



(12) 发明专利

(10) 授权公告号 CN 110287491 B

(45) 授权公告日 2024.01.12

| | |
|---|------------------------------|
| (21) 申请号 201910556230.9 | CN 108460150 A, 2018.08.28 |
| (22) 申请日 2019.06.25 | CN 108470026 A, 2018.08.31 |
| (65) 同一申请的已公布的文献号 申请公布号 CN 110287491 A | CN 108491512 A, 2018.09.04 |
| (43) 申请公布日 2019.09.27 | CN 109582949 A, 2019.04.05 |
| (73) 专利权人 北京百度网讯科技有限公司 地址 100085 北京市海淀区上地十街10号 百度大厦2层 | JP 2001052032 A, 2001.02.23 |
| (72) 发明人 韩翠云 陈玉光 刘远圳 潘禄 施茜 | JP 2001337945 A, 2001.12.07 |
| (74) 专利代理机构 北京同立钧成知识产权代理 有限公司 11205 专利代理师 朱颖 刘芳 | US 2015339288 A1, 2015.11.26 |
| (51) Int. Cl. | CN 105824803 A, 2016.08.03 |
| G06F 40/211 (2020.01) | CN 109710710 A, 2019.05.03 |
| G06F 40/258 (2020.01) | CN 108763333 A, 2018.11.06 |
| (56) 对比文件 | CN 106202574 A, 2016.12.07 |
| CN 106383817 A, 2017.02.08 | CN 107766585 A, 2018.03.06 |
| CN 107832299 A, 2018.03.23 | CN 109800413 A, 2019.05.24 |
| CN 107977363 A, 2018.05.01 | CN 107239445 A, 2017.10.10 |
| | CN 108319668 A, 2018.07.24 |
| | CN 108959312 A, 2018.12.07 |

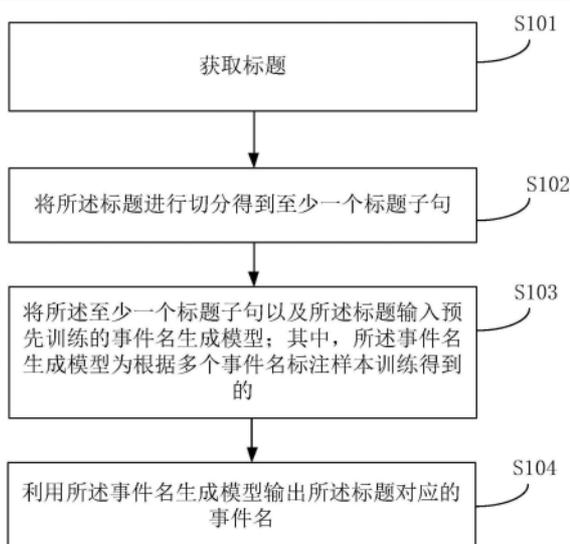
Naoya Okumura et al. Generating
Headline Candidates for News
Articles. 2016 IEEE 17th International
Conference on Information Reuse and
Integration. 2016, 355-340.

审查员 宁淑婷

权利要求书3页 说明书11页 附图2页

(54) 发明名称
 事件名生成方法及装置

(57) 摘要
 本发明实施例提供一种事件名生成方法及装置, 在将获取的标题切分为标题子句后, 进一步将该标题子句输入到预先训练的事件名生成模型, 事件名生成模型可以根据标题子句自动输出该标题对应的事件名, 整个过程不需要人工参与, 生成事件名的效率较高。



1. 一种事件名生成方法,其特征在于,所述方法包括:
 - 获取标题;
 - 将所述标题进行切分得到至少一个标题子句;
 - 将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的;
 - 利用所述事件名生成模型输出所述标题对应的事件名;
 - 所述事件名生成模型输出所述标题对应的事件名,包括:
 - 判断所述标题子句是否包括主谓结构;
 - 若所述标题子句不包括主谓结构,则根据所述标题对所述标题子句进行补充,得到第一补充子句;
 - 判断所述第一补充子句是否存在歧义;
 - 若所述第一补充子句存在歧义,则根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;
 - 判断所述第二补充子句的语义是否完整;
 - 若所述第二补充子句的语义不完整,判断所述第二补充子句中是否包含事件子句;
 - 若所述第二补充子句中包含所述事件子句,则对所述第二补充子句进行补充和/或改写,得到第三补充子句;确定所述第三补充子句为所述标题对应的事件名,并输出所述标题对应的事件名;
 - 若所述第二补充子句中不包含所述事件子句,则丢弃所述标题。
2. 根据权利要求1所述的方法,其特征在于,所述方法,还包括:
 - 在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。
3. 根据权利要求1所述的方法,其特征在于,所述方法,还包括:
 - 在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;
 - 在补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。
4. 根据权利要求1所述的方法,其特征在于,所述方法,还包括:
 - 在所述第一补充子句不存在歧义,且所述第一补充子句语义完整的情况下,确定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
5. 根据权利要求1所述的方法,其特征在于,所述方法,还包括:
 - 在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
6. 根据权利要求1-5任一项所述的方法,其特征在于,将所述标题进行切分得到至少一个标题子句,包括:
 - 在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。
7. 一种事件名生成装置,其特征在于,包括:

