



(12)发明专利

(10)授权公告号 CN 106202476 B

(45)授权公告日 2017.06.06

(21)申请号 201610565229.9

(56)对比文件

(22)申请日 2016.07.14

CN 105677783 A, 2016.06.15, 全文.

CN 105677795 A, 2016.06.15, 全文.

(65)同一申请的已公布的文献号

CN 104471568 A, 2015.03.25, 全文.

申请公布号 CN 106202476 A

审查员 孙昌璐

(43)申请公布日 2016.12.07

(73)专利权人 广州安望信息科技有限公司

地址 510000 广东省广州市天河区体育东路122号之二3304室

(72)发明人 黄明新

(74)专利代理机构 广州市越秀区哲力专利商标

事务所(普通合伙) 44288

代理人 陈振楔 李悦

(51)Int.Cl.

G06F 17/30(2006.01)

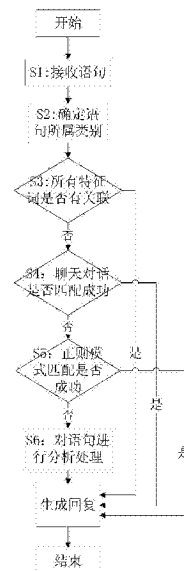
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种基于知识图谱的人机对话的方法及装置

(57)摘要

本发明公开了一种基于知识图谱的人机对话的方法及装置,该方法包括以下步骤:S1:接收用户发送的语句,并获取该语句的上文语句类别;S2:确定该语句最终所属类别;S3:通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联;S4:根据聊天对话库对该语句进行对话匹配;S5:对该语句进行正则模式匹配;S6:根据语句的类别进行分析处理并生成回复。本发明在一定程度上控制语句的类别,并且使常识的知识问答和定制的开放领域问答同一在一个流程中,有别于现存的自动问答系统中只依靠聊天库和分类模型的进行问题答案的搜索,并且在发明中加入了模板匹配和知识图谱搜索的技术,能够使得人机对话更加的丰富。



1. 一种基于知识图谱的人机对话的方法,其特征在于,包括以下步骤:

S1:接收用户发送的语句,获取该语句的上文语句,并确定该语句和上文语句的类别;

S2:根据该语句的类别和上文语句的类别来确定该语句的最终所属类别;步骤S2具体包括以下子步骤:

S21:根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级;

S22:判断所述语句的类别等级是否大于上文语句的类别等级,如果是,则选用所述语句的类别作为该语句的最终所属类别,如果否,则选用上文语句的类别作为该语句的最终所属类别;

S3:通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联,如果有,则根据特征词的关联生成回复,如果没有,则执行步骤S4;

S4:根据聊天对话库对该语句进行对话匹配,如果匹配成功,则生成回复,如果失败,则执行步骤S5;

S5:对该语句进行正则模式匹配,如果匹配成功,则生成相应的回复,如果匹配失败,则执行步骤S6;

S6:根据该语句的最终所属类别进行分析处理并生成回复,步骤S6具体包括以下子步骤:

S61:根据该语句的最终所属类别抽取与该语句的最终所属类别相关的属性特征;

S62:根据所得属性特征的抽取程序,抽取与属性特征对应的属性;

S63:根据所得属性生成回复。

2. 如权利要求1所述的基于知识图谱的人机对话的方法,其特征在于,步骤S4具体包括以下子步骤:

S41:根据聊天对话库对该语句进行全文搜索;

S42:通过Jaccard Similarity算法计算聊天对话库中候选语句与输入语句的相似度;

S43:判断所得相似度是否大于一阈值,如果大于该阈值,则选择相应语句进行回复,如果小于该阈值,则执行步骤S5。

3. 如权利要求2所述的基于知识图谱的人机对话的方法,其特征在于,所述阈值为0.7。

4. 一种基于知识图谱的人机对话的装置,其特征在于,包括以下模块:

接收模块:用于接收用户发送的语句,并获取该语句的上文语句,并确定该语句和上文语句的类别;

