



(12) 发明专利

(10) 授权公告号 CN 112183075 B

(45) 授权公告日 2023.06.30

(21) 申请号 202011080687.6

G06F 40/242 (2020.01)

(22) 申请日 2020.10.10

G06F 16/332 (2019.01)

(65) 同一申请的已公布的文献号

G06F 16/36 (2019.01)

申请公布号 CN 112183075 A

G06N 3/0442 (2023.01)

(43) 申请公布日 2021.01.05

(56) 对比文件

(73) 专利权人 网易(杭州)网络有限公司

CN 110175223 A, 2019.08.27

地址 310052 浙江省杭州市滨江区长河街  
道网商路599号4幢7层

CN 109918680 A, 2019.06.21

CN 105589848 A, 2016.05.18

(72) 发明人 张林箭 张聪 衣景龙 范长杰  
胡志鹏

CN 109727041 A, 2019.05.07

CN 111078844 A, 2020.04.28

CN 111427996 A, 2020.07.17

(74) 专利代理机构 北京超成律师事务所 11646  
专利代理师 裴素英

US 2019370327 A1, 2019.12.05

WO 2020177592 A1, 2020.09.10

审查员 喻阳

(51) Int. Cl.

G06F 40/247 (2020.01)

G06F 40/295 (2020.01)

权利要求书5页 说明书22页 附图7页

(54) 发明名称

一种对话中的槽提取方法、装置、电子设备  
及存储介质

(57) 摘要

本申请提供了一种对话中的槽提取方法、装置、电子设备及存储介质,该方法包括:在确定接收到的目标语料对应的目标意图后,基于目标语料和目标意图对应的第一槽位,确定目标意图对应的待填充槽位;每一个第一槽位对应至少一个第一槽值;基于目标意图对应的待填充槽位向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算第一回复语料分别与每一个第一槽值的目标相似度;根据第一回复语料分别与每一个第一槽值的目标相似度,从第一槽值中提取第一回复语料中匹配待填充槽位的第一目标槽值;这样,通过对话流程图来管理对话,提高了槽识别效率,通过相似度计算进行槽提取,动态拓展了NER模型的识别能力,提高了识别的准确度。

在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应至少一个第一槽值

S101

基于所述目标意图对应的待填充槽位向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度

S102

根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值

S103

1. 一种对话中的槽提取方法,其特征在于,所述槽提取方法包括:

在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应至少一个第一槽值;

基于所述目标意图对应的待填充槽向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度;

根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值;

所述基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位,包括:

对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,并确定所述第二槽值所属的第二槽位;根据所述目标意图对应的第一槽位以及所述第二槽值所属的第二槽位,确定所述目标意图对应的待填充槽位;

所述计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量;根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

2. 根据权利要求1所述的对话中的槽提取方法,其特征在于,所述提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,包括:

若所述第一回复语料中不包括特定类型的字符,则基于第一来自变换器的双向编码器表征量BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,并基于所述第三特征向量计算所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,并基于所述第四特征向量计算每一个第一槽值对应的第二特征向量;

若所述第一回复语料中包括特定类型的字符,则基于第一BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,基于所述第三特征向量,计算所述第一BERT模型下所述第一回复语料对应的第一特征向量;并基于第一通用语义表示模型ELMO模型提取所述第一回复语料中各个字符对应的第五特征向量,基于所述第五特征向量,计算第一ELMO模型下所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,基于所述第四特征向量,计算所述第一BERT模型下每一个第一槽值对应的第二特征向量;并基于第二ELMO模型提取每一个第一槽值中各个字符对应的第六特征向量,并基于所述第六特征向量,计算所述第二ELMO模型下每一个第一槽值对应的第二特征向量。

3. 根据权利要求2所述的对话中的槽提取方法,其特征在于,所述根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

若所述第一回复语料中不包括特定类型的字符,则计算第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向

量的相似度,得到所述第一回复语料分别与每一个第一槽值的目标相似度。

4. 根据权利要求2所述的对话中的槽提取方法,其特征在于,所述根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

若所述第一回复语料中包括特定类型的字符,则计算所述第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第一相似度;以及,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量分别与所述第二ELMO模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第二相似度;

基于BERT模型下的第一相似度和ELMO模型下的第二相似度,确定所述第一回复语料分别与每一个第一槽值的目标相似度。

5. 根据权利要求3或4所述的对话中的槽提取方法,其特征在于,计算所述第一回复语料对应的第一特征向量分别与每一个第一槽值对应的第二特征向量的相似度,包括:

针对每一个第一槽值,计算所述第一回复语料对应的第一特征向量与该槽值对应的第二特征向量之间的编辑距离、所述第一特征向量的分词与所述第二特征向量的分词之间的第三相似度、所述第一特征向量的字符与所述第二特征向量的字符之间的第四相似度;

对所述编辑距离、所述第三相似度以及所述第四相似度进行加权求和处理,得到所述第一回复语料与该第一槽值的相似度。

6. 根据权利要求1所述的对话中的槽提取方法,其特征在于,所述根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值,包括:

根据所述第一回复语料分别与每一个第一槽值的相似度,从所述第一槽值中选取与所述第一回复语料之间的相似度最大的候选槽值;

基于所述第一回复语料的长度,确定目标阈值;

若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值。

7. 根据权利要求6所述的对话中的槽提取方法,其特征在于,所述基于所述第一回复语料的长度,确定目标阈值,包括:

在所述第一回复语料的长度小于第一长度阈值时,确定预设的第一阈值为所述目标阈值;

在所述第一回复语料的长度大于第二长度阈值时,确定预设的第二阈值为所述目标阈值;其中,所述第二长度阈值大于等于所述第一长度阈值;所述第一阈值大于第二阈值。

8. 根据权利要求6所述的对话中的槽提取方法,其特征在于,通过如下方法确定预设的第二阈值:

获取所述第一回复语料中匹配任一第一槽值的第一分词的第三阈值,以及所述第一回复语料中除所述第一分词之外的第二分词与该任一第一槽值之间的第四阈值;

根据所述第三阈值和所述第四阈值,计算预设的第二阈值。

9. 根据权利要求6所述的对话中的槽提取方法,其特征在于,所述第一槽值包括规范槽值和所述规范槽值对应的同义词槽值;所述若所述候选槽值大于所述目标阈值,则提取所

述候选槽值为第一目标槽值,包括:

在所述候选槽值大于所述目标阈值时,若所述候选槽值为所述规范槽值,则提取所述候选槽值作为第一目标槽值;

若所述候选槽值为所述规范槽值对应的同义词槽值,则确定所述同义词槽值对应的规范槽值,并提取所述规范槽值作为所述第一目标槽值。

10. 根据权利要求1所述的对话中的槽提取方法,其特征在于,在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,所述槽提取方法还包括:

将所述第一目标槽值确定为最终槽值并填入匹配的待填充槽位中。

11. 根据权利要求1所述的对话中的槽提取方法,其特征在于,在接收到用户针对该第一反问的第一回复语料后,所述槽提取方法还包括:

对所述第一回复语料进行命名实体识别处理,得到所述第一回复语料中的第二目标槽值;

在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,所述槽提取方法还包括:

若所述第一目标槽值和所述第二目标槽值相同,则从所述第一目标槽值和所述第二目标槽值选取任一目标槽值作为最终槽值,并将该最终槽值填入匹配的待填充槽位中;

若所述第一目标槽值和所述第二目标槽值不同,则基于所述第一目标槽值和所述第二目标槽值分别对应的置信度,从所述第一目标槽值和所述第二目标槽值中选取最终槽值,并将所述最终槽值填入匹配的待填充槽位中。

12. 根据权利要求1所述的对话中的槽提取方法,其特征在于,所述对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

基于多模匹配算法AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值;

和/或,

将所述目标语料输入到预先训练好的命名实体识别NER模型中,提取所述目标语料中的第二槽值。

13. 根据权利要求12所述的对话中的槽提取方法,其特征在于,所述AC自动机中包括有限状态机和字典库;所述字典库中包括多个第三槽位且每个所述第三槽位下对应多个第三槽值;基于AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

基于所述有限状态机将所述目标语料与所述字典库进行匹配,提取所述目标语料中的第二槽值。

14. 根据权利要求12所述的对话中的槽提取方法,其特征在于,所述NER模型包括:特征提取器、循环神经网络LSTM和条件随机场算法CRF;将所述目标语料输入到预先训练好的NER模型中,提取所述目标语料中的第二槽值,包括:

将所述目标语料输入到所述特征提取器中,通过所述特征提取器提取所述目标语料中各个字符对应的第七特征向量,并基于所述第七特征向量,计算所述目标语料对应的第八特征向量;

将所述第八特征向量输入到LSTM中,并将LSTM的输出结果输入到CRF中,提取所述目标语料中的第二槽值。

15.根据权利要求12所述的对话中的槽提取方法,其特征在于,所述槽提取方法还包括:

若基于AC自动机得到的第二槽值与基于NER模型得到的第二槽值不同,则基于所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值分别对应的置信度,从所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值中选取最终第二槽值。

16.根据权利要求10或11所述的对话中的槽提取方法,其特征在于,所述槽提取方法还包括:

在检测到所述目标意图对应的第一槽位填入匹配的最终槽值后,基于所述目标意图以及所述目标意图在第一槽位下的各个最终槽值向用户进行第二反问;

在接收到用户针对该第二反问的第二回复语料后,对所述第二回复语料进行命名实体识别处理,得到识别结果;

若所述识别结果中包括所述第二回复语料对应第三槽值,则根据所述第三槽值从所述第一槽值中确定待更新槽位,并基于所述第三槽值更新所述待更新槽位。

17.一种对话中的槽提取装置,其特征在于,所述槽提取装置包括:

确定模块,用于在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应有至少一个第一槽值;

第一反问模块,用于基于所述目标意图对应的待填充槽向用户进行第一反问;

第一计算模块,用于在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度;

提取模块,用于根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值;

所述确定模块基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位,包括:

对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,并确定所述第二槽值所属的第二槽位;根据所述目标意图对应的第一槽位以及所述第二槽值所属的第二槽位,确定所述目标意图对应的待填充槽位;

所述第一计算模块计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量;根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

18.一种电子设备,其特征在于,包括:处理器、存储介质和总线,所述存储介质存储有所述处理器可执行的机器可读指令,当电子设备运行时,所述处理器与所述存储介质之间通过总线通信,所述处理器执行所述机器可读指令,以执行如权利要求1至16任一项所述的对话中的槽提取方法的步骤。

19. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器运行时执行如权利要求1至16任一项所述的对话中的槽提取方法的步骤。

## 一种对话中的槽提取方法、装置、电子设备及存储介质

### 技术领域

[0001] 本申请涉及游戏技术领域,具体而言,涉及一种对话中的槽提取方法、装置、电子设备及存储介质。

### 背景技术

[0002] 面向任务的对话系统旨在帮助用户完成某些任务(例如查找产品,预订住宿和餐馆),一般来说往往是“多轮”的,对话系统通过与用户进行多轮基于自然语言的对话来逐步收集与目标相关的信息,辅助用户成功获得某种服务。

[0003] 在任务型对话系统中,现有的任务型多轮对话的平台配置方案,往往需要上传很多的标注数据来训练模型,而这在实际的工业界场景中往往很难做到,一是数据量往往很难满足需求,二是标注成本高昂,并且,当需求及数据频繁变化时,需要根据新的需求利用新的数据,重新训练模型,费时费力,识别效率低。

