(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112948552 B
(45)授权公告日 2023.06.02

(21)申请号 202110217425.8

(22)申请日 2021.02.26

(65)同一申请的已公布的文献号
申请公布号 CN 112948552 A

(43)申请公布日 2021.06.11

(73)专利权人 北京信息科技大学
地址 100192 北京市海淀区清河小营东路12号

(72)发明人 赵刚 杨昊 王兴芬

(74)专利代理机构 北京天方智力知识产权代理事务所(普通合伙) 11719
专利代理师 路远

(51)Int.Cl.
G06F 16/332(2019.01)
G06F 16/35(2019.01)
G06F 16/36(2019.01)

(56)对比文件
CN 105844298 A,2016.08.10
CN 107633044 A,2018.01.26
CN 108763333 A,2018.11.06
CN 109977237 A,2019.07.05
CN 110134797 A,2019.08.16
CN 110968699 A,2020.04.07
CN 112241457 A,2021.01.19
US 2018013861 A1,2018.01.11
US 6118850 A,2000.09.12
WO 2020244262 A1,2020.12.10
单晓红 等.基于事理图谱的网络舆情演化路径分析——以医疗舆情为例.情报理论与实践.2019,第42卷(第09期),99-103+85.
庄文英 等.突发事件舆情演化与治理研究——基于拓展多意见竞争演化模型.情报杂志.2021,第40卷(第12期),127-134+185.
王军平 等.面向大数据领域的事理认知图谱构建与推断分析.中国科学:信息科学.2020,第50卷(第07期),988-1002.
Carpenter G A 等.ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition.Neural networks.1991,第4卷(第4期),493-504.
王兰成;娄国哲.基于知识图谱的网络舆情管理方法与实践研究.情报理论与实践.2019,43(第06期),97-101.
(续)

审查员 文兴丽

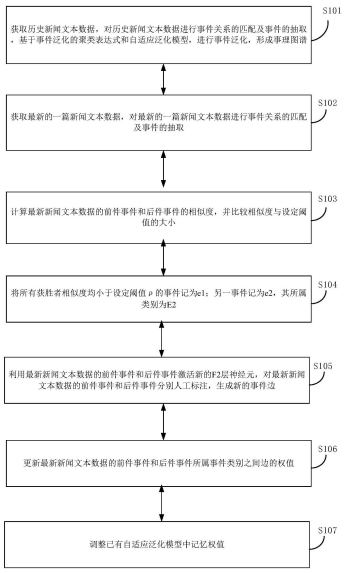
权利要求书4页 说明书8页 附图4页

(54)发明名称

一种事理图谱在线扩展方法及装置

(57)摘要

本发明公开了一种事理图谱在线扩展方法及装置,该方法包括以下步骤:获取历史新闻文本数据,构建事理图谱;获取最新新闻文本数据,对最新新闻文本数据进行事件关系的匹配及事件抽取;基于自适应泛化模型,利用抽取到的最新新闻文本数据的前件事件和后件事件,对构建的事理图谱进行在线扩展。该方法降低了人工成本,提高了扩展效率,增强了事理图谱的可移植性。



CN 112948552 B

[转续页]

[接上页]

(56)对比文件

单晓红;庞世红;刘晓燕;杨娟.基于事理图谱的网络舆情事件预测方法研究.情报理论与实践.2020,43(第10期),165-170+156.

张海涛;张连峰;王丹;刘健.基于自组织神经网络的图书馆关联知识聚合研究.情报理论与实践.2015,38(第09期),73-78.

1.一种事理图谱在线扩展方法,其特征是,该方法包括以下步骤:

获取历史新闻文本数据,构建事理图谱;

获取最新新闻文本数据,对最新新闻文本数据进行事件关系的匹配及事件抽取;

基于自适应泛化模型,利用抽取到的最新新闻文本数据的前件事件和后件事件,对构建的事理图谱进行在线扩展;

其中,所述事理图谱的构建方法为:

利用事件关系规则,构建因果事件规则库,根据因果事件规则库中事件关系规则,对历史新闻文本数据进行事件关系匹配,提取出历史新闻文本数据的前件和后件;

对提取出的历史新闻文本数据的前件和后件进行分词,抽取历史新闻文本数据的前件和后件中的事件,形成历史新闻文本数据的三元组;

基于事件泛化的聚类方法和自适应泛化模型,对历史新闻文本数据的三元组中前件事件和后件事件进行泛化,初步形成事理图谱,并保存自适应泛化模型中事理图谱的记忆权值;

其中,所述自适应泛化模型为:

$z_i = x_i + au_i$

$$q_i = \frac{z_i}{e + |Z|}$$

$v_i = f(q_i) + bf(s_i)$

$$f(x) = \begin{cases} 0 & -\theta \le x \le \theta \\ x & x > \theta 或 x < -\theta \end{cases}$$

$$u_i = \frac{v_i}{e + |V|}$$

$$g(y_j) = \begin{cases} d & j = I \\ 0 & else \end{cases}$$

$$p_i = u_i + \sum_{i=0}^{M-1} g(y_i) w_{ij}$$

$$\frac{dw_{ji}}{dt} = d(1-d)\left(\frac{u_i}{1-d} - w_{ji}\right)$$

$$\frac{dw_{ij}}{dt} = d(1-d)\left(\frac{u_i}{1-d} - w_{ij}\right)$$

$$r_i = \frac{u_i + cp_i}{e + |cp_i| + |U|}$$

$$|R| = \left[\sum_i (r_i)^2\right]^{1/2}$$

其中,a,b为正反馈系数,c为r向量的计算参数,d为调整的步幅值,e为弱归一化参数,ρ为设定的阈值,θ为门限值,I为获胜类别,x_i为输入变量,z_i为x向量的线性组合,|Z|为z向量的模长,q_i为z的归一化向量,s_i为p的归一化向量,f(x)为滤波函数,|V|为v向量的模长;y_j为输出向量,M为最大类别数,u_i为v的归一化向量,w为记忆权值,r_i为相似度向量,|U|为u向

