本从而得到问题的答案。本申请实施例中训练模型的大量训练语料都是根据意图模板中的例句组和例句组的组合规则生成的,因此训练得到的关键词提取模型和意图识别模型更能准确的提取关键词并进行意图归类,从而使得本实施例中生成的答案也能更准确。

[0399] 与上述本申请一种问句答案的生成方法实施例3所提供的方法相对应,参见图11,本申请还提供了一种问句答案的生成装置实施例3,在本实施例中,该生成装置可以集成于服务器端,例如问答引擎等,本实施例可以生成单轮对话形式中关于用户的问句的答案,该生成装置中的意图确定单元902可以包括:

[0400] 第二模型处理子单元1101,用于至少将所述词标注结果作为所述预先训练的意图

识别模型的输入,得到多个候选意图及其对应的可信度。

[0401] 第六确定子单元1102,用于将所述多个候选意图中可信度最高的候选意图确定为当前意图。

[0402] 第三判断子单元1103,用于判断所述当前意图的可信度是否大于预设可信度阈值。

[0403] 第七确定子单元1104,用于在所述第三判断子单元的结果为大于的情况下,将所述当前意图确定为所述目标意图。

[0404] 选取子单元1105,用于在所述第三判断子单元的结果为否的情况下,选取预设个数的前若干个候选意图作为待选择意图。

[0405] 发送子单元1106,用于将所述待选择意图发送至终端以供用户选择。

[0406] 第八确定子单元1107,用于将所述终端返回的、用户选择的意图确定为目标意图。

[0407] 第四判断子单元1108,用于依据所述目标意图对应的目标意图模板,判断所述词标注结果中各关键词是否完整。

[0408] 第九确定子单元1109,用于在所述第四判断子单元的结果为否的情况下,将不完整的关键词对应的意图参数确定为缺少的意图参数。

[0409] 问题生成子单元1110,用于依据所述缺少的意图参数和所述目标意图模板,生成所述目标意图相关的参数问题。

[0410] 问题返回子单元1111,用于将所述参数问题返回给客户端以便显示给用户;所述意图模板包括:所述意图参数和参数问题之间的对应关系。

[0411] 参数获取子单元1112,用于根据所述客户端返回的、用户输入的参数答案获取所述缺少的意图参数。

[0412] 其中,所述意图确定单元902还可以包括:

[0413] 保存子单元,用于将用户输入的对话内容保存至会话数据库中,所述对话内容包括:用户选择的意图和/或用户输入的参数答案。

[0414] 其中,所述意图确定单元902还可以包括:

[0415] 更新子单元,用于对所述会话数据库中的对话内容进行更新,或者,删除子单元,用于在到达预设的时间周期的时候,将用户输入的对话内容进行删除。

[0416] 其中,所述第二模型处理子单元1101具体可以包括:

[0417] 对话获取子单元,从会话数据库中获取所述对话内容;以及,输入子单元,用于将所述词标注结果和所述对话内容输入预先训练的意图识别模型的,得到多个候选意图及其对应的可信度。

[0418] 其中,所述第四判断子单元1108具体可以包括:

[0419] 参数获取子单元,用于获取所述目标意图模板中配置的全部意图参数;以及,赋值子单元,用于采用所述词标注结果中的各关键词的词类型,为所述目标意图模板中的全部意图参数进行赋值,赋值判断子单元,用于判断所述全部意图参数是否能赋值成功。

[0420] 采用本实施例中的生成装置,不仅会将用户最开始提问的问句作为关键词提取模型或意识识别模型的输入,还会在无法确定用户的意图或者能确定用户的意图可是意图参数不完整的情况下,向用户继续提问以使用户能够根据问题来选择自己的意图或者补充意图参数。因此,本实施例经过合理的推理就有可能能够明确用户的意图及获取到完整的意

图参数,从而实现与用户对话并且根据对话内容能够自动生成用户所需的答案,不仅提高了答案生成的准确率,还能使得生成的答案更符合用户的习惯,与用户的对话交互也非常友好,用户不需要去学习一种专门的交互方式,直接通过日常生活中的对话的方式就可以与服务器例如问答引擎交流。

[0421] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0422] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0423] 以上对本申请所提供的一种自然语言问句答案的生成方法、装置和服务器进行了详细介绍,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方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,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申请的限制。