### 发明内容

[0004] 有鉴于此,本申请的目的在于提供一种对话中的槽提取方法、装置、电子设备及存储介质,通过对话流程图来管理对话,提高了识别效率,通过相似度计算进行槽提取,动态拓展了NER模型的识别能力,提高了识别的准确度。

[0005] 第一方面,本申请实施例提供了一种对话中的槽提取方法,所述槽提取方法包括:

[0006] 在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应至少一个第一槽值;

[0007] 基于所述目标意图对应的待填充槽向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度;

[0008] 根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值。

[0009] 在一种可能的实施方式中,所述基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位,包括:

[0010] 对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,并确定所述第二槽值所属的第二槽位;

[0011] 根据所述目标意图对应的第一槽位以及所述第二槽值所属的第二槽位,确定所述目标意图对应的待填充槽位。

[0012] 在一种可能的实施方式中,所述计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0013] 提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量;

[0014] 根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

[0015] 在一种可能的实施方式中,所述提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,包括:

[0016] 若所述第一回复语料中不包括特定类型的字符,则基于第一来自变换器的双向编码器表征量BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,并基于所述第三特征向量计算所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,并基于所述第四特征向量计算每一个第一槽值对应的第二特征向量;

[0017] 若所述第一回复语料中包括特定类型的字符,则基于第一BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,基于所述第三特征向量,计算所述第一BERT模型下所述第一回复语料对应的第一特征向量;并基于第一通用语义表示模型ELMO模型提取所述第一回复语料中各个字符对应的第五特征向量,基于所述第五特征向量,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,基于所述第四特征向量,计算所述第一BERT模型下每一个第一槽值对应的第二特征向量;并基于第二ELMO模型提取每一个第一槽值中各个字符对应的第六特征向量,并基于所述第六特征向量,计算所述第二ELMO模型下每一个第一槽值对应的第二特征向量。

[0018] 在一种可能的实施方式中,所述根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0019] 若所述第一回复语料中不包括特定类型的字符,则计算第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的目标相似度。

[0020] 在一种可能的实施方式中,所述根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0021] 若所述第一回复语料中包括特定类型的字符,则计算所述第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第一相似度;以及,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量分别与所述第二ELMO模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第二相似度;

[0022] 基于BERT模型下的第一相似度和ELMO模型下的第二相似度,确定所述第一回复语料分别与每一个第一槽值的目标相似度。

[0023] 在一种可能的实施方式中,计算所述第一回复语料对应的第一特征向量分别与每一个第一槽值对应的第二特征向量的相似度,包括:

[0024] 针对每一个第一槽值,计算所述第一回复语料对应的第一特征向量与该槽值对应的第二特征向量之间的编辑距离、所述第一特征向量的分词与所述第二特征向量的分词之

间的第三相似度、所述第一特征向量的字符与所述第二特征向量的字符之间的第四相似度；

[0025] 对所述编辑距离、所述第三相似度以及所述第四相似度进行加权求和处理，得到所述第一回复语料与该第一槽值的相似度。

[0026] 在一种可能的实施方式中，所述根据所述第一回复语料分别与每一个第一槽值的目标相似度，从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值，包括：

[0027] 根据所述第一回复语料分别与每一个第一槽值的相似度，从所述第一槽值中选取与所述第一回复语料之间的相似度最大的候选槽值；

[0028] 基于所述第一回复语料的长度，确定目标阈值；

[0029] 若所述候选槽值大于所述目标阈值，则提取所述候选槽值为第一目标槽值。

[0030] 在一种可能的实施方式中，所述基于所述第一回复语料的长度，确定目标阈值，包括：

[0031] 在所述第一回复语料的长度小于第一长度阈值时，确定预设的第一阈值为所述目标阈值；

[0032] 在所述第一回复语料的长度大于第二长度阈值时，确定预设的第二阈值为所述目标阈值；其中，所述第二长度阈值大于等于所述第一长度阈值；所述第一阈值大于第二阈值。

[0033] 在一种可能的实施方式中，通过如下方法确定预设的第二阈值：

[0034] 获取所述第一回复语料中匹配任一第一槽值的第一分词的第三阈值，以及所述第一回复语料中除所述第一分词之外的第二分词与该任一第一槽值之间的第四阈值；

[0035] 根据所述第三阈值和所述第四阈值，计算预设的第二阈值。

[0036] 在一种可能的实施方式中，所述第一槽值包括规范槽值和所述规范槽值对应的同义词槽值；所述若所述候选槽值大于所述目标阈值，则提取所述候选槽值为第一目标槽值，包括：

[0037] 在所述候选槽值大于所述目标阈值时，若所述候选槽值为所述规范槽值，则提取所述候选槽值作为第一目标槽值；

[0038] 若所述候选槽值为所述规范槽值对应的同义词槽值，则确定所述同义词槽值对应的规范槽值，并提取所述规范槽值作为所述第一目标槽值。

[0039] 在一种可能的实施方式中，在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后，所述槽提取方法还包括：

[0040] 将所述第一目标槽值确定为最终槽值并填入匹配的待填充槽位中。

[0041] 在一种可能的实施方式中，在接收到用户针对该第一反问的第一回复语料后，所述槽提取方法还包括：

[0042] 对所述第一回复语料进行命名实体识别处理，得到所述第一回复语料中的第二目标槽值；

[0043] 在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后，所述槽提取方法还包括：

[0044] 若所述第一目标槽值和所述第二目标槽值相同，则从所述第一目标槽值和所述第

二目标槽值选取任一目标槽值作为最终槽值,并将该最终槽值填入匹配的待填充槽位中;

[0045] 若所述第一目标槽值和所述第二目标槽值不同,则基于所述第一目标槽值和所述第二目标槽值分别对应的置信度,从所述第一目标槽值和所述第二目标槽值中选取最终槽值,并将所述最终槽值填入匹配的待填充槽位中。

[0046] 在一种可能的实施方式中,所述对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0047] 基于多模匹配算法AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值;

[0048] 和/或,

[0049] 将所述目标语料输入到预先训练好的命名实体识别NER模型中,提取所述目标语料中的第二槽值。

[0050] 在一种可能的实施方式中,所述AC自动机中包括有限状态机和字典库;所述字典库中包括多个第三槽位且每个所述第三槽位下对应多个第三槽值;基于AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0051] 基于所述有限状态机将所述目标语料与所述字典库进行匹配,提取所述目标语料中的第二槽值。

[0052] 在一种可能的实施方式中,所述NER模型包括:特征提取器、循环神经网络LSTM和条件随机场算法CRF;将所述目标语料输入到预先训练好的NER模型中,提取所述目标语料中的第二槽值,包括:

[0053] 将所述目标语料输入到所述特征提取器中,通过所述特征提取器提取所述目标语料中各个字符对应的第七特征向量,并基于所述第七特征向量,计算所述目标语料对应的第八特征向量;

[0054] 将所述第八特征向量输入到LSTM中,并将LSTM的输出结果输入到CRF中,提取所述目标语料中的第二槽值。

[0055] 在一种可能的实施方式中,所述槽提取方法还包括:

[0056] 若基于AC自动机得到的第二槽值与基于NER模型得到的第二槽值不同,则基于所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值分别对应的置信度,从所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值中选取最终第二槽值。

[0057] 在一种可能的实施方式中,所述槽提取方法还包括:

[0058] 在检测到所述目标意图对应的第一槽位填入匹配的最终槽值后,基于所述目标意图以及所述目标意图在第一槽位下的各个最终槽值向用户进行第二反问;

[0059] 在接收到用户针对该第二反问的第二回复语料后,对所述第二回复语料进行命名实体识别处理,得到识别结果;

[0060] 若所述识别结果中包括所述第二回复语料对应第三槽值,则根据所述第三槽值从所述第一槽值中确定待更新槽位,并基于所述第三槽值更新所述待更新槽位。

[0061] 第二方面,本申请实施例还提供了一种对话中的槽提取装置,所述槽提取装置包括:

[0062] 确定模块,用于在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标

意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应至少有至少一个第一槽值;

[0063] 第一反问模块,用于基于所述目标意图对应的待填充槽向用户进行第一反问;

[0064] 第一计算模块,用于在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度;

[0065] 提取模块,用于根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值。

[0066] 第三方面,本申请实施例还提供了一种电子设备,包括:处理器、存储介质和总线,所述存储介质存储有所述处理器可执行的机器可读指令,当电子设备运行时,所述处理器与所述存储介质之间通过总线通信,所述处理器执行所述机器可读指令,以执行如第一方面任一项所述的对话中的槽提取方法的步骤。

[0067] 第四方面,本申请实施例还提供了一种计算机可读存储介质,所述计算机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运行时执行如第一方面任一项所述的对话中的槽提取方法的步骤。

[0068] 本申请实施例提供的一种对话中的槽提取方法及装置,所述槽提取方法包括:在确定接收到的目标语料对应的目标意图后,基于目标语料以及目标意图对应的第一槽位,确定目标意图对应的待填充槽位;其中,目标意图对应一对话流程图,该对话流程图中包括目标意图对应的第一槽位,且每一个第一槽位对应至少有至少一个第一槽值;基于目标意图对应的待填充槽向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算第一回复语料分别与每一个第一槽值的目标相似度;根据第一回复语料分别与每一个第一槽值的目标相似度,从第一槽值中提取第一回复语料中匹配待填充槽位的第一目标槽值。本申请中,通过对话流程图来管理对话,缓解了在真实的多轮对话的落地场景中人工标注数据缺少的问题,且在需求及数据频繁变化的场景下,无需更改代码,只需更新图结构和对应语料即可,缓解了工业界中需求及数据变化频繁的问题,提高了槽识别效率;并且,通过相似度计算的方法进行槽提取,动态拓展了命名实体识别NER模型的识别能力,且提高了槽识别的准确度。

## 附图说明

[0069] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0070] 图1a示出了本申请实施例提供的对话中的槽提取方法的流程图;

[0071] 图1b示出了不同阶段的对话中的槽提取方法的流程图;

[0072] 图2示出了本申请实施例提供的另一种对话中的槽提取方法的流程图;

[0073] 图3示出了本申请实施例提供的另一种对话中的槽提取方法的流程图;

[0074] 图4示出了本申请实施例提供的另一种对话中的槽提取方法的流程图;

[0075] 图5示出了本申请实施例提供的另一种对话中的槽提取方法的流程图;

[0076] 图6示出了本申请实施例提供的另一种对话中的槽提取方法的流程图;

- [0077] 图7示出了本申请实施例提供的另一种对话中的槽提取方法的流程图；
- [0078] 图8a示出了OA助手机器人的对话流程图；
- [0079] 图8b示出了每个槽位对应的规范槽值及其同义词表的示意图；
- [0080] 图9示出了本申请实施例提供的一种对话中的槽提取装置的结构示意图；
- [0081] 图10示出了本申请实施例提供的一种电子设备的结构示意图。

### 具体实施方式

[0082] 为使本申请实施例的目的、技术方案和优点更加清楚，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，应当理解，本申请中附图仅起到说明和描述的目的，并不用于限定本申请的保护范围。另外，应当理解，示意性的附图并未按实物比例绘制。本申请中使用的流程图示出了根据本申请的一些实施例实现的操作。应该理解，流程图的操作可以不按顺序实现，没有逻辑的上下文关系的步骤可以反转顺序或者同时实施。此外，本领域技术人员在本申请内容的指引下，可以向流程图添加一个或多个其他操作，也可以从流程图中移除一个或多个操作。

[0083] 另外，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。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围，而是仅仅表示本申请的选定实施例。基于本申请的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0084] 需要说明的是，本申请实施例中将会用到术语“包括”，用于指出其后所声明的特征的存在，但并不排除增加其它的特征。