量的模长，p_i为F_1层和F_2层交互向量，F_1为输入比较层，F_2为识别层，|R|为r向量的模长；当|R|+e≤ρ，则系统进入谐振，按照式$\frac{dw_{ji}}{dt}$和式$\frac{dw_{ij}}{dt}$更新权值。

2.根据权利要求1所述的事理图谱在线扩展方法，其特征是，所述事件泛化的聚类方法为：

统计新闻文本数据的三元组中前件事件和后件事件完全相同的元组数量；

将历史新闻文本数据的三元组中语义相似的事件聚为一类，并将这些事件所对应的元组数量值相加，得到元组数量总和；

根据每个事件的元组数量及元组数量总和，计算每个事件的概率。

3.根据权利要求2所述的事理图谱在线扩展方法，其特征是，所述每个事件的概率的计算方法为：

$$P_i=\frac{count_{\mathrm{i}}}{\sum_{\mathrm{k}=0}^{n}count_{\mathrm{k}}}$$

其中，i为事件，n为i事件的出度，$count_i$为事件i的元组数量。

4.根据权利要求1所述的事理图谱在线扩展方法，其特征是，所述对构建的事理图谱进行在线扩展的步骤包括：

将最新新闻文本数据的前件事件和后件事件进行向量化表示，并输入自适应泛化模型；

根据自适应泛化模型中事理图谱的记忆权值，计算得到最新新闻文本数据的前件事件和后件事件的竞争获胜者，并分别计算竞争获胜者与输入事件的相似度，将相似度与设定的阈值进行比较；

若最新新闻文本数据的前件事件和后件事件中至少一个事件的所有获胜者相似度均小于设定的阈值，利用最新新闻文本数据的前件事件和/或后件事件激活自适应泛化模型中新的计算单元，并根据计算单元所指类别中动词和名词出现的频率，人工标注新事件节点标签，生成新的事件边；

若最新新闻文本数据的前件事件和后件事件两者的所有竞争获胜者的相似度均大于设定的阈值，则调整自适应泛化模型中事理图谱的记忆权值。

5.根据权利要求4所述的事理图谱在线扩展方法，其特征是，所述利用最新新闻文本数据的前件事件和/或后件事件激活自适应泛化模型中新的计算单元的步骤包括：

若最新新闻文本数据的前件事件和后件事件中仅有一个事件的所有获胜者相似度均小于设定的阈值，利用所有竞争获胜者的相似度均小于设定的阈值的事件激活自适应泛化模型中新的计算单元，对新的计算单元按照其类别中动词和名词出现的频率进行人工标注标签，生成新事件节点；

若最新新闻文本数据的前件事件和后件事件两者的所有竞争获胜者的相似度均小于设定的阈值，利用最新新闻文本数据的前件事件和后件事件激活自适应泛化模型中新的计算单元，对最新新闻文本数据的前件事件和后件事件，分别按照其类别中动词和名词出现的频率进行人工标注，生成新的事件边，并增加两者的因果边赋予初始权值。

6.一种事理图谱在线扩展装置，其特征是，包括：

事理图谱初步构建模块，用于获取历史新闻文本数据，构建事理图谱；

数据获取模块，用于获取最新新闻文本数据；

事件抽取模块，用于对最新新闻文本数据进行事件关系的匹配及事件抽取；

事理图谱扩展模块，用于基于自适应泛化模型，利用抽取到的最新新闻文本数据的前件事件和后件事件，对构建的事理图谱进行在线扩展；

其中，所述事理图谱的构建方法为：

利用事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件；

对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组；

基于事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存自适应泛化模型中事理图谱的记忆权值；

其中，所述自适应泛化模型为：

$z_i = x_i + au_i$

$$q_i = \frac{z_i}{e + |Z|}$$

$v_i = f(q_i) + bf(s_i)$

$$f(x) = \begin{cases} 0 & -\theta \le x \le \theta \\ x & x > \theta 或 x < -\theta \end{cases}$$

$$u_i = \frac{v_i}{e + |V|}$$

$$g(y_j) = \begin{cases} d & j = I \\ 0 & else \end{cases}$$

$$p_i = u_i + \sum_{i=0}^{M-1} g(y_i) w_{ij}$$

$$\frac{dw_{ji}}{dt} = d(1-d)\left(\frac{u_i}{1-d} - w_{ji}\right)$$

$$\frac{dw_{ij}}{dt} = d(1-d)\left(\frac{u_i}{1-d} - w_{ij}\right)$$

$$r_i = \frac{u_i + cp_i}{e + |cp_i| + |U|}$$

$$|R| = \left[\sum_i (r_i)^2\right]^{1/2}$$

其中，a，b为正反馈系数，c为r向量的计算参数，d为调整的步幅值，e为弱归一化参数，ρ为设定的阈值，θ为门限值，I为获胜类别，x_i为输入变量，z_i为x向量的线性组合，|Z|为z向量的模长，q_i为z的归一化向量，s_i为p的归一化向量，f(x)为滤波函数，|V|为v向量的模长；y_j为输出向量，M为最大类别数，u_i为v的归一化向量，w为记忆权值，r_i为相似度向量，|U|为u向

量的模长，p_i为F_1层和F_2层交互向量，F_1为输入比较层，F_2为识别层，|R|为r向量的模长；当|R|+e≤ρ，则系统进入谐振，按照式$\frac{dw_{ji}}{dt}$和式$\frac{dw_{ij}}{dt}$更新权值；

所述事理图谱初步构建模块构建事理图谱的步骤包括：

利用事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件；

对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组；

基于事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存自适应泛化模型中事理图谱的记忆权值。

7.根据权利要求6所述的事理图谱在线扩展装置，其特征是，所述事理图谱扩展模块对构建的事理图谱进行在线扩展的步骤包括：

将最新新闻文本数据的前件事件和后件事件进行向量化表示，并输入自适应泛化模型；

根据自适应泛化模型中记忆权值，计算得到最新新闻文本数据的前件事件和后件事件的竞争获胜者，并分别计算竞争获胜者与输入事件的相似度，将相似度与设定的阈值进行比较；

若最新新闻文本数据的前件事件和后件事件中仅有一个事件的所有获胜者相似度均小于设定的阈值，利用所有竞争获胜者的相似度均小于设定的阈值的事件激活自适应泛化模型中新的计算单元，对新的计算单元按照其类别中动词和名词出现的频率进行人工标注标签，生成新事件节点；

