(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111401033 B
(45)授权公告日 2023.07.25

(21)申请号 202010195577.8

(22)申请日 2020.03.19

(65)同一申请的已公布的文献号
申请公布号 CN 111401033 A

(43)申请公布日 2020.07.10

(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

(72)发明人 潘禄 陈玉光 李法远 韩翠云 刘远圳 黄佳艳

(74)专利代理机构 北京银龙知识产权代理有限公司 11243
专利代理师 黄灿 胡永芳

(51)Int.Cl.
G06F 40/205(2020.01)
G06F 40/289(2020.01)
G06F 40/295(2020.01)

(56)对比文件
CN 110704598 A,2020.01.17
冯骁骋.基于标示学习的信息抽取技术研究.《中国博士论文全文数据库信息科技辑》.2019,全文.
郑文.面向交互式问答的人物事件关系抽取方法研究.《中国优秀硕士论文全文数据库信息科技辑》.2016,全文.
min-kyoung kim.design of question answering system with automated question generation.《4th international conference on networked compution & advanced information management》.2008,全文.

审查员 宋宇

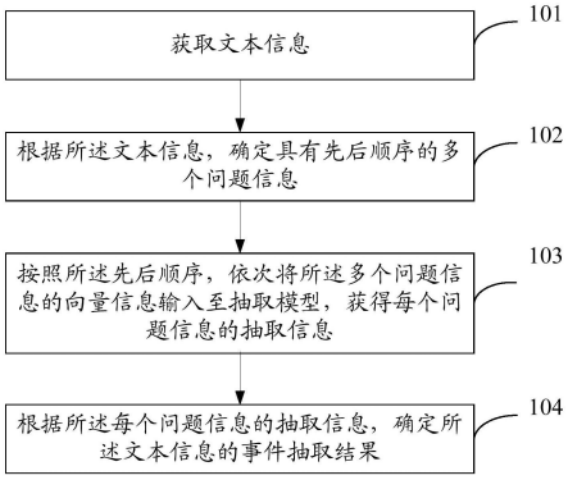
权利要求书4页 说明书16页 附图3页

(54)发明名称

事件抽取方法、事件抽取装置和电子设备

(57)摘要

本申请公开了事件抽取方法、事件抽取装置和电子设备，涉及计算机数据处理领域，尤其涉及知识图谱领域。具体实现方案为：获取文本信息；根据文本信息，确定具有先后顺序的多个问题信息；按照先后顺序，依次将多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息。在通过抽取模型对文本信息进行抽取时，依次对多个问题信息进行抽取，因为每个问题信息对应的答案不同，因此，在先前抽取的问题信息的答案基础上确定在后抽取的问题信息的答案时，可将先前抽取的问题信息的答案排除掉，缩小在后抽取的问题信息的抽取信息的范围，提高了信息抽取速度和准确率。



CN 111401033 B

1.一种事件抽取方法,其特征在于,包括:

获取文本信息;

根据所述文本信息,确定具有先后顺序的多个问题信息;

按照所述先后顺序,依次将所述多个问题信息的向量信息输入至抽取模型,获得每个问题信息的抽取信息,其中,所述向量信息包括答案标记向量;

根据所述每个问题信息的抽取信息及所述问题信息的先后顺序,依次获取各个问题信息的答案,以确定所述文本信息的事件抽取结果;其中,在确定在后问题信息的答案时,排除位于所述在后问题信息之前的问题信息的答案。

2.根据权利要求1所述的事件抽取方法,其特征在于,对于所述多个问题信息中排序在第一的第一问题信息,所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定;

对于所述多个问题信息中排序在第一问题信息之后的第二问题信息,所述第二问题信息的答案标记向量,根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。

3.根据权利要求1所述的事件抽取方法,其特征在于,所述向量信息还包括位置向量;

在所述根据所述文本信息,确定具有先后顺序的多个问题信息之后,所述按照所述先后顺序,依次将所述多个问题信息中的各个问题信息的向量信息输入至抽取模型,获得多个抽取信息之前,还包括:

对于所述多个问题信息的每一个问题信息,对所述问题信息进行分词处理,获得至少一个目标词;

获取所述至少一个目标词中每一个目标词的位置向量;

根据所述至少一个目标词中每一个目标词的位置向量,确定所述问题信息的位置向量。

4.根据权利要求3所述的事件抽取方法,其特征在于,所述获取所述至少一个目标词中每一个目标词的位置向量,包括:

若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,所述M和所述N均为正整数;

对于所述至少一个目标词中的每一个目标词,分别计算所述目标词到所述M个实体的M个第一相对位置,以及所述目标词到所述N个动词的N个第二相对位置;

分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上,获得M个第一位置向量和N个第二位置向量;

将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接,获得第一拼接向量;

将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接,获得第二拼接向量;

将所述第一拼接向量和所述第二拼接向量进行拼接,并将拼接结果作为所述目标词的位置向量。

5.根据权利要求4所述的事件抽取方法,其特征在于,所述若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中

的M个实体和N个动词,包括:

若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N,或者,若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M,则对所述问题信息进行句法依存分析,获得多个依存对;

选择所述多个依存对中包括在同一个依存对中的实体和动词,获得m个实体和n个动词,所述m和所述n均为正整数;

若所述m小于所述M,则从所述问题信息的所述m个实体之外的实体中,选择i个实体,以获得i个实体,其中,i为M与m的差值;

若所述n小于所述N,则从所述问题信息的所述n个动词之外的动词中,选择j个动词,以获得j个动词,其中,j为N与n的差值。

6.根据权利要求1所述的事件抽取方法,其特征在于,根据所述文本信息,确定具有先后顺序的多个问题信息,包括:

根据所述文本信息,确定所述文本信息的事件类型;

根据所述事件类型,确定多个论元角色;

分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接,确定多个问题;

分别将所述多个问题中的各个问题与所述文本信息进行拼接,获得所述多个问题信息;

根据所述事件类型中所述各论元角色的顺序,对所述各论元角色对应的问题信息进行排序,确定具有先后顺序的多个问题信息。

7.一种事件抽取装置,其特征在于,包括:

第一获取模块,用于获取文本信息;

第一确定模块,用于根据所述文本信息,确定具有先后顺序的多个问题信息;

第二获取模块,用于按照所述先后顺序,依次将所述多个问题信息的向量信息输入至抽取模型,获得每个问题信息的抽取信息,其中,所述向量信息包括答案标记向量;

第二确定模块,用于根据所述每个问题信息的抽取信息及所述问题信息的先后顺序,依次获取各个问题信息的答案,以确定所述文本信息的事件抽取结果;其中,在确定在后问题信息的答案时,排除位于所述在后问题信息之前的问题信息的答案。

8.根据权利要求7所述的事件抽取装置,其特征在于,对于所述多个问题信息中排序在第一的第一问题信息,所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定;

对于所述多个问题信息中排序在第一问题信息之后的第二问题信息,所述第二问题信息的答案标记向量,根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。

9.根据权利要求7所述的事件抽取装置,其特征在于,所述向量信息还包括位置向量;

所述装置还包括:

分词模块,用于对于所述多个问题信息的每一个问题信息,对所述问题信息进行分词处理,获得至少一个目标词;

第三获取模块,用于获取所述至少一个目标词中每一个目标词的位置向量;

第三确定模块,用于根据所述至少一个目标词中每一个目标词的位置向量,确定所述

问题信息的位置向量。

10.根据权利要求9所述的事件抽取装置,其特征在于,所述第三获取模块,包括:

第一获取子模块,用于若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,所述M和所述N均为正整数;

计算子模块,用于对于所述至少一个目标词中的每一个目标词,分别计算所述目标词到所述M个实体的M个第一相对位置,以及所述目标词到所述N个动词的N个第二相对位置;

映射子模块,用于分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上,获得M个第一位置向量和N个第二位置向量;

