一种新闻事件演化建模方法*

张 辉,李国辉,孙博良,贾 立
(国防科技大学 信息系统与管理学院,湖南 长沙 410073)

摘 要:为了准确地发现话题中事件间的潜在关系,提出一种新闻事件演化建模方法。该方法利用事件的时间关系、内容相似性、命名实体关联信息构建新的演化关系模型,并通过定义事件的五种演化模式,识别出演化过程中的开始、中间、结束事件,最后根据新演化模型及演化模式建立事件演化的有向无环图模型,揭示事件发展的潜在脉络结构。实验结果表明,本文方法能够有效检测事件演化,提升系统性能。

关键词:新闻事件演化;演化模式;TF·IEF模型;有向无环图
中图分类号:TP391 **文献标志码**:A **文章编号**:1001-2486(2013)04-0166-05

Modeling news event evolution

ZHANG Hui, *LI Guohui*, *SUN Boliang*, *JIA Li*
(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: A new method is proposed for modeling the news event evolution to precisely present the relationships between events. This method utilized the events timestamp, events content similarity, and events dependence between features to build a new event evolution model, and defined five different event evolution patterns to identify the seminal events, the intermediary and ending events. Ultimately, an event evolution graph model was constructed to present the underlying events relationship. Experiments were conducted, confirming that the proposed method is efficient for detecting event evolution, and improves performance of system.

Key words: news event evolution; evolution pattern; TF·IEF model; directed acyclic graph

通过信息检索工具,例如谷歌、百度,用户能够及时获取、更新不同网站关于同一话题的各种相关报道。但是,这种获取、更新过程产生了大量的新闻文本流,利用人工有效地管理、整合、分析这些文本流,从而得到话题内不同事件演化、发展的脉络是一个艰巨的任务。因此,从新闻事件的层次,利用计算机技术挖掘事件之间的演化关系,构建事件演化关系图,是话题演化研究的一个新的研究方向。

事件演化研究的主要任务是分析话题内事件间的关系,描述事件演化的来龙去脉。目前,直接从话题的层次对演化进行分析的研究比较多[1-2],而从事件层次对演化进行研究还不是特别充分和完善。Makkonen 等[3]首先提出将事件演化分析作为TDT的子课题来研究,并建立事件的地点、人名(组织)、时间、内容向量,计算事件间的相似性,使用图论知识来表示事件间的演化关系。Nallapati 等[4]提出事件线索的概念,将事件关系表示成树形结构,而不是图结构,分析事件之间的关系,并且提出了简单的事件模型。相对事件演化,Wei、Zhai 等[5]提出主题演化的概念,根据文本流的时间戳,发现文本主题演化的时序模式。Wei、Chang 等[6]提出事件片段的概念,将相关于同一事件的所有新闻报道根据时间分成多个片段,根据事件片段间的时序关系分析事件内部各片段间的演化关系。Yang 等[7-8]对同一话题下的相关事件演化进行研究,提出了事件演化的概念,利用事件内容相似性、事件时间接近度、事件报道分布接近度来共同识别事件间的演化关系。邱江涛等[9]提出了事件时序分析的概念,并提出以事件演化图的形式来描述事件的演化关系。邓镭等[10]提出了原子事件的概念,通过分析原子事件间的关系来对事件演化进行研究。Li 等[11]提出构建时序事件图来对事件演化进行研究,分析不同事件间的关系。

上述研究中都使用了内容精确匹配建立事件间关系,当两事件中不包含相同特征时,则认为两事件关系为零,研究没有考虑两事件间不同特征的关联性,特别是命名实体特征。例如一个事件

* **稿日期**:2013-04-01
基金项目:国家自然科学基金项目(61170158);国家部委资助项目;湖南省自然科学基金项目(12JJ5028)
作者简介:张辉(1983—),男,湖南湘潭人,博士研究生,E-mail:zhanghui@nudt.edu.cn;
李国辉(通信作者),男,教授,博士,博士生导师,E-mail:guohli@nudt.edu.cn