分类模块:用于根据该语句的类别和上文语句的类别来确定该语句的最终所属类别;分类模块具体包括以下子模块:

等级确定模块:根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级;

第一判断模块:判断所述语句的类别等级是否大于上文语句的类别等级,如果是,则选用所述语句的类别作为该语句的最终所属类别,如果否,则选用上文语句的类别作为该语句的最终所属类别;

知识图谱模块:用于通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联,如果有,则根据特征词的关联生成回复,如果没有,则执行聊天匹配模块;

聊天匹配模块:用于根据聊天对话库对该语句进行对话匹配,如果匹配成功,则生成回复,如果失败,则执行正则匹配模块;

正则匹配模块:用于对该语句进行正则模式匹配,如果匹配成功,则生成相应的回复,如果匹配失败,则执行分析处理模块;

分析处理模块:用于根据该语句的最终所属类别进行分析处理并生成回复,分析处理模块具体包括以下子模块:

属性抽取模块:用于根据该语句的最终所属类别抽取与该语句的最终所属类别相关的属性特征;

属性匹配模块:用于根据所得属性特征的抽取程序,抽取与属性特征对应的属性;

回复模块:用于根据所得属性生成回复。

5. 如权利要求4所述的基于知识图谱的人机对话的装置,其特征在于,聊天匹配模块具体包括以下子模块:

全文搜索模块:用于根据聊天对话库对该语句进行全文搜索;

计算模块:用于通过Jaccard Similarity算法计算聊天对话库中候选语句与输入语句的相似度;

第二判断模块:用于判断所得相似度是否大于一阈值,如果大于该阈值,则选择相应语句进行回复,如果小于该阈值,则执行正则匹配模块。

6. 如权利要求5所述的基于知识图谱的人机对话的装置,其特征在于,所述阈值为0.7。

## 一种基于知识图谱的人机对话的方法及装置

### 技术领域

[0001] 本发明属于人机问答的数据处理,尤其涉及一种基于知识图谱的人机对话的方法及装置。

### 背景技术

[0002] 智能问答系统以一问一答形式,精确的定位网站用户所需要的提问知识,通过与网站用户进行交互,为网站用户提供个性化的信息服务。

[0003] 智能问答系统是将积累的无序语料信息,进行有序和科学的整理,并建立基于知识的分类模型;这些分类模型可以指导新增加的语料咨询和服务信息,节约人力资源,提高信息处理的自动性,降低网站运行成本。基于对网站多年积累的关于政府和企业的的基本情况常见问题及其解答,整理为规范的问答库形式,以支撑各种形式问题的智能问答。方便了用户,提高了办事效率,提升了企业形象。

[0004] 现在智能问答系统通常会借助于自然语言处理和全文搜索技术对语句进行处理,单一的方式通常在某一些的准确率会比较低。以往的智能问答系统只会针对于某一个领域的问答,对于开放域的问答系统,现在还没有一种通用的,而且扩展性比较好的处理流程的出现。

[0005] 问答系统主要包括句子分类,信息检索,答案抽取,语义理解,知识图谱等方面的技术。

### 发明内容

[0006] 为了克服现有技术的不足,本发明的目的之一在于提供一种基于知识图谱的人机对话的方法,其能对常识和开放领域的问题进行回答。

[0007] 本发明的目的之二在于提供一种基于知识图谱的人机对话的装置,其能对常识和开放领域的问题进行回答。

[0008] 本发明的目的之一采用以下技术方案实现:

[0009] 一种基于知识图谱的人机对话的方法,包括以下步骤:

[0010] S1:接收用户发送的语句,并获取该语句的上文语句,并确定该语句和上文语句的类别;

[0011] S2:根据该语句的类别和上文语句的类别来确定该语句的最终所属类别;

[0012] S3:通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联,如果有,则根据特征词的关联生成回复,如果没有,则执行步骤S4;

[0013] S4:根据聊天对话库对该语句进行对话匹配,如果匹配成功,则生成回复,如果失败,则执行步骤S5;

[0014] S5:对该语句进行正则模式匹配,如果匹配成功,则生成相应的回复,如果匹配失败,则执行步骤S6;