- 获取模块,用于获取标题;
- 切分模块,用于将所述标题进行切分得到至少一个标题子句;
- 输入模块,用于将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的;
- 输出模块,用于利用所述事件名生成模型输出所述标题对应的事件名;
- 所述输出模块,具体用于:
- 判断所述标题子句是否包括主谓结构;
- 若所述标题子句不包括主谓结构,则根据所述标题对所述标题子句进行补充,得到第一补充子句;
- 判断所述第一补充子句是否存在歧义;
- 若所述第一补充子句存在歧义,则根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;
- 判断所述第二补充子句的语义是否完整;
- 若所述第二补充子句的语义不完整,判断所述第二补充子句中是否包含事件子句;
- 若所述第二补充子句中包含所述事件子句,则对所述第二补充子句进行补充和/或改写,得到第三补充子句;确定所述第三补充子句为所述标题对应的事件名,并输出所述标题对应的事件名;
- 若所述第二补充子句中不包含所述事件子句,则丢弃所述标题。
8. 根据权利要求7所述的装置,其特征在于,所述输出模块,还包括:
- 第一输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。
9. 根据权利要求7所述的装置,其特征在于,所述输出模块还包括:
- 第二输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;在补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。
10. 根据权利要求7所述的装置,其特征在于,所述输出模块,还包括:
- 第三输出单元,用于在所述第一补充子句不存在歧义,且所述第一补充子句语义完整的情况下,确定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
11. 根据权利要求7所述的装置,其特征在于,所述输出模块,还包括:
- 第四输出单元,用于在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
12. 根据权利要求7-11任一项所述的装置,其特征在于,所述切分模块包括:
- 切分子模块,用于在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。
13. 一种电子设备,其特征在于,包括:
- 处理器,存储器以及计算机程序;其中,所述计算机程序被存储在所述存储器中,并且

被配置为由所述处理器执行,所述计算机程序包括用于执行如权利要求1-6任一项所述的方法的指令。

14.一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存储有计算机程序,所述计算机程序被执行时,实现如权利要求1-6任一项所述的方法。

事件名生成方法及装置

技术领域

[0001] 本发明涉及深度学习技术领域,尤其涉及事件名生成方法及装置。

背景技术

[0002] 随着互联网的发展,新闻事件得到越来越多的关注,在新闻推送网站中,通常需要根据新闻对应的事件实体生成事件名,通过事件名为用户提供新闻引导。

[0003] 现有技术在对事件实体生成事件名时,首先通过生成模型对事件实体进行切分等操作,得到关键字,但是关键字通常在语流畅性和语义完整性上有欠缺,不能作为事件名,因此还需要通过人工编辑的方式对关键字编辑,才能得到事件名。

[0004] 但是现有生成事件名的方式需要人工参与,导致生成事件名的效率很低。

发明内容

[0005] 本发明实施例提供事件名生成方法及装置,以解决得到事件名的效率低的技术问题。

[0006] 本发明实施例第一方面提供一种事件名生成方法,包括:

[0007] 获取标题;

[0008] 将所述标题进行切分得到至少一个标题子句;

[0009] 将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的;

[0010] 利用所述事件名生成模型输出所述标题对应的事件名。

[0011] 可选的,所述事件名生成模型输出所述标题对应的事件名,包括:

[0012] 在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名。

[0013] 可选的,所述在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名,包括:

[0014] 在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0015] 可选的,所述在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名,包括:

[0016] 在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;

[0017] 在所述补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0018] 可选的,所述在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名,包括:

- [0019] 在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;
- [0020] 在所述第一补充子句的不存在歧义,且所述第一补充子句语义完整的情况下,确定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
- [0021] 可选的,所述在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名,包括:
- [0022] 在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;
- [0023] 在所述第一补充子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;
- [0024] 在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
- [0025] 可选的,所述事件名生成模型输出所述标题对应的事件名,包括:
- [0026] 在所述事件名生成模型根据所述至少一个标题子句不能得到事件名的情况下,对所述至少一个标题子句进行补充;
- [0027] 根据对所述至少一个标题子句的补充输出所述标题对应的事件名。
- [0028] 可选的,所述对所述至少一个标题子句进行补充,包括:
- [0029] 在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第三补充子句;
- [0030] 在所述第三补充子句存在歧义的情况下,根据所述标题或所述标题对应的正文对所述第三补充子句进行补充,得到第四补充子句。
- [0031] 可选的,所述根据对所述至少一个标题子句的补充输出所述标题对应的事件名,包括:
- [0032] 在所述第四补充子句语义不完整,且所述第四补充子句中包含事件子句的情况下,对所述第四补充子句进行补充和/或改写,得到第五补充子句;
- [0033] 确定所述第五补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。
- [0034] 可选的,将所述标题进行切分得到至少一个标题子句,包括:
- [0035] 在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。
- [0036] 本发明实施例第二方面提供一种事件名生成装置,包括:
- [0037] 获取模块,用于获取标题;
- [0038] 切分模块,用于将所述标题进行切分得到至少一个标题子句;
- [0039] 输入模块,用于将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的;
- [0040] 输出模块,用于利用所述事件名生成模型输出所述标题对应的事件名。
- [0041] 可选的,所述输出模块包括:
- [0042] 第一输出子模块,用于在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名。