[0085] 在现有的任务型多轮对话的平台配置方案，往往需要上传很多的标注数据来训练模型，而这在实际的工业界场景中往往很难做到，一是数据量往往很难满足需求，二是标注成本高昂，并且，当需求及数据频繁变化时，需要根据新的需求利用新的数据，重新训练模型，费时费力，识别效率低。并且，当用户只输入一个词或者很简短的文本时，现有的利用完整文本训练出的序列标注模型很难提取出其中的词槽，识别能力较差。

[0086] 基于此，本申请实施例提供了一种对话中的槽提取方法、装置、电子设备及存储介质，在通过对话流程图来管理对话，缓解了在真实的多轮对话的落地场景中人工标注数据缺少的问题，且在需求及数据频繁变化的场景下，无需更改代码，只需更新对话流程图的图结构和对应语料即可，缓解了工业界中需求及数据变化频繁的问题，提高了识别效率；并且，通过相似度计算的方法进行槽提取，动态拓展了命名实体识别命名实体识别 (Named Entity Recognition,NER) 模型的识别能力，且提高了识别的准确度。

[0087] 首先为了解决在真实的多轮对话的落地场景中人工标注数据缺少的问题，本申请实施例改变了业界流行的序列标注类的槽提取方法，改为改造传统NER的识别算法，并结合创新性的相似度匹配的算法来做槽提取任务。

[0088] 其次，为了解决需求及数据频繁变化的问题，本申请实施例采用了可拓展性的架构，利用自主设计对话流程图来管理整个对话过程，解析流程图可以获得意图信息、意图与槽的对应信息、槽值的同义词典等，当需求变化或数据变化时，本申请实施例无需更改代码，只需更新图结构和对应语料即可。当意图下有对应词槽时，本申请实施例要至少编写一

条相关语料来明确其槽值。

[0089] 为了充分的利用上下文信息让系统在用户输入非完整的但包括所需槽位的句子时也能提取出槽位,并在少样本的情况下做出更好地效果,本申请实施例使用流程图管理的方式来缩小槽的判断范围,同时使用多模匹配算法(Aho-Corasick automaton,AC)自动机辅助以长短期记忆人工神经网络(Long Short-Term Memory,LSTM)和条件随机场(Conditional Random Field,CRF)来判断槽值是否在输入的utterance中。此外,本申请实施例采用预训练模型改进了相似度匹配算法并融合进了传统算法中,利用BERT(创新性的用siamese架构进行了重新训练)和ELMO来提取语义信息,利用杰卡德Jaccard相似度、字符串相似度Levenshtein来提取表层的字符级信息。

[0090] 为了解决当用户输入过长文本时提取到的词义和语义信息与真正的槽信息产生偏移的问题,对此,本申请实施例创造性的提出了一个惩罚因子,用于惩罚过长长度对相似度的影响,利用此因子来调整阈值。

[0091] 为便于对本申请实施例进行理解,下面对本申请实施例提供的一种对话中的槽提取方法、装置、电子设备及存储介质进行详细介绍。

[0092] 如图1a所示,为本申请实施例提供的一种对话中的槽提取方法,可以应用于服务器,所述槽提取方法包括:

[0093] S101、在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应有至少一个第一槽值。

[0094] S102、基于所述目标意图对应的待填充槽向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

[0095] S103、根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值。

[0096] 本申请实施例提供了一种对话中的槽提取方法,在通过对话流程图来管理对话,缓解了在真实的多轮对话的落地场景中人工标注数据缺少的问题,且在需求及数据频繁变化的场景下,无需更改代码,只需更新对话流程图的图结构和对应语料即可,缓解了工业界中需求及数据变化频繁的问题,提高了识别效率;并且,通过相似度计算的方法进行槽提取,动态拓展了命名实体识别NER模型的识别能力,且提高了识别的准确度。

[0097] 下面分别对本申请实施例的上述示例性的各步骤分别进行说明:

[0098] S101、在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应有至少一个第一槽值。

[0099] 本申请实施例中,利用对话流程图的方式来管理对话流程,即每个对话任务对应一个对话流程图,每个对话流程图中包括一级意图(即与起始点连接的意图)和该一级意图对应的槽位,还可以包括一级意图对应的二级意图和该二级意图对应的槽位;其中,二级意图可以为多个。上述每张对话流程图中存储了有向图的结构,通过解析流程图可以获得

每个意图下所对应的槽位,每个槽位所对应的槽值,该槽值包括规范槽值和该规范槽值对应的同义词表,在本申请实施例中的NER算法和相似度算法中,若提取到了相应的词在同义词表中,都会将其对齐为所对应的槽值。

[0100] 本申请实施例中,当接收到目标语料后,如果该目标语料为第一轮语料,那么在识别到该目标语料对应的目标意图(即一级意图)后,预加载该目标意图对应的对话流程图,该对话流程图中包括该目标意图对应的第一槽位,该第一槽位可以为一个或者多个,且每一个第一槽位对应至少有一个第一槽值,之后,从上述对话流程图中获取该目标意图对应的第一槽位。如果该目标语料为第二轮语料,那么在识别到该目标语料对应的目标意图(即二级意图)后,从已加载好的该目标意图对应的对话流程图中,获取该目标意图对应的第一槽位。

[0101] 其中,在接收到目标语料后,还会对目标语料进行NER处理,提取该目标语料中的第二槽值,并确定该第二槽值所属的第二槽位;然后,根据目标意图对应的第一槽位以及第二槽值所属的第二槽位,确定目标意图对应的待填充槽位。

[0102] 比如,目标意图为订餐,订餐对应的第一槽位为三个,分别包括时间、地点和菜品;这里,第二槽位为从目标语料中已提取的第二槽值,比如,目标语料为“我要预定烤鱼”,第二槽值为烤鱼,对应的第二槽位为菜品;这样,目标意图对应的待填充槽位为:时间和地点。

[0103] S102、基于所述目标意图对应的待填充槽向用户进行第一反问,在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

[0104] 在本申请实施例设计的对话流程图中当需要对用户的一句话(即语料)进行槽提取时,分为以下几种情况:

[0105] 1)当确定了目标意图对应的待填充槽位(即获得此时要获取的槽位)后,便对用户发出反问,得到用户回答后,服务器调用的本申请实施例提供算法对用户的回答进行识别,以得到槽位信息。

[0106] 本步骤对应该种情况,在确定目标意图对应的待填充槽后,向用户进行第一反问,比如,为“请问是预定什么时间的烤鱼”或者“请问是预定今天的烤鱼还是明天的烤鱼”等,在接收到用户的回答(即第一回复语料)后,服务器基于相似度匹配算法从第一回复语料中提取匹配该待填充槽位的第一目标槽值。

[0107] 在基于相似度匹配算法从第一回复语料中提取匹配该待填充槽位的第一目标槽值时,首先计算第一回复语料分别与每一个第一槽值的目标相似度,具体包括:提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量;根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

[0108] 2)当所有槽位信息获得之后,在执行任务之前需要先向用户进行确认,得到用户的回答后,判断是否有该意图下别的槽值信息,若有,说明用户可能更改了部分槽值,则重新填槽。

[0109] 比如,填槽结果为“预定今天晚上一园区的烤鱼”,需要向用户确认“请问您是要预定今天晚上一园区的烤鱼吗”,基于用户的回答(第二回复语料),确认用户是否有槽值更改,如果有,基于更改后的槽值更新第一槽位中已填好的槽值。

[0110] 3) 若对话系统还没有进入填槽阶段,需要判断出用户的问题是何种意图,同时检测是否含有该意图下匹配第一槽位的第二槽值。

[0111] 对应步骤101中对“接收到的目标语料进行命名实体识别命名实体识别 (Named Entity Recognition,NER) 处理的过程”,在该过程中,除了要识别意图,也要识别匹配第一槽位的第二槽值,并进行填槽。

[0112] 4) 若不是上述三种情况,则需判断当前用户回答是否是一级意图,若是(满足一定阈值)则需进行意图跳出。

[0113] 该种情况下,当目标语料为第二轮语料时,需要基于当前所在对话流程图对应的二级意图识别模型对该目标语料进行意图识别,同时,还需要根据通用的一级识别模型对该目标语料进行意图识别,如果最终确定该目标语料对应的意图为一级意图,说明用户在说另外一件事,此时进行意图跳出,重新按照步骤101至步骤104的方案处理该目标语料。

[0114] 比如,当前处于订餐的对话流程图下,该流程图下的二级意图为需要停车;当识别目标语料对应的意图为查询银行,相应的,可以确定用户在说一件新事情,因此,跳出订餐的对话流程图,进入查询银行的新对话流程图。

[0115] 本申请实施例包括以下技术关键点:

[0116] 当服务器(也即系统)调用本申请实施例提供的算法模块时,通过以下字段判断当前对话所属的情况:

参数名	是否必选	数据类型	Example	说明
task_skill_target	否	String	"CF-701565169033019392"	流程图编号
slot_is_filling	否	String	"园区"	要填的槽
target_intent	否	String	"IN-701565365909454848"	要填的槽所属意图
nlu_core.cls	否	String	"CF-701565169033019392"	NLU 模块识别的意图
nlu_core.lv1_cls	否	String	"IN-697744800505397248"	NLU 模块识别的一级意图

[0119] 其中,task\_skill\_target是流程图id,用来标识不同的对话流程,slot\_is\_filling在不为空时,是当前要填的槽位,target\_intent是对话流程当前所在的目标意图,nlu\_core.cls(其中,“.”表示上下级关系)则是NLU模块对目标语料识别出的目标意图,该字段只在流程图尚未进入某个意图时生效,nlu\_core.lv1\_cls则是NLU模块在判断第二轮语料为一级意图时,供对话流程跳出当前对话流程图使用。

[0120] 如图1b所示,本申请实施例中包括以下流程:

[0121] 当slot\_is\_filling有值且target\_intent有值时,表明当前对话流程正处于上述第一种情况,即:当根据上文已经获得此时要获取的槽位(即目标意图对应的待填充槽位),对用户发出反问后得到用户回答(第一回复语料),此时,既用本申请实施例提供的NER的方法来对该目标意图所有的第一槽位进行识别,同时,创造性的提出利用相似度匹配的算法与该对话流程图中该第一槽位下所有的第一槽值(包括规范槽值及其同义词)做相似度匹配,得到相似度最高的作为捕获到的第一目标槽值;否则,当target\_intent有值时,代表对话流程正处于第二种情况,即:当获取目标意图对应的所有第一槽位对应的第一槽值之后,在执行目标意图指令(包括槽位和槽值)任务之前需要先向用户进行确认,得到用户的回答后,利用NER算法判断是否有该目标意图下别的槽值,若有,说明用户可能更改了部分第一槽位下的部分槽值,则更新第一槽位下的槽值并重新询问用户;否则当nlu\_core.cls\_confidence(代表NLU模块判断该句话是某种意图的可能性)有值且大于第一预设阈值(比如0.5),说明此时是第三种情况,即系统还没有进入填槽阶段,此时,NLU判断出用户的问话是某种意图,于是用NER算法来抽取相应的槽位信息,若用户在发问的同时便告知了部分槽位信息,那此时便可避免向用户重复发问的问题;否则当nlu\_core.lv1\_cls\_confidence(代表NLU模块判断该句话是某种意图的可能性)大于等于第二预设阈值(该第二预设阈值可以与第一预设阈值相同,比如0.5),说明是第四种情况,此时,说明NLU模块检测到了意图跳出,那么,便可以用NER算法来抽取相应意图的槽位信息,同样可以避免向用户重复发问的问题。

