



## (12)发明专利申请

(10)申请公布号 CN 110968700 A

(43)申请公布日 2020.04.07

(21)申请号 201911059404.7

(22)申请日 2019.11.01

(71)申请人 数地科技(北京)有限公司

地址 100080 北京市海淀区中关村大街1号  
17层V21号

(72)发明人 不公告发明人

(74)专利代理机构 北京君尚知识产权代理有限公司 11200

代理人 邱晓锋

(51)Int.Cl.

G06F 16/36(2019.01)

G06F 40/247(2020.01)

G06N 5/02(2006.01)

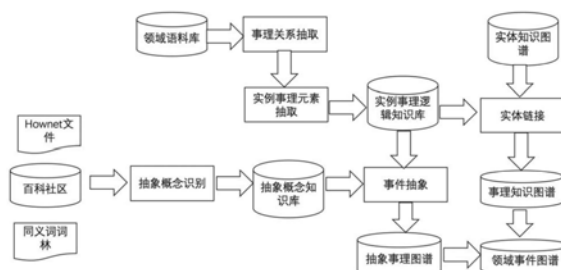
权利要求书2页 说明书8页 附图2页

### (54)发明名称

一种融合多类事理与实体知识的领域事件图谱构建方法和装置

### (57)摘要

本发明涉及一种融合多类事理与实体知识的领域事件图谱构建方法和装置。该方法包括：对领域语料库进行事理关系抽取和实例事理元素抽取，形成实例事理逻辑知识库；构建具有层级性的抽象概念知识库；利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象，形成抽象事理图谱；利用实体知识图谱，采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接，将事件知识与实体知识融合，形成事理知识图谱；将抽象事理图谱与事理知识图谱进行合并，形成领域事件图谱。本发明构建的领域事件图谱，能够综合静态实体性知识以及动作事件性知识两者的优点，可拓宽知识问答的应用范围，并可作为常识知识库扩充领域语言资源。



1. 一种融合多类事理与实体知识的领域事件图谱构建方法,其特征在于,包括以下步骤:

对领域预料库进行事理关系抽取,并利用事件论元抽取技术进行实例事理元素抽取,形成实例事理逻辑知识库;

构建具有层级性的抽象概念知识库;

利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象,形成抽象事理图谱;

利用实体知识图谱,采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接,实现事件知识与实体知识的融合,形成事理知识图谱;

将抽象事理图谱与事理知识图谱进行合并,形成领域事件图谱。

2. 根据权利要求1所述的方法,其特征在于,所述事理关系抽取,包括抽取以下八种事理关系:因果事理、条件事理、上下位事理、并列事理、互斥事理、顺承事理、转折事理、组成事理。

3. 根据权利要求1所述的方法,其特征在于,所述事理关系抽取的方法包括:基于关系词限定、词性过滤与事件成立性的抽取方法;基于关系词与依存句法规则的抽取方法;基于序列标注的事理抽取方法。

4. 根据权利要求1所述的方法,其特征在于,所述实例事理元素抽取,包括抽取以下事理元素:核心动作,指的是事件中的核心动词;动作主体,指围绕核心动词的施事者;动作客体,指围绕核心动词的受事者;事件发生时间,指事件发生时的背景时间;事件发生地点,指事件发生时的背景地点;事件关联人物,指事件发生的关联人物;事件关联公司,指事件发生的关联公司;事件关联机构,指事件发生的机构实体;事件极性,指事件本身的情绪;事件确定性,指事件发生的确定性;事件权威性,指因果事件的权威性;事件发生时态,指事件发生的时态,包括已经发生,将要发生,正在发生三种状态;事件来源,指事件抽取所在的文本来源。

5. 根据权利要求1所述的方法,其特征在于,所述抽象概念知识库是存储概念之间上下位层级的知识库,由名词性实体抽象知识库、性状类抽象知识库以及动作性事件抽象知识库三个知识库构成;通过融合HowNet、同义词词林、百科类层级分类树、百科词条义项和领域词表,形成具有层级性的抽象概念知识库。

6. 根据权利要求1所述的方法,其特征在于,所述抽象事理图谱,是运用事件抽象的方法对实例事理逻辑知识库的事件进行抽象,最终形成具有抽象层级性的抽象事理图谱,包括事件预处理、事件成分抽象和抽象成分融合三个步骤。

7. 根据权利要求1所述的方法,其特征在于,所述事理知识图谱是事件知识与实体知识之间的融合,通过将实例事理逻辑知识库与实体知识图谱之间进行链接,形成从事件知识到实体知识的通路。