若最新新闻文本数据的前件事件和后件事件两者的所有竞争获胜者的相似度均小于设定的阈值，利用最新新闻文本数据的前件事件和后件事件激活自适应泛化模型中新的计算单元，对最新新闻文本数据的前件事件和后件事件，分别按照该类别中动词和名词出现的频率进行人工标注，生成新的事件边，并增加两者的因果边赋予初始权值；

若最新新闻文本数据的前件事件和后件事件两者的所有竞争获胜者的相似度均大于设定的阈值，则调整自适应泛化模型的记忆权值。

一种事理图谱在线扩展方法及装置

技术领域

[0001] 本发明涉及事理图谱在线扩展技术领域，特别地涉及一种基于自适应泛化模型的事理图谱在线扩展方法及装置。

背景技术

[0002] 事理图谱是继知识图谱之后，以(前件事件，关系，后件事件)作为三元组所形成的事理知识库。与知识图谱所不同，事理图谱能够描绘出事件之间的演化规律和模式，可以应用于基于事理的问答，事件预测等。现有的事理图谱的构造方式都是基于大数据直接生成，实则为静态事理图谱。

[0003] 现有的事理图谱的生成方法为：事件关系抽取，事件的抽取，事件泛化，可视化。在事理图谱的在线扩展方面研究尤其罕见。在事件泛化上，现有的技术主要分为有监督学习和无监督学习两种方式：其中有监督学习为利用提前标注好的事件种子集作为训练集，通过特征提取配合深度学习进行分类任务，完成事件泛化。无监督学习主要是利用基于K-means的改良，利用欧式距离进行聚类。

[0004] 上述的利用有监督的方式进行事件泛化，其需要大量的训练样本集，目前并无完善统一的训练样本，故需要根据自身需求花费大量人力资源，进行标注，并且深度学习模型的训练时长较长，需要很高的时间成本。

[0005] 上述的无监督事件泛化的方式虽然降低了人力要求，但是依旧无法在线扩展节点，只能生成特定领域的静态事理图谱，可移植性，可扩展性差，不能够识别未知事件，仅能根据先验知识进行手工扩展。

发明内容

[0006] 有鉴于此，本发明提出一种基于自适应泛化模型的事理图谱在线扩展方法及装置，降低了人工成本，提高了扩展效率，增强了事理图谱的可移植性。

[0007] 本发明第一方面提供一种事理图谱在线扩展方法，该方法包括以下步骤：

[0008] 获取历史新闻文本数据，构建事理图谱；

[0009] 获取最新新闻文本数据，对最新新闻文本数据进行事件关系的匹配及事件抽取；

[0010] 基于自适应泛化模型，利用抽取到的最新新闻文本数据的前件事件和后件事件，对构建的事理图谱进行在线扩展。

[0011] 进一步地，所述事理图谱的构建方法为：

[0012] 利用事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件；

[0013] 对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组；

[0014] 基于事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存自适应泛化模型中事理图谱的记

忆权值。

[0015] 进一步地，所述对构建的事理图谱进行在线扩展的步骤包括：

[0016] 将最新新闻文本数据的前件事件和后件事件进行向量化表示，并输入自适应泛化模型；

[0017] 根据自适应泛化模型中事理图谱的记忆权值，计算得到最新新闻文本数据的前件事件和后件事件的竞争获胜者，并分别计算竞争获胜者与输入事件的相似度，将相似度与设定的阈值进行比较；

[0018] 若最新新闻文本数据的前件事件和后件事件中至少一个事件的所有获胜者相似度均小于设定的阈值，利用最新新闻文本数据的前件事件和/或后件事件激活自适应泛化模型中新的计算单元，并根据计算单元所指类别中动词和名词出现的频率，人工标注新事件节点标签，生成新的事件边；

[0019] 若最新新闻文本数据的前件事件和后件事件两者的所有获胜者的相似度均大于设定的阈值，则调整事理图谱的记忆权值。

[0020] 本发明第二方面提供一种事理图谱在线扩展装置，该装置包括：

[0021] 事理图谱初步构建模块，用于获取历史新闻文本数据，构建事理图谱；

[0022] 数据获取模块，用于获取最新新闻文本数据；

[0023] 事件抽取模块，用于对最新新闻文本数据进行事件关系的匹配及事件抽取；

[0024] 事理图谱扩展模块，用于基于自适应泛化模型，利用抽取到的最新新闻文本数据的前件事件和后件事件，对构建的事理图谱进行在线扩展。

[0025] 进一步地，所述事理图谱初步构建模块构建事理图谱的步骤包括：

[0026] 利用事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件；

[0027] 对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组；

[0028] 基于事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存自适应泛化模型中事理图谱的记忆权值。

[0029] 进一步地，所述事理图谱扩展模块对构建的事理图谱进行在线扩展的步骤包括：

[0030] 将最新新闻文本数据的前件事件和后件事件进行向量化表示，并输入自适应泛化模型；

[0031] 根据自适应泛化模型中记忆权值，计算得到最新新闻文本数据的前件事件和后件事件的竞争获胜者，并分别计算竞争获胜者与输入事件的相似度，将相似度与设定的阈值进行比较；

[0032] 若最新新闻文本数据的前件事件和后件事件中仅有一个事件的所有获胜者相似度均小于设定的阈值，利用所有获胜者的相似度均小于设定的阈值的事件激活自适应泛化模型中新的计算单元，对新的计算单元按照该类别中动词和名词出现的频率进行人工标注标签，生成新事件节点；

[0033] 若最新新闻文本数据的前件事件和后件事件两者的所有获胜者的相似度均小于设定的阈值，利用最新新闻文本数据的前件事件和后件事件激活自适应泛化模型中新的计

算单元,对最新新闻文本数据的前件事件和后件事件,分别按照该类别中动词和名词出现的频率进行人工标注,生成新的事件边,并增加两者的因果边赋予初始权值;