第二获取子模块,用于将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接,获得第一拼接向量;

第三获取子模块,用于将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接,获得第二拼接向量;

第四获取子模块,用于将所述第一拼接向量和所述第二拼接向量进行拼接,并将拼接结果作为所述目标词的位置向量。

11.根据权利要求10所述的事件抽取装置,其特征在于,所述第一获取子模块,包括:

第一获取单元,用于若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N,或者,若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M,则对所述问题信息进行句法依存分析,获得多个依存对;

第二获取单元,用于选择所述多个依存对中包括在同一个依存对中的实体和动词,获得m个实体和n个动词,所述m和所述n均为正整数;

第三获取单元,用于若所述m小于所述M,则从所述问题信息的所述m个实体之外的实体中,选择i个实体,以获得i个实体,其中,i为M与m的差值;

第四获取单元,用于若所述n小于所述N,则从所述问题信息的所述n个动词之外的动词中,选择j个动词,以获得j个动词,其中,j为N与n的差值。

12.根据权利要求7所述的事件抽取装置,其特征在于,所述第一获取模块,包括:

第一确定子模块,用于根据所述文本信息,确定所述文本信息的事件类型;

第二确定子模块,用于根据所述事件类型,确定多个论元角色;

第三确定子模块,用于分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接,确定多个问题;

拼接子模块,用于分别将所述多个问题中的各个问题与所述文本信息进行拼接,获得所述多个问题信息;

第四确定子模块,用于根据所述事件类型中所述各论元角色的顺序,对所述各论元角色对应的问题信息进行排序,确定具有先后顺序的多个问题信息。

13.一种电子设备,其特征在于,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处

理器执行,以使所述至少一个处理器能够执行权利要求1-6中任一项所述的方法。

14.一种存储有计算机指令的非瞬时计算机可读存储介质,其特征在于,所述计算机指令用于使所述计算机执行权利要求1-6中任一项所述的方法。

事件抽取方法、事件抽取装置和电子设备

技术领域

[0001] 本申请涉及计算机技术领域中的数据处理技术，尤其涉及一种事件抽取方法、事件抽取装置和电子设备。

背景技术

[0002] 信息抽取在海量的数据处理中有着非常重要的作用，其中，事件抽取是信息抽取领域的一个重要研究方向，事件抽取任务是从文本中抽取结构化的事件信息，包括事件的类型、触发词、事件的论元角色。事件抽取应用很广，在金融领域，可以定位、定量、定性分析金融中的各项活动，极大的解决人力资源；在医疗领域，通过诊断说明书以及病患的症状表述，可以很快锁定病患的疾病情况，可以让患者对病状的了解更加明确。

[0003] 但是目前的事件抽取方法，抽取效果差，事件抽取获得的结构化细信息准确率低。

发明内容

[0004] 本申请实施例提供一种事件抽取方法、事件抽取装置和电子设备，以解决事件抽取获得的结构化细信息准确率低的问题。

[0005] 为解决上述技术问题，本申请是这样实现的：

[0006] 本申请第一方面提供一种事件抽取方法，包括：

[0007] 获取文本信息；

[0008] 根据所述文本信息，确定具有先后顺序的多个问题信息；

[0009] 按照所述先后顺序，依次将所述多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息，其中，所述向量信息包括答案标记向量；

[0010] 根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果。

[0011] 进一步的，对于所述多个问题信息中排序在第一的第一问题信息，所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定；对于所述多个问题信息中排序在第一问题信息之后的第二问题信息，所述第二问题信息的答案标记向量，根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。

[0012] 进一步的，所述向量信息还包括位置向量；

[0013] 在所述根据所述文本信息，确定具有先后顺序的多个问题信息之后，所述按照所述先后顺序，依次将所述多个问题信息中的各个问题信息的向量信息输入至抽取模型，获得多个抽取信息之前，还包括：

[0014] 对于所述多个问题信息的每一个问题信息，对所述问题信息进行分词处理，获得至少一个目标词；

[0015] 获取所述至少一个目标词中每一个目标词的位置向量；

[0016] 根据所述至少一个目标词中每一个目标词的位置向量，确定所述问题信息的位置向量。

[0017] 进一步的，所述获取所述至少一个目标词中每一个目标词的位置向量，包括：

[0018] 若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,所述M和所述N均为正整数;

[0019] 对于所述至少一个目标词中的每一个目标词,分别计算所述目标词到所述M个实体的M个第一相对位置,以及所述目标词到所述N个动词的N个第二相对位置;

[0020] 分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上,获得M个第一位置向量和N个第二位置向量;

[0021] 将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接,获得第一拼接向量;

[0022] 将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接,获得第二拼接向量;

[0023] 将所述第一拼接向量和所述第二拼接向量进行拼接,并将拼接结果作为所述目标词的位置向量。

[0024] 进一步的,所述若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,包括:

[0025] 若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N,或者,若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M,则对所述问题信息进行句法依存分析,获得多个依存对;

[0026] 选择所述多个依存对中包括在同一个依存对中的实体和动词,获得m个实体和n个动词,所述m和所述n均为正整数;

[0027] 若所述m小于所述M,则从所述问题信息的所述m个实体之外的实体中,选择i个实体,以获得i个实体,其中,i为M与m的差值;

[0028] 若所述n小于所述N,则从所述问题信息的所述n个动词之外的动词中,选择j个动词,以获得j个动词,其中,j为N与n的差值。

[0029] 进一步的,根据所述文本信息,确定具有先后顺序的多个问题信息,包括:

[0030] 根据所述文本信息,确定所述文本信息的事件类型;

[0031] 根据所述事件类型,确定多个论元角色;

[0032] 分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接,确定多个问题;

[0033] 分别将所述多个问题中的各个问题与所述文本信息进行拼接,获得所述多个问题信息;

[0034] 根据所述事件类型中所述各论元角色的顺序,对所述各论元角色对应的问题信息进行排序,确定具有先后顺序的多个问题信息。

[0035] 本申请第二方面提供一种事件抽取装置,包括:

[0036] 第一获取模块,用于获取文本信息;

[0037] 第一确定模块,用于根据所述文本信息,确定具有先后顺序的多个问题信息;

[0038] 第二获取模块,用于按照所述先后顺序,依次将所述多个问题信息的向量信息输入至抽取模型,获得每个问题信息的抽取信息,其中,所述向量信息包括答案标记向量;

[0039] 第二确定模块,用于根据所述每个问题信息的抽取信息,确定所述文本信息的事件抽取结果。

[0040] 进一步的,对于所述多个问题信息中排序在第一的第一问题信息,所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定;对于所述多个问题信息中排序在第一问题信息之后的第二问题信息,所述第二问题信息的答案标记向量,根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。

[0041] 进一步的,所述向量信息还包括位置向量;

[0042] 所述装置还包括:

[0043] 分词模块,用于对于所述多个问题信息的每一个问题信息,对所述问题信息进行分词处理,获得至少一个目标词;

[0044] 第三获取模块,用于获取所述至少一个目标词中每一个目标词的位置向量;

[0045] 第三确定模块,用于根据所述至少一个目标词中每一个目标词的位置向量,确定所述问题信息的位置向量。

[0046] 进一步的,所述第三获取模块,包括:

[0047] 第一获取子模块,用于若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,所述M和所述N均为正整数;

[0048] 计算子模块,用于对于所述至少一个目标词中的每一个目标词,分别计算所述目标词到所述M个实体的M个第一相对位置,以及所述目标词到所述N个动词的N个第二相对位置;

[0049] 映射子模块,用于分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上,获得M个第一位置向量和N个第二位置向量;

[0050] 第二获取子模块,用于将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接,获得第一拼接向量;

[0051] 第三获取子模块,用于将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接,获得第二拼接向量;