[0122] S103、根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值。

[0123] 本申请实施例中,在得到第一回复语料分别与每一个第一槽值的目标相似度以后,根据该相似度,从所述第一槽值中选取与所述第一回复语料之间的相似度最大的候选槽值;之后,基于所述第一回复语料的长度,确定目标阈值;若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值。

[0124] 其中,若第一回复语料较长,则目标阈值较小;若第一回复语料较短,则目标阈值较大。

[0125] 进一步的,本申请实施例提供的对话中的槽提取方法中,所述提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,包括以下两种情况:

[0126] 第一种情况:若所述第一回复语料中不包括特定类型的字符,则基于第一来自变换器的双向编码器表征量BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,并基于所述第三特征向量计算所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,并基于所述第四特征向量计算每一个第一槽值对应的第二特征向量。

[0127] 可选的,上述特定类型的字符可以为数字,比如,为阿拉伯数字,1、2、3等。基于第一回复语料中各个字符对应的第三特征向量,计算第一回复语料对应的第一特征向量,包括:计算各个字符对应的第三特征向量的和值,并计算该和值与第一回复语料中字符数量的比值,得到第一回复语料对应的第一特征向量。

[0128] 如果第一回复语料中不包括阿拉伯数字,则基于第一来自变换器的双向编码器表征量(Bidirectional Encoder Representations from Transformers,BERT)模型提取第

一回复语料对应的第一特征向量。其中,用BERT模型提取特征向量的方法为:调用BERT模型对第一回复语料进行inference(即预测),由于BERT的底层往往提取的是词法和句法信息,而高层向量则往往表征语义等信息,而第三层在判断一个词是否在某句话中的任务里表现最好,因此,取BERT模型第三层的第一回复语料中每个字符的表示向量累加,然后取平均作为整个句子的特征向量。同时,对所有第一槽值也做类似的处理,提取出每个第一槽值的第二特征向量。

[0129] 第二种情况:若所述第一回复语料中包括特定类型的字符,则基于第一BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,基于所述第三特征向量,计算所述第一BERT模型下所述第一回复语料对应的第一特征向量;并基于第一通用语义表示模型ELMO模型提取所述第一回复语料中各个字符对应的第五特征向量,基于所述第五特征向量,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,基于所述第四特征向量,计算所述第一BERT模型下每一个第一槽值对应的第二特征向量;并基于第二ELMO模型提取每一个第一槽值中各个字符对应的第六特征向量,并基于所述第六特征向量,计算所述第二ELMO模型下每一个第一槽值对应的第二特征向量。

[0130] 可选的,如果第一回复语料中包括阿拉伯数字,则首先基于第一来自变换器的双向编码器表征量(Bidirectional Encoder Representations from Transformers,BERT)模型提取第一回复语料对应的第一特征向量。其中,用BERT模型提取特征向量的方法为:调用BERT模型对第一回复语料进行inference(即预测处理),由于BERT的底层往往提取的是词法和句法信息,而高层向量则往往表征语义等信息,而第三层在判断一个词是否在某句话中的任务里表现最好,因此,取BERT模型第三层的第一回复语料中每个字符的表示向量累加,然后取平均作为整个句子的特征向量。同时,对所有第一槽值也做类似的处理,提取出每个第一槽值的第二特征向量。

[0131] 然后,用通用语义表示模型(Embeddings from Language Model,ELMO)模型提取特征向量的方法为:调用ELMO模型对句子(即第一回复语料)进行inference(即预测处理),取该ELMO模型三层的特征向量进行加权求和(其中,ELMO模型的每一层的权重由训练所得),获得第一回复语料中每个字符的第六特征向量,然后取平均值得到每个句子(即第一回复语料)的特征表示。

[0132] 进一步的,本申请实施例提供的对话中的槽提取方法中,所述根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括以下两种情况:

[0133] 第一、若所述第一回复语料中不包括特定类型的字符,则计算第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的目标相似度。

[0134] 第二、若所述第一回复语料中包括特定类型的字符,则计算所述第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第一相似度;以及,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量分别与所述第二ELMO模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与

每一个第一槽值的第二相似度;之后,基于BERT模型下的第一相似度和ELMO模型下的第二相似度,确定所述第一回复语料分别与每一个第一槽值的目标相似度。

[0135] 如图2所示,其中,任意两个特征向量,即第一特征向量和第二特征向量的相似度的计算包括:

[0136] S201、针对每一个第一槽值,计算所述第一回复语料对应的第一特征向量与该槽值对应的第二特征向量之间的编辑距离、所述第一特征向量的分词与所述第二特征向量的分词之间的第三相似度、所述第一特征向量的字符与所述第二特征向量的字符之间的第四相似度。

[0137] S202、对所述编辑距离、所述第三相似度以及所述第四相似度进行加权求和处理,得到所述第一回复语料与该第一槽值的相似度。

[0138] 结合上述第一回复语料中是否包括特定类型的字符计算相似度的方法和步骤201和步骤202的相似度计算方法,进行具体说明:

[0139] 本申请实施例中,创造性的将相似度匹配算法引入了槽提取算法中,并同时引入了预训练模型来捕捉语义信息。但根据研究当前的BERT等词级预训练模型的阿拉伯数字捕捉能力都有所欠缺。在所有嵌入中,字符级方法要比词级、字级方法具备更强大的识数能力,而ELMO在综合任务上直接学的字符级模型整体上是最强的,因此,本申请中创造性的提出加入ELMO模型分数来调节模型的判断的算法。但利用BERT(创新性的用siamese架构进行了重新训练)和ELMO来提取语义信息还会产生语义误差的问题,比如:我生病请假,正确的槽值应为“病假”,但此时预训练模型可能认为“病”与“丧”词义很接近,反而倾向于“丧假”,由此带来不好的用户体验。对此,本申请中通过引入Levenshtein、Jaccard等指标来纠正BERT语义偏差的问题,取得了更好的效果。以下先介绍如何将语义信息和字符级信息统一整合的相似度算法,然后分别介绍BERT和ELMO的特征向量提取的方法及集成的语义信息相似度匹配算法流程。

[0140] 相应的,将语义信息和字符级信息统一整合的相似度算法。其中,为了避免预训练语言模型产生语义误差的问题,将Levenshtein、Jaccard指标引入了相似度计算公式。

[0141] Utterance(比如为第一回复语料)与第*i*个槽值(含有同义词典,最终规范化为规范槽值)的相似度计算公式为:

[0142]  $Sim_i = a_0 \times Levenshtein(s_i, u) + a_1 \times Jaccard_{word}(s_i, u) + a_0 \times Jaccard_{char}(s_i, u) + a_0 \times similarity_{semantic}(s_i, u);$

[0143] 其中, $Sim_i$ 为utterance与第*i*个槽值的相似度。Levenshtein为编辑距离, $Jaccard_{word}$ 为分词之后的Jaccard相似度, $Jaccard_{char}$ 为字符级的Jaccard相似度。 $a_0, a_1, a_2, a_3$ 均为权重参数,可根据不同的数据特点进行调整。针对本申请是实力中数据集的特点,目前定为 $\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{4}{7}$ 。 $similarity_{semantic}(s_i, u)$ 为utterance与第*i*个槽值的词嵌入相似度。

[0144] 在上述第一回复语料中不包括特定类型的字符时,目标相似度的计算公式如下:

[0145]  $similarity_{semantic}(s_i, u) = Similarity_{(s_i, u)}^{bert};$

[0146] 在上述第一回复语料中不包括特定类型的字符时,目标相似度的计算公式如下:

[0147]  $similarity_{semantic}(s_i, u) = 0.5 \times Similarity_{(s_i, u)}^{elmo} + 0.5 \times Similarity_{(s_i, u)}^{bert}。$

[0148] 其中,相似度计算方法:计算第一特征向量和第二特征向量的余弦相似度,即:

$$[0149] \quad \text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

;

[0150] 其中,此处, $A_i, B_i$ 分别代表向量A和B的各分量;向量A表示第一特征向量,向量B表示第二特征向量; $\cos(\theta)$ 表示第一特征向量和第二特征向量的余弦相似度, $n$ 表示向量的个数。

[0151] 本申请实施例中,集成算法步骤如下:

[0152] (1) 判断utterance (即第一回复语料) 中是否含有数字,若含有数字跳转到下述(2),若不含有数字则调用BERT模型提取句子特征向量,对所有可能的槽值也提取特征向量,计算出utterance与所有范围内的词槽的值的余弦相似度。

[0153] (2) 本申请实施例中,利用BERT模型提取结果的同时也利用ELMO模型提取特征向量,并计算和每个槽值的相似度。同时,将每个槽值对应的两个模型计算所得的相似度先乘0.5然后相加,即:

$$[0154] \quad \text{Score}_i = 0.5 \times \text{Similarity}_i^{\text{elmo}} + 0.5 \times \text{Similarity}_i^{\text{bert}};$$

[0155]  $\text{Score}_i$ 代表第*i*个槽值的得分,最后返回一个列表。

[0156] 本申请实施例中,在计算出utterance与所有的合理的槽值的相似度之后,获取最大相似度的值,找到该值的原型(即在哪个槽值的同义词典中)将其填入槽值对中,调用阈值计算模块,计算阈值,将槽值对与阈值一并返回。

[0157] 进一步的,如图3所示,本申请实施例提供的对话中的槽提取方法中,所述根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值,包括:

[0158] S301、根据所述第一回复语料分别与每一个第一槽值的相似度,从所述第一槽值中选取与所述第一回复语料之间的相似度最大的候选槽值。

[0159] S302、基于所述第一回复语料的长度,确定目标阈值。

[0160] 本申请实施例中,在所述第一回复语料的长度小于第一长度阈值时,确定预设的第一阈值为所述目标阈值;在所述第一回复语料的长度大于第二长度阈值时,确定预设的第二阈值为所述目标阈值;其中,所述第二长度阈值大于等于所述第一长度阈值;所述第一长度阈值大于第二长度阈值。

[0161] 可选的,第一长度阈值和第二长度阈值相同,为第一槽值的长度。比如,当utterance (比如第一回复语料) 的长度小于等于slot (第一槽值) 的长度时,目标阈值便设置为一个较高的数值,统计分析所得为0.8;当utterance (比如第一回复语料) 的长度大于等于slot (第一槽值) 的长度时,目标阈值便设置为一个较小的数值,统计分析所得为0.6;

[0162] S303、若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值。

[0163] 结合步骤301至步骤303,本申请实施例中,当用长文本来对槽值等短文本进行相似度匹配时,提取到的词义和语义信息必然会与真正的槽信息产生偏移,那此时的置信度就不可与短文本进行统一度量,对此,本申请实施例中创造性的提出了一个惩罚因子,用于惩罚过长长度对相似度的影响,利用此因子来调整阈值。

[0164] 阈值计算公式如下：

$$[0165] \quad Confidence = \begin{cases} 0.8 & \text{if } l_u \leq l_s \\ 0.3 + \frac{0.5}{\frac{l_u}{l_s} + 10^{-8}} & \text{if } l_u > l_s \end{cases}$$

[0166] 进一步的,如图4所示,本申请实施例提供的对话中的槽提取方法中,通过如下方法确定预设的第二阈值:

[0167] S401、获取所述第一回复语料中匹配任一第一槽值的第一分词的第三阈值,以及所述第一回复语料中除所述第一分词之外的第二分词与该任一第一槽值之间的第四阈值;