[0043] 可选的,所述第一输出子模块包括:

[0044] 第一输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0045] 可选的,所述第一输出子模块包括:

[0046] 第二输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;在所述补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0047] 可选的,所述第一输出子模块包括:

[0048] 第三输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;

[0049] 在所述第一补充子句的不存在歧义,且所述第一补充子句语义完整的情况下,确定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0050] 可选的,第一输出子模块包括:

[0051] 第四输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;在所述第一补充子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0052] 可选的,输出模块包括:

[0053] 第二输出子模块,用于在所述事件名生成模型根据所述至少一个标题子句不能得到事件名的情况下,对所述至少一个标题子句进行补充;根据对所述至少一个标题子句的补充输出所述标题对应的事件名。

[0054] 可选的,所述第二输出子模块,包括:

[0055] 第五输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第三补充子句;在所述第三补充子句存在歧义的情况下,根据所述标题或所述标题对应的正文对所述第三补充子句进行补充,得到第四补充子句。

[0056] 可选的,所述第二输出子模块,包括:

[0057] 第五输出单元,用于在所述第四补充子句语义不完整,且所述第四补充子句中不包含事件子句的情况下,对所述第四补充子句进行补充和/或改写,得到第五补充子句;确定所述第五补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0058] 可选的,所述切分模块包括:

[0059] 切分子模块,用于在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。

[0060] 本发明实施例第三方面提供一种电子设备,包括:处理器,存储器以及计算机程序;其中,所述计算机程序被存储在所述存储器中,并且被配置为由所述处理器执行,所述计算机程序包括用于执行如前述第一方面任一项所述的方法的指令。

[0061] 本发明实施例第四方面提供一种计算机可读存储介质,所述计算机可读存储介质

存储有计算机程序,所述计算机程序被执行时,实现如前述第一方面任一项所述的方法。

[0062] 本发明实施例相对于现有技术的有益效果:

[0063] 本发明实施例中提供了一种事件名生成方法及装置,在将获取的标题切分为标题子句后,进一步将该标题子句输入到预先训练的事件名生成模型,事件名生成模型可以根据标题子句自动输出该标题对应的事件名,整个过程不需要人工参与,生成事件名的效率较高。

附图说明

[0064] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术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发明的一些实施例,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。

[0065] 图1为本发明实施例提供的事件名生成方法的流程示意图;

[0066] 图2为本发明实施例提供的事件名生成方法的事件名标注示意图;

[0067] 图3为本发明实施例提供的事件名生成方法的另一流程示意图;

[0068] 图4为本发明提供的事件名生成装置实施例的结构示意图。

具体实施方式

[0069] 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,都属于本发明保护的范围。

[0070] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0071] 应当明确,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前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0072] 在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

[0073] 应当理解,本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0074] 应当理解,尽管在本发明实施例中可能采用术语第一、第二、第三等来描述XXX,但这些XXX不应限于这些术语。这些术语仅用来将XXX彼此区分开。例如,在不脱离本发明实施例范围的情况下,第一XXX也可以被称为第二XXX,类似地,第二XXX也可以被称为第一XXX。

[0075] 取决于语境,如在此所使用的词语“如果”、“若”可以被解释成为“在……时”或

“当……时”或“响应于确定”或“响应于检测”。类似地,取决于语境,短语“如果确定”或“如果检测(陈述的条件或事件)”可以被解释成为“当确定时”或“响应于确定”或“当检测(陈述的条件或事件)时”或“响应于检测(陈述的条件或事件)”。

[0076] 还需要说明的是,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商品或者系统不仅包括那些要素,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商品或者系统所固有的要素。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商品或者系统中还存在另外的相同要素。

[0077] 本发明实施例中提供了一种事件名生成方法和装置,在将获取的标题切分为标题子句后,进一步将该标题子句输入到预先训练的事件名生成模型,事件名生成模型可以根据标题子句自动输出该标题对应的事件名,整个过程不需要人工参与,生成事件名的效率较高。