[0052] 第四获取子模块,用于将所述第一拼接向量和所述第二拼接向量进行拼接,并将拼接结果作为所述目标词的位置向量。

[0053] 进一步的,所述第一获取子模块,包括:

[0054] 第一获取单元,用于若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N,或者,若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M,则对所述问题信息进行句法依存分析,获得多个依存对;

[0055] 第二获取单元,用于选择所述多个依存对中包括在同一个依存对中的实体和动词,获得m个实体和n个动词,所述m和所述n均为正整数;

[0056] 第三获取单元,用于若所述m小于所述M,则从所述问题信息的所述m个实体之外的实体中,选择i个实体,以获得i个实体,其中,i为M与m的差值;

[0057] 第四获取单元,用于若所述n小于所述N,则从所述问题信息的所述n个动词之外的动词中,选择j个动词,以获得j个动词,其中,j为N与n的差值。

[0058] 进一步的,所述第一获取模块,包括:

[0059] 第一确定子模块,用于根据所述文本信息,确定所述文本信息的事件类型;

[0060] 第二确定子模块，用于根据所述事件类型，确定多个论元角色；

[0061] 第三确定子模块，用于分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接，确定多个问题；

[0062] 拼接子模块，用于分别将所述多个问题中的各个问题与所述文本信息进行拼接，获得所述多个问题信息；

[0063] 第四确定子模块，用于根据所述事件类型中所述各论元角色的顺序，对所述各论元角色对应的问题信息进行排序，确定具有先后顺序的多个问题信息。

[0064] 本申请第三方面提供一种电子设备，其特征在于，包括：

[0065] 至少一个处理器；

[0066] 以及与所述至少一个处理器通信连接的存储器；

[0067] 其中，所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行第一方面所述的方法。

[0068] 本申请第四方面提供一种存储有计算机指令的非瞬时计算机可读存储介质，所述计算机指令用于使所述计算机执行第一方面所述的方法。

[0069] 上述申请中的一个实施例具有如下优点或有益效果：

[0070] 在通过抽取模型对文本信息进行抽取时，依次对多个问题信息进行抽取，因为每个问题信息对应的答案不同，因此，在先前抽取的问题信息的答案基础上确定在后抽取的问题信息的答案时，可将先前抽取的问题信息的答案排除掉，缩小在后抽取的问题信息的抽取信息的范围，提高了信息抽取速度和准确率。

[0071] 在通过抽取模型对文本信息进行抽取，获得第二问题信息的抽取信息时，考虑了排序在第二问题信息之前的至少一个问题信息的抽取信息，因为每个问题信息对应的答案不同，因此，在其他问题信息的答案基础上确定第二问题信息时，可将其他问题信息的答案排除掉，缩小为第二问题信息确定抽取信息时的范围，提高了信息抽取速度和准确率。

[0072] 在根据所述文本信息，确定具有先后顺序的多个问题信息时，根据文本信息的事件类型包括的多个论元角色，来构建多个问题信息，并根据多个论元角色在事件类型中的顺序，来确定多个问题信息的顺序，便于后续按照多个问题信息的顺序，依次确定问题信息的答案，并在确定当前的问题信息的答案时，排除掉排在该问题信息之前的其他问题信息的答案，缩小当前的问题信息的答案的查找范围，提高效率和准确率。

[0073] 向量信息还包括问题信息的位置向量，对于所述多个问题信息的每一个问题信息，对所述问题信息进行分词处理，获得至少一个目标词；获取所述至少一个目标词中每一个目标词的位置向量；根据所述至少一个目标词中每一个目标词的位置向量，确定所述问题信息的位置向量。问题信息的位置向量中，包含了问题信息的每一个目标词分别与问题信息的实体和动词之间的相对位置，充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0074] 目标词的位置向量中，包含了目标词分别与问题信息的实体和动词之间的相对位置，使得问题信息的位置向量可充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0075] 在所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N，或者，若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实

体个数大于或等于所述M,则对所述问题信息进行句法依存分析,获得多个依存对,并优先选择所述多个依存对中包括在同一个依存对中的实体和动词,以提高后续信息抽取的准确率。

[0076] 上述可选方式所具有的其他效果将在下文中结合具体实施例加以说明。

附图说明

[0077] 附图用于更好地理解本方案,不构成对本申请的限定。其中:

[0078] 图1是本申请实施例提供的事件抽取方法的流程图;

[0079] 图2是本申请实施例提供的抽取模型的各层结构示意图;

[0080] 图3是本申请提供的根据论元角色确定的多个问题信息的处理顺序示意图;

[0081] 图4是本申请实施例提供的事件抽取装置的结构图;

[0082] 图5是用来实现本申请实施例的事件抽取方法的电子设备的框图。

具体实施方式

[0083] 以下结合附图对本申请的示范性实施例做出说明,其中包括本申请实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本申请的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0084] 参见图1,图1是本申请实施例提供的事件抽取方法的流程图,如图1所示,本实施例提供一种事件抽取方法,应用于电子设备,包括以下步骤:

[0085] 步骤101、获取文本信息。

[0086] 文本信息可为用户输入的信息,也可为搜索引擎根据用户输入的查询信息进行查询后获得的搜索结果信息。文本信息可为一段文字。

[0087] 步骤102、根据所述文本信息,确定具有先后顺序的多个问题信息。

[0088] 事件抽取一般包括对触发词(event trigger)的抽取和对论元(event argument)的抽取。通过文本信息的触发词,可以定义事件类型。每种事件类型下会有相应的角色(role),即论元角色。

[0089] 根据文本信息,确定多个问题信息,例如,可根据文本信息的事件类型中的论元角色来确定问题信息,每个论元角色确定一个问题信息,根据事件类型中各个论元角色的顺序,确定各个问题信息的顺序。

[0090] 步骤103、按照所述先后顺序,依次将所述多个问题信息的向量信息输入至抽取模型,获得每个问题信息的抽取信息,所述向量信息包括答案标记向量。

[0091] 对于所述多个问题信息中排序在第一的第一问题信息,所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定;对于所述多个问题信息中排序在第一问题信息之后的第二问题信息,所述第二问题信息的答案标记向量,根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。这样,获得第二问题信息的抽取信息时,考虑了排序在第二问题信息之前的至少一个问题信息的抽取信息,因为每个问题信息对应的答案不同,因此,在其他问题信息的答案基础上确定第二问题信息时,可将其他问题信息的答案排除掉,缩小为第二问题信息确定抽取信息时的范围,提高了信息抽取速度和准确率。

[0092] 问题信息的答案标记向量根据排在该问题信息前的其他问题信息的抽取信息确定。根据问题信息对应的抽取信息，可确定问题信息对应的答案，该答案为文本信息中的文字。

[0093] 由于多个问题信息具有先后顺序，因此排序中从第二位开始的答案标记向量可由排在该问题信息前的其他问题信息的抽取信息确定。例如多个问题信息为依次排序的第一问题信息、第二问题信息和第三问题信息。第二问题信息的答案标记向量根据第一问题信息的抽取信息确定，第三问题信息的答案标记向量根据第一问题信息和第二问题信息的抽取信息确定。

[0094] 对于排序在首位的问题信息的答案标记向量，可根据文本信息的初始标记确定，例如，文本信息中各个字的初始标记为0，根据文本信息的初始标记可确定第一问题信息的答案标记向量。若某个字属于排在第二问题信息之前的问题信息的答案，则将文本信息中该字标记为1(不属于排在第二问题信息之前的问题信息的答案部分仍旧标记为0)，然后根据标记，确定第二问题信息的答案标记向量。也就是说，依次将多个问题信息的向量信息输入至抽取模型，每次向抽取模型输入一个问题信息的向量信息，抽取模型输出该问题信息的抽取信息，这可视为一轮抽取。抽取信息可包括答案在文本信息中的开始位置和结束位置，根据抽取信息定位到文本信息的具体位置，获得抽取内容(包括字、词组等)，抽取内容可视为问题信息的答案。