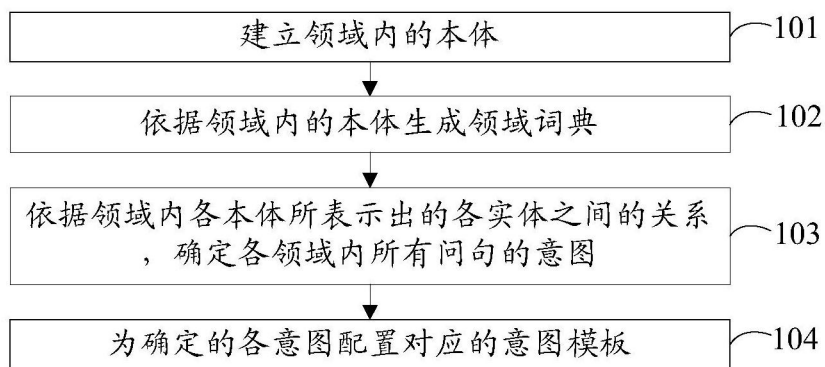


图1

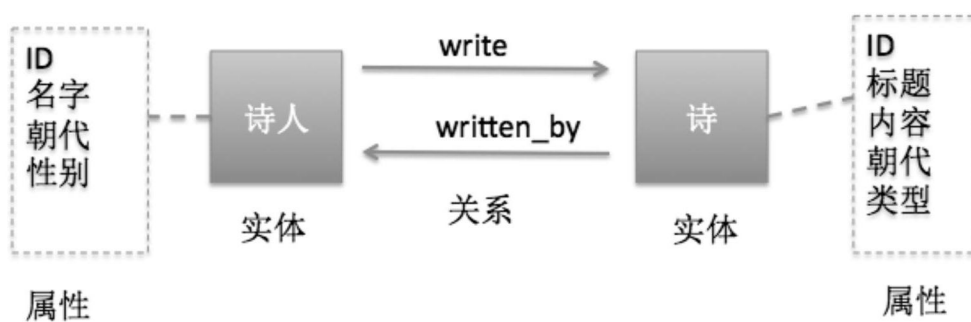


图2

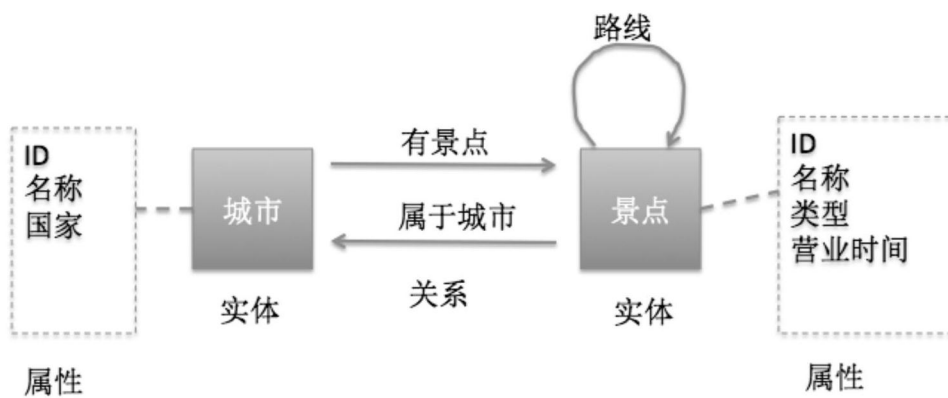