中出现“美国”,另一事件中出现“华盛顿”;一个事件中出现“奥巴马”,另一事件中出现“希拉里”等。因此,本文在事件内容精确匹配的同时,对事件间不同命名实体的关联信息进行模糊匹配。

1　事件演化相关概念及定义

1.1　事件时间

由事件定义[7-8]可知,每一个事件都有一个时间特征,通过时间可以确定事件时序关系。事件时间主要是指事件发生的时间,这个时间可能是一个时间点,例如朝鲜金正日去世;也可能是一段时间,如重庆搜捕周克华。本文统一用标记(t)表示事件时间,$t=[st, et]$,st($start\ time$)表示事件开始的时间点,选取事件中最早报道的发布时间表示;et($end\ time$)表示事件结束的时间点,选取事件中最晚报道的发布时间表示。当 st 与 et 相同时,表示 t 是一个时间点,当不同时,则表示时间段。利用文献[13]中的方法对事件时间进行提取与识别,将其映射到形式化的日历中。在事件时序关系中,相对于某事件,时间靠前的事件为前序事件,靠后的事件为后序事件。

1.2　事件演化图

定义 1(事件演化):话题中事件随时间产生、发展、变化、消亡的过程。

定义 2(事件演化关系):事件间的有向逻辑依赖关系,或事件间的有向关联。

根据定义 2 可知,存在演化关系的两个事件,应满足两个条件:1)时序上具有先后关系;2)内容上具有一定的关联。时间上的先后关系,表示演化的方向。如果事件 E_A 到事件 E_B 存在演化关系,则事件 E_A 的时间 t_A 必先于事件 E_B 的时间 t_B,用符号表示为 $t_A \rightarrow t_B$,否则表示为 $t_A \nmapsto t_B$。因此,依据演化条件,基于图论知识中的有向无环图[5-7]建立事件演化关系图模型,表示为 $G=\{V, L\}$,其中,$V=\{E_1^{t_1}, E_2^{t_2}, \cdots, E_n^{t_n}\}$ 表示图中的节点集合,每个节点代表一个事件,$L=\{(E_i^{t_i}, E_j^{t_j})\}$ 表示图中有向弧集合,$(E_i^{t_i}, E_j^{t_j})$ 表示事件 $E_i^{t_i}$ 到事件 $E_j^{t_j}$ 存在演化关系,且 $i, j \in \{1,2,\cdots,n\}$, $i \neq j$,绘制事件演化图示例如图 1 所示。

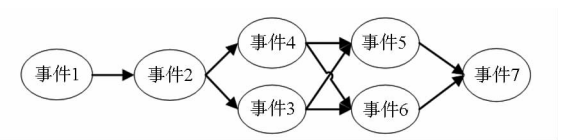

图 1　事件演化图示例

Fig. 1　Sample of event evolution graph

事件演化关系图 G 采用邻接矩阵存储,邻接矩阵表示事件之间关系,矩阵如式(1)所示。

$$A=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{1}$$

式中:$a_{ij}=\begin{cases} 1, & E_i \text{ 到 } E_j \text{ 存在演化关系} \\ 0, & E_i \text{ 到 } E_j \text{ 不存在演化关系} \\ 1, & \text{当 } i=j \text{ 时,事件自相关} \end{cases}$

1.3　事件演化模式

从图 1 可看出,对于一个完整演化过程,应该包含开始、中间、结束等事件,中间事件演变可能是多个事件合并成一个事件,可能是一个事件分裂成多个事件,也可能是由一个事件发展成另一个事件。因此,将事件演化类型分为五种模式,分别是开始演化模式、一对一演化模式、一对多演化模式、多对一演化模式、结束演化模式。

设话题 $T=\{E_1^{t_1}, E_2^{t_2}, \cdots, E_n^{t_n}\}$,其中 $E_i^{t_i}$ 表示 T 中的第 i 个事件,t_i 表示事件 $E_i^{t_i}$ 的时间,$t_1, t_2, \cdots, t_n$ 表示事件的先后顺序,λ 表示演化关系阈值。