[0168] S402、根据所述第三阈值和所述第四阈值,计算预设的第二阈值。

[0169] 结合步骤401至步骤402,当utterance的长度小于等于slot(槽值)的长度时,阈值便设置为一个较高的数值(统计分析所得)0.8,当utterance的长度大于slot(槽值)的长度时,认为utterance中匹配s的词的相似度仍然需要大于等于0.8,而其他的词与slot的平均

相似度为0.3,所以可得到此时的阈值计算公式  $\frac{0.8+0.3(\frac{l_u}{l_s}-1)}{\frac{l_u}{l_s}} \approx 0.3 + \frac{0.5}{\frac{l_u}{l_s} + 10^{-8}}$ 。

[0170] 进一步的,如图5所示,本申请实施例提供的对话中的槽提取方法,所述第一槽值包括规范槽值和所述规范槽值对应的同义词槽值;所述若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值,包括:

[0171] S501、在所述候选槽值大于所述目标阈值时,若所述候选槽值为所述规范槽值,则提取所述候选槽值作为第一目标槽值。

[0172] S502、若所述候选槽值为所述规范槽值对应的同义词槽值,则确定所述同义词槽值对应的规范槽值,并提取所述规范槽值作为所述第一目标槽值。

[0173] 结合步骤501和步骤502,比如,待填充槽位为餐时,相应的第一槽值中,规范槽值为午餐:同义词为午饭、中午、中餐、中饭。当检测候选槽值为“午餐”时,直接将“午餐”作为第一目标槽值。当检测候选槽值为“中午”时,根据“中午”确定“午餐”,并将“午餐”作为第一目标槽值。

[0174] 进一步的,本申请实施例中提供的对话中的槽提取方法,在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,所述槽提取方法还包括以下情况:

[0175] 第一,将所述第一目标槽值确定为最终槽值并填入匹配的待填充槽位中。

[0176] 比如,待填充槽位为餐时,基于相似度匹配算法确定“午餐”作为第一目标槽值,则将“午餐”填入“餐时”的槽位中。

[0177] 第二、在接收到用户针对该第一反问的第一回复语料后,所述槽提取方法还包括:

[0178] 对所述第一回复语料进行命名实体识别处理,得到所述第一回复语料中的第二目标槽值。

[0179] 本申请实施例,除了通过上述相似度匹配算法提取第一回复语料中的第一目标槽值,还可以通过基于AC自动机对第一回复语料进行命名实体识别处理,得到所述第一回复语料中的第二目标槽值;或者,将第一回复语料输入到预先训练好的NER模型中,得到所述第一回复语料中的第二目标槽值。或者,基于AC自动机和NER模型分别对第一回复语料进行命名实体识别处理,得到第一回复语料中的第二目标槽值。

[0180] 在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,所述槽提取方法还包括:

[0181] 若所述第一目标槽值和所述第二目标槽值相同,则从所述第一目标槽值和所述第二目标槽值选取任一目标槽值作为最终槽值,并将该最终槽值填入匹配的待填充槽位中。

[0182] 若所述第一目标槽值和所述第二目标槽值不同,则基于所述第一目标槽值和所述第二目标槽值分别对应的置信度,从所述第一目标槽值和所述第二目标槽值中选取最终槽值,并将所述最终槽值填入匹配的待填充槽位中。

[0183] 这里,当通过相似度匹配算法、AC自动机和NER模型得到的命名实体识别结果不同时,基于各自结果的置信度,选取最终槽值并将最终槽值填入匹配的待填充槽位中。

[0184] 进一步的,本申请实施例提供的对话中的槽提取方法,所述对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0185] 基于多模匹配算法AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值;

[0186] 和/或,

[0187] 将所述目标语料输入到预先训练好的命名实体识别NER模型中,提取所述目标语料中的第二槽值。

[0188] 本申请实施例中,针对目标语料,也要基于AC自动机和/或NER模型对目标语料进行命名实体识别处理。

[0189] 下面分别对基于AC自动机和/或NER模型对目标语料进行命名实体识别处理的过程进行说明:

[0190] 第一、所述AC自动机中包括有限状态机和字典库;所述字典库中包括多个第三槽位且每个所述第三槽位下对应多个第三槽值;基于AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0191] 基于所述有限状态机将所述目标语料与所述字典库进行匹配,提取所述目标语料中的第二槽值。

[0192] AC自动机用于在输入的一串字符串中匹配有限组“字典”中的子串。它与普通字符串匹配的不同点在于同时与所有字典串进行匹配。算法均摊情况下具有近似于线性的时间复杂度,约为字符串的长度加所有匹配的数量。

[0193] 该算法主要依靠构造一个有限状态机(比如,在一个trie树中添加失配指针)来实现。这些额外的失配指针允许在查找字符串失败时进行回退(例如设Trie树的单词cat匹配失败,但是在Trie树中存在另一个单词cart,失配指针就会指向前缀ca),转向某前缀的其他分支,免于重复匹配前缀,提高算法效率。

[0194] 第二、如图6所示,所述NER模型包括:特征提取器、循环神经网络LSTM和条件随机场算法CRF;将所述目标语料输入到预先训练好的NER模型中,提取所述目标语料中的第二槽值,包括:

[0195] S601、将所述目标语料输入到所述特征提取器中,通过所述特征提取器提取所述目标语料中各个字符对应的第七特征向量,并基于所述第七特征向量,计算所述目标语料对应的第八特征向量。

[0196] S602、将所述第八特征向量输入到LSTM中,并将LSTM的输出结果输入到CRF中,提

取所述目标语料中的第二槽值。

[0197] 结合步骤601至步骤602,简单的LSTM的优点是能够通过双向的设置学习到观测序列(输入的字)之间的依赖,在训练过程中,LSTM能够根据目标(比如识别实体)自动提取观测序列的特征,但是缺点是无法学习到状态序列(输出的标注)之间的关系,在命名实体识别任务中,标注之间是有一定的关系的,比如B类标注(表示某实体的开头)后面不会再接一个B类标注,所以LSTM在解决NER这类序列标注任务时,虽然可以省去很繁杂的特征工程,但是也存在无法学习到标注上下文的缺点。

[0198] 而CRF的优点就是能对隐含状态建模,学习状态序列的特点,但其需要手动提取序列特征。所以本申请实施例中在LSTM后面再加一层CRF,以获得两者的优点。

[0199] 本申请实施例中,采用AC自动机来检索在词典和词库中的槽值,采用在内部知识图谱数据集上训练出的LSTM+CRF算法来提取可能的实体(槽值)。

[0200] 本申请实施例会根据步骤103的解释中提到的四种情况根据不同的需要来灵活的调用相似度匹配算法和NER算法,然后输出相应的结果。其中,在反问填槽阶段(即第一种情况下)通过相似度匹配算法、AC自动机和LSTM+CRF三种方式进行槽值提取,在其他三种情况下,通过AC自动机和LSTM+CRF两种方式进行槽值提取。

[0201] 进一步的,本申请实施例提供的对话中的槽提取方法,所述槽提取方法还包括:

[0202] 若基于AC自动机得到的第二槽值与基于NER模型得到的第二槽值不同,则基于所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值分别对应的置信度,从所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值中选取最终第二槽值。

[0203] 进一步的,如图7所示,本申请实施例提供的对话中的槽提取方法,所述槽提取方法还包括:

[0204] S701、在检测到所述目标意图对应的第一槽位填入匹配的最终槽值后,基于所述目标意图以及所述目标意图在第一槽位下的各个最终槽值向用户进行第二反问;

[0205] S702、在接收到用户针对该第二反问的第二回复语料后,对所述第二回复语料进行命名实体识别处理,得到识别结果;

[0206] S703、若所述识别结果中包括所述第二回复语料对应第三槽值,则根据所述第三槽值从所述第一槽值中确定待更新槽位,并基于所述第三槽值更新所述待更新槽位。

[0207] 结合步骤701至步骤703,对应步骤103的解释中的第二种情况,比如,填槽结果为“预定今天晚上一园区的烤鱼”,需要向用户确认“请问您是要预定今天晚上一园区的烤鱼吗”,基于用户的回答(第二回复语料),确认用户是否有槽值更改,如果有,基于更改后的槽值更新第一槽位中已填好的槽值。

[0208] 为了更好地理解对话流程,本申请实施例基于具体实例对上述对话中的槽提取方法进行说明。

[0209] 例如,现在需要新建一个OA助手机器人,对话流程图定义如图8a所示。

[0210] 针对订餐意图进行举例。

[0211] 订餐意图下有三个槽:园区、餐时、菜谱。

[0212] 每个槽都有自己的规范槽值及其同义词表,如图8b所示。

[0213] 设置用户发言为Ut,robot(即助手机器人)回复为Rt,括号内为注释,描述了此时robot应该做何种判断,执行何种操作。当用户想要订餐时,发生如下对话:

- [0214] U0:你好(注:系统初始状态,需要反问用户意图);
- [0215] R0:你好,请问有什么可以帮您的吗?我可以帮你办理一些流程,比如考勤申请、补签卡、访客申请、预定餐饮、查询菜单等。我也能回答关于伏羲实验室的一些小问题(注:反问话术);
- [0216] U1:预订餐饮(注:属于步骤103的解释中的第三种情况。调用NLU模块,识别意图为:预订餐饮;然后调用槽提取模块,提取出的槽值未达到阈值,故视为没有槽值);
- [0217] R1:请问你要预订一园区还是二园区(注:由图8b所知,预订餐饮意图需要园区、餐时、菜谱共三个槽。所以对用户发问,来获取“预订餐饮”所必需的槽信息,先发问“园区”槽位);
- [0218] U2:1(注:属于步骤103的解释中的第一种情况。调用槽提取模块,因为文本太短,NER算法未识别出槽值,调相似度匹配算法按照上述描述,计算出匹配最高的槽值为“一园区”,且置信度大于阈值,将槽位“园区”填槽位“一园区”);
- [0219] R2:请问中餐还是晚餐(注:此时还缺少“餐时”和“菜谱”信息,反问逻辑发问“餐时”);
- [0220] U3:中餐烤鱼(注:属于步骤103的解释中的第一种情况。此时槽提取模块的NER算法不仅提取出了“餐时”,也提取出了“烤鱼”,都在该意图下的同义词典中);
- [0221] R3:请确认信息是否正确:预订一园区午餐的自制烤鱼。如果不正确请提供正确信息;
- [0222] U4:哎,算了,还是羊蝎子吧(注:属于步骤103的解释中的第二种情况。此时NER算法提取出了槽位“菜谱”);
- [0223] R4:请确认信息是否正确:预订一园区晚餐的羊蝎子。如果不正确请提供正确信息(注:对话系统更新槽值,重新反问);
- [0224] U5:正确(注:属于步骤103的解释中的第二种情况。此时NER算法和相似度匹配算法没有提取出槽位或者不达标值。);
- [0225] R4:已经为您预订了一园区中餐的羊蝎子。
- [0226] 至此,便完成了用户的一次订餐任务,且可判断出语料中不存在的槽值,良好的应对了长短文本的提取问题,并避免了对用户重复发问的缺陷。
- [0227] 通过上面的技术方案,解决了以下问题:
- [0228] 第一、本申请实施例提出的通过流程图来管理对话进程,可以帮助缩小槽的判断范围,一定程度上解决了到数据稀疏问题及OOV问题,缓解了在真实的多轮对话的落地场景中人工标注数据缺少的问题。
- [0229] 第二、本申请实施例中相似度匹配算法、AC自动机和NER模型(LSTM+CRF)的算法结合第一步中的流程图管理使得无需更改代码,只需更新图结构和对应语料即可。缓解了工业界中需求及数据变化频繁变化的问题。
- [0230] 第三、本申请实施例中创造性的提出了一个惩罚因子,用于惩罚过长长度对相似度的影响,解决当用户输入过长文本时提取到的词义和语义信息与真正的槽信息产生偏移的问题。
- [0231] 第四、使用AC自动机辅助以LSTM和CRF来判断槽值是否在输入的utterance中。利用BERT和ELMO来提取语义信息,利用Jaccard相似度、Levenshtein来提取表层的字符级信