[0095] 对文本信息中前一轮抽取中获得的答案进行标记，例如，对于文本信息中属于前轮问题信息的答案的词(例如在词的下方标记)，标记为1，不属于前轮问题信息的答案标记为0，根据标记后的文本信息，获得本来问题新的答案标记向量；

[0096] 抽取信息可为问题信息的答案在文本信息中的开始位置和结束位置，根据抽取信息可确定文本信息中的具体位置，从而确定出问题信息的答案。

[0097] 步骤104、根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果。

[0098] 根据每个问题信息的抽取信息，获取到问题信息的答案，各个问题信息的答案即为文本信息的事件抽取结果。

[0099] 抽取模型的训练样本可为根据训练语料获得的向量信息，利用根据训练语料获得的向量信息对神经网络模型进行训练，获得抽取模型。根据训练语料获得向量信息的方式，与本申请中根据文本信息获得向量信息的方式一致，在此不做赘述。

[0100] 本实施例中，获取文本信息；根据所述文本信息，确定具有先后顺序的多个问题信息；按照所述先后顺序，依次将所述多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息，其中，所述向量信息包括答案标记向量；根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果。在通过抽取模型对文本信息进行抽取，依次对多个问题信息进行抽取，因为每个问题信息对应的答案不同，因此，在先前抽取的问题信息的答案基础上确定在后抽取的问题信息的答案时，可将先前抽取的问题信息的答案排除掉，缩小在后抽取的问题信息的抽取信息的范围，提高了信息抽取速度和准确率。

[0101] 在本申请一个实施例中，根据所述文本信息，确定具有先后顺序的多个问题信息，包括：

[0102] 根据所述文本信息，确定所述文本信息的事件类型；

[0103] 根据所述事件类型，确定多个论元角色；

[0104] 分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接，确定多个问题；

[0105] 分别将所述多个问题中的各个问题与所述文本信息进行拼接，获得所述多个问题信息；

[0106] 根据所述事件类型中所述各论元角色的顺序，对所述各论元角色对应的问题信息进行排序，确定具有先后顺序的多个问题信息。

[0107] 本实施例提供了根据文本信息确定具有选后顺序的多个问题信息的一种实现方式。

[0108] 首先，根据所述文本信息，确定所述文本信息的事件类型，通过文本信息的触发词，可以定义事件类型。在ACE中大概定义了7个大的事件类型和30个事件子类型(subtype)，而在TAC中定义了8个大的事件类型和15个事件子类型(subtype)。每种事件类型下会有相应的角色(role)，即论元角色。

[0109] 根据每个论元角色可确定一个问题，例如，将事件类型分别与各论元角色进行拼接，获得多个问题。在拼接时，可将事件类型的名称与一个论元角色的名称进行拼接，获得一个问题。进一步的，再将各个问题分别与文本信息进行拼接，获得多个问题信息。在每个问题信息中，在问题的尾部和文本信息的尾部均可采用预设字符进行标识，例如，预设字符可采用[SEP]，以标识末尾位置。问题信息可视为一段文本。

[0110] 根据事件类型确定的多个论元角色之间具有先后顺序，这样，可根据事件类型中所述各论元角色的顺序，对所述各论元角色对应的问题信息进行排序，确定具有先后顺序的多个问题信息。例如，事件类型为A，A有3个论元角色，顺序依次是论元角色1，论元角色2和论元角色3，分别根据论元角色1，论元角色2和论元角色3可构建3个问题信息，分别是问题信息1、问题信息2、问题信息3，那么这3个问题信息的顺序分别为问题信息1、问题信息2、问题信息3。

[0111] 本实施例中，在根据所述文本信息，确定具有先后顺序的多个问题信息时，根据文本信息的事件类型包括的多个论元角色，来构建多个问题信息，并根据多个论元角色在事件类型中的顺序，来确定多个问题信息的顺序，便于后续按照多个问题信息的顺序，依次确定问题信息的答案，并在确定当前的问题信息的答案时，排除掉排在该问题信息之前的其他问题信息的答案，缩小当前的问题信息的答案的查找范围，提高效率和准确率。

[0112] 在本申请一个实施例中，所述向量信息还包括位置向量；

[0113] 在所述根据所述文本信息，确定具有先后顺序的多个问题信息之后，所述按照所述先后顺序，依次将所述多个问题信息中的各个问题信息的向量信息输入至抽取模型，获得多个抽取信息之前，还包括：

[0114] 对于所述多个问题信息的每一个问题信息，对所述问题信息进行分词处理，获得至少一个目标词；

[0115] 获取所述至少一个目标词中每一个目标词的位置向量；

[0116] 根据所述至少一个目标词中每一个目标词的位置向量，确定所述问题信息的位置向量。

[0117] 问题信息可视为一段文本，对于每个问题信息，都可以获取其位置向量，获取方法

为:首先,对问题信息进行分词,可获得一个或多个词,分词处理后获得的每个词都可为目标词。例如,若问题信息为“小明来到唐人街”,分词处理获得三个词:“小明”、“来到”、“唐人街”,这三个词即为三个目标词。然后,对每个目标词均获取位置向量,一个目标词获取一个位置向量。目标词的位置向量根据目标词分别与问题信息中的实体和动词确定,例如在问题信息中,目标词与实体之间的距离,目标词与动词之间的距离。问题信息中表示人物、组织、地点或者机构等的词可视为实体。最后,将每一个目标词的位置向量进行拼接,可获得一个问题信息的位置向量。例如,问题信息包括2个目标词,这2个目标词的位置向量分别为A和B,将A和B进行字符拼接,可获得问题信息的位置向量,此处采用符号A和B表示两个位置向量,并不限定A和B为位置向量的表达式。对于每一个问题信息,重复上述获取过程,可获得每一个问题信息对应的位置向量。

[0118] 问题信息的向量信息还可以包括词向量和词性向量。词向量的获取过程包括:将目标词输入至无监督模型中,获得目标词的词向量,无监督模型的训练样本可包括新闻标题和正文,然后将各个目标词的词向量进行拼接,获得问题信息的词向量。词性向量(POS Embedding)是指将目标词的词性映射为一个多维向量,相同的词性使用相同的向量初始化,在识别模型训练中,词性向量会根据训练语料和目标不同进行值的优化,然后将各个目标词的词性向量进行拼接,获得问题信息的词性向量。

[0119] 进一步的,问题信息的向量信息还可以包括名词向量和指代词向量。通过语言工具提取出名词,名词可以为实体名词,如人物、机构、地方;指代词通过规则获取,如他、她、它等。

[0120] 本实施例中,向量信息还包括问题信息的位置向量,对于所述多个问题信息的每一个问题信息,对所述问题信息进行分词处理,获得至少一个目标词;获取所述至少一个目标词中每一个目标词的位置向量;根据所述至少一个目标词中每一个目标词的位置向量,确定所述问题信息的位置向量。问题信息的位置向量中,包含了问题信息的每一个目标词分别与问题信息的实体和动词之间的相对位置,充分利用目标词分别与实体和动词之间的关系,可进一步提高信息抽取的准确率。

[0121] 在本申请一个实施例中,所述获取所述至少一个目标词中每一个目标词的位置向量,包括:

[0122] 若所述问题信息包括的实体个数大于或等于M,且所述问题信息包括的动词个数大于或等于N,则获取所述问题信息中的M个实体和N个动词,所述M和所述N均为正整数;

[0123] 对于所述至少一个目标词中的每一个目标词,分别计算所述目标词到所述M个实体的M个第一相对位置,以及所述目标词到所述N个动词的N个第二相对位置;

[0124] 分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上,获得M个第一位置向量和N个第二位置向量;

[0125] 将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接,获得第一拼接向量;

[0126] 将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接,获得第二拼接向量;