[0015] S6:根据该语句的最终所属类别进行分析处理并生成回复。

- [0016] 优选地,步骤S2具体包括以下子步骤:
- [0017] S21:根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级;
- [0018] S22:判断所述语句的类别等级是否大于上文语句的类别等级,如果是,则选用所述语句的类别作为该语句的最终所属类别,如果否,则选用上文语句的类别作为该语句的最终所属类别。能够进一步公开步骤S2的实现方式。
- [0019] 优选地,步骤S4具体包括以下子步骤:
- [0020] S41:根据聊天对话库对该语句进行全文搜索;
- [0021] S42:通过Jaccard Similarity算法计算聊天对话库中候选语句与输入语句的相似度;
- [0022] S43:判断所得相似度是否大于一阈值,如果大于该阈值,则选择相应语句进行回复,如果小于该阈值,则执行步骤S5。其能进一步公开步骤S4的具体实现方式。
- [0023] 优选地,所述阈值为0.7。其能进一步公开阈值数值的设置。
- [0024] 优选地,步骤S6具体包括以下子步骤:
- [0025] S61:根据该语句的最终所属类别抽取与该语句的最终所属类别相关的属性特征;
- [0026] S62:根据所得属性特征的抽取程序,抽取与属性特征对应的属性;
- [0027] S63:根据所得属性生成回复。其能进一步公开步骤S6的实现方式。
- [0028] 本发明的目的之二采用以下技术方案实现:
- [0029] 一种基于知识图谱的人机对话的装置,包括以下模块:
- [0030] 接收模块:用于接收用户发送的语句,并获取该语句的上文语句,并确定该语句和上文语句的类别;
- [0031] 分类模块:用于根据该语句的类别和上文语句的类别来确定该语句的最终所属类别;
- [0032] 知识图谱模块:用于通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联,如果有,则根据特征词的关联生成回复,如果没有,则执行聊天匹配模块;
- [0033] 聊天匹配模块:用于根据聊天对话库对该语句进行对话匹配,如果匹配成功,则生成回复,如果失败,则执行正则匹配模块;
- [0034] 正则匹配模块:用于对该语句进行正则模式匹配,如果匹配成功,则生成相应的回复,如果匹配失败,则执行分析处理模块;
- [0035] 分析处理模块:用于根据该语句的最终所属类别进行分析处理并生成回复。
- [0036] 优选地,分类模块具体包括以下子模块:
- [0037] 等级确定模块:根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级;
- [0038] 第一判断模块:判断所述语句的类别等级是否大于上文语句的类别等级,如果是,则选用所述语句的类别作为该语句的最终所属类别,如果否,则选用上文语句的类别作为该语句的最终所属类别。能够进一步确定分类模块所包含的具体子模块。
- [0039] 优选地,聊天匹配模块具体包括以下子模块:
- [0040] 全文搜索模块:用于根据聊天对话库对该语句进行全文搜索;
- [0041] 计算模块:用于通过Jaccard Similarity算法计算聊天对话库中候选语句与输入

语句的相似度；

[0042] 第二判断模块：用于判断所得相似度是否大于一阈值，如果大于该阈值，则选择相应语句进行回复，如果小于该阈值，则执行正则匹配模块。其能进一步公开聊天匹配模块所包含的具体模块。

[0043] 优选地，所述阈值为0.7。其能进一步公开阈值的数值。

[0044] 优选地，分析处理模块具体包括以下子模块：

[0045] 属性抽取模块：用于根据该语句的最终所属类别抽取与该语句的最终所属类别相关的属性特征；

[0046] 属性匹配模块：用于根据所得属性特征的抽取程序，抽取与属性特征对应的属性；

[0047] 回复模块：用于根据所得属性生成回复。其能进一步公开分析处理模块所包含的子模块。

[0048] 相比现有技术，本发明的有益效果在于：

[0049] 本发明在一定程度上控制语句的类别，并且使常识的知识问答和定制的开放领域问答同在一个流程中，有别于现存的自动问答系统中只依靠聊天库和分类模型的进行问题答案的搜索，并且在发明中加入了模板匹配和知识图谱搜索的技术，能够使得人机对话更加的丰富。