8. 根据权利要求1所述的方法,其特征在于,所述领域事件图谱是一个有向有环图,节点代表事件,有向边代表事件之间的事理关系;所述领域事件图谱既关注事件外部之间的演化逻辑关系,又关注事件本身的丰富属性信息,是抽象事件与事件知识库两者的统一结合体。

9. 一种融合多类事理与实体知识的领域事件图谱构建装置,其特征在于,包括:

实例事理逻辑知识库构建模块,负责对领域预料库进行事理关系抽取,并利用事件论元抽取技术进行实例事理元素抽取,形成实例事理逻辑知识库;

抽象概念知识库构建模块,负责构建具有层级性的抽象概念知识库;

抽象事理图谱构建模块,负责利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象,形成抽象事理图谱;

事理知识图谱构建模块,负责利用实体知识图谱,采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接,实现事件知识与实体知识的融合,形成事理知识图谱;

领域事件图谱构建模块,负责将抽象事理图谱与事理知识图谱进行合并,形成领域事件图谱。

10. 一种计算机,其特征在于,包括存储器和处理器,所述存储器存储计算机程序,所述计算机程序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权利要求1~8中任一权利要求所述方法的指令。

## 一种融合多类事理与实体知识的领域事件图谱构建方法和装置

### 技术领域

[0001] 本发明涉及一种融合多类事理与实体知识的领域事件图谱构建方法和装置,属于自然语言处理及知识工程领域,属于一种领域知识库构建方法。

### 背景技术

[0002] 目前,以“事件”为核心,以描述事件之间的逻辑演化关系的逻辑事理图谱正受到广泛关注,这类图谱一方面关注事件之间的空间语义和时间语义状态的转移,能够表达出一种较相关性语义更强、更为直接的关联性;另一方面,知识承载能力更强,事件中的动作具有关联名词性主体和名词性客体,名词性的主客体本身具有丰富的实体知识信息。这两方面的特征决定了事理图谱能够支持事件预测、逻辑问答、逻辑检验等多种应用。

[0003] 高质量的领域事件图谱包括抽象事理图谱、事理知识图谱两个部分组成,在实现技术上涉及到自然语言处理中的事件表示、事件论元抽取、事件关系识别、实体链接等多个方面。其中,抽象事理图谱中的事理较为抽象,旨在挖掘具有普遍推广意义的事件以及事件之间的转移,与此相关的工作包括脚本Script以及叙事性事件链条narrative event chain的自动抽取,并形成了CausalBank以及TimeBank标注语料库。在事件论元抽取上,相关工作主要有ACE (Automatic Content Extraction) 自动内容抽取相关的事件抽取任务。其中论元,指的是事件的组成要素,如“结婚”这一事件中的“结婚时间”、“结婚地点”、“结婚人物”为该事件的论元。

[0004] 受限于当前自然语言处理技术的水平,目前事件逻辑知识库的构建和应用上还存在以下不足:

[0005] 1、全自动抽取性能不高。目前基于全自动方式构建提取的事件及事件关系数据精度不高。以依存句法或正则规则模板抽取的方法能够得到较高的召回率,但错误率较高。以深度学习序列标注方法进行抽取能够保证较高的准确率,但需要大规模的标注语料。因此需要提出一种能够规避三种方法不足的方法。

[0006] 2、事件抽取中的事件元素抽取往往需要提前对元素类别进行预设,不利于大规模事件表示的知识库的构建,一种可接受且语义较为丰富的事件通用论元有提出的必要。

[0007] 3、事件抽取中的关系抽取目前主要集中在因果和顺承两种类型,缺乏对其他类型的抽取尝试和应用点探索。探索更为全面的事件类型体系及其应用点具有实际意义。

[0008] 4、有关事件性知识和实体性知识的融合及链接的工作还相对空白。将实体性知识链接到事件性知识中的事件当中,可进一步丰富事件的语义,拓展应用场景。

### 发明内容

[0009] 针对现有技术中存在的上述技术问题,本发明的目的在于提出一种融合多类事理与实体知识的领域事件图谱构建方法和装置。

[0010] 本发明采用的技术方案如下:

- [0011] 一种融合多类事理与实体知识的领域事件图谱构建方法,包括以下步骤:
- [0012] 对领域预料库进行事理关系抽取,并利用事件论元抽取技术进行实例事理元素抽取,形成实例事理逻辑知识库;
- [0013] 构建具有层级性的抽象概念知识库;
- [0014] 利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象,形成抽象事理图谱;
- [0015] 利用实体知识图谱,采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接,完成事理知识与实体知识的融合,形成事理知识图谱;
- [0016] 将抽象事理图谱与事理知识图谱进行合并,最终形成领域事件图谱。
- [0017] 进一步地,通过设定事理关系,借助依存规则、正则模板的方式抽取事理逻辑,并设定通用事件论元类型,利用事件论元抽取技术进行论元抽取(即进行实例事理元素抽取),形成实例事理逻辑知识库。
- [0018] 进一步地,所述领域事件图谱是一个有向有环图,节点代表事件,有向边代表事件之间的因果、条件、互斥、组成、转折、并发、上下位等逻辑关系,即事理关系。领域事件图谱既关注事件外部之间的演化逻辑关系,又关注事件本身的丰富属性信息,是抽象事件与事件知识库两者的统一结合体。
- [0019] 进一步地,所述领域事件图谱中的多类事理关系包括八种,具体为因果事理、条件事理、上下位事理、并列事理、互斥事理、顺承事理、转折事理、组成事理。
- [0020] 进一步地,所述事理关系的抽取方法指根据指定的事理关系类型,采用基于关系词限定、词性过滤与事件成立性的抽取方法;基于关系词与依存句法规则的抽取方法;基于序列标注的事理抽取方法。
- [0021] 进一步地,所述实例事理元素抽取指的是对事件进行结构化表示。结构化表示的结果为抽取与事件相关的十二种通用论元,论元即事理元素,包括事件核心动作、事件主体、事件客体、事件发生时间、事件发生地点、事件关联实体(关联人物、关联公司、关联机构)、事件情绪、事件时态、事件确定性。
- [0022] 进一步地,所述抽象概念知识库,指的是存储概念之间上下位层级的知识库,由名词性实体抽象知识库、性状类抽象知识库以及动作性事件抽象知识库三个知识库构成,构建思想为:融合HowNet、同义词词林、百科类层级分类树、百科词条义项、领域词表等抽象概念。
- [0023] 进一步地,所述抽象事理图谱,指的是运用事件抽象的方法对实例事理逻辑知识库的事件进行抽象,最终形成具有抽象层级性的事理图谱,包括事件预处理、成分抽象和抽象成分融合三个步骤。
- [0024] 进一步地,所述事理知识图谱是事件知识与实体知识之间的融合,事件知识与实体知识融合,指的是将实例知识图谱和实体知识图谱之间进行链接,形成从事件知识到实体知识的通路,包括待链接实体获取以及实体链接两个部分组成。
- [0025] 基于同一发明构思,本发明还提供一种融合多类事理与实体知识的领域事件图谱构建装置,其包括:
- [0026] 实例事理逻辑知识库构建模块,负责对领域预料库进行事理关系抽取,并利用事件论元抽取技术进行实例事理元素抽取,形成实例事理逻辑知识库;

- [0027] 抽象概念知识库构建模块,负责构建具有层级性的抽象概念知识库;
- [0028] 抽象事理图谱构建模块,负责利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象,形成抽象事理图谱;
- [0029] 事理知识图谱构建模块,负责利用实体知识图谱,采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接,实现事件知识与实体知识的融合,形成事理知识图谱;
- [0030] 领域事件图谱构建模块,负责将抽象事理图谱与事理知识图谱进行合并,形成领域事件图谱。
- [0031] 与现有技术相比,本发明具有以下优点:
- [0032] 1、本发明提出了一种融合多类事理与实体知识的领域事件图谱构建装置,该装置从抽象事理、实例事理、知识融合事件三个角度出发进行图谱构建,可搭建一个层次分明、体系相对完善的事件图谱。
- [0033] 2、本发明提出了多种事件关系,给出了可行的抽取方案,并对应用点进行了论述,为后续的应用提供了参考。
- [0034] 3、本发明提出了一种融合现有概念性层知识库资源的有效方法,可构建起一个质量较高的层次性抽象知识库。
- [0035] 4、本发明提出了一种针对实例事件进行层次性抽象的方法,能够对实例性事件抽象成具有明显阶梯性的抽象事件图谱,可支持不同层级的应用。
- [0036] 5、本发明提出了一种事件性知识与实体性知识链接和融合的方法,有利于进一步拓展事件图谱的语义信息和丰富度。
- [0037] 6、采用本发明方法构建的领域事件图谱,能够综合静态实体性知识以及动作事件性知识两者的优点,可进一步拓宽知识问答的应用范围,并提供基于事件知识的预警和推理服务,并可作为常识知识库扩充领域语言资源。

## 附图说明

- [0038] 图1为融合多类事理与实体知识的领域事件图谱构建方法的流程图。
- [0039] 图2为实例事理逻辑知识库构建框架图。
- [0040] 图3为抽象概念知识库构建框架图。

