



(12) 发明专利

(10) 授权公告号 CN 107291783 B

(45) 授权公告日 2021.04.30

(21) 申请号 201610227718.3

审查员 李萌

(22) 申请日 2016.04.12

(65) 同一申请的已公布的文献号
申请公布号 CN 107291783 A

(43) 申请公布日 2017.10.24

(73) 专利权人 芋头科技(杭州)有限公司
地址 310000 浙江省杭州市余杭区五常街
道西溪艺术集合村莲公荡路10号101
室

(72) 发明人 陈见耸 高鹏

(74) 专利代理机构 北京中原华和知识产权代理
有限责任公司 11019
代理人 寿宁

(51) Int. Cl.
G06F 16/9032 (2019.01)

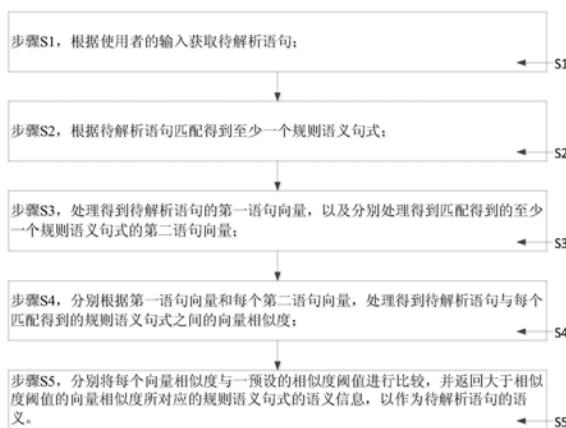
权利要求书2页 说明书8页 附图4页

(54) 发明名称

一种语义匹配方法及智能设备

(57) 摘要

本发明公开了一种语义匹配方法及智能设备;方法包括:根据使用者的输入获取待解析语句;根据待解析语句匹配得到至少一个规则语义句式;处理得到待解析语句的第一语句向量,以及分别处理得到匹配得到的至少一个规则语义句式的第二语句向量;分别根据第一语句向量和每个第二语句向量,处理得到待解析语句与每个匹配得到的规则语义句式之间的向量相似度;分别将每个向量相似度与一预设的相似度阈值进行比较,并返回大于相似度阈值的向量相似度所对应的规则语义句式的语义信息,以作为待解析语句的语义。上述技术方案的有益效果是:解决语义匹配需要人工撰写大量的语义句式规则,降低语义匹配操作的复杂度,并且大幅提升语义匹配的准确度。



1. 一种语义匹配方法,适用于智能设备;其特征在于,于所述智能设备中预设多个规则语义句式,还包括:

步骤S1,根据使用者的输入获取待解析语句;

步骤S2,根据所述待解析语句匹配得到至少一个所述规则语义句式;

步骤S3,处理得到所述待解析语句的第一语句向量,以及分别处理得到匹配得到的至少一个所述规则语义句式的第二语句向量;

步骤S4,分别根据所述第一语句向量和每个所述第二语句向量,处理得到所述待解析语句与每个匹配得到的所述规则语义句式之间的向量相似度;

步骤S5,分别将每个所述向量相似度与一预设的相似度阈值进行比较,并返回大于所述相似度阈值的所述向量相似度所对应的所述规则语义句式的语义信息,以作为所述待解析语句的语义;

在上述步骤之前,预设所述规则语义句式,并建立关联于所述规则语义句式的索引的方法,具体包括:

步骤A1,分别以使用者预先设置的不同类型的标签替代所述规则语义句式中所应类型的关键信息;

步骤A2,将每个所述标签视为一个字,并以每个字为一个索引单元,建立对于所述规则语义句式的索引。

2. 如权利要求1所述的语义匹配方法,其特征在于,所述步骤A2具体包括:

步骤A21,采用哈希倒排索引方式罗列在所有所述规则语义句式中所出现的所述索引单元;

步骤A22,在每个所述索引单元后分别链接关联于所述索引单元的每个所述规则语义句式的序号。

3. 如权利要求1所述的语义匹配方法,其特征在于,在执行所述步骤S2之前,首先以不同类型的所述标签替代所述待解析语句中所应类型的关键信息;

所述步骤S2具体包括:

步骤S21,将每个所述标签视为一个字,并以所述待解析语句中的每个字作为一个检索单元,依据所述规则语义句式的所述索引,分别检索得到匹配于所述待解析语句的至少一个所述规则语义句式;

步骤S22,分别处理得到每个检索得到的所述规则语义句式与所述待解析语句之间的匹配度;

步骤S23,分别将关联于每个检索得到的所述规则语义句式的所述匹配度与一预设的匹配度阈值进行比较,保留大于所述匹配度阈值的所述匹配度所对应的至少一个所述规则语义句式;

步骤S24,输出被保留的至少一个所述规则语义句式,以作为匹配得到的所述规则语义句式。

4. 如权利要求3所述的语义匹配方法,其特征在于,所述步骤S22中,依照下述公式计算得到所述匹配度:

$$S = (S_1 + S_2) / 2;$$

其中,S表示所述匹配度;

S_1 表示所述待解析语句与所述规则语义句式之间的匹配部分占所述待解析语句的比例；

S_2 表示所述待解析语句与所述规则语义句式之间的匹配部分占所述规则语义句式之间的比例。

5. 如权利要求1所述的语义匹配方法,其特征在於,预先训练形成一向量处理模型;

所述步骤S3中,计算得到所述第一语句向量的方法包括:

步骤S31a,将一条所述待解析语句进行分词处理;

步骤S32a,将经过所述分词处理的所述待解析语句中的每个词输入至所述向量处理模型中,以分别得到关联于每个词的词向量;