## 附图说明

[0050] 图1为本发明一种基于知识图谱的人机对话的方法的流程图；

[0051] 图2为本发明一种基于知识图谱的人机对话的装置的结构框图；

[0052] 图3为本发明的语句分类的分类表。

## 具体实施方式

[0053] 下面，结合附图以及具体实施方式，对本发明做进一步描述：

[0054] 如图1所示，本发明提供了一种基于知识图谱的人机对话的方法，包括以下步骤：

[0055] S1：接收用户发送的语句，并获取该语句的上文语句，并确定该语句和上文语句的类别；

[0056] S2：根据该语句的类别和上文语句的类别来确定该语句的最终所属类别；步骤S2具体包括以下子步骤：

[0057] S21：根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级；

[0058] S22：判断所述语句的类别等级是否大于上文语句的类别等级，如果是，则选用所述语句的类别作为该语句的最终所属类别，如果否，则选用上文语句的类别作为该语句的最终所属类别。在此步骤是为了确定在后续步骤中对该语句以哪种分类来继续进行处理，分类结果的确定需要通过对上下文中的当前分类来比对，然后才确定选用那种分类，比对的方式主要采用而类别的等级比较，以高等级的类别优先，如图3所示，1级为最低等级，46级为最高级别，从1级到46级等级依次上升，例如，用户说“我要导航”的时候，这一句子无疑是“导航”分类的，但是，导航需要有一个目的地，所以，后端会生成“请问您的目的地是哪”以询问用户的目的地，当客户回应是，例如用户说“广州”，“广州”这个短语进入分类时，它

可能会被分类为“什么”这个类别,而如果一旦客户回复的短语被分类至其他类别时,导航这个目的就无法完成了,所以需要类别优先级做修正。由于“导航”这个分类在用户的会话中未被关闭,所以当“广州”被分类成“什么”时,需要与“导航”进行等级比较,由于“导航”的等级比“什么”的高,所以“广州”这个短语需要服从“导航”这个分类,如此,多轮对话就得以完成。

[0059] 但是还有例外的情况,例如上一句话中是为天气分类的,然后下一句话是没有分类的,则该句子当做天气分类的处理;

[0060] S3:通过知识图谱抽取该语句中的特征词,并判断所有特征词是否有关联,如果有,则根据特征词的关联生成回复,如果没有,则执行步骤S4;通过将抽取到的特征词作为知识图谱的搜索的起点,进行知识的关联搜索,如果全部特征词都有关联,则生成答案。

[0061] 例如,在知识图谱中包含有节点广州,然后有属性市花,属性值为木棉花,则形成三元组(广州,市花,木棉花)。当句子“广州的市花是什么呢?”或者“木棉花是哪里的市花?”进入图谱搜索的时候,由于知识图谱中包含广州、市花和木棉花,则通过对词语的提取,然后找到对应的三元组,生成回复“广州的市花是木棉花。”。

[0062] S4:根据聊天对话库对该语句进行对话匹配,如果匹配成功,则生成回复,如果失败,则执行步骤S5;步骤S4具体包括以下子步骤:

[0063] S41:根据聊天对话库对该语句进行全文搜索;

[0064] S42:通过Jaccard Similarity算法计算聊天对话库中候选语句与输入语句的相似度;

[0065] S43:判断所得相似度是否大于一阈值,如果大于该阈值,则选择相应语句进行回复,如果小于该阈值,则执行步骤S5。所述阈值为0.7。该步骤主要是匹配聊天的语句,例如当对话中出现“你好”之类的日常问候聊天的语句的时候,则可以通过该步骤完成日常的聊天功能。

[0066] S5:对该语句进行正则模式匹配,如果匹配成功,则生成相应的回复,如果匹配失败,则执行步骤S6;该步骤主要应用在系统功能,敏感词和部分特殊句式的处理,例如,当用户问“你是谁”之类的句子的时候,这个问句相当于app中的about的功能,匹配与系统功能相关的信息,生成回复。