[0034] 若最新新闻文本数据的前件事件和后件事件两者的所有获胜者的相似度均大于设定的阈值,则调整事理图谱的记忆权值。

[0035] 上述的基于自适应泛化模型的事理图谱在线扩展方法,基于自适应泛化模型,利用网络结构的记忆性,进行事件泛化,能够在较少的人为干预下,在线生成新的事件节点,完成事理图谱的在线扩展,能够动态生成事理图谱,增强了事理图谱在不同领域应用的可移植性,可扩展性。

附图说明

[0036] 为了说明而非限制的目的,现在将根据本发明的优选实施例、特别是参考附图来描述本发明,其中:

[0037] 图1是实施例一提供的事理图谱在线扩展方法的流程图。

[0038] 图2(a)、2(b)和2(c)是事件聚类示意图。

[0039] 图3是自适应泛化模型的结构示意图。

[0040] 图4是实施例二提供的事理图谱在线扩展装置的结构框图。

具体实施方式

[0041] 为了能够更清楚地理解本发明的上述目的、特征和优点,下面结合附图和具体实施例对本发明进行详细描述。需要说明的是,在不冲突的情况下,本发明的实施例及实施例中的特征可以相互组合。

[0042] 在下面的描述中阐述了很多具体细节以便于充分理解本发明,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。

[0043] 除非另有定义,本文所使用的所有的技术和科学术语与属于本发明的技术领域的技术人员通常理解的含义相同。本文中在本发明的说明书中所使用的术语只是为了描述具体的实施例的目的,不是旨在于限制本发明。

[0044] 实施例一

[0045] 图1是本发明实施例一提供的一种基于自适应泛化模型的事理图谱在线扩展方法的流程图。

[0046] 在本实施例中,所述事理图谱在线扩展方法可以应用于计算机装置中,对于需要进行事理图谱在线扩展的计算机装置,可以直接在计算机装置上集成本发明的方法所提供的用于事理图谱在线扩展的功能,或者以软件开发工具包(Software Development Kit,SDK)的形式运行在计算机装置上。

[0047] 如图1所示,所述事理图谱在线扩展方法具体包括以下步骤,根据不同的需求,该流程图中步骤的顺序可以改变,某些步骤可以省略。

[0048] 本实施例中,所述计算机装置可以为个人电脑、服务器、智能电视、便携式电子设备如手机、平板电脑等设备。

[0049] 步骤S101、所述计算机装置获取历史新闻文本数据,对历史新闻文本数据进行事

件关系的匹配及事件的抽取，基于事件泛化的聚类方法和自适应泛化模型，进行事件泛化，形成事理图谱，并保存自适应泛化模型中事理图谱的记忆权值w_{ij}和w_{ji}。

[0050] 上述步骤S1中所述计算机装置获取历史新闻文本数据，对历史新闻文本数据进行事件关系的匹配及事件抽取的步骤包括：

[0051] 首先，所述计算机装置利用现有语言领域的事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件。

[0052] 接着，所述计算机装置对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组<Pre,r,Post>，其中Pre和Post分别表示历史新闻文本数据的前件事件与后件事件，r指历史新闻文本数据的前件事件和后件事件之间的因果关系。所述计算机装置定义事件为$E=\{x|x=V_{max}\cup N_{max}\}$，其中$V_{max}$为前/后件中出现次数最多的动词，$N_{max}$为前件事件和后件事件中出现次数最多的词语。

[0053] 上述的步骤S101中事件泛化的聚类方法为：

[0054] 所述计算机装置统计历史新闻文本数据的三元组中前件事件和后件事件完全相同的元组数量，并记为count，形成多个Pre——count——Post的图谱形式。如图2(a)所示，统计三元组e1中前件事件和后件事件完全相同的元组数量。

[0055] 接着，所述计算机装置将历史新闻文本数据的三元组中语义相似的事件聚为一类，并将这些事件所对应的count值相加。如图2(b)所示，对语义相似的事件e2、e4进行聚类形成事件e'。

[0056] 所述计算机装置根据每个事件的元组数量及元组数量总和，计算每个事件的概率p_i，如图2(c)所示。其中，概率p_i的计算表达式如下：

[0057] $$p_i=\frac{count_i}{\sum_{k=0}^{n}count_k} \quad (1)$$

[0058] 式中，n为i节点的出度。

[0059] 所述自适应泛化模型是对自适应共振网络的改进，应用于事件泛化领域。自适应共振网络是采用自稳机制和竞争学习的一种自组织学习。其结构如图3所示，其中，F_1为输入比较层，可以抑制噪声；F_2为识别层，可以输出分类；空心箭头表示兴奋激励，实心箭头表示抑制激励。所述自适应泛化模型的具体公式如下：

[0060] $z_i=x_i+au_i$ (2)

[0061] $$q_i=\frac{z_i}{e+|Z|} \quad (3)$$

[0062] $v_i=f(q_i)+bf(s_i)$ (4)

[0063] $$f(x)=\begin{cases}0 & -\theta\le x\le\theta\\ x & \theta\le x或x\le-\theta\end{cases} \quad (5)$$

[0064] $$u_i=\frac{v_i}{e+|V|} \quad (6)$$

[0065] 其中，x_i为输入变量，z_i为x向量的线性组合，|Z|为z向量的模长，q_i为z的归一化向

量，s_i为p的归一化向量，f(x)为滤波函数，u_i为v的归一化向量，|V|为v向量的模长。

[0066] 由于word2vec向量化后含有负向量，故f(x)改进为式(5)，a,b>0，式(3)和式(6)可看做z_i和v_i的归一化处理，其中e为极小的正数，e<<1。

[0067] $$g(y_j)=\begin{cases}d & j=I\\ 0 & else\end{cases} \quad (7)$$

[0068] $$p_i=u_i+\sum_{i=0}^{M-1}g(y_i)w'_{ij} \quad (8)$$

由底向上 $$\frac{dw_{ji}}{dt}=d(1-d)\left(\frac{u_i}{1-d}-w_{ji}\right) \quad (9)$$

[0069] 由顶向下 $$\frac{dw_{ij}}{dt}=d(1-d)\left(\frac{u_i}{1-d}-w_{ij}\right) \quad (10)$$