[0127] 将所述第一拼接向量和所述第二拼接向量进行拼接,并将拼接结果作为所述目标词的位置向量。

[0128] 本实施例中，M和N为预设值，可预先设置，例如将M设置为2，N设置为1。优选的M为3，N为2。若问题信息包括的实体个数大于或等于M，且所述问题信息包括的动词个数大于或等于N，也就是说，问题信息包括的实体个数和动词个数都不小于各自的预设值，那么可从问题信息中获取到M个实体和N个动词。

[0129] 对于所述至少一个目标词中的每一个目标词，分别计算所述目标词到所述M个实体的M个第一相对位置，以及所述目标词到所述N个动词的N个第二相对位置。例如，若至少一个目标词包括第一目标词和第二目标词，实体包括第一实体和第二实体，动词包括第一动词，则计算第一目标词与第一实体之间的第一相对位置，以及第一目标词与第二实体之间的第二相对位置，获得2个第一相对位置；计算第一目标词与第一动词之间的第二相对位置，获得1个第二相对位置。

[0130] 同样的，对于第二目标词，计算第二目标词与第一实体之间的第一相对位置，以及第二目标词与第二实体之间的第二相对位置，获得2个第一相对位置；计算第二目标词与第一动词之间的第二相对位置，获得1个第二相对位置。

[0131] 然后对于每个目标词，分别将目标词对应的所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上，获得M个第一位置向量和N个第二位置向量。预设维度可根据实际情况进行设置，在此不做限定。

[0132] 进一步的，将目标词对应的所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接，获得第一拼接向量；将目标词对应的所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接，获得第二拼接向量。拼接可理解为对字符串的拼接，即将M个第一位置向量按照字符串的方式进行首尾拼接。

[0133] 然后，将目标词对应的所述第一拼接向量和所述第二拼接向量进行拼接，并将拼接结果作为所述目标词的位置向量。在本申请中，拼接可理解为将第一拼接向量和第二拼接向量以字符串的形式进行首尾拼接。

[0134] 本实施例中，从问题信息中选择M个实体和N个动词，然后对于至少一个目标词中的每一个目标词，分别计算所述目标词到所述M个实体的M个第一相对位置，以及所述目标词到所述N个动词的N个第二相对位置；然后分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上，获得M个第一位置向量和N个第二位置向量；再将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接，获得第一拼接向量；将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接，获得第二拼接向量；最后将所述第一拼接向量和所述第二拼接向量进行拼接，并将拼接结果作为所述目标词的位置向量。这样，目标词的位置向量中，包含了目标词分别与问题信息的实体和动词之间的相对位置，使得问题信息的位置向量可充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0135] 在本申请一个实施例中，所述若所述问题信息包括的实体个数大于或等于M，且所述问题信息包括的动词个数大于或等于N，则获取所述问题信息中的M个实体和N个动词，包括：

[0136] 若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N，或者，若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M，则对所述问题信息进行句法依存分析，获得多个依存对；

[0137] 选择所述多个依存对中包括在同一个依存对中的实体和动词，获得m个实体和n个动词，所述m和所述n均为正整数；

[0138] 若所述m小于所述M，则从所述问题信息的所述m个实体之外的实体中，选择i个实体，以获得i个实体，其中，i为M与m的差值；

[0139] 若所述n小于所述N，则从所述问题信息的所述n个动词之外的动词中，选择j个动词，以获得j个动词，其中，j为N与n的差值。

[0140] 本实施例中，在问题信息的实体个数大于M，且动词个数不小于N，或者，问题信息的动词个数大于N，且实体个数不小于M时，需要从问题信息的实体和动词中选出M个实体和N个动词。

[0141] 在选择时，优先选择在同一个依存对中的实体和动词，即实体与动词之间直接发生依存关系，构成一个依存对。例如，张三喊李四，让李四呼叫王五，“张三”和“喊”之间有直接关系，在同一个依存对中，“张三”和“呼叫”之间没有直接关系，则优先选择位于同一依存对中的实体“张三”和动词“喊”。

[0142] 在选择完所有位于同一依存对中的实体和动词后，若实体的个数小于M，则从问题信息剩余的实体中选择i个实体，以使得最终选中的实体总个数有M个。在从问题信息剩余的实体中选择i个实体时，可根据剩余实体的重要性来进行选择，或者根据剩余实体在问题信息中的先后顺序来进行选择，在此不做限定。

[0143] 若动词的个数小于N，则从问题信息剩余的动词中选择j个动词，以使得最终选中的动词总个数有N个。在从问题信息剩余的动词中选择j个动词时，可根据剩余动词的重要性得分来进行选择，或者根据剩余动词在问题信息中的先后顺序来进行选择，在此不做限定。

[0144] 本实施例中，在所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N，或者，若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M，则对所述问题信息进行句法依存分析，获得多个依存对，并优先选择所述多个依存对中包括在同一个依存对中的实体和动词，以提高后续信息抽取的准确率。

[0145] 在本申请一个实施例中，获取所述至少一个目标词中每一个目标词的位置信息，包括：

[0146] 对于所述至少一个目标词中的每一个目标词，若所述问题信息包括的实体个数U小于M，则获得所述目标词到所述U个实体的U个第一相对位置，其中，所述U和所述M均为正整数；

[0147] 将所述U个第一相对位置采用0向量初始化，以获得M个第一相对位置；

[0148] 若所述问题信息包括的动词个数V小于N，则获取所述目标词到所述V个动词的V个第二相对位置，其中，所述V和所述N均为正整数；

[0149] 将所述V个第二相对位置采用0向量初始化，以获得N个第二相对位置；

[0150] 分别将所述M个第一相对位置、所述N个第二相对位置映射到所述正态分布向量上，获得M个第一位置向量和N个第二位置向量。

[0151] 将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接，获得第一拼接向量；

[0152] 将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接，获得第二拼接向量；

[0153] 将所述第一拼接向量和所述第二拼接向量进行拼接，并将拼接结果作为所述目标词的位置向量。

[0154] 本实施例为问题信息包括的实体个数或者动词个数小于预设值时的情况。M和N为预设值，可预先设置，优选的M为3，N为2。

[0155] 若所述问题信息包括的实体个数U小于M，则获得所述目标词到所述U个实体的U个第一相对位置，然后将所述U个第一相对位置采用0向量初始化，以获得M个第一相对位置，在初始化时，可采用一个或多个0向量对U个第一相对位置进行填充，获得M个第一相对位置。一个0向量的长度和一个位置向量的长度一样。若所述问题信息包括的实体个数V小于N，则获得所述目标词到所述V个实体的V个第二相对位置，然后将所述V个第二相对位置采用0向量初始化，以获得N个第二相对位置，在初始化时，可采用一个或多个0向量对V个第二相对位置进行填充，获得N个第二相对位置。一个0向量的长度和一个位置向量的长度一样。最后，分别将所述M个第一相对位置、所述N个第二相对位置映射到所述正态分布向量上，获得M个第一位置向量和N个第二位置向量。将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接，获得第一拼接向量；将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接，获得第二拼接向量；将所述第一拼接向量和所述第二拼接向量进行拼接，并将拼接结果作为所述目标词的位置向量。对于问题信息中的所有目标词，均可采用上述方式进行处理，获得各目标词对应的位置向量。

[0156] 本实施例中，在问题信息包括的实体个数或者动词个数小于预设值时，采用0向量对U个第一相对位置或者V个第二相对位置进行初始化，以获得M个第一相对位置和所述N个第二相对位置，并最终获得目标词对应的位置向量，问题信息的位置向量中，包含了问题信息的每一个目标词分别与问题信息的实体和动词之间的相对位置，充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0157] 本实施例中，根据问题信息的获取词向量、位置向量以及答案标记向量，可应用到抽取模型的训练阶段中。图2为抽取模型的各层结构示意图，如图2所示：