息并组合计算相似度的。动态拓展了NER模型的识别能力,并克服了语义提取有偏差、数字识别能力不足等缺点。

[0232] 基于同一发明构思,本申请实施例中还提供了与对话中的槽提取方法对应的对话中的槽提取装置,由于本申请实施例中的装置解决问题的原理与本申请实施例上述对话中的槽提取方法相似,因此装置的实施可以参见方法的实施,重复之处不再赘述。

[0233] 参照图9所示,为本申请实施例提供的一种对话中的槽提取装置,所述槽提取装置包括:

[0234] 确定模块901,用于在确定接收到的目标语料对应的目标意图后,基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位;其中,所述目标意图对应一对话流程图,该对话流程图中包括所述目标意图对应的第一槽位,且每一个所述第一槽位对应至少一个第一槽值;

[0235] 第一反问模块902,用于基于所述目标意图对应的待填充槽向用户进行第一反问;

[0236] 第一计算模块903,用于在接收到用户针对该第一反问的第一回复语料后,计算所述第一回复语料分别与每一个第一槽值的目标相似度;

[0237] 提取模块904,用于根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值。

[0238] 在一种可能的实施方式中,所述确定模块901基于所述目标语料以及所述目标意图对应的第一槽位,确定所述目标意图对应的待填充槽位,包括:

[0239] 对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,并确定所述第二槽值所属的第二槽位;

[0240] 根据所述目标意图对应的第一槽位以及所述第二槽值所属的第二槽位,确定所述目标意图对应的待填充槽位。

[0241] 在一种可能的实施方式中,所述第一计算模块903计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0242] 提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量;

[0243] 根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度。

[0244] 在一种可能的实施方式中,所述第一计算模块903提取所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,包括:

[0245] 若所述第一回复语料中不包括特定类型的字符,则基于第一来自变换器的双向编码器表征量BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,并基于所述第三特征向量计算所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,并基于所述第四特征向量计算每一个第一槽值对应的第二特征向量;

[0246] 若所述第一回复语料中包括特定类型的字符,则基于第一BERT模型提取所述第一回复语料中各个字符对应的第三特征向量,基于所述第三特征向量,计算所述第一BERT模型下所述第一回复语料对应的第一特征向量;并基于第一通用语义表示模型ELMO模型提取所述第一回复语料中各个字符对应的第五特征向量,基于所述第五特征向量,计算所述第

一ELMO模型下所述第一回复语料对应的第一特征向量;以及,基于第二BERT模型提取每一个第一槽值中各个字符对应的第四特征向量,基于所述第四特征向量,计算所述第一BERT模型下每一个第一槽值对应的第二特征向量;并基于第二ELMO模型提取每一个第一槽值中各个字符对应的第六特征向量,并基于所述第六特征向量,计算所述第二ELMO模型下每一个第一槽值对应的第二特征向量。

[0247] 在一种可能的实施方式中,所述第一计算模块903根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0248] 若所述第一回复语料中不包括特定类型的字符,则计算第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的目标相似度。

[0249] 在一种可能的实施方式中,所述第一计算模块903根据所述第一回复语料对应的第一特征向量以及每一个第一槽值对应的第二特征向量,计算所述第一回复语料分别与每一个第一槽值的目标相似度,包括:

[0250] 若所述第一回复语料中包括特定类型的字符,则计算所述第一BERT模型下所述第一回复语料对应的第一特征向量分别与所述第二BERT模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第一相似度;以及,计算所述第一ELMO模型下所述第一回复语料对应的第一特征向量分别与所述第二ELMO模型下每一个第一槽值对应的第二特征向量的相似度,得到所述第一回复语料分别与每一个第一槽值的第二相似度;

[0251] 基于BERT模型下的第一相似度和ELMO模型下的第二相似度,确定所述第一回复语料分别与每一个第一槽值的目标相似度。

[0252] 在一种可能的实施方式中,所述第一计算模块903计算所述第一回复语料对应的第一特征向量分别与每一个第一槽值对应的第二特征向量的相似度,包括:

[0253] 针对每一个第一槽值,计算所述第一回复语料对应的第一特征向量与该槽值对应的第二特征向量之间的编辑距离、所述第一特征向量的分词与所述第二特征向量的分词之间的第三相似度、所述第一特征向量的字符与所述第二特征向量的字符之间的第四相似度;

[0254] 对所述编辑距离、所述第三相似度以及所述第四相似度进行加权求和处理,得到所述第一回复语料与该第一槽值的相似度。

[0255] 在一种可能的实施方式中,所述提取模块904根据所述第一回复语料分别与每一个第一槽值的目标相似度,从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值,包括:

[0256] 根据所述第一回复语料分别与每一个第一槽值的相似度,从所述第一槽值中选取与所述第一回复语料之间的相似度最大的候选槽值;

[0257] 基于所述第一回复语料的长度,确定目标阈值;

[0258] 若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值。

[0259] 在一种可能的实施方式中,所述提取模块904基于所述第一回复语料的长度,确定目标阈值,包括:

[0260] 在所述第一回复语料的长度小于第一长度阈值时,确定预设的第一阈值为所述目标阈值;

[0261] 在所述第一回复语料的长度大于第二长度阈值时,确定预设的第二阈值为所述目标阈值;其中,所述第二长度阈值大于等于所述第一长度阈值;所述第一阈值大于第二阈值。

[0262] 在一种可能的实施方式中,所述槽提取装置包括:

[0263] 获取模块,用于获取所述第一回复语料中匹配任一第一槽值的第一分词的第三阈值,以及所述第一回复语料中除所述第一分词之外的第二分词与该任一第一槽值之间的第四阈值;

[0264] 第二计算模块,用于根据所述第三阈值和所述第四阈值,计算预设的第二阈值。

[0265] 在一种可能的实施方式中,所述第一槽值包括规范槽值和所述规范槽值对应的同义词槽值;所述提取模块904若所述候选槽值大于所述目标阈值,则提取所述候选槽值为第一目标槽值,包括:

[0266] 在所述候选槽值大于所述目标阈值时,若所述候选槽值为所述规范槽值,则提取所述候选槽值作为第一目标槽值;

[0267] 若所述候选槽值为所述规范槽值对应的同义词槽值,则确定所述同义词槽值对应的规范槽值,并提取所述规范槽值作为所述第一目标槽值。

[0268] 在一种可能的实施方式中,所述槽提取装置还包括:

[0269] 第一填充模块,用于在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,将所述第一目标槽值确定为最终槽值并填入匹配的待填充槽位中。

[0270] 在一种可能的实施方式中,所述槽提取装置还包括:

[0271] 第一识别处理模块,用于在接收到用户针对该第一反问的第一回复语料后,对所述第一回复语料进行命名实体识别处理,得到所述第一回复语料中的第二目标槽值;

[0272] 第二填充模块,用于在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,若所述第一目标槽值和所述第二目标槽值相同,则从所述第一目标槽值和所述第二目标槽值选取任一目标槽值作为最终槽值,并将该最终槽值填入匹配的待填充槽位中;

[0273] 第三填充模块,用于在从所述第一槽值中提取所述第一回复语料中匹配所述待填充槽位的第一目标槽值之后,若所述第一目标槽值和所述第二目标槽值不同,则基于所述第一目标槽值和所述第二目标槽值分别对应的置信度,从所述第一目标槽值和所述第二目标槽值中选取最终槽值,并将所述最终槽值填入匹配的待填充槽位中。

[0274] 在一种可能的实施方式中,所述确定模块901对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0275] 基于多模匹配算法AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值;

[0276] 和/或,

[0277] 将所述目标语料输入到预先训练好的命名实体识别NER模型中,提取所述目标语料中的第二槽值。

[0278] 在一种可能的实施方式中,所述AC自动机中包括有限状态机和字典库;所述字典库中包括多个第三槽位且每个所述第三槽位下对应多个第三槽值;所述确定模块901基于AC自动机对所述目标语料进行命名实体识别处理,提取所述目标语料中的第二槽值,包括:

[0279] 基于所述有限状态机将所述目标语料与所述字典库进行匹配,提取所述目标语料中的第二槽值。

[0280] 在一种可能的实施方式中,所述NER模型包括:特征提取器、循环神经网络LSTM和条件随机场算法CRF;所述确定模块901将所述目标语料输入到预先训练好的NER模型中,提取所述目标语料中的第二槽值,包括:

[0281] 将所述目标语料输入到所述特征提取器中,通过所述特征提取器提取所述目标语料中各个字符对应的第七特征向量,并基于所述第七特征向量,计算所述目标语料对应的第八特征向量;

[0282] 将所述第八特征向量输入到LSTM中,并将LSTM的输出结果输入到CRF中,提取所述目标语料中的第二槽值。

[0283] 在一种可能的实施方式中,所述槽提取装置还包括:

[0284] 选取模块,用于若基于AC自动机得到的第二槽值与基于NER模型得到的第二槽值不同,则基于所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值分别对应的置信度,从所述AC自动机对应的第二槽值和所述NER模型对应的第二槽值中选取最终第二槽值。

[0285] 在一种可能的实施方式中,所述槽提取装置还包括:

[0286] 第二反问模块,用于在检测到所述目标意图对应的第一槽位填入匹配的最终槽值后,基于所述目标意图以及所述目标意图在第一槽位下的各个最终槽值向用户进行第二反问;

[0287] 第二识别处理模块,用于在接收到用户针对该第二反问的第二回复语料后,对所述第二回复语料进行命名实体识别处理,得到识别结果;

[0288] 更新模块,用于若所述识别结果中包括所述第二回复语料对应第三槽值,则根据所述第三槽值从所述第一槽值中确定待更新槽位,并基于所述第三槽值更新所述待更新槽位。

[0289] 本申请实施例提供的一种对话中的槽提取装置,通过对话流程图来管理对话,缓解了在真实的多轮对话的落地场景中人工标注数据缺少的问题,且在需求及数据频繁变化的场景下,无需更改代码,只需更新图结构和对应语料即可,缓解了工业界中需求及数据变化频繁的问题,提高了槽识别效率;并且,通过相似度计算的方法进行槽提取,动态拓展了命名实体识别NER模型的识别能力,且提高了槽识别的准确度。

[0290] 如图10所示,本申请实施例提供的一种电子设备1000,包括:处理器1001、存储器1002和总线,所述存储器1002存储有所述处理器1001可执行的机器可读指令,当电子设备运行时,所述处理器1001与所述存储器1002之间通过总线通信,所述处理器1001执行所述机器可读指令,以执行如上述对话中的槽提取方法的步骤。

[0291] 具体地,上述存储器1002和处理器1001能够为通用的存储器和处理器,这里不做具体限定,当处理器1001运行存储器1002存储的计算机程序时,能够执行上述对话中的槽提取方法。

[0292] 对应于上述对话中的槽提取方法,本申请实施例还提供了一种计算机可读存储介质,所述计算机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运行时执行上述对话中的槽提取方法的步骤。

[0293] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统 and 装置的具体工作过程,可以参考方法实施例中的对应过程,本申请中不再赘述。在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,又例如,多个模块或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口,装置或模块的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0294] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0295] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0296] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个处理器可执行的非易失的计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0297] 以上仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