[0070] $$r_i=\frac{u_i+cp_i}{e+|cP|+|U|} \quad (11)$$

[0071] $$|R|=\left[\sum_i (r_i)^2\right]^{1/2} \quad (12)$$

[0072] 其中，a,b,c,d,e,ρ,θ为自适应泛化模型的超参数，其中a,b为正反馈系数，c为r向量的计算参数，d为调整的步幅值，e为弱归一化参数，ρ为设定的阈值，θ为门限值，I为获胜类别，y_j为输出向量，M为最大类别数，u_i为v的归一化向量，p_i为F_1层和F_2层交互向量，w为记忆权值，r_i为相似度向量，|U|为u向量的模长，|R|为r向量的模长。

[0073] 当|R|+e≤ρ，则系统进入谐振，按照式(9)和式(10)更新权值；否则F_2重置。其中，I为事件类别号，0≤d≤1，cd/(1-d)≤1，上述a,b,c,d,e,ρ,θ均为自适应泛化模型的超参数，其会具体影响泛化效果，可以利用各种参数调节方法，如遗传算法等提前计算得出。

[0074] 所述计算机装置基于上述的事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存事理图谱的记忆权值w_{ij}和w_{ji}。

[0075] 步骤S102、所述计算机装置获取最新的一篇新闻文本数据，对最新的一篇新闻文本数据进行事件关系的匹配及事件的抽取。

[0076] 上述步骤S102中所述计算机装置对最新的一篇新闻文本数据进行事件关系的匹配及事件抽取的步骤包括：

[0077] 首先，所述计算机装置根据因果事件规则库中事件关系规则，对最新的一篇新闻文本数据进行事件关系匹配，提取出最新新闻文本数据的前件和后件。

[0078] 接着，所述计算机装置对提取出的最新新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成最新新闻文本数据的三元组<P′re,r′,P′ost>，其中P′re和P′ost分别表示最新新闻文本数据的前件事件与后件事件，r′指最新新闻文本数据的前件事件和后件事件之间的因果关系。

[0079] 步骤S103、所述计算机装置基于自适应泛化模型，计算最新新闻文本数据的前件事件和后件事件的相似度，并比较相似度与设定阈值的大小。

[0080] 所述计算机装置分别将最新新闻文本数据的前件事件P′re和后件事件P′ost利用word2vec向量化表示，传入自适应泛化模型中F1层，依据记忆权值w_{ij}计算，逐次得到自适应

泛化模型中F2层中的竞争获胜者，逐个计算竞争获胜者与输入事件的相似度，将相似度与设定阈值ρ进行比较。若最新新闻文本数据的前件事件P′re和后件事件P′ost中仅有一个的所有获胜者相似度均小于设定阈值ρ，则转至步骤S4，若最新新闻文本数据的前件事件P′re和后件事件P′ost两者的所有获胜者的相似度均小于设定的阈值ρ，转至步骤S105；否则转至步骤S106。

[0081] 本实施例基于自适应泛化模型，与传统的K-means及其改进相比较，自适应泛化模型具备记忆性，并且能够对非平稳，有噪声环境进行学习，具备更优的泛化效果。

[0082] 本实施例使用自适应泛化模型进行事件扩展，可以对新事件进行在线生成，通过比较|R|+e≤ρ判断是否在F2层激活新的神经元，来判断所输入事件是否为新增事件，从而决定对后续的事理图谱扩展节点还是动态调整权值。

[0083] 步骤S104、所述计算机装置将所有获胜者相似度均小于设定阈值ρ的事件记为e1；另一事件记为e2，其所属类别为E2。e1会激活自适应泛化模型中新的F2层神经元，对新神经元按照该类别中动词和名词出现的频率进行人工标注标签，生成新事件节点E1。事理图谱中增加E1与E2因果边并赋予初始权值，转至步骤S7。

[0084] 本实施例在事理图谱在新增事件节点的过程，仅在为事件类别打上可视化标签需要人工参与，其余部分完全由算法完成，降低了人工的成本，提高了效率。

[0085] 步骤S105、所述计算机装置利用最新新闻文本数据的前件事件P′re和后件事件P′ost激活新的F2层神经元，对最新新闻文本数据的前件事件P′re和后件事件P′ost分别按照该类别中动词和名词出现的频率人工标注，生成新的事件边，并增加两者的因果边赋予初始权值，转至步骤S7。

[0086] 步骤S106、所述计算机装置更新最新新闻文本数据的前件事件P′re和后件事件P′ost所属事件类别之间边的权值，转至步骤S107。

[0087] 步骤S107、所述计算机装置调整已有自适应泛化模型中事理图谱的记忆权值w_{ij}和w_{ji}，转至步骤S102，依次循环，实现事理图谱在线扩展。

[0088] 本实施例提出的事理图谱在线扩展方法，基于自适应泛化模型可以在非平稳的环境下进行无监督的学习的特点，利用自适应泛化模型进行事件泛化，不需要逐个事件分类标注，大大地降低了人工标注成本。

[0089] 本实施例提出的事理图谱在线扩展方法所采用的自适应泛化模型应用了记忆权值，其中蕴含已泛化事件信息，具有长期记忆性，故每次发现新事件时仅需要为新事件标签，不需要重新训练已有的事件，其应用在事理图谱扩展上提高了工作的效率。

[0090] 利用本实施例提出的事理图谱在线扩展方法所提出的自适应泛化模型进行事理图谱扩展，可将事理图谱在水平领域进行应用，从而完成更多事件预测等事理图谱下游任务，增强了事理图谱的可移植性。

[0091] 实施例二

[0092] 图4是本发明实施例二提供的基于自适应泛化模型的事理图谱在线扩展装置20的结构框图。

[0093] 在本实施例中，所述事理图谱在线扩展装置20可以应用于计算机装置中，所述事理图谱在线扩展装置20可以包括多个由程序代码段所组成的功能模块。所述事理图谱在线扩展装置20中的各个程序段的程序代码可以存储于计算机装置的存储器中，并由所述计算