[0158] 输入层：输入的是根据构造的问题文档对<问题，文档>获取的词向量、位置向量和答案标记向量。其中，问题由训练语料的事件类型和一个论元角色的名拼接而成的，文档(即训练语料)是潜在包含事件论元答案的内容。一个问题和文档拼接成一个句子(即问题信息)，用[SEP]标识问题和文档的末尾位置。拼接成的句子中，并对句子进行分词处理，获得目标词，每个目标词都有词向量和位置向量，分别根据每个目标词的词向量和位置向量，可获得句子的词向量和位置向量。另外，还获取句子的答案标记向量，即前一轮问题回答的答案在文档中的位置，已回答过的位置标1，未回答标0，然后将标记后的文档转化为向量作为答案标记向量。

[0159] 模型网络：可采用基础的神经网络模型。

[0160] 输出层：本轮问题信息的答案在文档中的开始位置和结束位置。

[0161] 如3所示为事件类型下根据论元角色确定的多个问题信息的处理示意图，如图3所示，问题询问顺序：先询问事件的触发词，再询问事件所有的论元角色，角色顺序是固定的；在当前轮询问时，需要将之前询问过输出的答案进行整合，生成当前询问条件下的历史回

答标记，如果训练语料(或者文本信息)中某个字符为之前询问过程中的答案，则在此字符位置标记为1，否则为0。

[0162] 从图3可以看出，排序在前的论元角色(具体为根据论元角色确定的问题信息)的输出结果(即抽取信息)，影响排序在后的论元角色的输出结果。

[0163] 本申请通过构造以论元角色为基础的问题，利用阅读理解技术学习目标答案，同时，由于相同的答案不会分配多个角色，在获取不同的角色的答案时，将已回答的答案作为特征的一部分，可以帮助抽取模型减少候选项，进一步提升模型效果。通过对文本信息进行事件抽取，得到结构化信息，可以提升电子设备理解文本内容的能力，帮助减少大量信息，进一步提升人工效率。

[0164] 参见图4，图4是本申请实施例提供的事件抽取装置的结构图，如图4所示，本实施例提供一种事件抽取装置400，包括：

[0165] 第一获取模块401，用于获取文本信息；

[0166] 第一确定模块402，用于根据所述文本信息，确定具有先后顺序的多个问题信息；

[0167] 第二获取模块403，用于按照所述先后顺序，依次将所述多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息，其中，所述向量信息包括答案标记向量；

[0168] 第二确定模块404，用于根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果。

[0169] 在本申请一个实施例中，对于所述多个问题信息中排序在第一的第一问题信息，所述第一问题信息的答案标记向量根据所述文本信息的初始标记确定；对于所述多个问题信息中排序在第一问题信息之后的第二问题信息，所述第二问题信息的答案标记向量，根据排在所述第二问题信息之前的至少一个问题信息的抽取信息确定。

[0170] 在本申请一个实施例中，所述向量信息还包括位置向量；

[0171] 所述装置还包括：

[0172] 分词模块，用于对于所述多个问题信息的每一个问题信息，对所述问题信息进行分词处理，获得至少一个目标词；

[0173] 第三获取模块，用于获取所述至少一个目标词中每一个目标词的位置向量；

[0174] 第三确定模块，用于根据所述至少一个目标词中每一个目标词的位置向量，确定所述问题信息的位置向量。

[0175] 在本申请一个实施例中，所述第三获取模块，包括：

[0176] 第一获取子模块，用于若所述问题信息包括的实体个数大于或等于M，且所述问题信息包括的动词个数大于或等于N，则获取所述问题信息中的M个实体和N个动词，所述M和所述N均为正整数；

[0177] 计算子模块，用于对于所述至少一个目标词中的每一个目标词，分别计算所述目标词到所述M个实体的M个第一相对位置，以及所述目标词到所述N个动词的N个第二相对位置；

[0178] 映射子模块，用于分别将所述M个第一相对位置、所述N个第二相对位置映射到预设维度的正态分布向量上，获得M个第一位置向量和N个第二位置向量；

[0179] 第二获取子模块，用于将所述M个第一位置向量按照所述M个实体在所述问题信息中的先后顺序进行拼接，获得第一拼接向量；

[0180] 第三获取子模块，用于将所述N个第二位置向量按照所述N个动词在所述问题信息中的先后顺序进行拼接，获得第二拼接向量；

[0181] 第四获取子模块，用于将所述第一拼接向量和所述第二拼接向量进行拼接，并将拼接结果作为所述目标词的位置向量。

[0182] 在本申请一个实施例中，所述第一获取子模块，包括：

[0183] 第一获取单元，用于若所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N，或者，若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M，则对所述问题信息进行句法依存分析，获得多个依存对；

[0184] 第二获取单元，用于选择所述多个依存对中包括在同一个依存对中的实体和动词，获得m个实体和n个动词，所述m和所述n均为正整数；

[0185] 第三获取单元，用于若所述m小于所述M，则从所述问题信息的所述m个实体之外的实体中，选择i个实体，以获得i个实体，其中，i为M与m的差值；

[0186] 第四获取单元，用于若所述n小于所述N，则从所述问题信息的所述n个动词之外的动词中，选择j个动词，以获得j个动词，其中，j为N与n的差值。

[0187] 在本申请一个实施例中，所述第一获取模块，包括：

[0188] 第一确定子模块，用于根据所述文本信息，确定所述文本信息的事件类型；

[0189] 第二确定子模块，用于根据所述事件类型，确定多个论元角色；

[0190] 第三确定子模块，用于分别将所述事件类型与所述多个论元角色中的各论元角色进行拼接，确定多个问题；

[0191] 拼接子模块，用于分别将所述多个问题中的各个问题与所述文本信息进行拼接，获得所述多个问题信息；

[0192] 第四确定子模块，用于根据所述事件类型中所述各论元角色的顺序，对所述各论元角色对应的问题信息进行排序，确定具有先后顺序的多个问题信息。

[0193] 事件抽取装置400能够实现图1所示的方法实施例中电子设备实现的各个过程，为避免重复，这里不再赘述。

[0194] 本申请实施例的事件抽取装置400，获取文本信息；根据所述文本信息，确定具有先后顺序的多个问题信息；按照所述先后顺序，依次将所述多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息，其中，所述向量信息包括答案标记向量；根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果。在通过抽取模型对文本信息进行抽取时，依次对多个问题信息进行抽取，因为每个问题信息对应的答案不同，因此，在先前抽取的问题信息的答案基础上确定在后抽取的问题信息的答案时，可将先前抽取的问题信息的答案排除掉，缩小在后抽取的问题信息的抽取信息的范围，提高了信息抽取速度和准确率。

[0195] 根据本申请的实施例，本申请还提供了一种电子设备和一种可读存储介质。

[0196] 如图5所示，是根据本申请实施例的事件抽取方法的电子设备的框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装

置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本申请的实现。

[0197] 如图5所示，该电子设备包括：一个或多个处理器501、存储器502，以及用于连接各部件的接口，包括高速接口和低速接口。各个部件利用不同的总线互相连接，并且可以被安装在公共主板上或者根据需要以其它方式安装。处理器可以对在电子设备内执行的指令进行处理，包括存储在存储器中或者存储器上以在外部输入/输出装置(诸如，耦合至接口的显示设备)上显示GUI的图形信息的指令。在其它实施方式中，若需要，可以将多个处理器和/或多条总线与多个存储器和多个存储器一起使用。同样，可以连接多个电子设备，各个设备提供部分必要的操作(例如，作为服务器阵列、一组刀片式服务器、或者多处理器系统)。图5中以一个处理器501为例。

[0198] 存储器502即为本申请所提供的非瞬时计算机可读存储介质。其中，所述存储器存储有可由至少一个处理器执行的指令，以使所述至少一个处理器执行本申请所提供的事件抽取方法。本申请的非瞬时计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行本申请所提供的事件抽取方法。