## 具体实施方式

- [0041] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面通过具体实施例和附图,对本发明做进一步详细说明。
- [0042] 本发明在实施过程中主要包括实例事理逻辑知识库构建、抽象概念知识库构建、抽象事理图谱构建、事件知识与实体知识的融合三个步骤,总体流程如图1所示,具体包括以下步骤:
- [0043] 一.实例事理逻辑知识库构建
- [0044] 图2为实例事理逻辑知识库构建框架图,包括事理关系抽取、实例事理元素抽取等步骤。
- [0045] 1、界定领域事件图谱的概念

[0046] 领域事件图谱是一个有向有环图,节点代表事件,有向边代表事件之间的因果、条件、互斥、组成、转折、上下位等逻辑关系。领域事件图谱既关注事件外部之间的演化逻辑关系,又关注事件本身的丰富属性信息,是抽象事件与事件知识库两者的统一结合体。

[0047] 2、界定领域事件图谱中的事理关系

[0048] 本发明中将所述事件之间的逻辑关系类型分成八种,类型包括:

[0049] 1) 因果事理。用于描述某一事件导致某一事件发生,如<智利地震,有色金属上涨>;

[0050] 2) 条件事理。表示某一事件条件下另一事件发生,如<中美贸易达成协议,中美贸易战缓和>;

[0051] 3) 上下位事理。某事件是另一事件的上层抽象,如<人民币贬值,货币贬值>

[0052] 4) 并列事理。表示某事件与另一事件同时发生,如<奏国歌,升国旗>

[0053] 5) 互斥事理。某一事件与另一事件形成对立,如<美国攻打伊拉克,美国拯救伊拉克>

[0054] 6) 顺承事理。表示某一时间紧接着另一事件发生,如<到售票口,刷身份证>

[0055] 7) 转折事理。表示本该发生的事件未能发生,如<刻苦学习,成绩下降>。

[0056] 8) 组成事理。表示某事件是另一事件的组成部分,如<奏国歌,升国旗仪式>

[0057] 3、事理关系抽取

[0058] 事理关系抽取的方法指根据指定的事理关系类型,采用基于关系词限定、词性过滤与事件成立性的抽取方法;基于关系词与依存句法规则的抽取方法;基于序列标注的事理抽取方法。

[0059] 1) 基于关系词限定的、词性过滤与事件成立性的方法。关系限定词指能够明显显示事理成分的关联指示词,如“导致”,“因为-所以”等因果关联词,其余七种事理关联词同理。通过关系词限定,抽取出事件描述字段;词性过滤指通过使用中文分词器进行分词和词性标注后设定需要保留的词性类别;事件成立性指事件需包含特定的事件成分,如事件动作、事件主体、客体等。

[0060] 2) 基于关系词与依存句法规则的抽取方法。具体包括通过对关系限定词对候选事件进行确定,借助依存句法分析手段,确定关联词对应的事件主体和事件客体,并通过修饰词之间的语义关系对事件进行扩展。

[0061] 3) 基于序列标注的方法,包括利用BIO或BIOES等标记集对起始事件和结尾事件描述进行标注,最后训练序列标注模型进行标注。

[0062] 4、实例事理元素抽取

[0063] 实例事理元素抽取是对事件的结构化表示。结构化表示的结果为抽取出与事件相关的通用论元,即事理元素。所述通用论元,包括:核心动作,指的是事件中的核心动词;动作主体,指围绕核心动词的施事者;动作客体,指围绕核心动词的受事者;事件发生时间,指事件发生时的背景时间;事件发生地点,指事件发生时的背景地点;事件关联人物,指事件发生的关联人物;事件关联公司,指事件发生的关联公司;事件关联机构,指事件发生的机构实体;事件极性,指事件本身的情绪;事件确定性,指事件发生的确定性;事件权威性,指因果事件的权威性;事件发生时态,指事件发生的时态,包括已经发生,将要发生,正在发生三种状态;事件来源,指事件抽取所在的文本来源。

[0064] 4.1事件核心动作的抽取

[0065] 所述事件核心动作,指事件描述能够明显描述事件动作的词语,通常为动词或者动名词。

[0066] 提取方法为:

[0067] 1) 对事件描述进行依存句法分析,获取与ROOT节点存在HED关系的词语作为核心动作,若该词语存在C00关系,则将C00关系关联的词语也当作核心动作。

[0068] 2) 对事件描述进行句子成分分析,获取整棵句法树最外层VP成分作为核心动作。

[0069] 3) 将方法1) 和方法2) 中抽取得到的词语作为事件的核心动作。

[0070] 4.2事件动作主体的抽取

[0071] 所述事件动作主体的抽取,指的是抽取与该事件相关的动作发出者。具体提取方法为:

[0072] 1) 对事件描述进行依存句法分析,遍历整个事件核心动作词表,以每个核心动作为核心,找到与该动作词语存在SBV的词语,并对SBV的词语与ATT修饰成分进行组合,形成事件主体;