机装置的至少一个处理器所执行，以实现(详见图1描述)事理图谱在线扩展功能。

[0094] 本实施例中，所述事理图谱在线扩展装置20根据其所执行的功能，可以被划分为多个功能模块。所述功能模块可以包括：事理图谱初步构建模块201、数据获取模块202、事件抽取模块203以及事理图谱扩展模块204。本发明所称的模块是指一种能够被至少一个处理器所执行并且能够完成固定功能的一系列计算机程序段，其存储在存储器中。在本实施例中，关于各模块的功能将在后续的实施例中详述。

[0095] 所述事理图谱初步构建模块201，用于获取历史新闻文本数据，对历史新闻文本数据进行事件关系的匹配及事件的抽取，基于事件泛化的聚类方法和自适应泛化模型，进行事件泛化，初步形成事理图谱，并保存自适应泛化模型中事理图谱的记忆权值为w_{ij}和w_{ji}。

[0096] 所述事理图谱初步构建模块201获取历史新闻文本数据，对历史新闻文本数据进行事件关系的匹配及事件抽取的步骤包括：

[0097] 首先，利用现有语言领域的事件关系规则，构建因果事件规则库，根据因果事件规则库中事件关系规则，对历史新闻文本数据进行事件关系匹配，提取出历史新闻文本数据的前件和后件。

[0098] 接着，对提取出的历史新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成历史新闻文本数据的三元组＜Pre,r,Post＞，其中Pre和Post分别表示历史新闻文本数据的前件事件与后件事件，r指历史新闻文本数据的前件事件和后件事件之间的因果关系。

[0099] 所述事理图谱初步构建模块201基于事件泛化的聚类方法和自适应泛化模型，对历史新闻文本数据的三元组中前件事件和后件事件进行泛化，初步形成事理图谱，并保存事理图谱的记忆权值为w_{ij}和w_{ji}。

[0100] 所述数据获取模块202，用于获取最新的一篇新闻文本数据。

[0101] 所述事件抽取模块203，用于对最新新闻文本数据进行事件关系的匹配及事件的抽取。

[0102] 上述事件抽取模块203对最新的一篇新闻文本数据进行事件关系的匹配及事件抽取的步骤包括：

[0103] 首先，根据因果事件规则库中事件关系规则，对最新的一篇新闻文本数据进行事件关系匹配，提取出最新新闻文本数据的前件和后件。

[0104] 接着，对提取出的最新新闻文本数据的前件和后件进行分词，抽取历史新闻文本数据的前件和后件中的事件，形成最新新闻文本数据的三元组＜P′re,r′,P′ost＞，其中P′re和P′ost分别表示最新新闻文本数据的前件事件与后件事件，r′指最新新闻文本数据的前件事件和后件事件之间的因果关系。

[0105] 所述事理图谱扩展模块204，用于基于自适应泛化模型，计算最新新闻文本数据的前件事件和后件事件的相似度，并比较相似度与设定阈值的大小，根据相似度与设定阈值的比较结果，利用自适应泛化模型对初步形成的事理图谱进行在线扩展。

[0106] 所述事理图谱扩展模块204分别将最新新闻文本数据的前件事件P′re和后件事件P′ost利用word2vec向量化表示，传入自适应泛化模型中F1层，依据参数w_{ij}计算，逐次得到自适应泛化模型中F2层中的竞争获胜者，逐个计算竞争获胜者与输入事件的相似度，将相似度与设定阈值ρ进行比较，便于后续事理图谱扩展。

[0107] 本实施例基于自适应泛化模型，与传统的K-means及其改进相比较，自适应泛化模型具备记忆性，并且能够对非平稳，有噪声环境进行学习，具备更优的泛化效果。

[0108] 本实施例使用自适应泛化模型进行事件扩展，可以对新事件进行在线生成，通过比较|R|+e≤ρ判断是否在F2层激活新的神经元，来判断所输入事件是否为新增事件，从而决定对后续的事理图谱扩展节点还是动态调整权值。

[0109] 上述的事理图谱扩展模块205根据相似度与设定阈值的比较结果，利用自适应泛化模型对初步形成的事理图谱进行在线扩展的具体实现过程包括：

[0110] 若最新新闻文本数据的前件事件P′re和后件事件P′ost中仅有一个的所有获胜者相似度均小于设定阈值ρ，则将所有获胜者相似度均小于设定阈值ρ的事件记为e1；另一事件记为e2，其所属类别为E2。e1会激活自适应泛化模型中新的F2层神经元，对新神经元进行人工标注标签，生成新事件节点E1。事理图谱中增加E1与E2因果边并赋予初始权值，调整已有自适应泛化模型中事理图谱的记忆权值w_{ij}和w_{ji}。

[0111] 若最新新闻文本数据的前件事件P′re和后件事件P′ost两者的所有获胜者的相似度均小于ρ，利用最新新闻文本数据的前件事件P′re和后件事件P′ost激活新的F2层神经元，对最新新闻文本数据的前件事件P′re和后件事件P′ost分别人工标注，生成新的事件边，并增加两者的因果边赋予初始权值，调整已有自适应泛化模型中事理图谱的记忆权值w_{ij}和w_{ji}。

[0112] 本实施例提出的事理图谱在线扩展装置，基于自适应泛化模型可以在非平稳的环境下进行无监督的学习的特点，利用自适应泛化模型进行事件泛化，不需要逐个事件分类标注，大大地降低了人工标注成本。

[0113] 本实施例提出的事理图谱在线扩展装置所采用的自适应泛化模型应用了记忆权值，其中蕴含已泛化事件信息，具有长期记忆性，故每次发现新事件时仅需要为新事件标签，不需要重新训练已有的事件，其应用在事理图谱扩展上提高了工作的效率。

[0114] 利用本实施例提出的事理图谱在线扩展装置所提出的自适应泛化模型进行事理图谱扩展，可将事理图谱在水平领域进行应用，从而完成更多事件预测等事理图谱下游任务，增强了事理图谱的可移植性。

[0115] 上述具体实施方式，并不构成对本发明保护范围的限制。本领域技术人员应该明白的是，取决于设计要求和其他因素，可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等，均应包含在本发明保护范围之内。