步骤S33a,根据所有所述词向量处理得到所述待解析语句的所述第一语句向量。

6. 如权利要求1所述的语义匹配方法,其特征在於,预先训练形成一向量处理模型;

所述步骤S3中,计算得到所述第二语句向量的方法包括:

步骤S31b,将一条所述规则语义句式进行分词处理;

步骤S32b,将经过所述分词处理的所述规则语义句式中的每个词输入至所述向量处理模型中,以分别得到关联于每个词的词向量;

步骤S33b,根据所有所述词向量处理得到所述规则语义句式的所述第二语句向量。

7. 如权利要求5或6所述的语义匹配方法,其特征在於,训练形成所述向量处理模型的方法包括:

步骤B1,获取预设的多个语料信息;

步骤B2,分别对每个所述语料信息进行分词处理;

步骤B3,分别将每个经过所述分词处理的所述语料信息作为所述向量处理模型的输入,根据所述向量处理模型输出对应不同词的所述词向量;

步骤B4,经过多个所述语料信息的训练,最终训练形成所述向量处理模型;

每个所述语料信息中包括:

一条语句内容;或者

一段语句内容;或者

多段语句内容。

8. 如权利要求1所述的语义匹配方法,其特征在於,所述步骤S4中,处理得到所述向量相似度的方法包括:

采用余弦相似度计算方法直接度量得到所述向量相似度;

或者

首先计算得到所述第一语句向量与对应的所述第二语句向量之间的向量距离,随后将所述向量距离转换为对应的所述向量相似度。

9. 一种智能设备,其特征在於,采用如权利要求1-8之一所述的语义匹配方法。

一种语义匹配方法及智能设备

技术领域

[0001] 本发明涉及语义分析技术领域,尤其涉及一种语义匹配方法及智能设备。

背景技术

[0002] 随着智能设备的普及,人与智能设备之间如何进行更直接友好的信息交互称为一个比较重要的问题。现阶段对于人与智能设备之间信息交互的方式大体可以包括:直接通过输入设备(例如键盘或者鼠标)的方式,通过识别使用者的手势动作进行信息交互的方式,以及通过识别使用者的语音信息进行信息交互的方法等。在实践中,由于自然语言(即口头语言)对于使用者具有天然的便捷性和友好性,因此基于自然语言的语义匹配和识别的信息交互方式自然需要重点发展,以期待给使用者带来较佳的使用体验。

[0003] 但是现有技术中,智能设备对自然语言进行语义分析从而支持实现人机交互的前提是需要手动输入大量的语句规则来支持语义匹配的过程,这会给使用者或者开发者带来极大的麻烦,因此降低了语义分析的效率;并且,通常的语义分析方法是将语句规则与待判断的语句进行一一对应的匹配,匹配到则返回该语句规则的语义,匹配不到返回匹配失败的结果,因此语义分析的准确性依赖于使用者或开发者手动输入的语义规则的数量,即语义规则数据库的规模大小,由于使用者或开发者手动输入的语义规则非常有限,因此通常导致语义分析的结果并不准确,从而影响语音交互方式的使用体验。

发明内容

[0004] 根据现有技术中存在的上述问题,现提供一种语义匹配方法及智能设备的技术方案,旨在解决传统语义匹配需要事先人工撰写大量的语义句式规则,降低语义匹配操作的复杂度,并且大幅提升语义匹配的准确度。

[0005] 上述技术方案具体包括:

[0006] 一种语义匹配方法,适用于智能设备;其中,于所述智能设备中预设多个规则语义句式,还包括:

[0007] 步骤S1,根据使用者的输入获取待解析语句;

[0008] 步骤S2,根据所述待解析语句匹配得到至少一个所述规则语义句式;

[0009] 步骤S3,处理得到所述待解析语句的第一语句向量,以及分别处理得到匹配得到的至少一个所述规则语义句式的第二语句向量;

[0010] 步骤S4,分别根据所述第一语句向量和每个所述第二语句向量,处理得到所述待解析语句与每个匹配得到的所述规则语义句式之间的向量相似度;

[0011] 步骤S5,分别将每个所述向量相似度与一预设的相似度阈值进行比较,并返回大于所述相似度阈值的所述向量相似度所对应的所述规则语义句式的语义信息,以作为所述待解析语句的语义。

[0012] 优选的,该语义匹配方法,其中,预设所述规则语义句式,并建立关联于所述规则语义句式的索引的方法包括:

[0013] 步骤A1,分别以使用者预先设置的不同类型的标签替代所述规则语义句式中所相应类型的键信息;

[0014] 步骤A2,将每个所述标签视为一个字,并以每个字为一个索引单元,建立对于所述规则语义句式的索引。

[0015] 优选的,该语义匹配方法,其中,所述步骤A2具体包括:

[0016] 步骤A21,采用哈希倒排索引方式罗列在所有所述规则语义句式中所出现的所述索引单元;

[0017] 步骤A22,在每个所述索引单元后分别链接关联于所述索引单元的每个所述规则语义句式的序号。

[0018] 优选的,该语义匹配方法,其中,在执行所述步骤S2之前,首先以不同类型的所述标签替代所述待解析语句中所相应类型的键信息;

[0019] 所述步骤S2具体包括:

[0020] 步骤S21,将每个所述标签视为一个字,并以所述待解析语句中的每个字作为一个检索单元,依据所述规则语义句式的所述索引,分别检索得到匹配于所述待解析语句的至少一个所述规则语义句式;

[0021] 步骤S22,分别处理得到每个检索得到的所述规则语义句式与所述待解析语句之间的匹配度;

[0022] 步骤S23,分别将关联于每个检索得到的所述规则语义句式的所述匹配度与一预设的匹配度阈值进行比较,保留大于所述匹配度阈值的所述匹配度所对应的至少一个所述规则语义句式;