[0073] 2) 对事件描述进行句子成分分析,遍历整个事件核心动作词表,以每个核心动作为核心,找到位于该词VP成分之前连接的NP短语作为事件主体;

[0074] 3) 若步骤1) 与步骤2) 中的事件均为找到主体,那将该词之前的描述字符串均作为事件主体

[0075] 4.3事件动作客体的抽取

[0076] 所述事件动作客体的抽取,指的是抽取与该事件相关的动作受事者。具体提取方法为:

[0077] 1) 对事件描述进行依存句法分析,遍历整个事件核心动作词表,以每个核心动作为核心,找到与该动作词语存在VOB关系的词语,并对VOB的词语与ATT及CMP关系修饰成分进行组合,形成事件主体;

[0078] 2) 对事件描述进行句子成分分析,遍历整个事件核心动作词表,以每个核心动作为核心,找到位于该词VP成分之后连接的NP短语作为事件客体;

[0079] 3) 若步骤1) 与步骤2) 中的事件均为找到主体,那将该词之后的描述字符串均作为事件主体

[0080] 4.4事件发生时间的抽取

[0081] 所述事件发生时间的抽取,指的是抽取给定上下文该事件的发生时间。具体抽取方法为:

[0082] 1) 对事件描述、事件上下文(事件描述句左右窗口为N的长句)、事件来源文本发布时间,进行时间词识别,并对识别结果进行统一标准化处理,即处理形成“年-月-日”的标准形式,形成对应的时间实体列表。

[0083] 2) 以1) 中识别的结果,按事件描述、事件来源文本创建时间的优先级进行选取,得到最终的事件发生时间。

[0084] 4.5事件发生地点的抽取

[0085] 所述事件发生地点的抽取,指的是抽取给定上下文该事件发生的地点。具体抽取方法为:



[0086] 1) 对事件描述、事件上下文(事件描述句左右窗口为N的长句)进行地点词识别。

[0087] 2) 以1)中的结果,按照事件描述、事件上下文的优先级次序进行排序,得到最终的事件发生地点,若事件发生地点缺失,则留空。

[0088] 4.6事件关联实体的抽取

[0089] 所述事件关联实体的抽取,指的是抽取给定上下文中与该事件相关的实体。实体类型包括行业领域相关的实体类型,如常见的公司实体、人物实体、机构实体。具体抽取方法为:

[0090] 1) 利用开源的带有实体识别的自然语言处理工具进行常规实体的识别,得到人物实体和结构实体;

[0091] 2) 收集领域实体词表,采用实体词匹配的方式,完成领域实体词的识别。

[0092] 3) 标注领域实体识别语料库,训练领域实体序列标注模型,完成关联实体的抽取。

[0093] 4) 将三种方法得到的实体进行联合汇总,得到最终的关联实体列表。

[0094] 4.7事件情绪的计算

[0095] 所述事件情绪的计算,指对事件描述本身所反映出来的情感方向及情感强度进行计算。事件情感方向包括正向、中性、负向三种;情感强度用于衡量情绪影响的程度。具体计算方法为:

[0096] 1) 利用情感词表,并设定情感计算规则,考虑窗口、否定词、程度副词以及依存规则,进行情绪的加权计算,得到事件的情绪得分。

[0097] 2) 将事件的情绪得分进行整理,将情绪得分的绝对值作为事件情绪的强度,以事件情绪数值前的符号作为情绪方向,但情绪为负时,为负相,情绪值为零时,情绪为中性,反之为负。

[0098] 4.8事件确定性的计算

[0099] 所述事件确定性的计算,指对事件描述本身发生的确定性进行计算,并得到事件确定性的等级,等级包括不可能、可能以及肯定三个层级。具体计算方法为:

[0100] 1) 收集整理出汉语中表达情态的词语集合,包括表否定的否定副词,表祈使语气的状s态词,形成事件确定性词表。

[0101] 2) 针对事件确定词,设定确定性计算规则,得到事件确定性等级。

[0102] 4.9事件时态性的计算

[0103] 所述事件时态性的计算,指的是对事件描述本身发生的时态性进行计算,得到事件时态性的类别,包括过去,现在以及未来三个时态。具体计算方法为:

[0104] 1) 收集整理出汉语中表达时态的词语集合,包括三个时态分别对应的特征词,形成事件时态词语集合;

[0105] 2) 基于事件时态词,设定时态计算规则,得到事件时态的类别。

[0106] 二. 抽象概念知识库构建

[0107] 所述抽象概念知识库,指的是存储概念之间上下位层级的知识库,由名词性实体抽象知识库、性状类抽象知识库以及动作性事件抽象知识库三个知识库构成。其中:

[0108] 所述名词性实体抽象知识库,指具有明显上下位语义名词性实体层级的实体知识库,如“苹果”和“水果”,“水果”和“植物”分别描述了一对上下位抽象关系,“苹果”可以往上抽象成“水果”,“水果”可以进一步抽象成“植物”。大量具有这样上下位抽象关系的实体对,

共同组成一个庞大的知识网络。

[0109] 所述性状类抽象知识库,指具有明显上下位语义形容词或副词性词语组成的知识库,如“悲伤”和“消极”,“消极”和“变坏”分别描述了一对性状类的抽象层级关系。

[0110] 所述动作类抽象知识库,指具有明显组成或上下位语义动词组成的知识库,如“杀人”和“犯罪”,“犯罪”和“作恶”分别描述了一堆事件性动词的抽象层级关系。

[0111] 在具体实施上,遵循图3所示的流程,包括以下步骤:

[0112] 1) 获取Hownet抽象层级。以Hownet (知网中文词库) 中的Glossary文件作为基本的概念层级体系文件。Glossary文件中描述了概念之间的上下位层级关系。

[0113] 2) 获取百科类抽象层级。采集百度百科、互动百科、维基百科中的百科分类层级,将三个分类体系进行融合,并设置根节点为“实体”。

[0114] 3) 融合抽象层级。将百科类概念体系与Hownet概念层级进行融合,形成最终的概念层级体系。

[0115] 4) 训练领域词向量。利用词向量工具训练领域行业语料,得到领域词向量文件以及领域词表。

[0116] 5) 同义词拓展。以领域词向量中的所有词为种子词,使用领域词向量中进行相似度计算,设定相似度阈值,将与种子词相似度大于指定阈值的词作为高度相似词,并进行存储。

[0117] 6) 同义词融合。利用5) 得到的同义词以及同义词词林中的同义词汇总,得到最终的同义词扩展词典。

[0118] 7) 百科词语开放分类获取。以步骤4) 中得到的所有词在百度百科、互动百科以及维基百科中进行检索,得到每个词的开放分类,并将其所属概念分类作为该词的上位概念。

[0119] 8) HowNet词语义项获取。以Hownet中的Whole.DAT文件作为基础概念词汇文件。所述Whole.DAT文件中对超过6万个词语进行了细粒度义项的解释。同时以该词条后的义项作为上位概念,形成词条的上位概念字典。

[0120] 9) 词语上位概念融合。将步骤8) 得到的词条及其所属概念分类,与7) 中的基础概念词汇文件进行融合,形成最终的词语上位概念词典。

[0121] 10) 词语概念描述融合。基于6) 中得到的同义词拓展词典,对9) 中形成最终的概念词汇文件进行概念扩展,若同义词中的某个词在概念词汇文件中,将其余同义词的概念描述都表示成该词的描述。

[0122] 11) 词语抽象路径生成。根据词语上位概念词典以及概念词汇文件,构建词条与概念的上下位有向图,利用最短路径算法,得到每个词条的最长的最短路径作为该词条的上下位抽象路径。

[0123] 三. 抽象事理图谱的构建

[0124] 所述抽象事理图谱,指的是运用事件抽象的方法对实例事理逻辑知识库中的实例型事件图谱中的事件进行抽象,最终形成具有抽象层级性的事理图谱。抽象的过程遵循以下几个步骤:

[0125] 1) 事件预处理。对输入的实例事理图谱中的事件进行分词和词性标注,仅保留其中的名词性成分以及动词性成分。

[0126] 2) 事件成分抽象。利用抽象图谱,对结果1) 中的名词性成分和动词性成分进行抽

象,对名词性成分往上抽象,形成多层次(如图3中的)的抽象结果。

[0127] 3) 抽象成分融合。利用2)得到的结果,得到不同层级下的具有统一抽象层级的名词性成分和动词性成分,形成整体性的事件多层次抽象结果。

[0128] 四. 事理知识图谱的构建

[0129] 所述事理知识图谱的构建,即事件知识与实体知识之间的融合,指的是将实例事理逻辑知识库与实体知识图谱之间进行链接,形成从事件知识到实体知识的通路。融合的过程遵循以下步骤:

[0130] 1) 待链接实体获取。给定知识图谱,根据实例事理逻辑知识库中的关联实体类型进行筛选,得到需要链接的知识实体列表。

[0131] 2) 实体链接。应用实体链接的方法,利用实例事理逻辑知识库中关联实体的上下信息作为特征,将事件图中不同类型的实体连接到1)中得到的待链接知识实体列表当中。

[0132] 五. 领域事件图谱的构建

[0133] 所述领域事件图谱的构建,指将抽象事理图谱与事理知识图谱进行合并形成一个大的领域图谱的过程,具体合并方式为,针对实例事理逻辑知识库中的每一个事件,将该事件关联的抽象事理图谱中的抽象事件和事理知识图谱中的知识进行组合,形成一个大的图谱。