图3

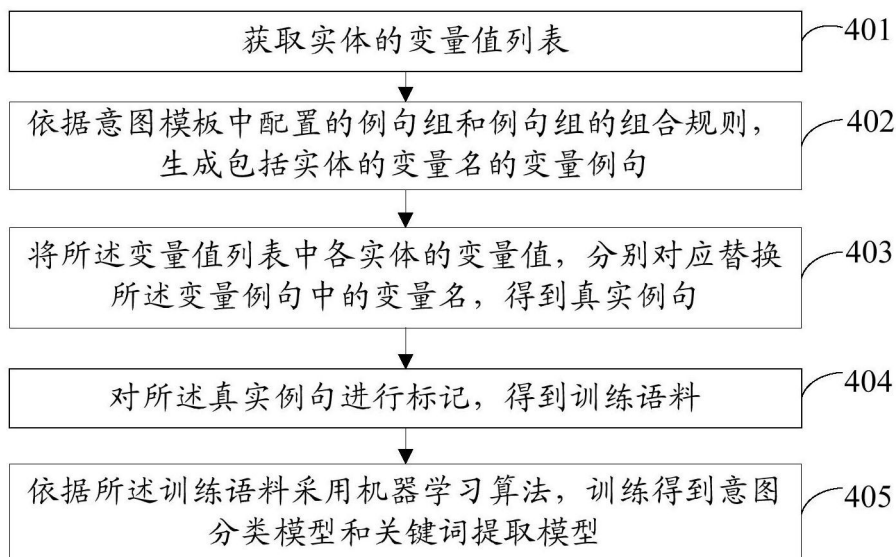


图4

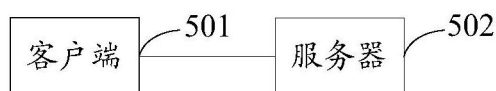


图5

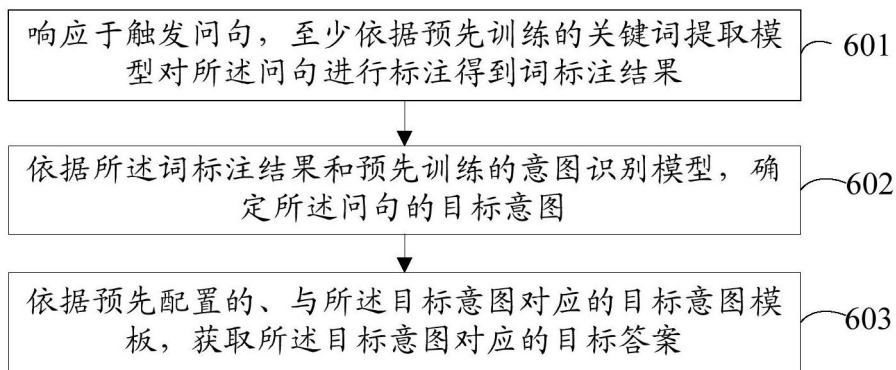