[0078] 本发明实施例可以应用于终端,终端可以包括:手机、平板电脑、笔记本电脑、台式电脑或者服务器等可以运行事件名生成模型的电子设备;本发明实施例对此不作具体限定。

[0079] 本发明实施例所描述的标题可以是新闻网站等中包括的新闻标题等;本发明实施例对此不作具体限定。

[0080] 本发明实施例所描述的标题子句可以是标题中去除对于事件描述没有关键作用的内容之后剩下的内容,其中,对于事件描述没有关键作用的内容可以包括形容词、副词和标点等,本发明实施例对此不作具体限定。

[0081] 本发明实施例所描述的事件名可以是至少具备主语和谓语的、且能清楚表达一个事件的词组或句子,本发明实施例对事件名的具体形式不作限定。

[0082] 本发明实施例所描述的事件子句可以是至少包含一个动词,或者比较明确的事件类型的词组。示例的,明确的事件类型可以为地震、车祸等,本发明实施例对此不作具体限定。

[0083] 本发明实施例所描述的事件名生成模型可以是基于神经网络等深度学习算法训练的模型,事件名生成模型中可以包括多个卷积层等,本发明实施例对此不作具体限定。

[0084] 如图1所示,图1为本发明实施例提供的事件名生成方法的流程示意图。

[0085] 该方法具体可以包括:

[0086] 步骤S101:获取标题。

[0087] 本发明实施例中,可以通过在新闻网站中抓取标题的方式获取标题;也可以通过在标题数据库中获取标题的方式获取标题;还可以根据实际的应用场景,采用其他方式获取标题,本发明实施例对此不作具体限定。

[0088] 步骤S102:将所述标题进行切分得到至少一个标题子句。

[0089] 本发明实施例中,可以根据实际的应用场景,确定适应的切分策略,将标题切分为至少一个标题子句。示例的,切分策略可以包括基于正则匹配切的切分策略等,本发明实施例对将标题进行切分得到至少一个标题子句的方法不作具体限定。

[0090] 可选的,步骤S102包括:在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。

[0091] 本发明实施例中,对标题中所包括的事件进行判定,在标题中包括的事件大于两个的情况下,可以认为该标题是盘点类新闻标题,不对应具体的事件,对于该类盘点类新闻标题生成事件名后,不能为用户提供较好的事件名引导,且会对计算资源造成占用,因此,可以将包括的事件大于两个的标题去除,保留包括的事件不大于两个的标题,进而根据包括的事件不大于两个的标题生成事件名,从而可以避免对计算资源的占用。

[0092] 步骤S103:将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的。

[0093] 本发明实施例中,事件名生成模型是预先训练的,训练该事件名生成模型时,具体可以包括:样本标注和模型调优两个步骤。

[0094] 具体应用中,样本标注步骤中,可以首先获取样本,在对样本进行标注。示例的,可以从事件图谱资源库中召回各领域近一年的新闻资讯,再随机选取5000个事件,将其整理为事件、新闻连接、标题和正文的格式,然后将标题切分得到标题子句,建立标题和标题子句的映射关系,再由人工标注标题子句对应的实际事件名,从而得到事件名标注样本。参照图2,示出了一种事件名标注的示意图,可以对标题进行人工标注事件名,还可以对该事件名进行打分,通过事件名的打分反映该事件名的准确程度。

[0095] 在模型调优步骤中,可以事件名标注样本输入神经网络模型中,根据该神经网络模型输出的事件名与实际事件名的差异,调整神经网络模型中各节点的参数,直到神经网络模型输出的事件名与实际事件名的差异小于预设差异值,可以认为训练得到事件名生成模型。

[0096] 步骤S104:利用所述事件名生成模型输出所述标题对应的事件名。

[0097] 具体应用中,将至少一个标题子句以及标题输入该预先训练的事件名生成模型后,该事件名生成模型就可以按照训练中学习的算法等,自动生成以及输出该标题对应的事件名。

[0098] 可以理解,实际应用中,可能存在着根据对至少一个标题子句的简单处理就能得到事件名的情况,也能存在根据对至少一个标题子句的简单处理不能得到事件名的情况。

[0099] 可选的,步骤S104可以包括在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名。

[0100] 具体应用中,根据该至少一个标题子句得到事件名可以是,对该至少一个标题子句进行简单的补充后,就可以得到事件名。可选的,根据该至少一个标题子句得到事件名的情况下,输出该标题对应的事件名可以通过四种实现方式。

[0101] 第一种实现方式为:在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0102] 本发明实施例中,标题子句中包括主谓结构,且该标题子句的不存在歧义,且该标题子句语义完整,可以认为该标题子句符合事件名的定义,可以将该标题子句作为标题对应的事件名进行输出。

[0103] 第二种实现方式为:在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;在所述补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,