[0134] 基于同一发明构思,本发明的另一个实施例提供一种融合多类事理与实体知识的领域事件图谱构建装置,其包括:

[0135] 实例事理逻辑知识库构建模块,负责对领域预料库进行事理关系抽取,并利用事件论元抽取技术进行实例事理元素抽取,形成实例事理逻辑知识库;

[0136] 抽象概念知识库构建模块,负责构建具有层级性的抽象概念知识库;

[0137] 抽象事理图谱构建模块,负责利用抽象概念知识库对实例事理逻辑知识库中的实例事件进行实体词抽象和谓词性抽象,形成抽象事理图谱;

[0138] 事理知识图谱构建模块,负责利用实体知识图谱,采用实体链接技术对实例事理逻辑知识库中的实例事件进行实体链接,实现事件知识与实体知识的融合,形成事理知识图谱;

[0139] 领域事件图谱构建模块,负责将抽象事理图谱与事理知识图谱进行合并,形成领域事件图谱。

[0140] 基于同一发明构思,本发明的另一个实施例提供一种计算机/服务器,其包括存储器和处理器,所述存储器存储计算机程序,所述计算机程序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本发明方法中各步骤的指令。

[0141] 基于同一发明构思,本发明的另一个实施例提供一种计算机可读存储介质(如ROM/RAM、磁盘、光盘),所述计算机可读存储介质存储计算机程序,所述计算机程序被计算机执行时,实现本发明方法的各个步骤。

[0142] 本发明未详细阐述的部分属于本领域技术人员的公知技术。

[0143] 以上实施例仅用以说明本发明的技术方案而非对其进行限制,本领域的普通技术人员可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明的原理和范围,本发明的保护范围应以权利要求书所述为准。