[0067] S6:根据该语句的最终所属类别进行分析处理并生成回复;步骤S6具体包括以下子步骤:

[0068] S61:根据该语句的最终所属类别抽取与该语句的最终所属类别相关的属性特征;

[0069] S62:根据所得属性特征的抽取程序,抽取与属性特征对应的属性;

[0070] S63:根据所得属性生成回复。在本步骤中主要处理前述步骤处理不了的对话,例如:用户输入“广州今天冷不冷?”,这个句子的类别是天气,对于天气关注的属性主要有四个时间、地点、详细类别和程度,通过为天气语句做的定义属性抽取程序能够提取到三个属性,时间为今天,地点为广州,程度是冷,然后根据时间和城市去查询获取天气数据,再根据天气数据中的气温数据来判断今天是否冷,从而完成对该语句的回答。在该步骤中,通常会需要调用其他的模块来查询相关的信息从而来生成相应的答案。

[0071] 如图2所示,本发明提供了一种基于知识图谱的人机对话的装置,包括以下模块:

[0072] 接收模块:用于接收用户发送的语句,并获取该语句的上文语句,并确定该语句和

上文语句的类别；

[0073] 分类模块：用于根据该语句的类别和上文语句的类别来确定该语句的最终所属类别；分类模块具体包括以下子模块：

[0074] 等级确定模块：根据该语句的类别和上文语句的类别来确定所述语句的类别等级和上文语句的类别等级；

[0075] 第一判断模块：判断所述语句的类别等级是否大于上文语句的类别等级，如果是，则选用所述语句的类别作为该语句的最终所属类别，如果否，则选用上文语句的类别作为该语句的最终所属类别。

[0076] 知识图谱模块：用于通过知识图谱抽取该语句中的特征词，并判断所有特征词是否有关联，如果有，则根据特征词的关联生成回复，如果没有，则执行聊天匹配模块；

[0077] 聊天匹配模块：用于根据聊天对话库对该语句进行对话匹配，如果匹配成功，则生成回复，如果失败，则执行正则匹配模块；聊天匹配模块具体包括以下子模块：

[0078] 全文搜索模块：用于根据聊天对话库对该语句进行全文搜索；

[0079] 计算模块：用于通过Jaccard Similarity算法计算聊天对话库中候选语句与输入语句的相似度；

[0080] 第二判断模块：用于判断所得相似度是否大于一阈值，如果大于该阈值，则选择相应语句进行回复，如果小于该阈值，则执行正则匹配模块。所述阈值为0.7。

[0081] 正则匹配模块：用于对该语句进行正则模式匹配，如果匹配成功，则生成相应的回复，如果匹配失败，则执行分析处理模块；

[0082] 分析处理模块：用于根据该语句的最终所属类别进行分析处理并生成回复。分析处理模块具体包括以下子模块：

[0083] 属性抽取模块：用于根据该语句的最终所属类别抽取与该语句类别相关的属性特征；

[0084] 属性匹配模块：根据所得属性特征的抽取程序，抽取与属性特征对应的属性；

[0085] 回复模块：根据所得属性生成回复。

[0086] 对本领域的技术人员来说，可根据以上描述的技术方案以及构思，做出其它各种相应的改变以及形变，而所有的这些改变以及形变都应该属于本发明权利要求的保护范围之内。