图6

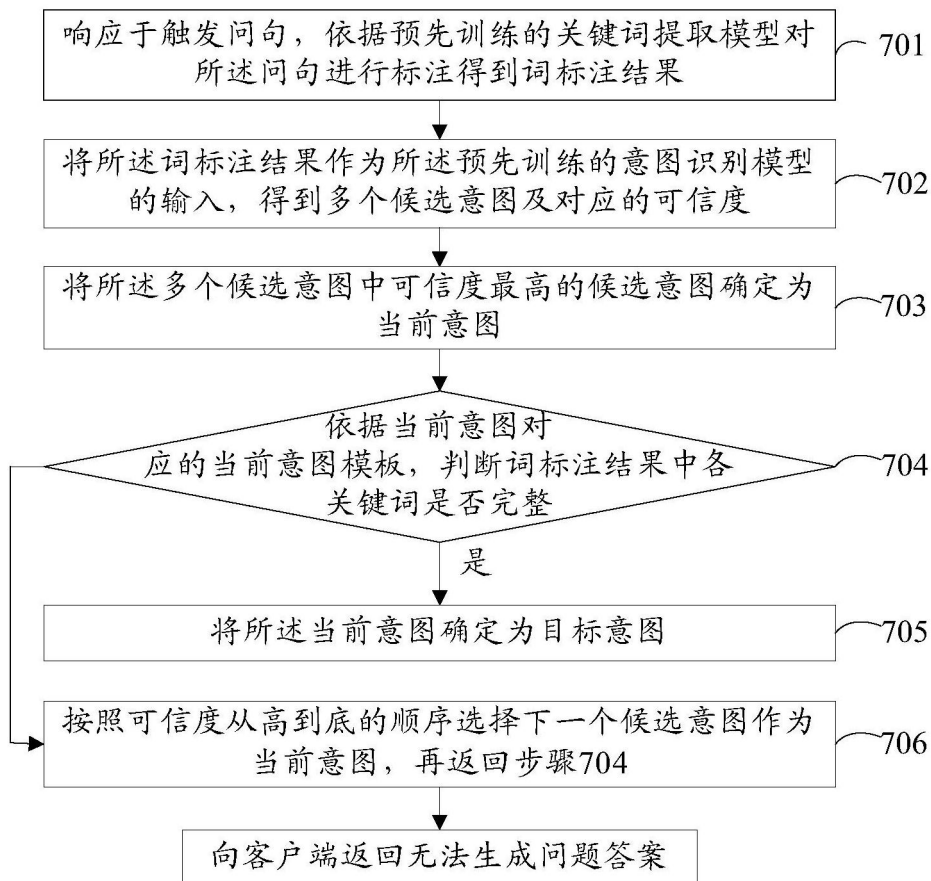


图7

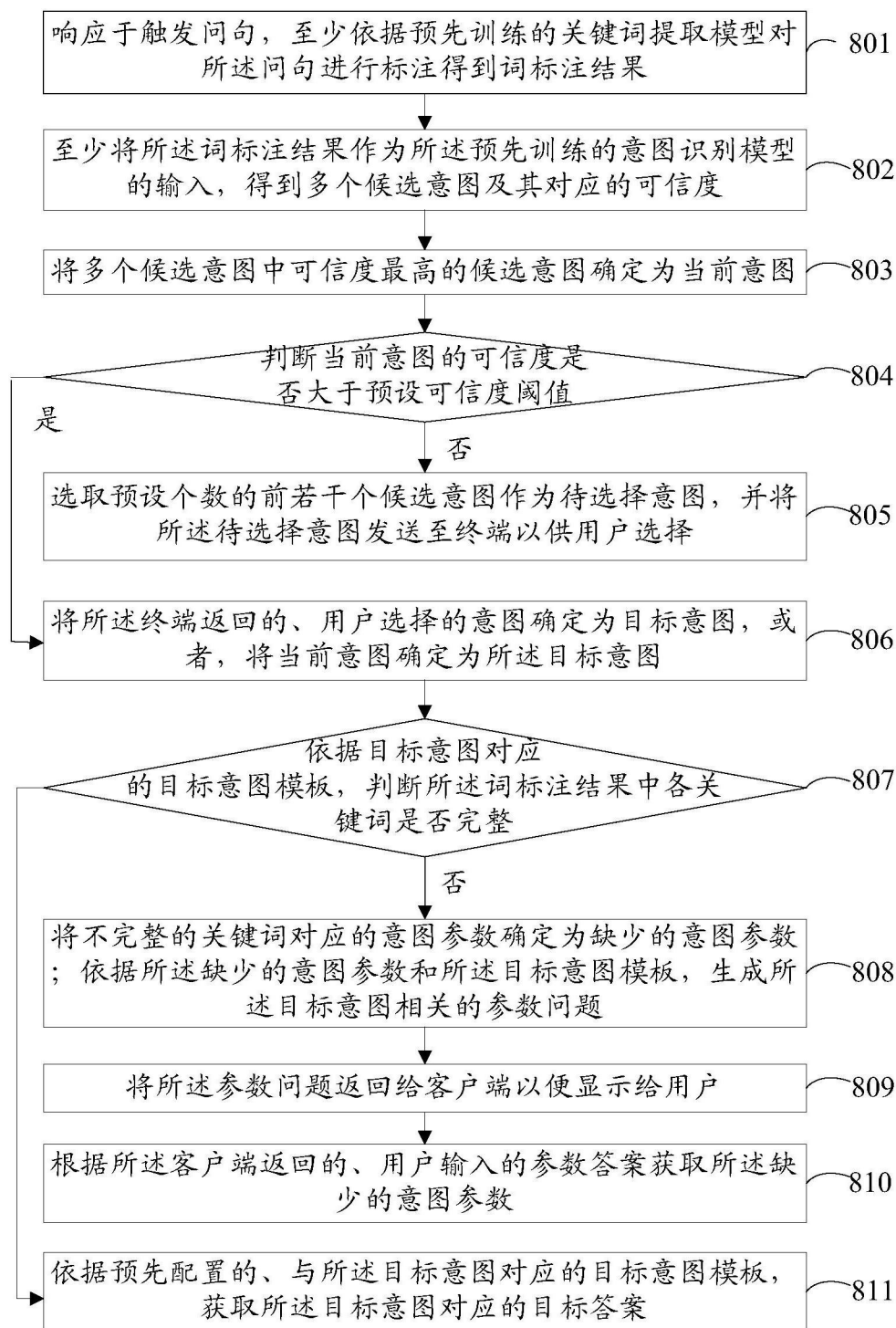


图8