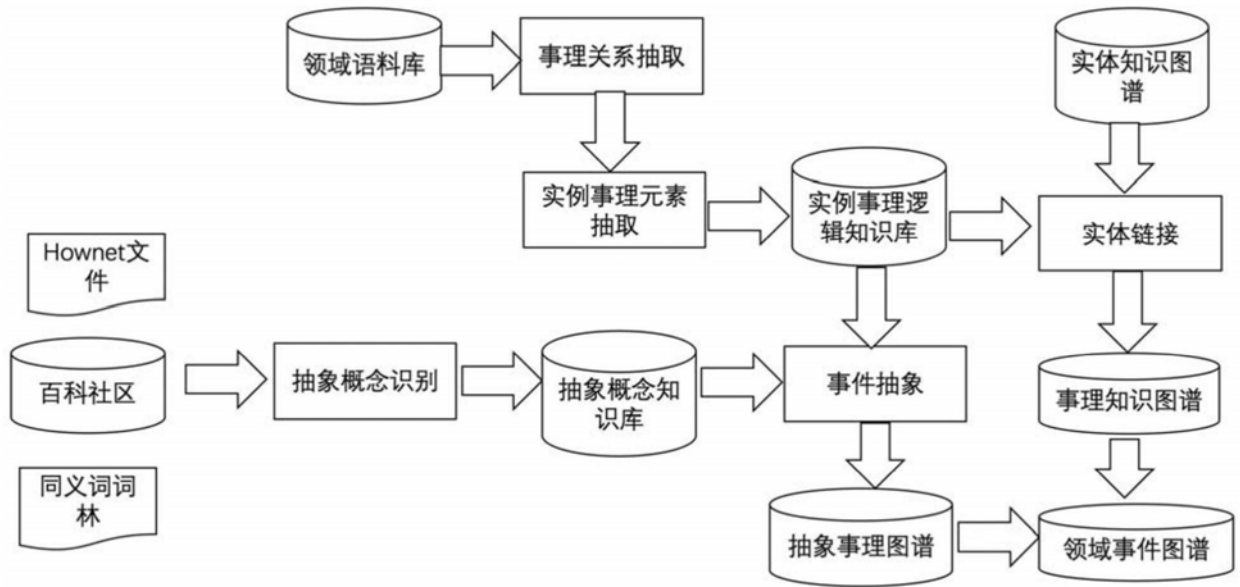


图1

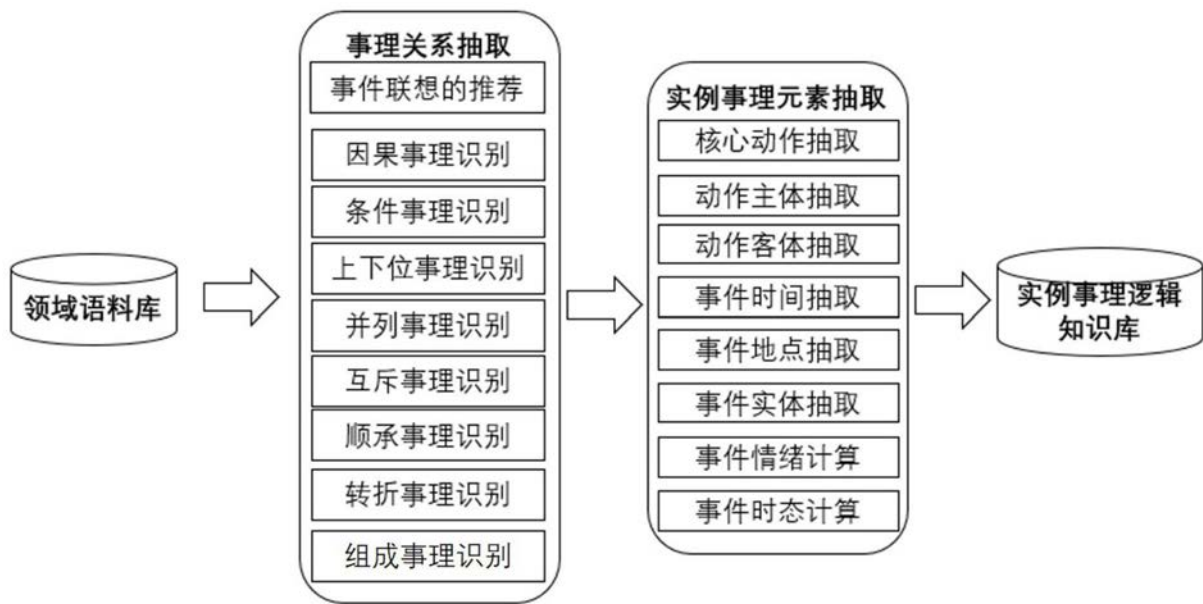


图2

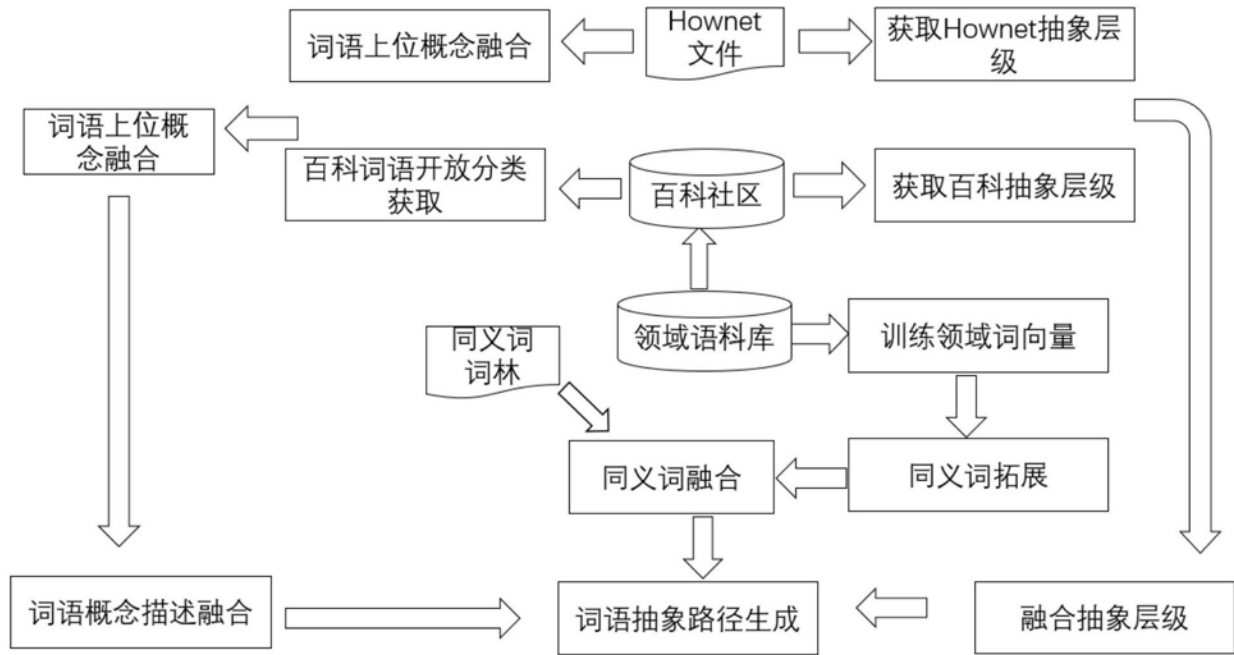


图3