定义 3(前序关联事件):话题中,事件 $E_i^{t_i}$ 与前序事件 $E_j^{t_j}(t_j \in \{t_1, t_2, \cdots, t_{i-1}\})$ 的相似性最大,且 $sim(E_i^{t_i}, E_j^{t_j}) \geqslant \lambda$,则 $E_j^{t_j}$ 称为 $E_i^{t_i}$ 的前向关联事件,记为 $prior(E_i^{t_i})$。

定义 4(后序关联事件):话题中,事件 $E_i^{t_i}$ 与后序事件 $E_k^{t_k}(t_k \in \{t_{i+1}, t_{i+2}, \cdots, t_n\})$ 的相似性最大,且 $sim(E_i^{t_i}, E_k^{t_k}) \geqslant \lambda$,则 $E_k^{t_k}$ 称为 $E_i^{t_i}$ 的后向关联事件,记为 $post(E_i^{t_i})$。

根据前、后序关联事件的定义,对五种事件演化模式进行定义。

定义 5(开始演化模式):对于事件 $E_i^{t_i}$,如果话题中 $prior(E_i^{t_i})$ 不存在,且 $post(E_i^{t_i})$ 存在,则 $E_i^{t_i}$ 为开始事件,表示开始演化模式。

定义 6(一对一演化模式):对于事件 $E_i^{t_i}$,如果有 $E_j^{t_j}=prior(E_i^{t_i})$,且 $E_i^{t_i}=post(E_j^{t_j})$,则 $E_j^{t_j}$ 到 $E_i^{t_i}$ 是一对一演化模式。

定义 7(一对多演化模式):对于事件 $E_i^{t_i}$,如果有 $E_j^{t_j}=prior(E_i^{t_i})$,但 $E_i^{t_i} \neq post(E_j^{t_j})$,且有 $E_j^{t_j}$ 的后序事件 $E_k^{t_k}$ 使得 $sim(E_j^{t_j}, E_k^{t_k}) \geqslant \lambda$,则事件 $E_j^{t_j}$ 到 $E_k^{t_k}$ 和 $E_i^{t_i}$ 是一对多演化模式。

定义 8(多对一演化模式):对于事件 $E_j^{t_j}$,如果有 $E_i^{t_i}=post(E_j^{t_j})$,但 $E_j^{t_j} \neq prior(E_i^{t_i})$,且有 $E_i^{t_i}$ 的前序事件 $E_k^{t_k}$,使得 $sim(E_i^{t_i}, E_k^{t_k}) \geqslant \lambda$,则事件 $E_k^{t_k}$ 和 $E_j^{t_j}$ 到 $E_i^{t_i}$ 是多对一演化模式。

定义 9(结束演化模式):对于事件 $E_i^{t_i}$,如果 $post(E_i^{t_i})$ 不存在,且 $prior(E_i^{t_i})$ 存在,则 $E_i^{t_i}$ 为结束事件,表示结束演化模式。

2 事件演化关系建模

2.1 事件建模

研究事件演化关系,首先要生成话题的事件集,每个事件至少由一篇新闻报道组成。事件集可以使用人工方式进行收集,也可以利用文献[15]中的方法对话题中新闻报道聚类生成事件集。本文主要研究事件演化关系建模方法,为了减少报道聚类误差对实验评价的影响,选择人工方式采集话题的事件集。

使用 TF · IEF 模型[12]创建事件的内容向量。设 $T=\{E_1^{t_1},E_2^{t_2},\cdots,E_n^{t_n}\}$ 表示一个话题事件集合,$\omega(f_j,E_i^{t_i})$ 表示第 i 个事件中第 j 个特征的权重值,$\{(f_j,\omega(f_j,E_i^{t_i}))\mid j=1,2,\cdots,k\}$ 表示 E_i 的 k 个特征及其权重值,TF · IEF 模型计算特征权重公式如下:

$$\omega(f_j,E_i^{t_i})=\frac{[1+\log_2 TF(f_j,E_i^{t_i})]\cdot IEF(f_j)}{\sqrt{\sum_{j=1}^{n}\left\{[1+\log_2 TF(f_j,E_i^{t_i})]\cdot IEF(f_j)\right\}^2}} \tag{2}$$

$$IEF(f_j)=\log_2\frac{|T|+1}{|EF(f_j)|+0.5} \tag{3}$$

其中,$TF(f_j,E_i^{t_i})$ 是特征 f_j 在事件 $E_i^{t_i}$ 出现的频次,$|EF(f_j)|$ 是出现特征词 f_j 的事件数;$|T|$ 是 T 中总的事件数。