并输出所述标题对应的事件名。

[0104] 本发明实施例中,标题子句中包括主谓结构,但是该标题子句存在歧义,则该存在歧义的标题子句不能作为事件名,可以通过标题和/或标题对应的正文对该标题子句进行补充,若补充后的标题子句语义完整,则可以认为补充后的标题子句符合事件名的定义,可以将补充后的标题子句作为标题对应的事件名进行输出。本发明实施例中,通过标题和/或标题对应的正文对该标题子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题和/或标题对应的正文对该标题子句进行适当的补充。

[0105] 第三种实现方式为:在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;在所述第一补充子句的不存在歧义,且所述第一补充子句语义完整的情况下,确定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0106] 本发明实施例中,标题子句中不包括主谓结构,可以根据标题对该标题子句进行补充,得到第一补充子句,若第一补充子句不存在歧义且语义完整,则可以认为第一补充子句符合事件名的定义,可以将第一补充子句作为标题对应的事件名进行输出。本发明实施例中,通过标题对该标题子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题对该标题子句进行适当的补充。

[0107] 第四种实现方式为:在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;在所述第一补充子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0108] 本发明实施例中,标题子句中不包括主谓结构,可以根据标题对该标题子句进行补充,得到第一补充子句,若第一补充子句存在歧义,可以根据标题和/或该标题对应的正文对第一补充子句进行补充,得到第二补充子句,若第二补充子句语义完整,则可以认为第二补充子句符合事件名的定义,可以将第二补充子句作为标题对应的事件名进行输出。本发明实施例中,通过标题对该标题子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题对该标题子句进行适当的补充,得到第一补充子句。通过标题和/或标题对应的正文对该第一补充子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题和/或标题对应的正文对该第一补充子句进行适当的补充。

[0109] 可选的,步骤S104可以包括在所述事件名生成模型根据所述至少一个标题子句不能得到事件名的情况下,对所述至少一个标题子句进行补充;根据对所述至少一个标题子

句的补充输出所述标题对应的事件名。

[0110] 具体应用中,根据该至少一个标题子句不能得到事件名可以是,对该至少一个标题子句进行简单的补充后,不能得到事件名,需要进行进一步的操作才能得到事件名。

[0111] 可选的,所述对所述至少一个标题子句进行补充,包括:在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第三补充子句;在所述第三补充子句存在歧义的情况下,根据所述标题或所述标题对应的正文对所述第三补充子句进行补充,得到第四补充子句。

[0112] 本发明实施例中,标题子句中不包括主谓结构,可以根据标题对该标题子句进行补充,得到第三补充子句,若第三补充子句存在歧义,可以根据标题和/或该标题对应的正文对第三补充子句进行补充,得到第四补充子句。本发明实施例中,通过标题对该标题子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题对该标题子句进行适当的补充,得到第三补充子句。通过标题和/或标题对应的正文对该第三补充子句进行补充的具体方式可以根据实际的应用场景确定,示例的,可以根据标题对应事件的领域等,按照需求获取事件名标注样本,然后根据该事件名标注样本训练符合要求的事件名生成模型,使得该事件名生成模型能够通过标题和/或标题对应的正文对该第三补充子句进行适当的补充,得到第四补充子句。

[0113] 可选的,所述根据对所述至少一个标题子句的补充输出所述标题对应的事件名,包括:在所述第四补充子句语义不完整,且所述第四补充子句中包含事件子句的情况下,对所述第四补充子句进行补充和/或改写,得到第五补充子句;确定所述第五补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0114] 本发明实施例中,在第四补充子句语义不完完整的情况下,可以进一步判定该第四补充子句是否包含事件子句,在第四补充子句包含事件子句的情况下,可以认为该第四补充子句还有生成事件名的可能,则进一步对第四补充子句进行补充和/或改写,得到第五补充子句,将该第五补充子句作为该标题对应的事件名进行输出。

[0115] 可以理解,若第四补充子句不包含事件子句,则可以认为该第四补充子句没有生成事件名的可能,因此可以丢弃该标题,不再进行后续的操作。

[0116] 综上所述,本发明实施例中提供了一种事件名生成方法及装置,在将获取的标题切分为标题子句后,进一步将该标题子句输入到预先训练的事件名生成模型,事件名生成模型可以根据标题子句自动输出该标题对应的事件名,整个过程不需要人工参与,生成事件名的效率较高。可以理解,因为本申请实施例的事件名生成方法不需要人工参与,所以可以适用于对新闻网站中的大量标题进行生成事件名的操作。

[0117] 参照图3,示出了本发明实施例中生成事件名的一种实现流程。如图3所示,该方法具体可以包括:

[0118] 在对标题切分得到标题子句后,判定该标题子句是否包括主谓结构,若该标题子句不包括主谓结构,则根据标题进行补充,得到第一补充内容,并进一步判定第一补充内容是否存在歧义;若该标题子句包括主谓结构,则进一步判定该标题子句是否存在歧义,若该标题子句或第一补充内容存在歧义,则根据标题和/或标题对应的正文进一步补充,得到第

二补充内容,并进一步判定第二补充内容语义是否完整;若该标题子句或第一补充内容不存在歧义,则进一步判定该标题子句或第一补充内容语义是否完整,在该标题子句或第一补充内容或第二补充内容语义完整的情况下,输出将语义完整的该标题子句或第一补充内容或第二补充内容作为该标题对应的事件名输出;在该标题子句或第一补充内容或第二补充内容语义不完整的情况下,进一步判定该语义不完整的标题子句或第一补充内容或第二补充内容是否包含事件子句,若该语义不完整的标题子句或第一补充内容或第二补充内容包含事件子句,则进一步对该语义不完整的标题子句或第一补充内容或第二补充内容进行补充和/或改写,将改写后的内容作为标题对应的事件名输出;若该语义不完整的标题子句或第一补充内容或第二补充内容不包含事件子句,则丢弃该标题,不根据该标题进行事件名生成。

[0119] 可以理解,上述步骤只是对本发明实施例的示例介绍,每个步骤不是必须的,可以根据实际的应用场景,采用其中的部分或全部步骤生成事件名,本发明实施例对此不作具体限定。

[0120] 本发明实施例中,先采用抽取策略从标题中切分得到标题子句,再根据生成策略对在抽取策略中不能得到事件名的内容进一步补充改写,进一步生成事件名,提升了根据标题生成事件名的生成概率。

[0121] 图4为本发明提供的事件名生成装置实施例的结构示意图。如图4所示,本实施例提供的事件名生成装置包括:

[0122] 获取模块410,用于获取标题;

[0123] 切分模块420,用于将所述标题进行切分得到至少一个标题子句;

[0124] 输入模块430,用于将所述至少一个标题子句以及所述标题输入预先训练的事件名生成模型;其中,所述事件名生成模型为根据多个事件名标注样本训练得到的;

[0125] 输出模块440,用于利用所述事件名生成模型输出所述标题对应的事件名。

[0126] 可选的,所述输出模块包括:

[0127] 第一输出子模块,用于在根据所述至少一个标题子句得到事件名的情况下,输出所述标题对应的事件名。

[0128] 可选的,所述第一输出子模块包括:

[0129] 第一输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句不存在歧义,且所述标题子句语义完整的情况下,确定所述标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0130] 可选的,所述第一输出子模块包括:

[0131] 第二输出单元,用于在所述标题子句中包括主谓结构,且所述标题子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述标题子句进行补充;在所述补充后的标题子句语义完整的情况下,确定所述补充后的标题子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0132] 可选的,所述第一输出子模块包括:

[0133] 第三输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;

[0134] 在所述第一补充子句的不存在歧义,且所述第一补充子句语义完整的情况下,确

定所述第一补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0135] 可选的,第一输出子模块包括:

[0136] 第四输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第一补充子句;在所述第一补充子句存在歧义的情况下,根据所述标题和/或所述标题对应的正文对所述第一补充子句进行补充,得到第二补充子句;在所述第二补充子句语义完整的情况下,确定所述第二补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0137] 可选的,输出模块包括:

[0138] 第二输出子模块,用于在所述事件名生成模型根据所述至少一个标题子句不能得到事件名的情况下,对所述至少一个标题子句进行补充;根据对所述至少一个标题子句的补充输出所述标题对应的事件名。

[0139] 可选的,所述第二输出子模块,包括:

[0140] 第五输出单元,用于在所述标题子句中不包括主谓结构的情况下,根据所述标题对所述标题子句进行补充,得到第三补充子句;在所述第三补充子句存在歧义的情况下,根据所述标题或所述标题对应的正文对所述第三补充子句进行补充,得到第四补充子句。

[0141] 可选的,所述第二输出子模块,包括:

[0142] 第五输出单元,用于在所述第四补充子句语义不完整,且所述第四补充子句中包含事件子句的情况下,对所述第四补充子句进行补充和/或改写,得到第五补充子句;确定所述第五补充子句为所述标题对应的事件名,并输出所述标题对应的事件名。

[0143] 可选的,所述切分模块包括:

[0144] 切分子模块,用于在所述标题内包括的事件不大于两个的情况下,将所述标题进行切分得到至少一个标题子句。

[0145] 本发明实施例中提供了一种事件名生成方法及装置,在将获取的标题切分为标题子句后,进一步将该标题子句输入到预先训练的事件名生成模型,事件名生成模型可以根据标题子句自动输出该标题对应的事件名,整个过程不需要人工参与,生成事件名的效率较高。