[0023] 步骤S24,输出被保留的至少一个所述规则语义句式,以作为匹配得到的所述规则语义句式。

[0024] 优选的,该语义匹配方法,其中,所述步骤S22中,依照下述公式计算得到所述匹配度:

$$[0025] \quad S = (S_1 + S_2) / 2;$$

[0026] 其中,S表示所述匹配度;

[0027] S_1 表示所述待解析语句与所述规则语义句式之间的匹配部分占所述待解析语句的比例;

[0028] S_2 表示所述待解析语句与所述规则语义句式之间的匹配部分占所述规则语义句式之间的比例。

[0029] 优选的,该语义匹配方法,其中,预先训练形成一向量处理模型;

[0030] 所述步骤S3中,计算得到所述第一语句向量的方法包括:

[0031] 步骤S31a,将一条所述待解析语句进行分词处理;

[0032] 步骤S32a,将经过所述分词处理的所述待解析语句中的每个词输入至所述向量处理模型中,以分别得到关联于每个词的词向量;

[0033] 步骤S33a,根据所有所述词向量处理得到所述待解析语句的所述第一语句向量。

[0034] 优选的,该语义匹配方法,其中,预先训练形成一向量处理模型;

[0035] 所述步骤S3中,计算得到所述第二语句向量的方法包括:

[0036] 步骤S31b,将一条所述规则语义句式进行分词处理;

- [0037] 步骤S32b,将经过所述分词处理的所述规则语义句式中的每个词输入至所述向量处理模型中,以分别得到关联于每个词的词向量;
- [0038] 步骤S33b,根据所有所述词向量处理得到所述规则语义句式的所述第二语句向量。
- [0039] 优选的,该语义匹配方法,其中,训练形成所述向量处理模型的方法包括:
- [0040] 步骤B1,获取预设的多个语料信息;
- [0041] 步骤B2,分别对每个所述语料信息进行分词处理;
- [0042] 步骤B3,分别将每个经过所述分词处理的所述语料信息作为所述向量处理模型的输入,根据所述向量处理模型输出对应不同词的所述词向量;
- [0043] 步骤B4,经过多个所述语料信息的训练,最终训练形成所述向量处理模型;
- [0044] 每个所述语料信息中包括:
- [0045] 一条语句内容;或者
- [0046] 一段语句内容;或者
- [0047] 多段语句内容。
- [0048] 优选的,该语义匹配方法,其中,所述步骤S4中,处理得到所述向量相似度的方法包括:
- [0049] 采用余弦相似度计算方法直接度量得到所述向量相似度;
- [0050] 或者
- [0051] 首先计算得到所述第一语句向量与对应的所述第二语句向量之间的向量距离,随后将所述向量距离转换为对应的所述向量相似度。
- [0052] 一种智能设备,其中,采用上述的语义匹配方法。
- [0053] 上述技术方案的有益效果是:提供一种语义匹配方法,能够解决传统语义匹配需要事先人工撰写大量的语义句式规则,降低语义匹配操作的复杂度,并且大幅提升语义匹配的准确度。

附图说明

- [0054] 图1是本发明的较佳的实施例中,一种语义匹配方法的总体流程示意图;
- [0055] 图2是本发明的较佳的实施例中,建立关联于规则语义句式的索引的流程示意图;
- [0056] 图3是本发明的较佳的实施例中,以每个字为索引单元建立索引的流程示意图;
- [0057] 图4是本发明的较佳的实施例中,根据待解析语句匹配得到至少一个规则语义句式的流程示意图;
- [0058] 图5是本发明的较佳的实施例中,计算得到第一语句向量的流程示意图;
- [0059] 图6是本发明的较佳的实施例中,计算得到第二语句向量的流程示意图;
- [0060] 图7是本发明的较佳的实施例中,训练形成向量处理模型的流程示意图。

具体实施方式

- [0061] 下面将结合本发明实施例中的附图,对本发明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发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动的前提下所获得的所有其

他实施例,都属于本发明保护的范围。

[0062] 需要说明的是,在不冲突的情况下,本发明中的实施例及实施例中的特征可以相互组合。

[0063] 下面结合附图和具体实施例对本发明作进一步说明,但不作为本发明的限定。

[0064] 本发明的较佳的实施例中,基于现有技术中存在的上述问题,现提供一种语义分配方法,适用于智能设备,例如适用于移动终端,或者适用于智能机器人等其他智能设备。该方法中,于上述智能设备中预设多个规则语义句式。每个规则语义句式均具有相同的预设格式。该预设格式在下文中会详述。

[0065] 则本发明的较佳的实施例中,如图1所示,上述语义分配方法具体包括:

[0066] 步骤S1,根据使用者的输入获取待解析语句;

[0067] 步骤S2,根据待解析语句匹配得到至少一个规则语义句式;

[0068] 步骤S3,处理得到待解析语句的第一语句向量,以及分别处理得到匹配得到的至少一个规则语义句式的第二语句向量;

[0069] 步骤S4,分别根据第一语句向量和每个第二语句向量,处理得到待解析语句与每个匹配得到的规则语义句式之间的向量相似度;

[0070] 步骤S5,分别将每个向量相似度与一预设的相似度阈值进行比较,并返回大于相似度阈值的向量相似度所对应的规则语义句式的语义信息,以作为待解析语句的语义。

[0071] 在一个具体实施例中,首先根据使用者的输入获取待解析语句。使用者的输入方式可以为通过设置在智能设备上的拾音器(例如麦克风)输入一句自然语言,则智能设备会将使用者输入的自然语言作为待解析语句。