2.2 演化关系建模

以往研究中事件演化关系值函数均使用事件内容向量进行精确匹配,计算相似度。事件向量相似度计算常使用余弦距离,表示如公式(4)所示:

$$sim(E_A,E_B)=\sum_{f\in E_A\cap E_B}\omega(f,E_A)\cdot\omega(f,E_B) \tag{4}$$

其中 $E_A\cap E_B$ 表示事件 E_A 和 E_B 的共有特征集合。

本文在精确匹配的基础上,使用一种模糊匹配策略对事件间命名实体特征的关联度进行计算。关联度是指两个特征在事件集中的关联程度,共同出现在一个事件中被认为是一次关联。在文本处理中,通常使用互信息来计算两个特征词在文本中的关联度,而本文中的特征关联度在计算目标上与互信息具有很大的相似性。因此,本文借鉴互信息的思想,把互信息中两个不同特征在一个文本中共同出现和独立出现的概率替换为它们在整个事件集中共现和独立出现的事件个数,同时考虑每个特征在事件中的权重值,具体计算如公式(5)所示:

$$rela(f_A,f_B)=\frac{\omega(f_A)\cdot\omega(f_B)\cdot cooc(f_A,f_B)}{sioc(f_A)\cdot sioc(f_B)} \tag{5}$$

$$sioc(f_A)=oc(f_A)-cooc(f_A,f_B) \tag{6}$$

$$sioc(f_B)=oc(f_B)-cooc(f_A,f_B) \tag{7}$$

其中,$\omega(f_A)$ 表示特征 f_A 在事件 A 中权重;$rela(f_A,f_B)$ 表示事件 A 中特征 f_A 与事件 B 中特征 f_B 的关联度;$cooc(f_A,f_B)$ 表示特征 f_A 与 f_B 共同出现的事件个数;$oc(f_A)$ 表示特征 f_A 出现的事件数;$sioc(f_A)$ 表示特征 f_A 单独出现的事件数。

事件关联度计算如公式(8)所示,首先对两事件向量中的不同命名实体特征计算两两关联度,然后进行求和并平均。

$$ass(E_A,E_B)=\frac{\sum_{f_A\in E_A}\sum_{f_B\in E_B}rela(f_A,f_B)}{|\overrightarrow{E_A}||\overrightarrow{E_B}|} \tag{8}$$

其中,$ass(E_A,E_B)$ 表示事件 A 与事件 B 的关联度;$|\overrightarrow{E_A}|$ 与 $|\overrightarrow{E_B}|$ 表示事件向量的势。

结合事件时间关系、事件向量精确匹配以及命名实体特征模糊匹配构建一种新的事件演化关系模型如下:

$$evo(E_A,E_B)=\begin{cases}0, & \text{if } t_A\mapsto t_B\\ sim(E_A,E_B)*ass(E_A,E_B), & \text{if } t_A\rightarrow t_B\end{cases} \tag{9}$$

2.3 演化关系图构建算法

利用本文建立的演化关系模型以及定义的事件演化的五种模式,构建事件演化关系图算法如下。

Input:话题事件集 T 及事件时间,演化关系阈值 λ。

Output:事件演化关系邻接矩阵 A。

Step1:根据事件时间顺序,将事件进行排序,得到 $T=\{E_1^{t_1},E_2^{t_2},\cdots,E_n^{t_n}\}$;

Step2:依事件排序,建立事件演化关系邻接矩阵 A,行、列顺序即事件时间的先后顺序,并初始化矩阵 A,将对角线元素置为1,其余元素置为0;

Step3:对 $\forall E_i^{t_i}\in T$,计算 $post(E_i^{t_i})$,如果 $post(E_i^{t_i})$ 不存在,但 $prior(E_i^{t_i})$ 存在,则 $E_i^{t_i}$ 为消亡事件,标记 $F(E_i^{t_i})=End$;如果存在,但 $E_i^{t_i}\neq prior(post(E_i^{t_i}))$,则 $F(E_i^{t_i})=Merge$;