[0199] 存储器502作为一种非瞬时计算机可读存储介质，可用于存储非瞬时软件程序、非瞬时计算机可执行程序以及模块，如本申请实施例中的事件抽取方法对应的程序指令/模块(例如，附图4所示的第一获取模块401、第一确定模块402、第二获取模块405和第二确定模块404)。处理器501通过运行存储在存储器502中的非瞬时软件程序、指令以及模块，从而执行服务器的各种功能应用以及数据处理，即实现上述方法实施例中的事件抽取方法。

[0200] 存储器502可以包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需要的应用程序；存储数据区可存储根据实现事件抽取方法的电子设备的使用所创建的数据等。此外，存储器502可以包括高速随机存取存储器，还可以包括非瞬时存储器，例如至少一个磁盘存储器件、闪存器件、或其他非瞬时固态存储器件。在一些实施例中，存储器502可选包括相对于处理器501远程设置的存储器，这些远程存储器可以通过网络连接至实现事件抽取方法的电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0201] 实现事件抽取方法的电子设备还可以包括：输入装置503和输出装置504。处理器501、存储器502、输入装置503和输出装置504可以通过总线或者其他方式连接，图5中以通过总线连接为例。

[0202] 输入装置503可接收输入的数字或字符信息，以及产生与实现事件抽取方法的电子设备的用户设置以及功能控制有关的键信号输入，例如触摸屏、小键盘、鼠标、轨迹板、触摸板、指示杆、一个或者多个鼠标按钮、轨迹球、操纵杆等输入装置。输出装置504可以包括显示设备、辅助照明装置(例如，LED)和触觉反馈装置(例如，振动电机)等。该显示设备可以包括但不限于，液晶显示器(LCD)、发光二极管(LED)显示器和等离子体显示器。在一些实施方式中，显示设备可以是触摸屏。

[0203] 此处描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、专用ASIC(专用集成电路)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用

或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0204] 这些计算程序(也称作程序、软件、软件应用、或者代码)包括可编程处理器的机器指令，并且可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。如本文使用的，术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如，磁盘、光盘、存储器、可编程逻辑装置(PLD))，包括，接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0205] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置(例如，CRT(阴极射线管)或者LCD(液晶显示器)监视器)；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0206] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)和互联网。

[0207] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。

[0208] 根据本申请实施例的技术方案，包括如下有益效果：

[0209] 在通过抽取模型对文本信息进行抽取时，依次对多个问题信息进行抽取，因为每个问题信息对应的答案不同，因此，在先前抽取的问题信息的答案基础上确定在后抽取的问题信息的答案时，可将先前抽取的问题信息的答案排除掉，缩小在后抽取的问题信息的抽取信息的范围，提高了信息抽取速度和准确率。

[0210] 在通过抽取模型对文本信息进行抽取，获得第二问题信息的抽取信息时，考虑了排序在第二问题信息之前的至少一个问题信息的抽取信息，因为每个问题信息对应的答案不同，因此，在其他问题信息的答案基础上确定第二问题信息时，可将其他问题信息的答案排除掉，缩小为第二问题信息确定抽取信息时的范围，提高了信息抽取速度和准确率。

[0211] 在根据所述文本信息，确定具有先后顺序的多个问题信息时，根据文本信息的事件类型包括的多个论元角色，来构建多个问题信息，并根据多个论元角色在事件类型中的顺序，来确定多个问题信息的顺序，便于后续按照多个问题信息的顺序，依次确定问题信息的答案，并在确定当前的问题信息的答案时，排除掉排在该问题信息之前的其他问题信息

的答案，缩小当前的问题信息的答案的查找范围，提高效率和准确率。

[0212] 向量信息还包括问题信息的位置向量，对于所述多个问题信息的每一个问题信息，对所述问题信息进行分词处理，获得至少一个目标词；获取所述至少一个目标词中每一个目标词的位置向量；根据所述至少一个目标词中每一个目标词的位置向量，确定所述问题信息的位置向量。问题信息的位置向量中，包含了问题信息的每一个目标词分别与问题信息的实体和动词之间的相对位置，充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0213] 目标词的位置向量中，包含了目标词分别与问题信息的实体和动词之间的相对位置，使得问题信息的位置向量可充分利用目标词分别与实体和动词之间的关系，可进一步提高信息抽取的准确率。

[0214] 在所述问题信息包括的实体个数大于所述M且所述问题信息包括的动词个数大于或等于所述N，或者，若所述问题信息包括的动词个数大于所述N且所述问题信息包括的实体个数大于或等于所述M，则对所述问题信息进行句法依存分析，获得多个依存对，并优先选择所述多个依存对中包括在同一个依存对中的实体和动词，以提高后续信息抽取的准确率。

[0215] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本申请公开的技术方案所期望的结果，本文在此不进行限制。

[0216] 上述具体实施方式，并不构成对本申请保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等，均应包含在本申请保护范围之内。

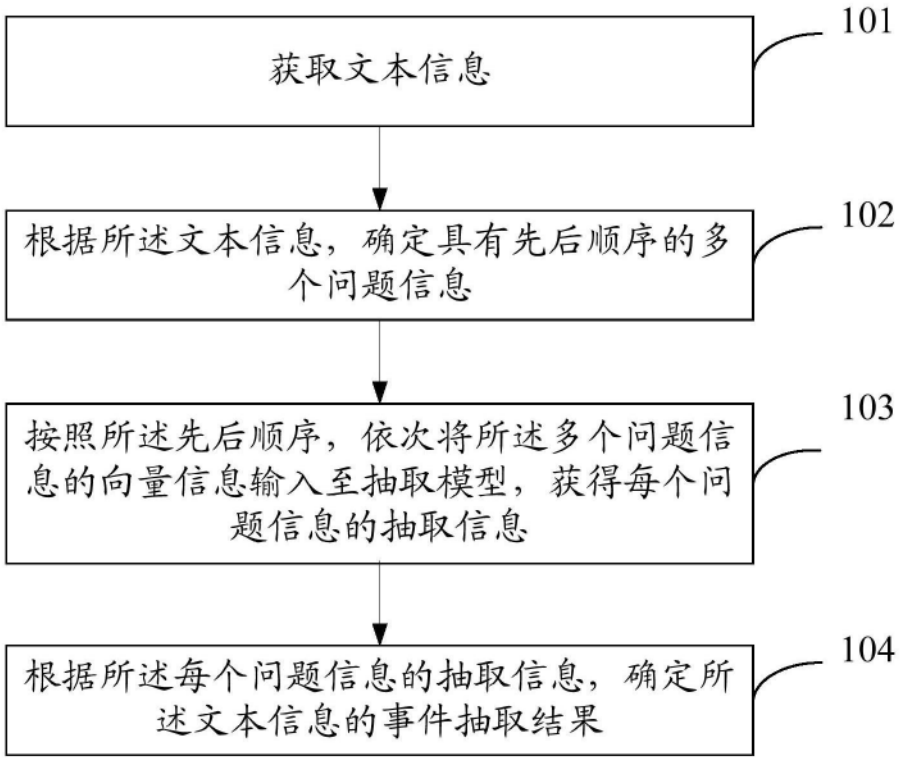

101
获取文本信息
102
根据所述文本信息，确定具有先后顺序的多个问题信息
103
按照所述先后顺序，依次将所述多个问题信息的向量信息输入至抽取模型，获得每个问题信息的抽取信息
104
根据所述每个问题信息的抽取信息，确定所述文本信息的事件抽取结果