[0072] 在该实施例中,在得到上述待解析语句之后,根据该待解析语句匹配得到至少一个规则语义句式。该匹配规则大致为:将待解析语句转换为上述预设格式,然后将经过转换的待解析语句分别与上述规则语义句式进行匹配,得到相匹配的至少一个规则语义句式。上述过程在下文中会详述。

[0073] 在该实施例中,匹配得到上述至少一个规则语义句式之后,分别处理得到待解析语句的第一语句向量,以及处理得到每个匹配得到的规则语义句式的第二语句向量,并且分别计算得到第一语句向量分别与每个第二语句向量之间的向量的相似度,将这个相似度作为该规则语义句式相对于待解析语句的相似度。

[0074] 最后,根据该相似度,确定最终匹配该待解析语句的规则语义句式,并将该规则语义句式的语义信息作为该待解析语句的语义信息,从而可以进行后续的语音交互操作。

[0075] 本发明的较佳的实施例中,根据向量相似度没有找到相匹配的规则语义句式(所有规则语义句式的向量相似度均不高于上述相似度阈值),则直接返回交互失败的提示信息。

[0076] 本发明的较佳的实施例中,在上述步骤之前预先设置上述规则语义句式,并且建立关联于规则语义句式的索引的方法如图2所示,具体包括:

[0077] 步骤A1,分别以使用者预先设置的不同类型的标签替代规则语义句式相应类型的键信息;

[0078] 步骤A2,将每个标签视为一个字,并以每个字为一个索引单元,建立对于规则语义句式的索引。

[0079] 换言之,本发明的较佳的实施例中,上述预设格式即为以规则语义句式中的每个字作为一个索引单元形成的语句格式。具体地,使用者首先预设多个不同类型的标签,然后以不同类型的标签替代规则语义句式中的相应的内容,最后将每个标签视为一个字,并且以一个字作为一个索引单元建立对于规则语义句式的索引。

[0080] 例如:对于飞机票、火车票、汽车票等类似的领域,可以设定一个标签名为“起始地”,设定另一个标签名为“目的地”,以及设定一个标签名为“出发时间”。

[0081] 则对于一个规则语义句式:某时某刻(出发时间)从起始地到目的地,则可以转换成预设格式的规则语义句式,即出发时间+从+起始地+到+目的地。

[0082] 进一步地,本发明的较佳的实施例中,如图3所示,上述步骤A2具体包括:

[0083] 步骤A21,采用哈希倒排索引方式罗列在所有规则语义句式出现的索引单元;

[0084] 步骤A22,在每个索引单元后分别链接关联于索引单元的每个规则语义句式的序号。

[0085] 具体地,本发明的较佳的实施例中,采用哈希倒排索引方式将出现在所有规则语义句式中的所有索引单元罗列出来,并在每个索引单元后链接包括该索引单元的每个规则语义句式的序号,从而构成一个完整的规则语义句式的索引目录。

[0086] 则在实际检索匹配的过程中,可以根据待解析语句中包括的索引单元,直接根据索引目录找到相匹配的所有规则语义句式。

[0087] 本发明的较佳的实施例中,在执行步骤S2之前,首先以不同类型的标签替代待解析语句中相应类型的关键信息;

[0088] 则上述步骤S2具体如图4所示,包括:

[0089] 步骤S21,将每个标签视为一个字,并以待解析语句中的每个字作为一个检索单元,依据规则语义句式的索引,分别检索得到匹配于待解析语句的至少一个规则语义句式;

[0090] 步骤S22,分别处理得到每个检索得到的规则语义句式与待解析语句之间的匹配度;

[0091] 步骤S23,分别将关联于每个检索得到的规则语义句式的匹配度与一预设的匹配度阈值进行比较,保留大于匹配度阈值的匹配度所对应的至少一个规则语义句式;

[0092] 步骤S24,输出被保留的至少一个规则语义句式,以作为匹配得到的规则语义句式。

[0093] 具体地,本发明的较佳的实施例中,为了便于待解析语句和规则语义句式进行匹配,在匹配之前首先同样需要将待解析语句转换成上述预设格式,即:

[0094] 首先,以不同类型的标签替换待解析语句中相应的关键信息。例如,对于一句待解析语句:15时30分从北京到上海的飞机,则这句可以被转换成:出发时间(15时30分)+从+出发地(北京)+到+目的地(上海)+的+交通工具(飞机)。其中出发时间、出发地、目的地和交通工具均为预先设置的标签。

[0095] 随后,根据已经转换成预设格式的待解析语句中的每个字作为相应的索引单元,以在上述已经形成的索引目录中进行检索,从而得到所有相匹配的规则语义句式。具体地,可以根据一个待解析语句中的每个索引单元进行逐个检索,检索得到每个包括在待解析语句中的索引单元所关联的所有规则语义句式并输出。

[0096] 本发明的较佳的实施例中,上述过程只是一个初步检索匹配的过程,该过程中检

索得到的规则语义句式可能会非常多。为了进一步缩小匹配的范围,对检索得到的规则语义句式需要执行下述的处理:

[0097] 计算得到每个规则语义句式和待解析语句的匹配度,并根据匹配度缩小匹配范围。例如,确定一个匹配度阈值,并保留匹配度高于该匹配度阈值的相应的规则语义句式。

[0098] 本发明的较佳的实施例中,可以依照下述公式计算得到匹配度:

$$[0099] \quad S = (S_1 + S_2) / 2; \quad (1)$$

[0100] 其中,S表示匹配度;

[0101] S_1 表示待解析语句与规则语义句式之间的匹配部分占待解析语句的比例;

[0102] S_2 表示待解析语句与规则语义句式之间的匹配部分占规则语义句式之间的比例。

[0103] 具体地,所谓待解析语句与规则语义句式之间的匹配部分占待解析语句的比例,例如:待解析语句中包括索引单元1+2+3+4+5,相应地,相匹配的规则语义句式中包括索引单元1+3+4+6+7+8+9,则上述匹配部分(1,3,4)占待解析语句的比例即为3/5。

