

一种基于扩展描述逻辑的事件实例检测方法^{*}

唐英英, 刘 炜[†], 苏小英, 刘菲京

(上海大学 计算机工程与科学学院, 上海 200444)

摘 要: 针对事件要素之间客观存在的语义关系以及事件与事件类之间的语义关系进行研究, 提出了一种基于扩展描述逻辑的事件实例检测方法。该方法利用事件中的时间、动作、环境要素对事件进行语义补充, 然后利用扩展描述逻辑中概念的可满足性和概念包含的推理功能对事件实例进行检测, 最后通过实验证明该方法可实现文本中的事件实例检测, 准确率达到85.12%。

关键词: 描述逻辑; 事件实例检测; 事件; 事件类

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2014)11-3285-06

doi:10.3969/j.issn.1001-3695.2014.11.020

Event instance checking method based on extended description logic

TANG Ying-ying, LIU Wei[†], SU Xiao-ying, LIU Fei-jing

(School of Computer Engineering & Science, Shanghai University, Shanghai 200444, China)

Abstract: By studying the features of event class and relations among event elements, this paper proposed an event instance checking method based on extended description logic. This method supplemented semantic information at first, and then it verified whether an event was an instance of a specific event class by using effective reasoning capability of extended description logic, such as concept subsumption and satisfiability. Finally, it carried out an experiment to evaluate this approach. The experimental results show that this method can achieve instance checking of event classes with an accuracy of 85.12 percents.

Key words: description logic; event instance checking; event; event class

0 引言

认知科学家认为, 人类的命题记忆是以事件为存储单位的。事件是随着时间变化而变化的具体事实, 事件与事件之间存在本质的内在联系。人类描述历史, 就是描述一连串的事件及事件之间的联系。同样, 在互联网上, 大量的信息(尤其是新闻类)也是由文字描述的事件组成的。如何在爆炸式增长的互联网信息中获得有价值的信息, 是语义 Web 技术领域一项富有挑战性的工作。在计算机中引入面向事件的知识, 将事件作为描述互联网文本信息的知识表示单元, 能促进计算机类似于人脑进行信息处理, 推动文本语义处理技术的发展。事件知识的处理主要包括事件知识的表示和事件知识的推理。事件知识的推理是多层次的, 根据研究对象的粒度主要分为两类: a) 针对事件类及事件实例的推理, 如事件类实例检测、事件要素的缺省查询、事件状态推导等问题, 这个层次的推理把事件要素作为主要的研究对象, 研究事件与事件要素之间的关系, 以及要素与要素之间的关系; b) 针对事件与事件之间关系的推理, 包括分类关系推理(一个事件类包含另一个事件类)和非分类关系推理(即由一个事件推出其他的事件), 这个层次的推理把事件作为研究的基本知识单元。

事件实例检测即判断一个具体事件是否属于某个事件类。例如, 由一组关于“汶川大地震”的事实公理, 推导出“汶川大

地震是地震事件类的实例”。它是事件完成各个层次推理的基础, 一些关于事件本体的推理可以通过事件实例检测间接来完成, 如事件要素的缺省查询。此外, 它对于基于事件本体的应用, 如从海量的网络文本进行突发事件自动识别、分析和预警, 具有重要的意义。描述逻辑被认为是以对象为中心的表示语言中最为重要的归一形式。将描述逻辑的概念扩充为事件类, 将描述逻辑个体扩充为事件个体, 以这种方式对描述逻辑进行扩展显得非常自然, 甚至可以将经典描述逻辑中的概念当做是一种特殊的事件类。除此之外, 描述逻辑属于一阶谓词逻辑的可判定子集, 其基本语言具有良好的可扩展性, 而且包括概念的可满足性检测、概念包含关系的检测、等价关系检测、ABox 的一致性检测和实例检测等推理服务。这些特点使得在描述逻辑基础上改造其推理算法实现事件实例检测成为可能。在传统描述逻辑中, Tableau 算法^[1]被广泛地用于描述逻辑中判定概念的可满足性或概念的包含关系。然而事件作为一种大粒度的知识单元, 它与概念有着本质的区别, 描述逻辑中的 Tableau 算法不能直接用来实现事件实例的检测, 需要先将事件分解为时间、动作、环境、对象等事件要素, 然后通过事件实例是否满足事件类的各个要素的定义来进行判断。