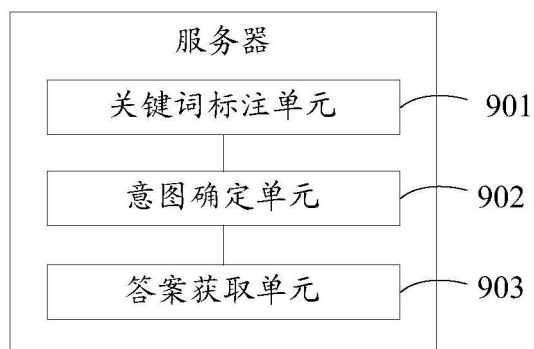


图9

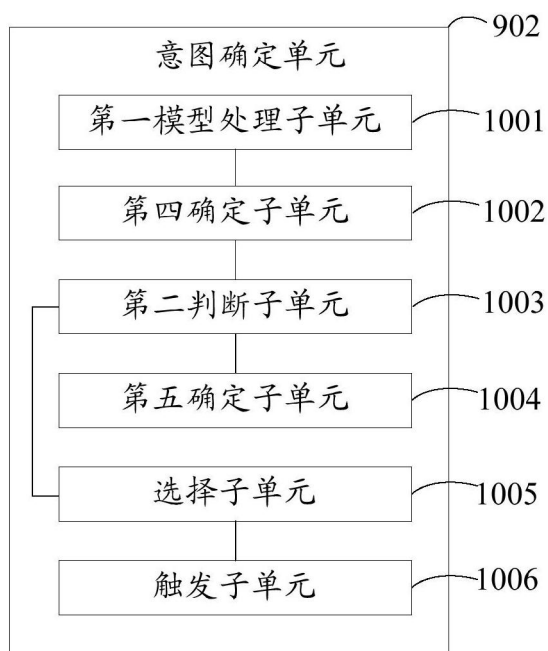


图10

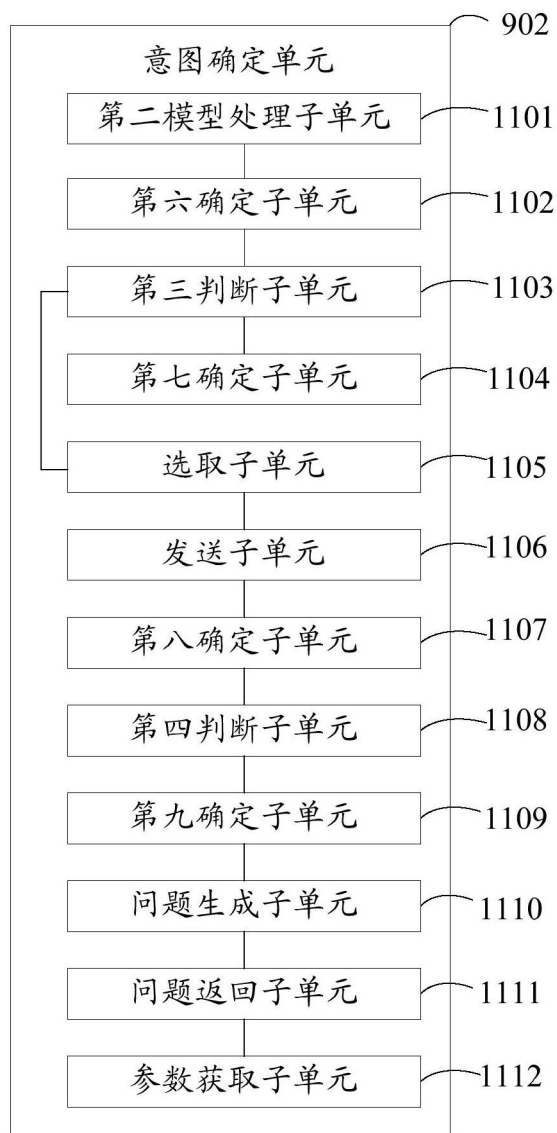


图11