[0104] 类似上文中所述,所谓待解析语句与规则语义句式之间的匹配部分占规则语义句式之间的比例,同样依据上述示例,匹配部分(1,3,4)占规则语义句式的比例即为3/7。

[0105] 则依照上述公式(1),最终的匹配度S就为 $(3/5+3/7)/2=18/35$ 。

[0106] 本发明的较佳的实施例中,计算得到匹配度之后,将该匹配度与一预设的匹配度阈值进行比较:若该匹配度高于匹配度阈值,则保留相应的规则语义句式;反之,忽略相应的规则语义句式。

[0107] 则经过上述处理,最终可以缩小匹配范围,保留至少一个规则语义句式。

[0108] 本发明的较佳的实施例中,在执行上述语义匹配方法之前,预先训练形成一向量处理模型。该向量处理模型用于处理得到不同词的词向量。

[0109] 则本发明的较佳的实施例中,上述步骤S3可以被划分为计算得到第一语句向量的部分,以及计算得到第二语句向量的部分。

[0110] 本发明的较佳的实施例中,如图5所示,上述计算得到第一语句向量的方法具体包括:

[0111] 步骤S31a,将一条待解析语句进行分词处理;

[0112] 步骤S32a,将经过分词处理的待解析语句中的每个词输入至向量处理模型中,以分别得到关联于每个词的词向量;

[0113] 步骤S33a,根据所有词向量处理得到待解析语句的第一语句向量。

[0114] 具体地,本发明的较佳的实施例中,所谓分词处理,是指将一条待解析语句划分成不同的词语,即将一条待解析语句转换成由不同的词语构成的组合结构。例如:从北京到上海的飞机,可以被划分为从+北京+到+上海+的+飞机。上述分词的规则在现有技术中已有较多实现方式,在此不再赘述。

[0115] 本发明的较佳的实施例中,经过分词的待解析语句可以为在词与词之间添加特殊标记的语句,例如“从’北京’到’上海’的’飞机”。两个特殊标记之间的即为一个词。

[0116] 本发明的较佳的实施例中,经过分词处理后,将待解析语句中的每个词都放入训练形成的向量处理模型中,作为该模型的输入量,以处理得到每个词的词向量。

[0117] 最后,本发明的较佳的实施例中,将每个词的词向量组合形成上述待解析语句的第一语句向量。

[0118] 本发明的较佳的实施例中,类似上文中处理得到第一语句向量的方法,上述步骤S3中,处理得到第二语句向量的方法如图6所示,具体包括:

[0119] 步骤S31b,将一条规则语义句式进行分词处理;

[0120] 步骤S32b,将经过分词处理的规则语义句式中的每个词输入至向量处理模型中,以分别得到关联于每个词的词向量;

[0121] 步骤S33b,根据所有词向量处理得到规则语义句式的第二语句向量。

[0122] 上述过程与上述步骤S31a-S33a类似,在此不再赘述。

[0123] 本发明的较佳的实施例中,训练形成向量处理模型的方法包括:

[0124] 步骤B1,获取预设的多个语料信息;

[0125] 步骤B2,分别对每个语料信息进行分词处理;

[0126] 步骤B3,分别将每个经过分词处理的语料信息作为向量处理模型的输入,根据向量处理模型输出对应不同词的词向量;

[0127] 步骤B4,经过多个语料信息的训练,最终训练形成向量处理模型。

[0128] 具体地,本发明的较佳的实施例中,上述每个语料信息中包括:一条语句内容;或者一段语句内容;或者多段语句内容。例如,从网络上随机搜索得到的一句话,或者一段话,或者一整篇文章。由于训练样本的数量决定了向量处理模型的准确程度,即训练样本越多,向量处理模型越精确。因此,可以在网络上随机搜索大量的语料信息,并作为训练向量处理模型的输入量。

[0129] 本发明的较佳的实施例中,同样地,对每个语料信息进行分词处理,包括在语料信息中的不同的词语输入到神经网络中,经过神经网络的处理得到相应的输出量。最终经过大量的语料信息中包括的词的训练,得到训练形成的向量处理模型。上述训练过程在现有技术中存在较多的实现方式,在此不再展开。

[0130] 本发明的较佳的实施例中,上述步骤S4中,处理得到向量相似度的方法包括:

[0131] 采用余弦相似度计算方法,根据上述第一语句向量和相应的第二语句向量,直接度量得到向量相似度;

[0132] 或者

[0133] 首先计算得到第一语句向量与对应的第二语句向量之间的向量距离,随后将向量距离转换为对应的向量相似度。

[0134] 本发明的较佳的实施例中,如上文中所述,在计算得到上述待解析语句的第一语句向量和对应的一个规则语义句式的第二语句向量的相似度之后,判断该向量相似度是否大于一预设的相似度阈值,并将向量相似度大于该相似度阈值的对应的规则语义句式的语义信息作为该待解析语句的语义信息,以作为依据进行后续的信息交互处理。

[0135] 本发明的一个较佳的实施例中,在信息交互的过程中,最佳的语义识别结果应该为一个确定的结果。而在上述过程中,可能存在多个规则语义句式的向量相似度大于相似度阈值而被保留。此时需要根据向量相似度进行排列,并获取向量相似度最高的一个规则语义句式,并将其语义信息作为待解析语句的语义信息。

[0136] 本发明的另一个较佳的实施例中,在信息交互的过程中,智能设备自动识别出的最佳选项可能并不是使用者所需的结果,因此可以允许存在多个语义识别结果供使用者选择。例如,在上述过程中,预先设定一个选项数目,例如4个选项。随后判断通过向量相似度