[0146] 本发明各实施例提供的事件名生成装置可用于执行如前述各对应的实施例所示的方法,其实现方式与原理相同,不再赘述。

[0147] 本发明实施例还提供一种电子设备,包括:处理器,存储器以及计算机程序;其中,所述计算机程序被存储在所述存储器中,并且被配置为由所述处理器执行,所述计算机程序包括用于执行如前述实施例中任一项所述的方法的指令。

[0148] 本发明实施例还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被执行时,实现如前述实施例中任一项所述的方法。

[0149] 本领域普通技术人员可以理解:实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时,执行包括上述各方法实施例的步骤;而前述的存储介质包括:ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0150] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依

然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

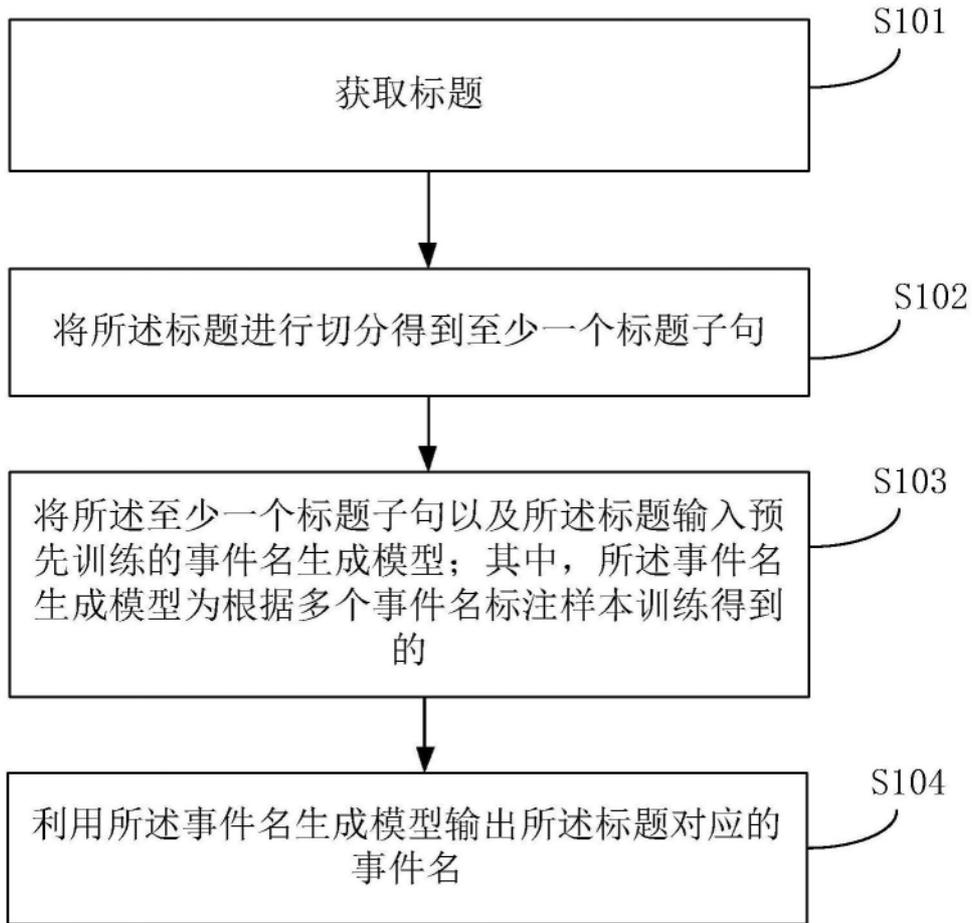


图1

事件名-标注

剩余事件: 11530个; 已审事件: 4367个; 非事件: 103个

当前事件@id: 48a44544eb7992f32ddf78c1d50b23fd

| 序号 | 标题 |
|----|---------------------------|
| 1 | 希腊山火已致91遇难 民众为活命被迫跳入海中 |
| 2 | 希腊山火已致91遇难 民众为活命被迫跳入海中(图) |

| 序号 | 候选事件名 | 评审 |
|----|-------------|--|
| 1 | 民众为活命被迫跳入海中 | <input type="radio"/> 1分 <input type="radio"/> 0.5分 <input type="radio"/> 0分 |
| 2 | 希腊山火已致91遇难 | <input type="radio"/> 1分 <input type="radio"/> 0.5分 <input type="radio"/> 0分 |