Step4:对 $\forall E_i^{t_i}\in T$,计算 $prior(E_i^{t_i})$,如果 $prior(E_i^{t_i})$ 不存在,但 $post(E_i^{t_i})$ 存在,则 $E_i^{t_i}$ 为开始事件,$F(E_i^{t_i})=Start$;如果 $post(prior(E_i^{t_i}))=E_i^{t_i}$,则 $F(E_i^{t_i})=Develop$;如果 $post(prior(E_i^{t_i}))\neq E_i^{t_i}$,则 $F(E_i^{t_i})=Split$;

Step5:如果 $F(E_i^{t_i}) = Develop$,则将 $prior(E_i^{t_i})$ 与 $E_i^{t_i}$ 对应的 A 中元素设为 1;如果 $F(E_i^{t_i}) = Split$,则将 $prior(E_i^{t_i})$ 及满足 $sim(piror(E_i^{t_i}), E_k^{t_k}) > \lambda$ 的后序事件 $E_k^{t_k}$ 对应的元素设为 1;如果 $F(E_i^{t_i}) = Merge$,则将事件 $post(E_i^{t_i})$ 及满足 $sim(post(E_i^{t_i}), E_l^{t_l}) \geqslant \lambda$ 的前序事件 $E_l^{t_l}$ 对应的元素设为 1;

Step5:输出演化关系邻接矩阵 A。

3　实验结果及分析

3.1　实验数据

实验数据是利用网络爬虫工具从中国新闻网国际专栏中采集的 2011 年 3 月 11 日至 5 月 11 日关于日本 2011 年"3.11 地震"的相关报道,共计 1982 篇。由于本文重点不是研究事件探测,所以直接通过人工方式对采集的报道所涉及的事件进行识别,选取 8 个事件共 270 篇报道,根据最早报道时间对事件进行排序,最后得到事件相关信息,见表 1。

表 1　2011 年日本"3.11 地震"话题涉及事件

Tab.1　Component events in "3.11 Earthquake"

事件	事件概要	报道数
E_1	日本东部近海发生地震	7
E_2	地震引发巨大海啸	17
E_3	地震、海啸带来巨大损失	89
E_4	日本自卫队抗震救灾	58
E_5	福岛核电站发生故障	43
E_6	核电站故障,东电公司限电	16
E_7	汽车产业停产、减产	8
E_8	日本经济遭受重创	32

根据表 1 及事件时间,邀请两名本领域专家建立真实的事件演化关系如图 2 所示,总共包含 16 条演化关系,用以验证系统探测的演化关系的正确性和完整性。从图中可以看出,E_1 为开始事件,E_4 和 E_8 为结束事件,其余事件为中间事件。

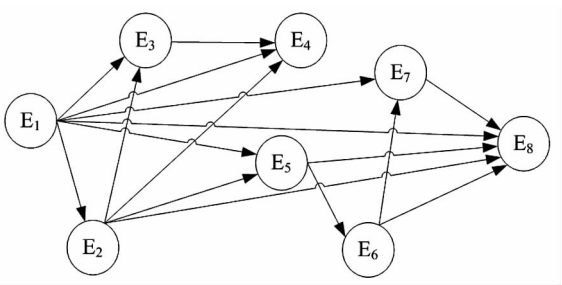

图 2　2011 年日本地震的事件演化图

Fig.2　Event evolution graph of topic"2011 Japan Earthquake"