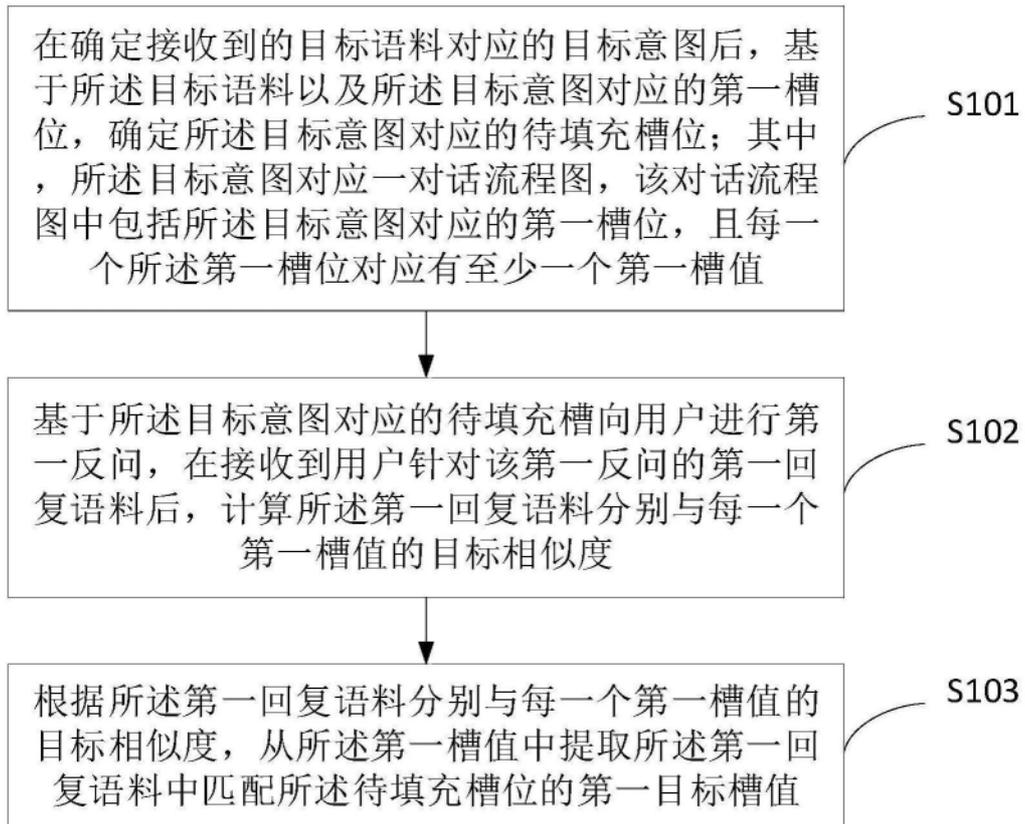


图1a

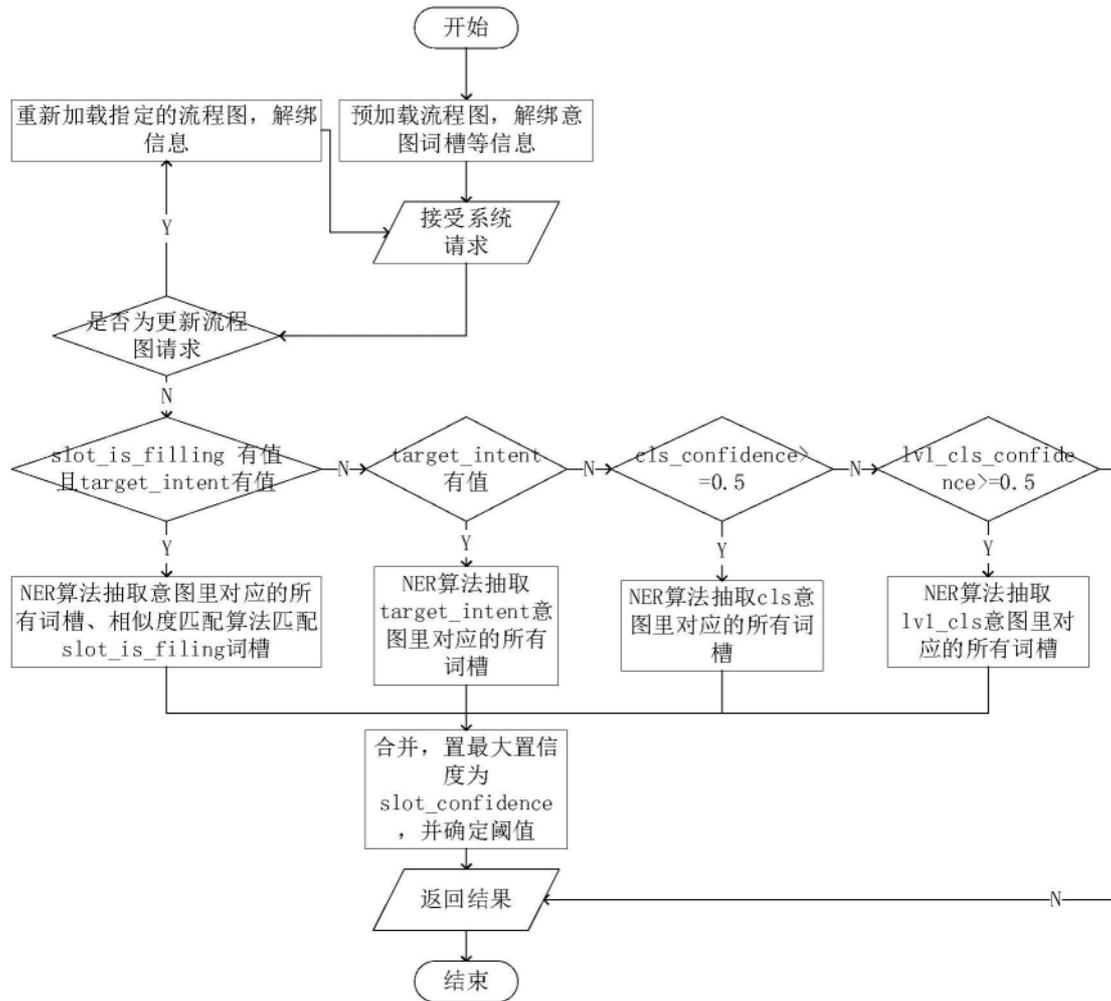


图1b

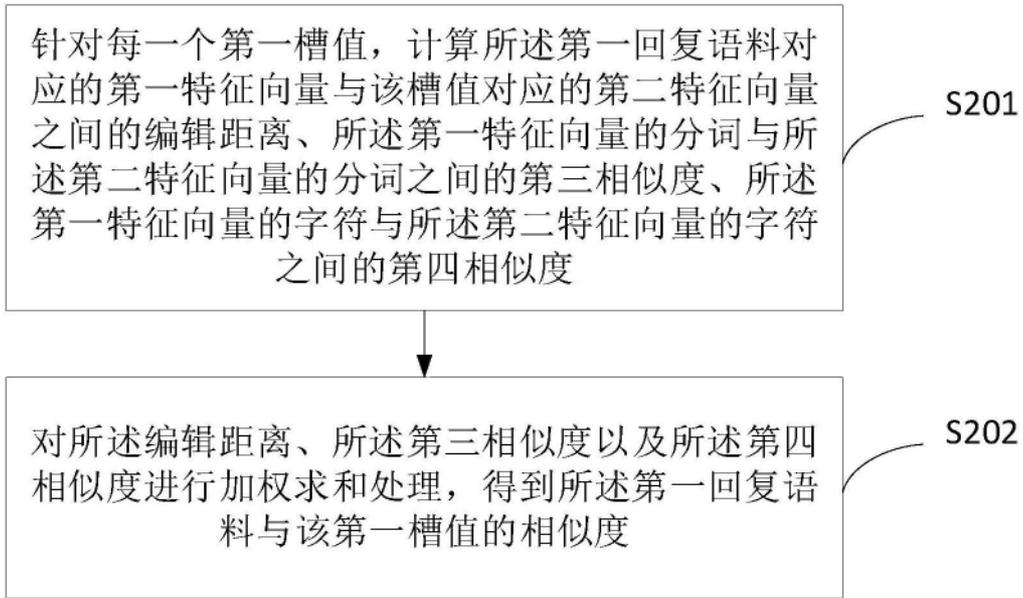


图2

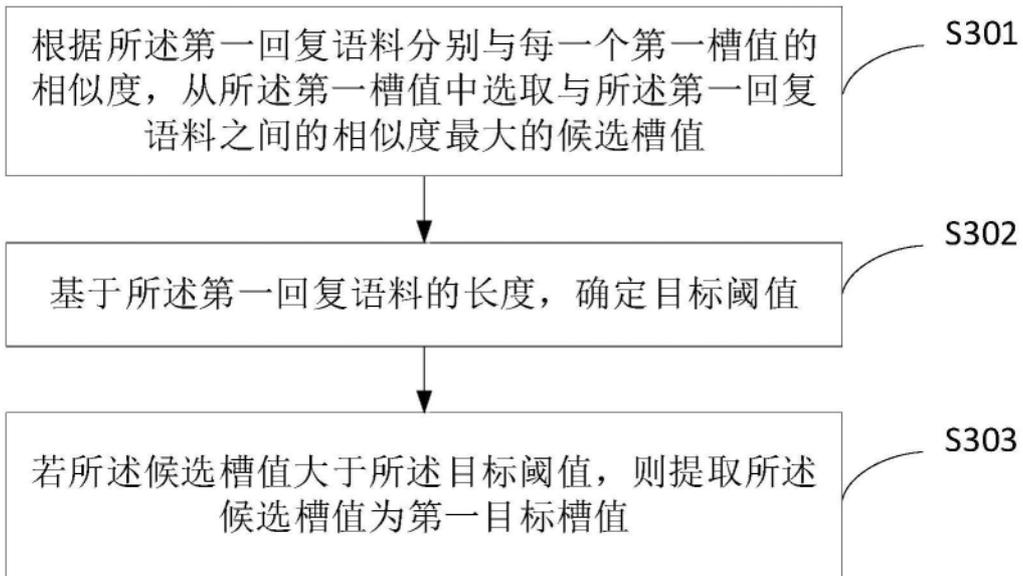


图3

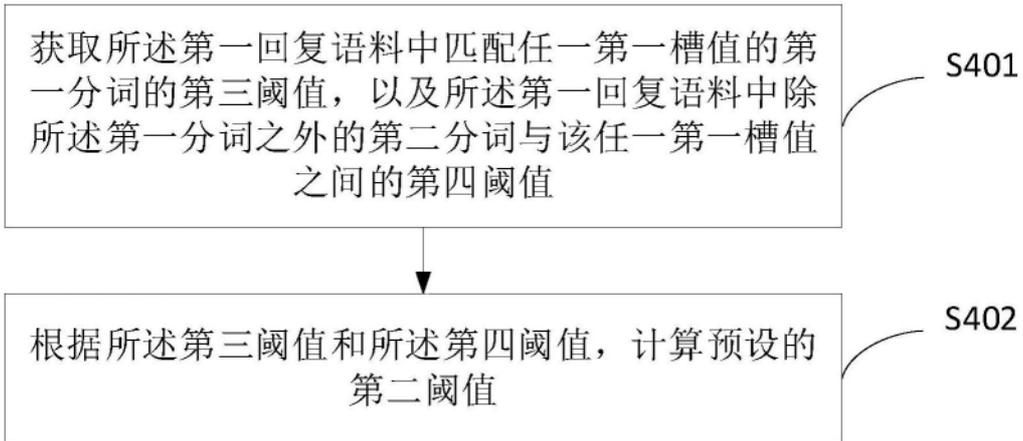


图4

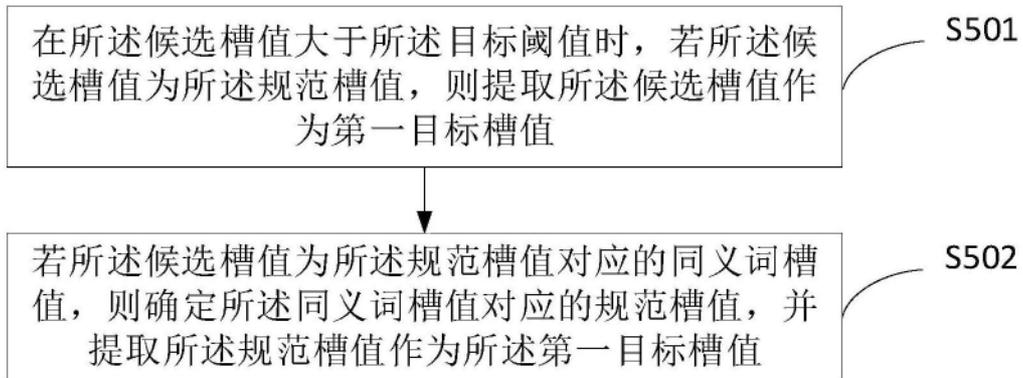


图5

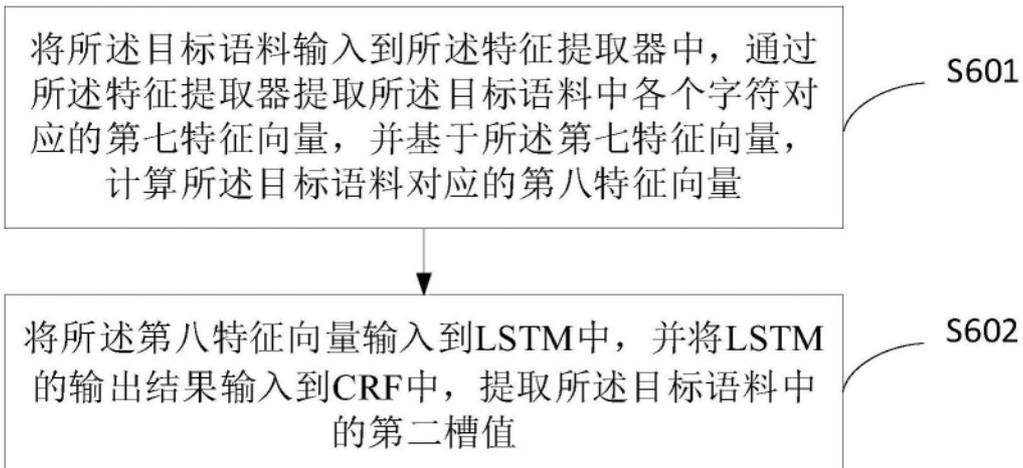