图1

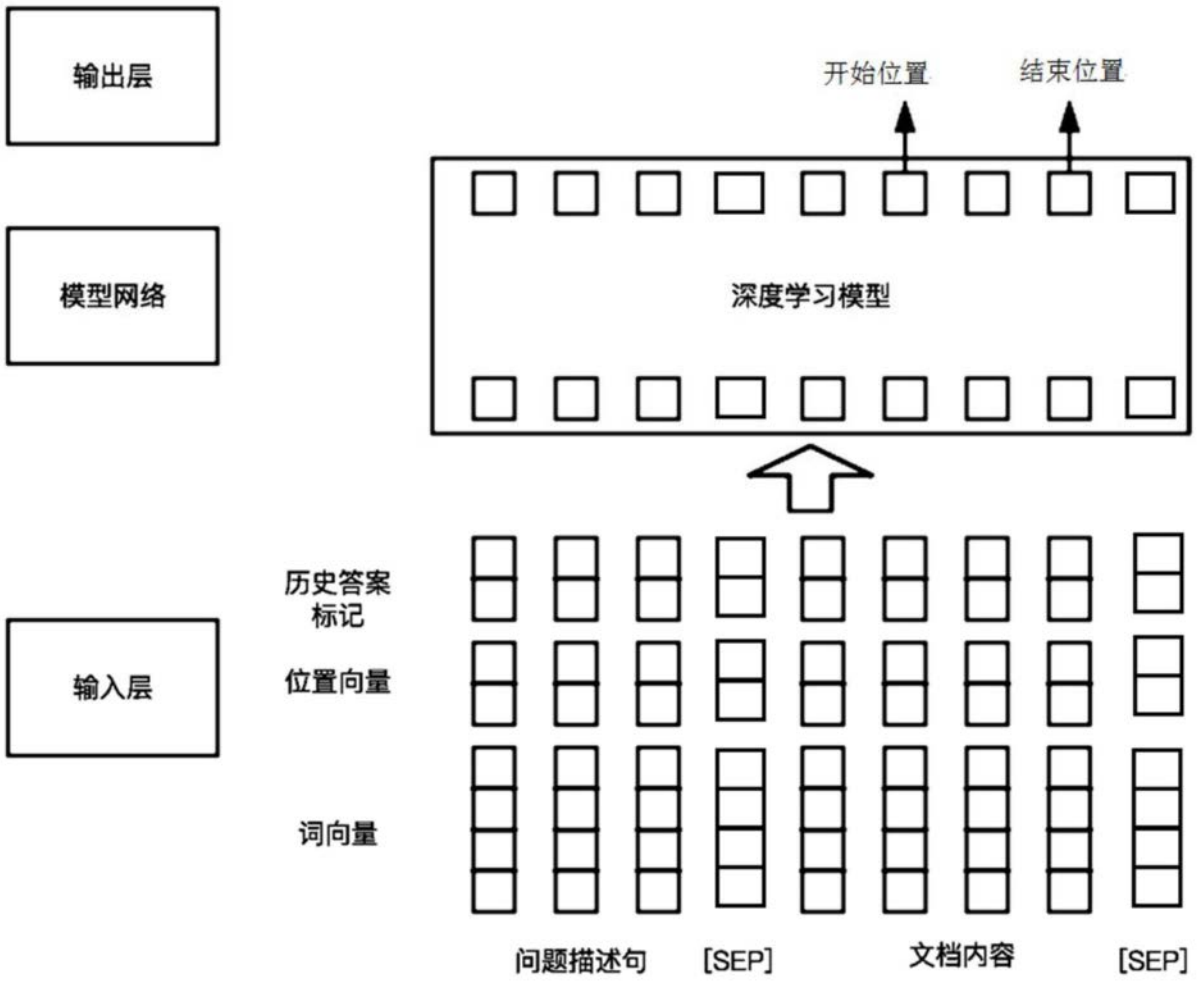

输出层
模型网络
输入层
开始位置
结束位置
深度学习模型
历史答案标记
位置向量
词向量
问题描述句
[SEP]
文档内容
[SEP]

图2

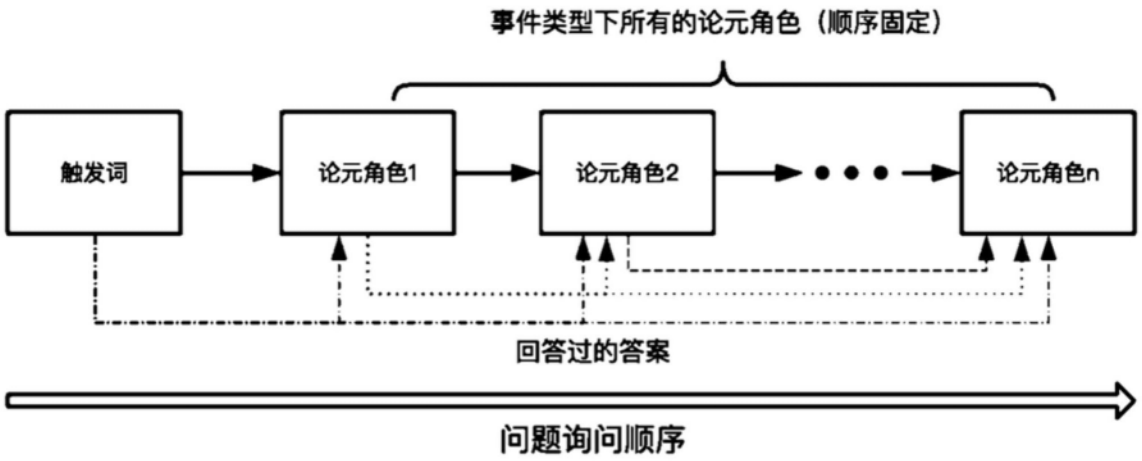

事件类型下所有的论元角色（顺序固定）
触发词
论元角色1
论元角色2
论元角色n
回答过的答案
问题询问顺序

图3

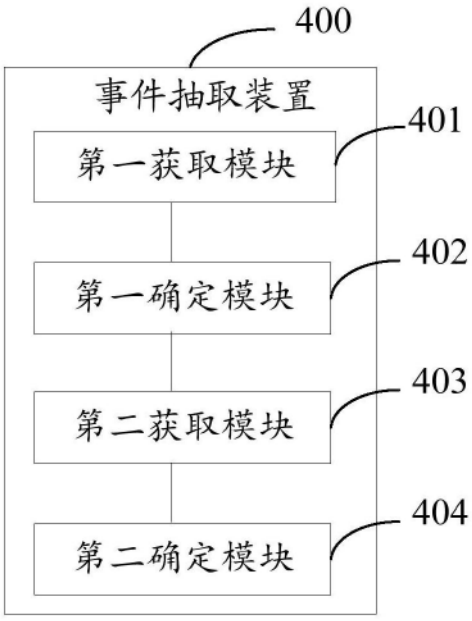

400
事件抽取装置
401
第一获取模块
402
第一确定模块
403
第二获取模块
404
第二确定模块

图4

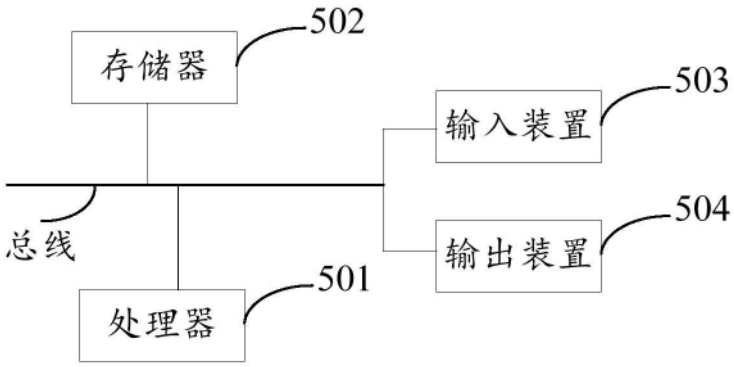

502
存储器
503
输入装置
504
输出装置
总线
501
处理器

图5