3.2　实验评价方法

人工创建真实的事件演化图 $G = \{T, L\}$,如图 2 所示;同时使用本文系统自动建立事件演化图 $G' = \{T, L'\}$,T 表示话题包含的所有事件集合。参考文献[8]的评价方法,实验评价主要是比较 L 与 L' 的不同,使用准确率 P 和召回率 R 两个指标对系统进行定量评价。假设事件演化关系集合 L'' 表示 L 与 L' 的公共部分,用数学表示为:

$$L'' = L \cap L' \tag{10}$$

则准确率 P 就是系统探测的正确的演化关系数与系统探测到的总的演化关系数的比值,表示为:

$$P = \frac{|L''|}{|L'|} \tag{11}$$

则召回率 R 就是系统探测的正确的演化关系数与真实演化关系图中总的关系数的比值,表示为:

$$R = \frac{|L''|}{|L|} \tag{12}$$

3.3　实验结果及分析

本文提出的演化关系模型主要考虑事件内容的相似性及命名实体特征关联度(CS * FA),实验首先对不同演化关系模型的性能进行比较,参与对比的有文献[8]中提出的两种演化关系模型:1)事件内容相似性(CS)模型,2)事件内容相似性及报道分布接近度(CS * DF)模型。利用三种模型系统自动构建事件演化关系,计算三种模型对应系统在不同阈值 λ 时的准确率和召回率,绘制系统召回率 - 准确率曲线[8],如图 3 所示,其中阈值变化范围为 0 ~ 0.15。

从图 3 中可以看出,本文提出的 CS * FA 模型要明显优于 CS * DF 和 CS 模型,而 CS * DF 模型要略优于 CS 模型。这说明演化关系模型中增加实体特征关联度及报道分布接近度均能提升系统性能,而本文提出的 CS * FA 模型对系统性能提升明显。

事件内容建模使用 TF · IEF 模型,该模型不同于 TF · IDF 和 TF[8],实验对比了三种事件模型对系统的性能影响,其中,演化关系模型均使用 CS * DF,实验结果如图 4 所示。从图 4 中可以看出,TF · IDF 方法的系统性能最优,而 TF · IEF 的系统性能不及 TF · IDF,但优于 TF。

利用本文提出的方法,自动探测演化关系,当 $\lambda = 0.04$ 时,对应的演化关系如图 5 所示。其中探测到 10 条正确的演化关系,三条虚假的演化关系,丢失 6 条演化关系。

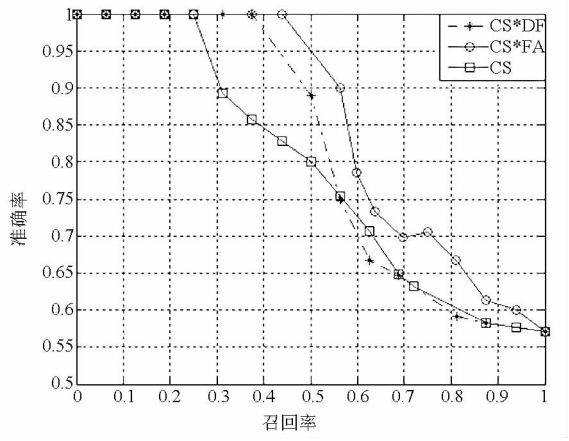

图3 不同事件演化关系值函数比较

Fig. 3 Comparison of different event evolution scoring functions

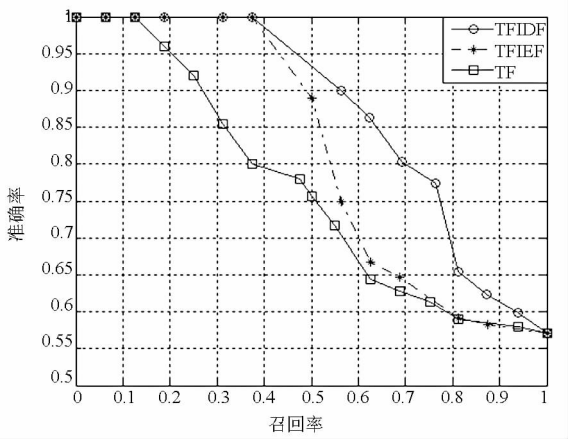

图4 不同的内容相似性计算方法比较

Fig. 4 Comparison of different similarity measures

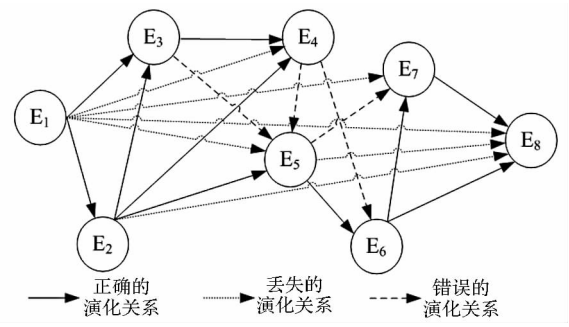

图5 2011年日本地震的事件演化图

Fig. 5 Event evolution graph of topic"2011 Japan Earthquake"