图6

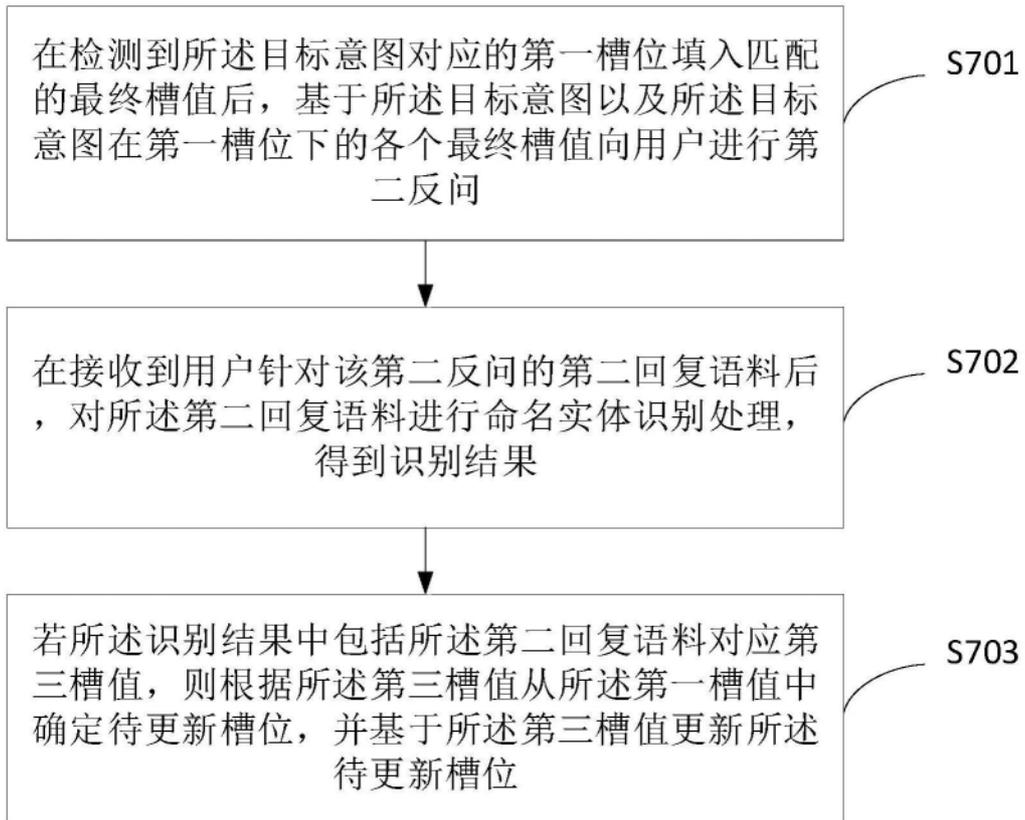


图7

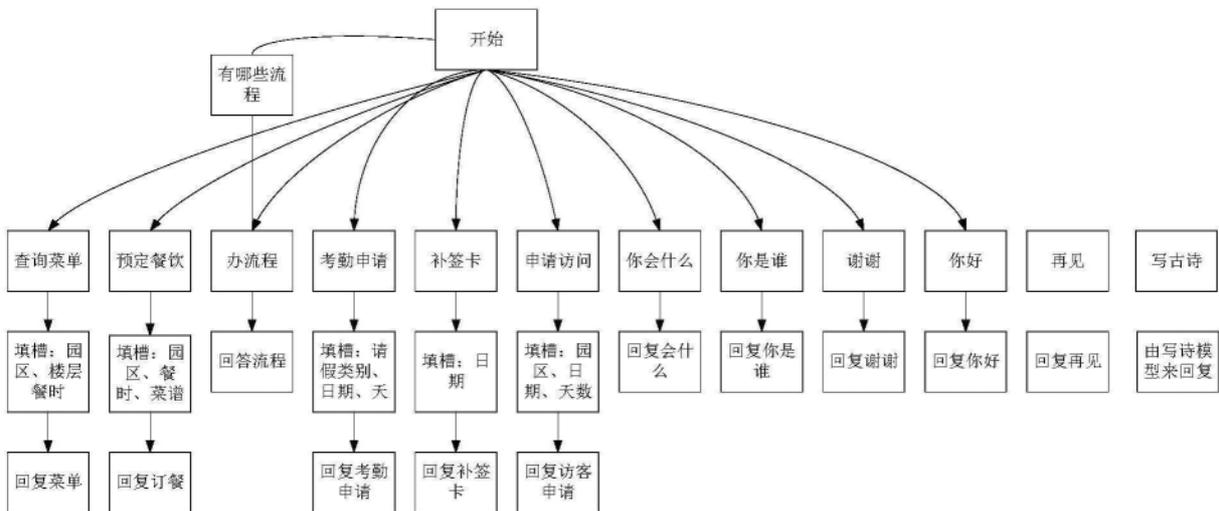


图8a

名称	同义词
园区	一园区: 一区 1区 1园区 二园区: 二区 2区 2园区
餐时	午餐: 午饭 中午 中餐 中饭 晚餐: 晚饭 晚上 早餐: 早饭 早上
菜谱	羊蝎子: 自制烤鱼: 烤鱼

图8b

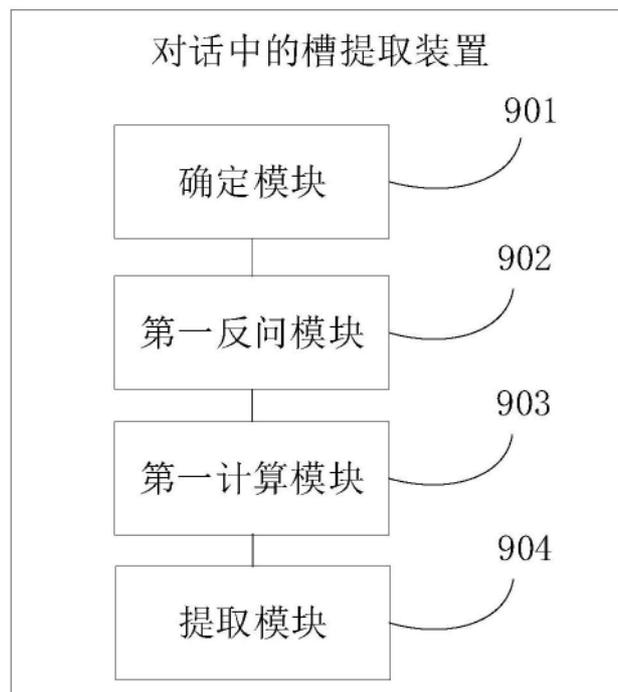


图9

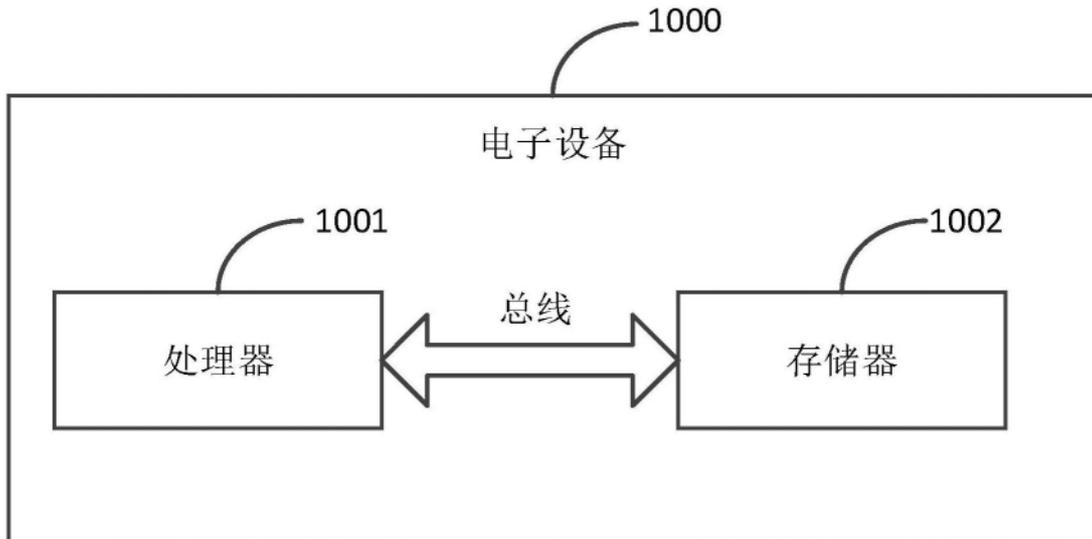


图10