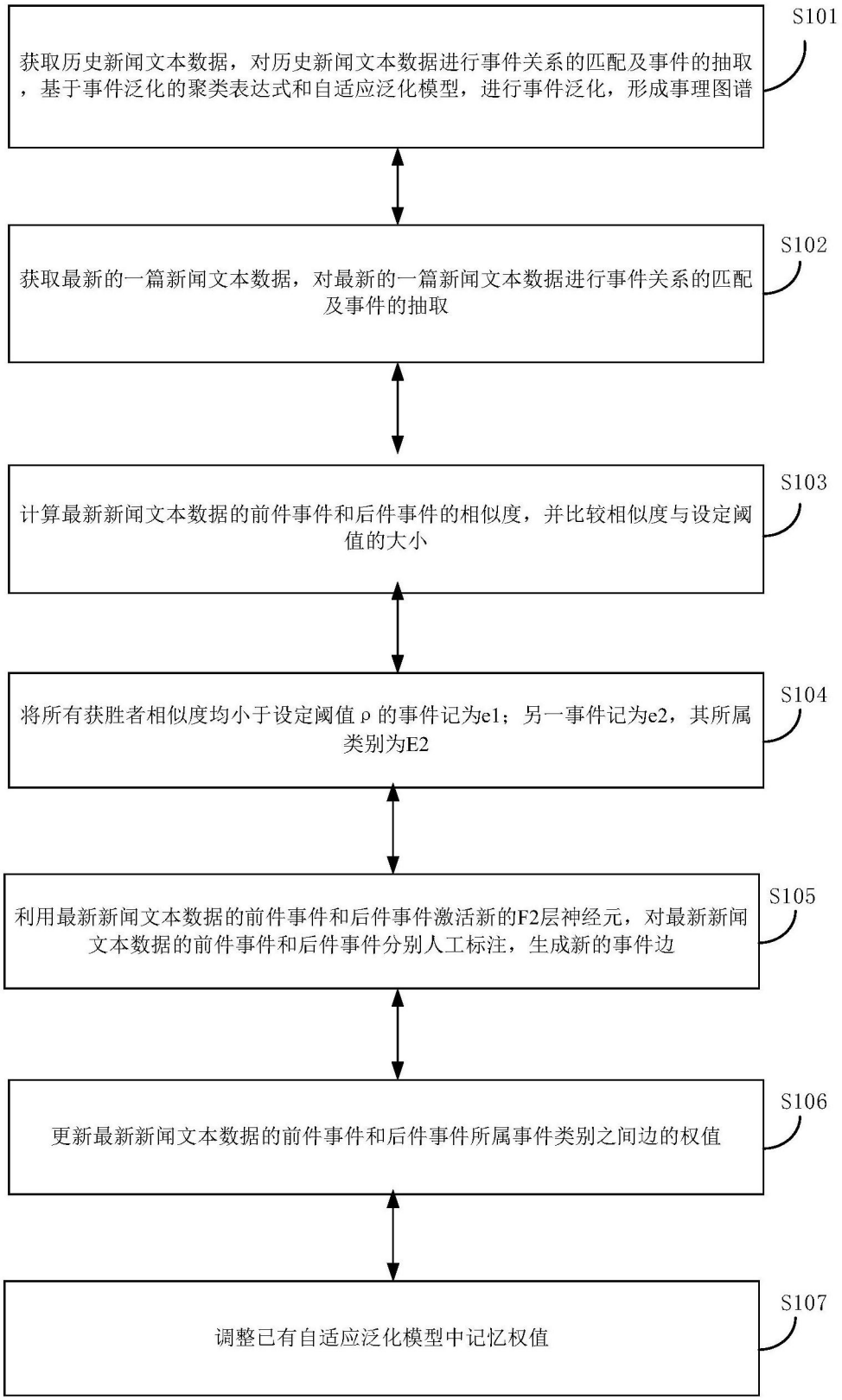

S101
获取历史新闻文本数据，对历史新闻文本数据进行事件关系的匹配及事件的抽取，基于事件泛化的聚类表达式和自适应泛化模型，进行事件泛化，形成事理图谱
S102
获取最新的一篇新闻文本数据，对最新的一篇新闻文本数据进行事件关系的匹配及事件的抽取
S103
计算最新新闻文本数据的前件事件和后件事件的相似度，并比较相似度与设定阈值的大小
S104
将所有获胜者相似度均小于设定阈值ρ的事件记为e1；另一事件记为e2，其所属类别为E2
S105
利用最新新闻文本数据的前件事件和后件事件激活新的F2层神经元，对最新新闻文本数据的前件事件和后件事件分别人工标注，生成新的事件边
S106
更新最新新闻文本数据的前件事件和后件事件所属事件类别之间边的权值
S107
调整已有自适应泛化模型中记忆权值

图1

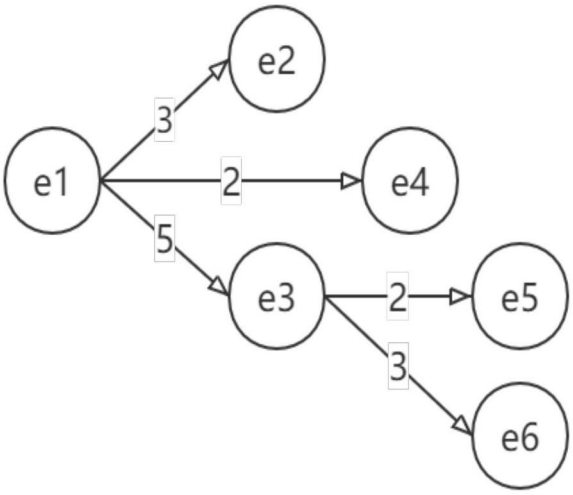


图2(a)

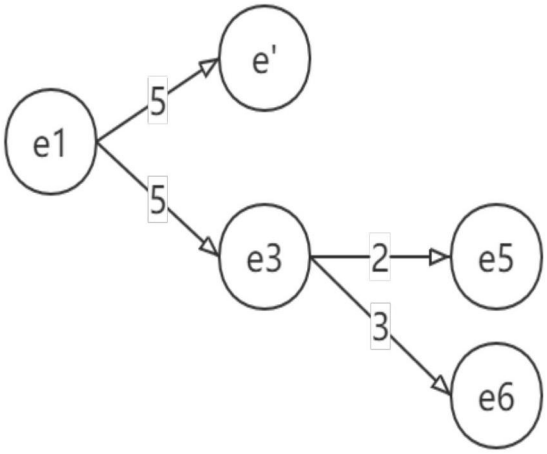


图2(b)