4 结论

每天网络都有大量关于新闻事件的报道,为了能够快速、有效地获取新闻事件来龙去脉,本文通过事件时间、事件内容相似性、事件特征关联信息三个要素构建事件演化图。事件演化图中的节点表示事件,图中的有向边表示事件间的依赖关系。本文提出的事件演化关系发现方法能够较好地呈现同一主题下事件间潜在的演化发展脉络。

参考文献(References)

[1] 胡艳丽, 白亮, 张维明. 网络舆情中一种基于OLDA的在线话题演化方法[J]. 国防科技大学学报, 2012, 34(1): 150-154.
HU Yanli, BAI Liang, ZHANG Weiming. OLDA-based method for online topic evolution in network public opinion analysis[J]. Journal of National University of Defense Technology, 2012, 34(1):150-154. (in Chinese)

[2] 胡艳丽, 白亮, 张维明. 一种话题演化建模与分析方法[J]. 自动化学报, 2012, 38(10):1690-1697.
HU Yanli, BAI Liang, ZHANG Weiming. Modeling and analyzing topic evolution[J]. Acta Automatica Sinica, 2012, 38(10):1690-1697. (in Chinese)

[3] Makkonen J. Investigations on event evolution in TDT[C]// Proceedings of HLT NAACL 2003 Student Research Workshop, 2003: 43-48.

[4] Nallapati R, Feng A, Peng F C, et al. Event threading within news topic[C]//Proceedings of the CIKM'04, 2004: 446-453.

[5] Wei Q Z, Zhai C X. Discovering evolutionary theme patterns from text: an exploration of temporal text mining[C]// Proceeding 11th ACM SIGKDD International Conference Knowledge Discovery Data Mining, 2005:198-207.

[6] Wei C P, Chang Y H. Discovering event evolution patterns from document sequences[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2007, 37(2):273-283.

[7] Yang C C, Shi X D, Wei C P. Tracing the event evolution of terror attacks from on-line news[C]//Proceedings of ISI 2006, San Diego: Springer Verlag, 2006: 343-354.

[8] Yang C C, Shi X D, Wei C P. Discoverying event evolution graphs from news corpora[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2009, 39(4): 850-863.

[9] Qiu J T, Li C, Qiao S J, et al. Timeline analysis of web news events[C]//Proceedings of the 4th International Conference on Advanced Data Mining and Applications, 2008: 123-134.

[10] Deng L, Ding Z Y, Xu B Y. Exploring event evolution patterns at the atomic level[C]//Proceedings of 2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2011:40-47.

[11] Li Q. Temporal event map construction for event search[R/OL]. 2012, http://ietfc.lzu.edu.cn/spe.html.

[12] 张辉, 李国辉, 贾立, 等. 一种基于TF·IEF模型的在线新闻事件探测方法[J]. 国防科技大学学报, 2013, 35(3).
ZHANG Hui., LI Guohui, JIA Li, et al. On-line news event detection based on TF·IEF model[J]. Journal of National University of Defense Technology, 2013, 35(3). (in Chinese)

[13] Schilder F, Habel C. From temporal expressions to temporal information: Semantic tagging of news messages[C]// Proceedings of ACL-2001 Workshop on Temporal and Spatial Information Processing, 2001.

[14] Lin Y R, Sundaram H, Chi Y, et al. Blog community discovery and evolution based on mutual awareness expansion[C]//Proceedings of the International Conference on Web Intelligence, 2007:48-56.

[15] Zhang H, Li G H, Xu X W. A on-line news documents clustering method[C]//Proceedings of the 8th International Conference on Active Media Technology, 2012:82-92.