的判断被保留的规则语义句式的数目：若大于4个，则保留向量相似度最高的四个规则语义句式；若不大于4个，则全部保留。随后将这些被保留的规则语义句式通过选项的形式显示，以供使用者选择。最后将被使用者选中的规则语义句式的语义信息作为待解析语句的语义信息，以进行后续的交互处理。

[0137] 本发明的较佳的实施例中，还提供一种智能设备，其中采用上文中所述的语义匹配方法。

[0138] 以上所述仅为本发明较佳的实施例，并非因此限制本发明的实施方式及保护范围，对于本领域技术人员而言，应当能够意识到凡运用本发明说明书及图示内容所作出的等同替换和显而易见的变化所得到的方案，均应当包含在本发明的保护范围内。

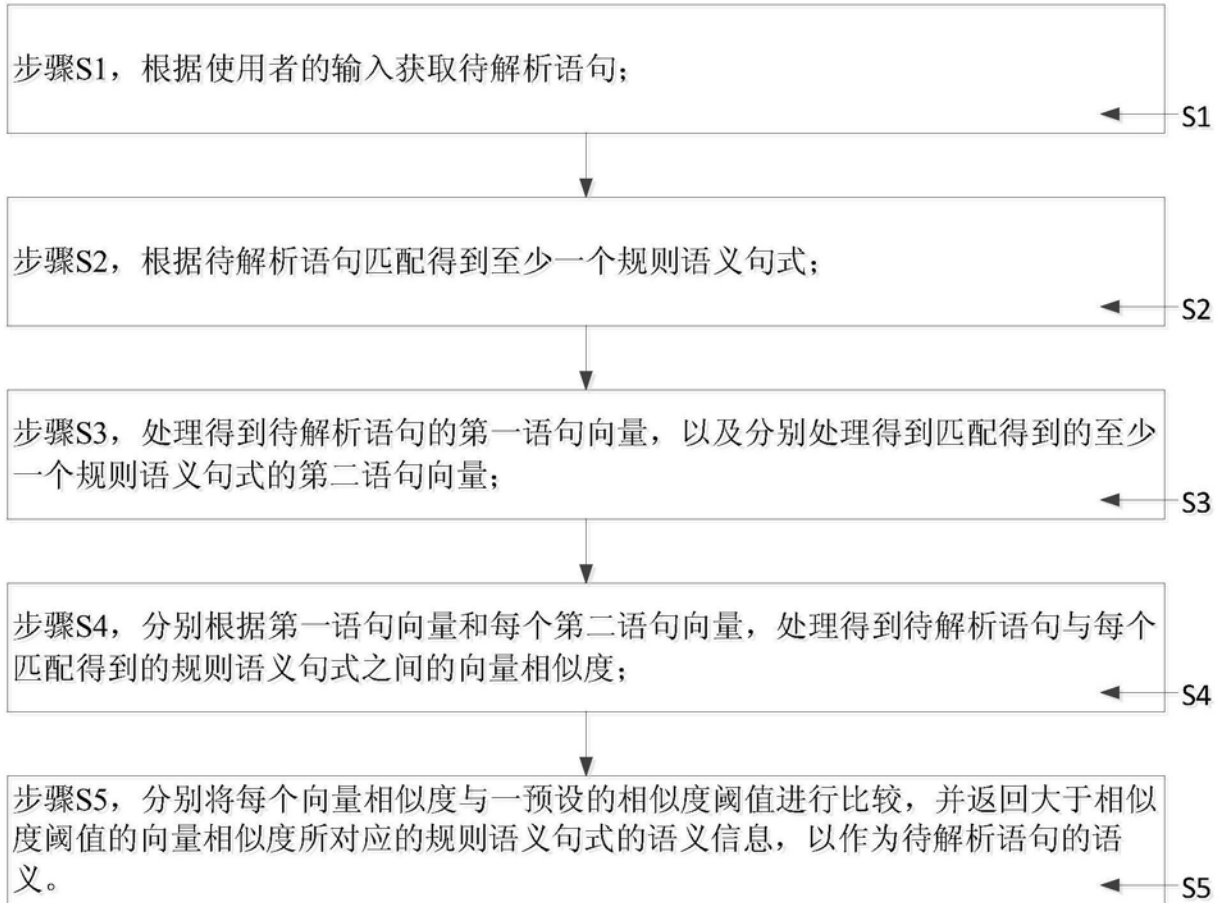


图1

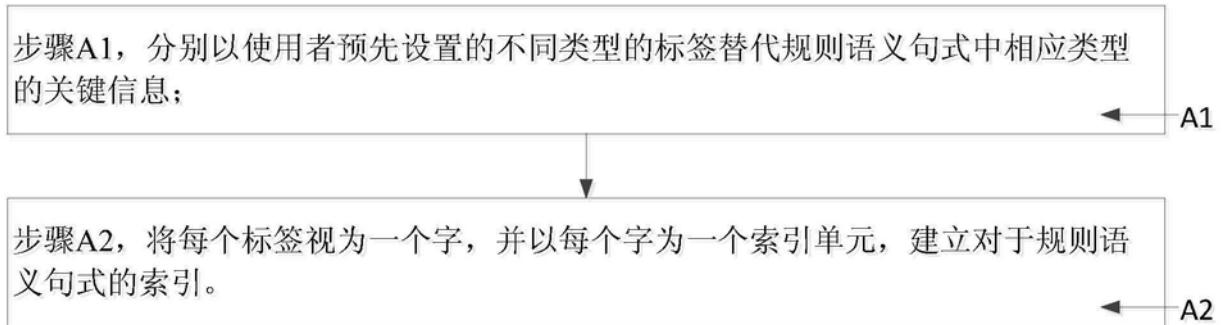


图2

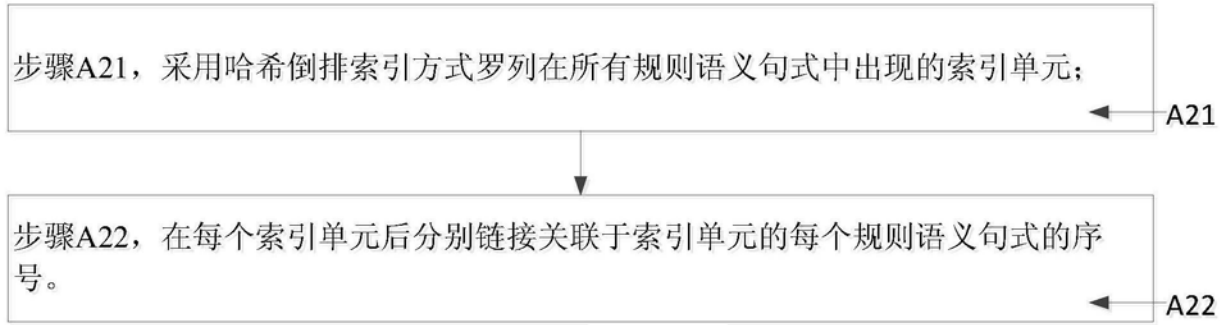