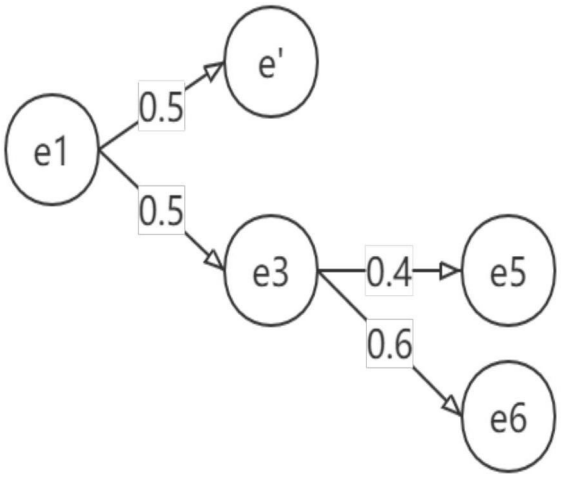

e'
0.5
e1
0.5
e3
0.4
e5
0.6
e6

图2(c)

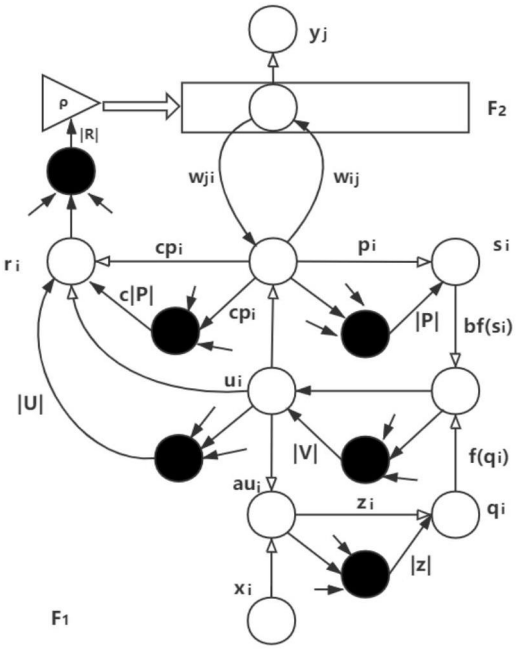

yj
ρ
F2
|R|
wji
wij
ri
cpi
pi
si
c|P|
cpi
|P|
bf(si)
ui
|U|
|V|
f(qi)
aui
zi
qi
|z|
xi
F1

图3

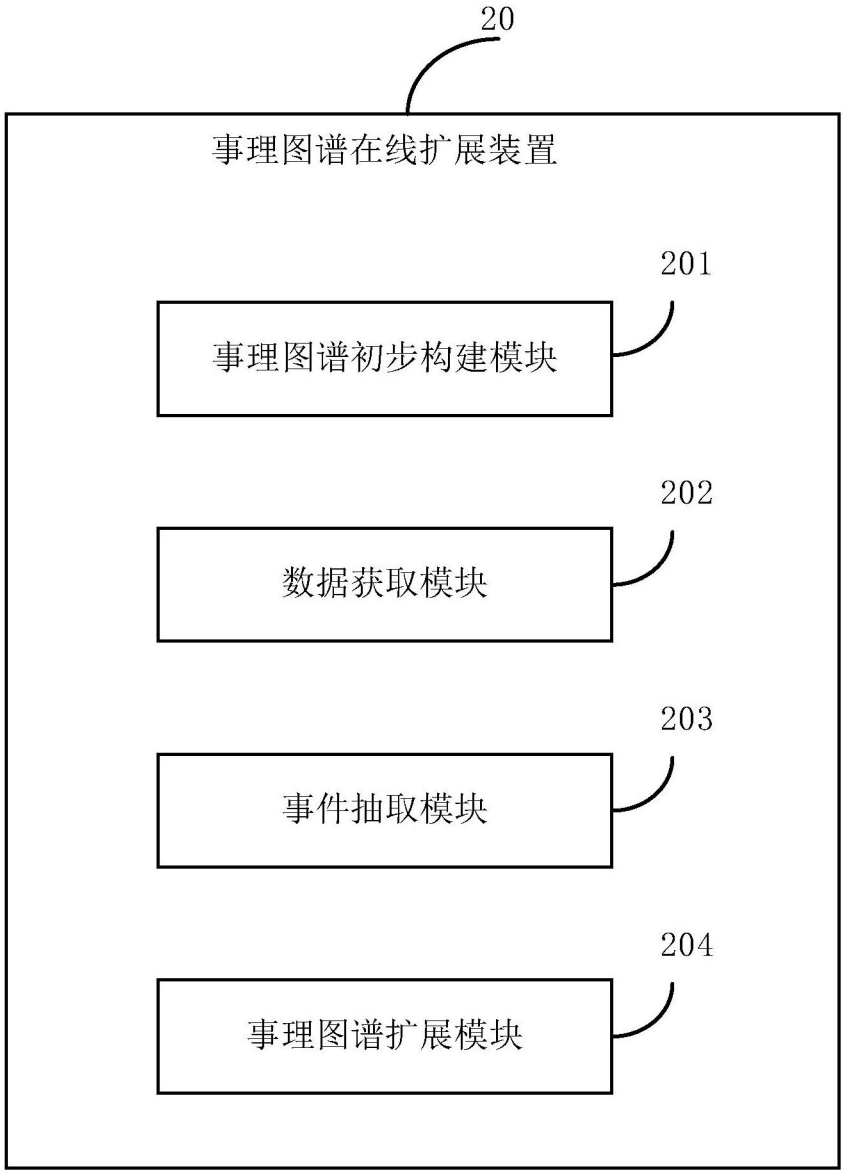

20
事理图谱在线扩展装置
201
事理图谱初步构建模块
202
数据获取模块
203
事件抽取模块
204
事理图谱扩展模块

图4