通过分析和总结人类认识事件的客观规律以及事件在文本中的表现规律, 本文提出了一种基于扩展描述逻辑的事件实例检测方法。首先简单分析事件在文本中的表现规律, 接着介绍对象要素匹配算法, 该算法用来解决识别事件的参与者在事

收稿日期: 2013-11-20; 修回日期: 2013-12-27 基金项目: 国家自然科学基金资助项目(61305053, 61273328); 上海市自然科学基金资助项目(12ZR1410900)

作者简介: 唐英英(1988-), 女, 浙江人, 硕士, 主要研究方向为语义本体与推理; 刘炜(1978-), 男(通信作者), 副研究员, 博士, 主要研究方向为语义本体、知识表示(liuw@shu.edu.cn); 苏小英(1980-), 女, 讲师, 博士研究生, 主要研究方向为数据挖掘、自然语言处理; 刘菲京(1989-), 女, 硕士研究生, 主要研究方向为事件本体形式化。

件类中身份的问题;然后介绍事件实例中时间要素推理需要遵守的规则,并指出在环境要素和动作要素中往往存在一些隐含信息,这些隐含信息对于判断事件所属的事件类有很重要的作用,并简单介绍了环境要素和动作要素补充的方法;最后对大量的事件实例进行检测来验证本文提出的方法,并对检测结果进行分析。

1 相关工作

1.1 事件知识处理研究

客观世界中的事件是可以进行分类的。ACE 测评会议将事件分为 8 大类和 33 个子类型。文献[2]中的上层本体模型提出将事件分为自然事件、人为事件以及动态状态、静态状态。该本体以序类逻辑作为理论基础,采用 n 元谓词表示事件并用事件语义函数对事件进行定义。事件与对象之间的二元关系包括对象之间的关系、事件之间的关系、因果关系,其中事件之间关系包括事件实例与事件类的关系。Nguyen 等人^[3]指出可以通过回答几个问题来判断^[2]事件属于哪个事件类。在疾病报道中自然语言的语言表现与专家知识具有鸿沟的背景下,文献[4]设计了事件本体。它根据自然语言的语言表现将事件类的上层分为事件发生状态类型的事件类和事件类型的事件类。状态类型的事件类又可继续分为状态和过程。事件类的下层根据专家知识模型通过添加约束条件来建立。该事件本体机制包含疾病之间关联的模型、新闻报道的事件表达式和事件的形式化表示三个方面。在疾病爆发的监测中,综合运用这三方面的知识对新闻报道进行监测。郝秀兰等人^[5]在 HowNet 定义的事件类的基础上,通过为事件类定义角色语义表将 HowNet 的事件类与语义解释联系起来,从而完成事件类与事件实例类的联系。

刘宗田等人^[6-8]基于事件的文本语义处理技术研究提出了一个通用的事件本体模型,给出了事件和事件类的六要素表示,这也是本文工作研究的理论基础。文献[6]将事件定义为在特定时间和环境下发生的、由若干角色参与的、表现出一些动作特征的一件事情。其中,对象、动作、环境是事件中的三个事件要素。一个事件类是指具有相同特征的事件集合,事件分类的依据是事件的属性。事件类中的对象包括对象类型和角色类型的定义;事件类中的环境要素刻画事件发生场所及其特征;事件类的动作用来刻画这个类事件的变化过程及其特征。

1.2 描述逻辑扩展研究

描述逻辑是知识表示和推理的形式化工具,具有清晰的模型—理论语义和有效的概念分层等推理服务功能,同时是 OWL(由 W3 制定的一种语义 Web 的本体语言)的理论基础。随着语义 Web 的发展,描述逻辑的扩展成为语义 Web 的研究热点。因此,本文在前面的工作中根据事件和事件类的特点对传统的描述逻辑进行扩展,并给出了扩展描述逻辑的语法和语义^[9]。