图3

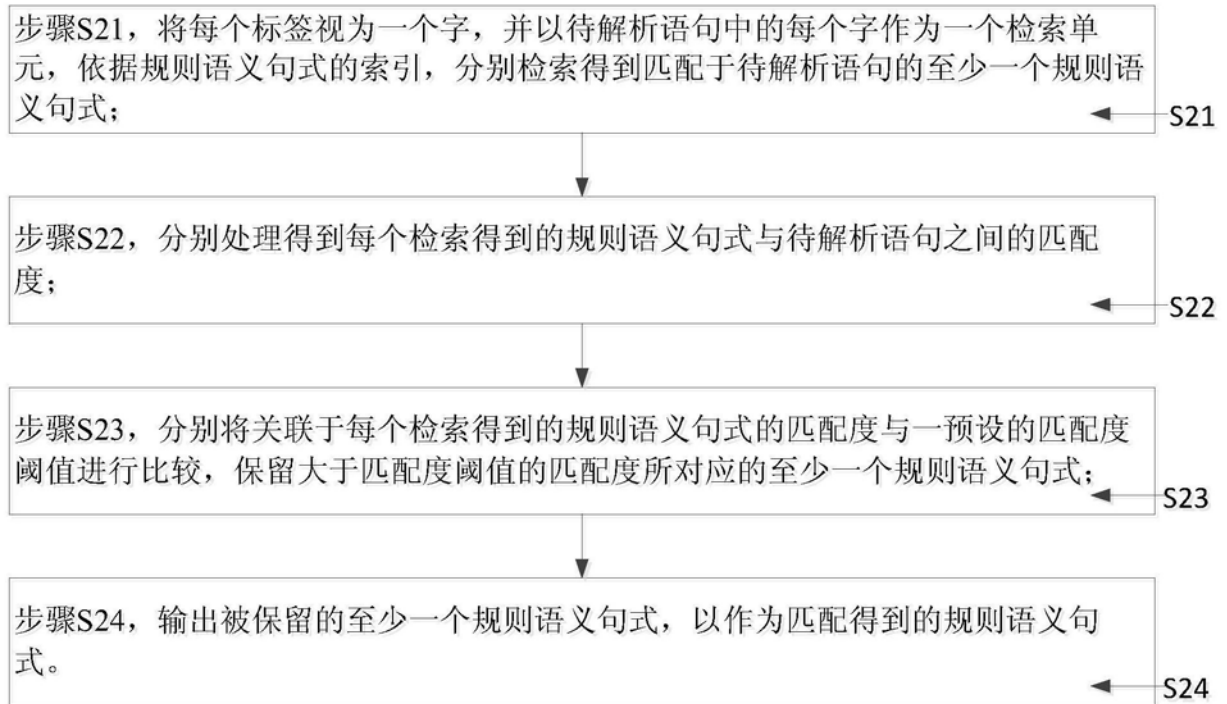


图4

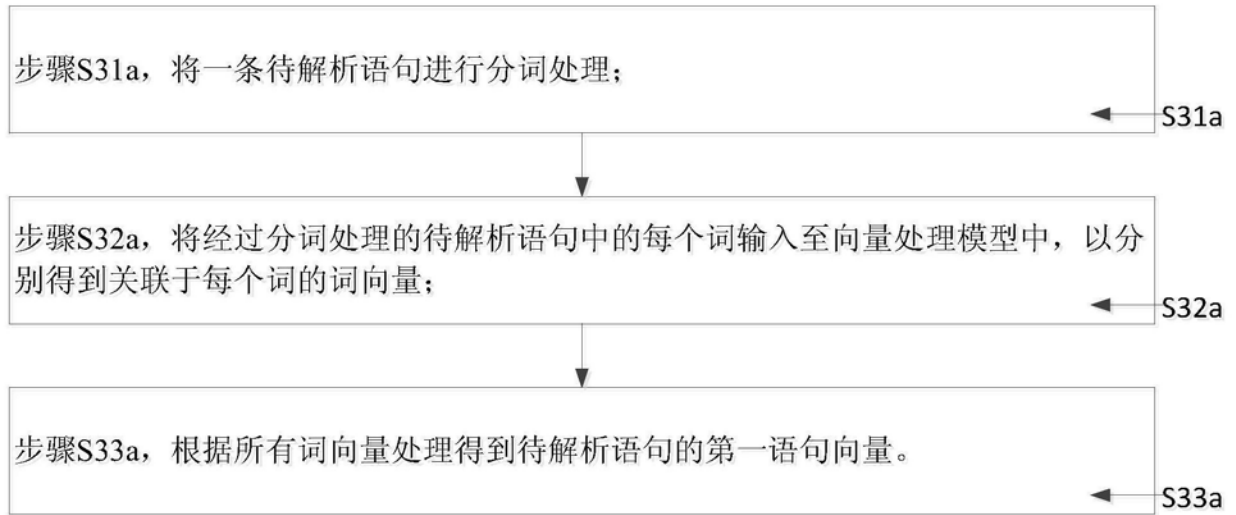


图5

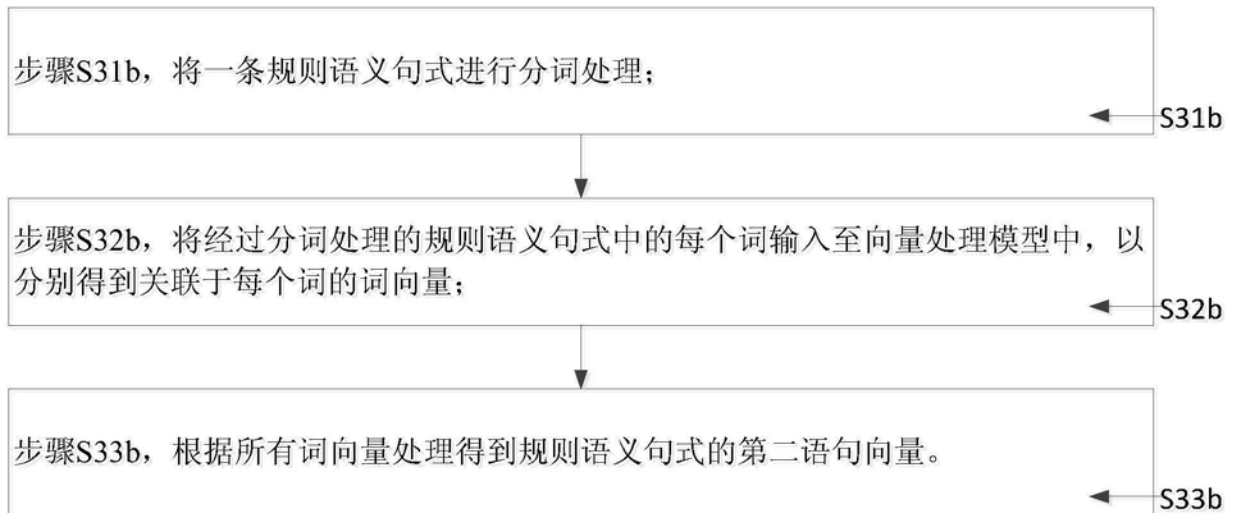


图6

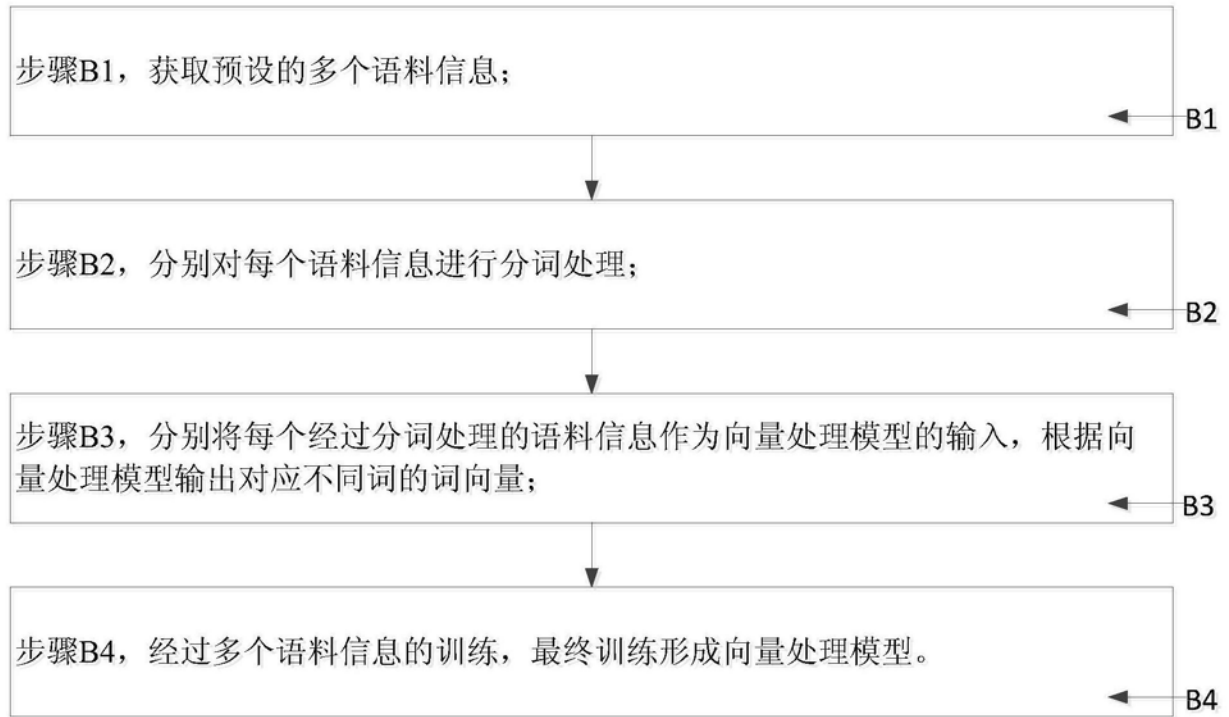


图7