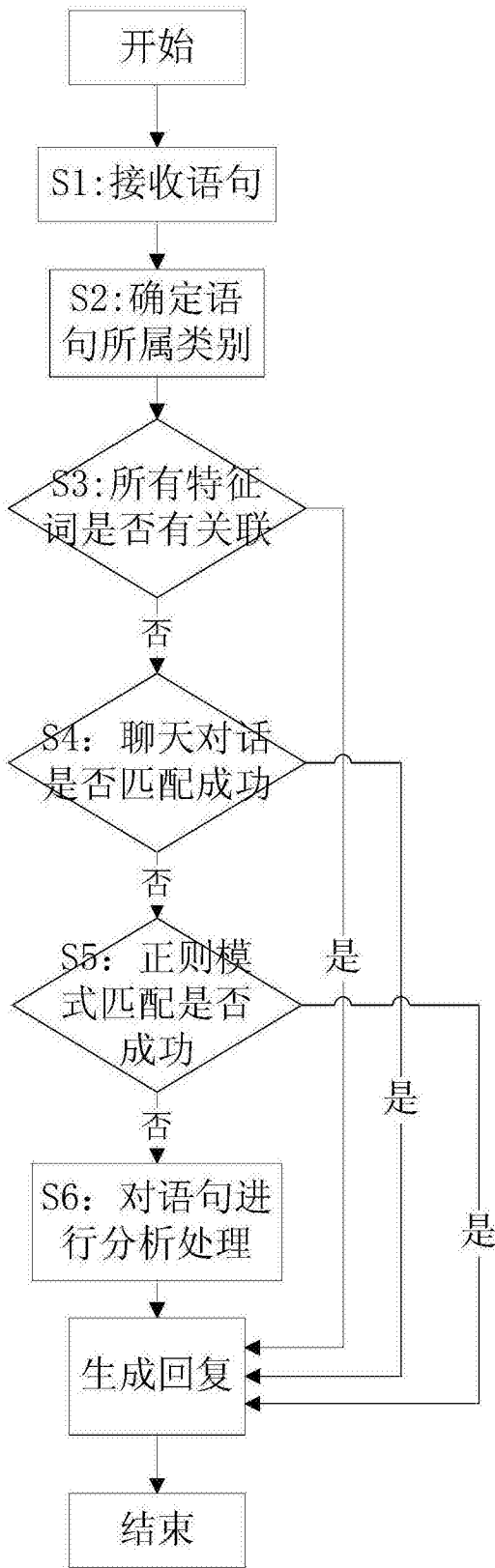


图1

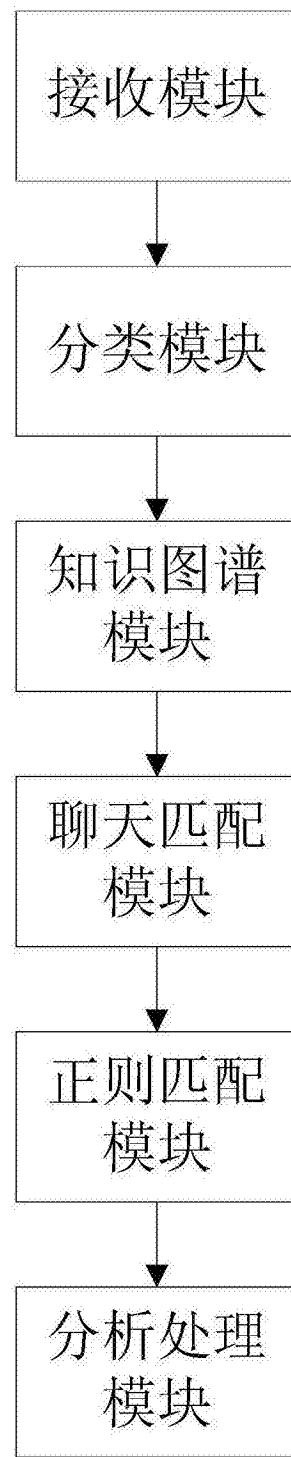


图2

1	买卖	17	笑话	33	IP 查询
2	人物	18	星座	34	电话号码
3	天气	19	日期	35	导航
4	附近	20	医药	36	交通
5	敏感词	21	菜谱	37	快递查询
6	工作	22	诗词	38	解梦
7	出租招租	23	歌词	39	定位
8	交友	24	故事	40	人的匹配
9	什么	25	旅游	41	闹钟提醒
10	如何	26	电影	42	公交线路
11	为什么	27	新闻	43	打电话
12	什么时候	28	算术	44	民俗文化
13	哪儿	29	图片	45	酒店问答
14	哪个	30	股票	46	医院信息
15	是非	31	汇率		
16	尺度	32	特殊技能		

图3