不是事件, 跳过 确定

图2

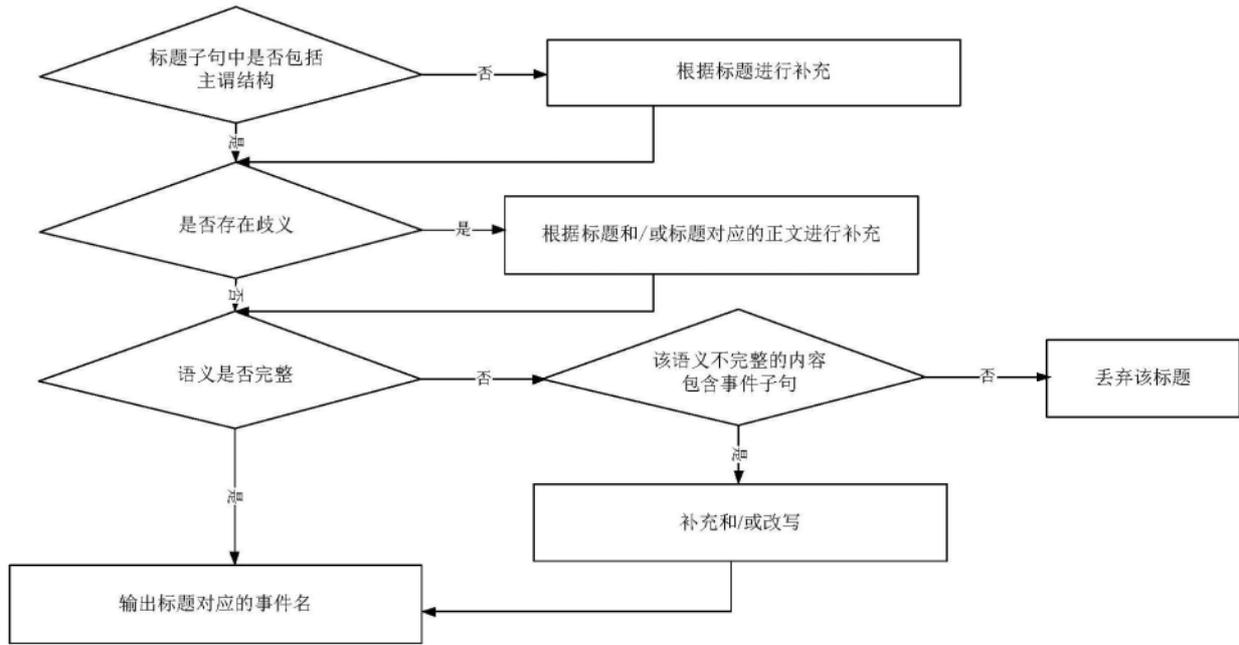


图3

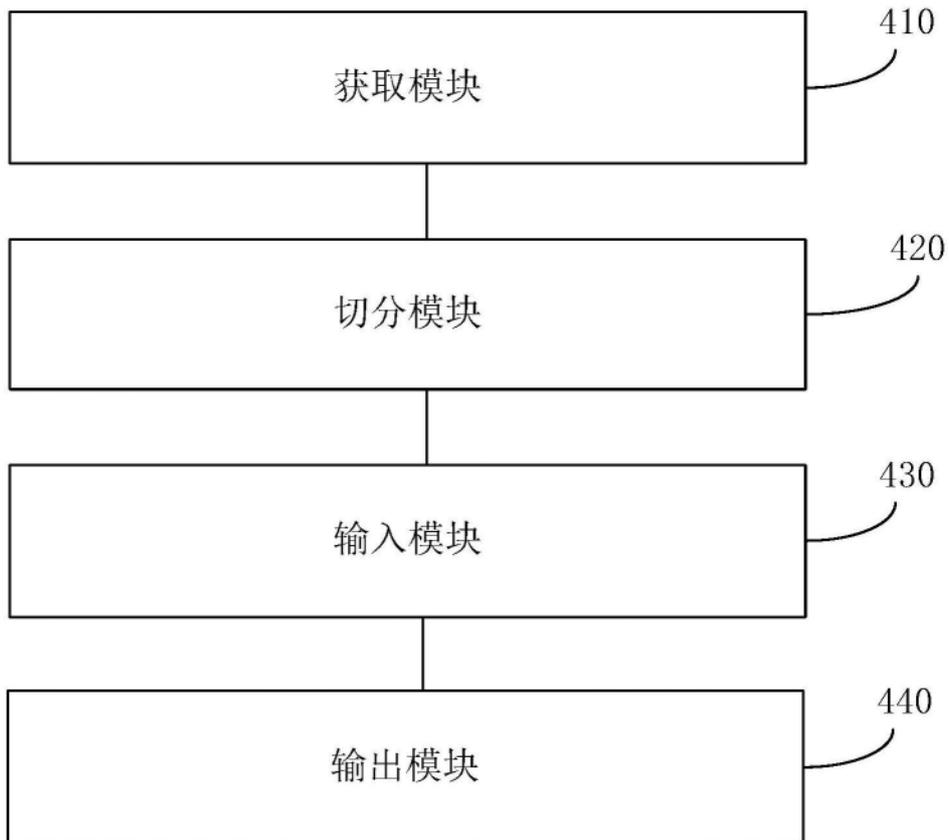


图4