在众多的传统描述逻辑中,SHOQ(D)具有较为强大的描述能力,它除了最简单的描述逻辑 ALC 具备的构子以外,还具备更多的性质。其中:S 命名可追溯到命题(多)模态逻辑,表示在 ALC 基础上允许部分角色具备传递性;H 代表角色可分层;O 代表枚举算子;Q 表示定性的数量限定;D 代表有型域,包括数值、字符串、时间等这类有型对象。因此,SHOQ(D)能够满足事件表示的部分需求,如事件类中的事件要素都是采用

SHOQ(D)概念进行定义的。

在传统的描述逻辑推理问题中,检测一个概念个体是否属于一个概念类称之为实例检测问题,该问题主要利用 Tableau 算法来解决。近年来,随着语义 Web 的发展,Tableau 的各种优化算法已经在实用推理机中得到实现。OWL Full、OWL DL 和 OWL Lite 是 OWL 的三种类型,OWL DL 和 OWL Lite 可多项式规约到 SHION(D)和 SHIF(D),描述逻辑推理机 FaCT++^[10]和 RACER^[11]已实现 OWL Lite 推理,Pellet^[12]则是一个较完善的 OWL DL 推理机。这些本体推理机在检测冲突、优化表达和实例的归类等方面发挥着重要的作用。

2 事件实例检测方法

事件知识的推理主要是指针对事件各个要素及事件与事件之间关系的推理问题,研究在事件粒度级别上的一些隐含相关的推理问题,以支持类似于人脑的联想与推理。事件实例检测属于事件推理中最基础的推理问题。本文将基于扩展描述逻辑的事件实例检测分为三个步骤:a)识别事件中参与者在指定事件类中的身份;b)先利用事件中时间的客观规律对事件进行推理,然后根据动作和环境要素在事件本体中找到对应的隐含信息,并将这些信息补充到事件中;c)综合利用描述逻辑的概念包含检测和概念可满足性检测,检测事件实例是否满足事件类的定义,当满足事件类定义时,则可认为是事件类的一个实例。

2.1 概念定义

人类认识事件主要是通过事件的发生过程以及发生产生的结果来区分,这些大量的事件通过大脑图像和视频记忆存储在大脑之中。因为信息量非常之大,大脑会将这些信息自动归纳,将具有相同的特征归为一类,例如当想到“地震事件”,脑子就会联想到大地摇晃的画面,当你眼睛看到类似画面时,大脑立马会想到可能地震要发生了。本文采用文献[6]中对事件的六要素定义,事件类则用各个要素的共同特性来表示。计算机识别事件可以学习人认识事件的方法,即通过检测事件的各个要素的可满足性来判定事件实例是否为某个事件类的实例。

定义 1 一个事件 e 是事件类 E 的实例当且仅当存在解释 I 使得事件类成立,记为 $\{e\} \models E$ 。

在文字或话语表述中,人们描述自己认知的事件有关的信息与知识,其中最一般的情况是表述一个事件发生或不发生。

定义 2 正事件表述是指一段文字或话语表达某个事件发生了。

正事件实例可以用 $E(e, s, p, m, q, t_1, t_2)$ 表示。其中: E 表示事件类名; e 表示事件名;是 E 的一个实例; s 为持久断言; p 为前置断言; m 为中间断言; q 为后置断言; t_1 为事件开始时间; t_2 为结束时间。例如有一个文字表述:8 月 28 日发生在北京 X 中学的学生集体中毒事件。可以表示为 $\text{Poison}(\text{student-poison}, \{\text{Student}(a), \text{study}(a, \text{school}), \text{name}(\text{school}, X), \text{School}(\text{school})\}, -, \{\text{Poisoned}(a), \text{AtPlace}(\text{school})\}, -, 28/8\ 00:00, 29/8\ 00:00)$ 。其中:Poison 是事件类;student-poison 指具体事件; a 、school 是个体;Student 是关于对象类型的谓词;study 是一个动作谓词;School 是一个环境谓词;name(school, X) Poisoned(a) AtPlace(school) 是一般的断言。

定义 3 负事件表述是指一段文字或话语表达了某个事件未发生。

负事件类实例可以用 $\neg E(e, s, p, m, q, t_1, t_2)$ 表示。

定义 4 半事件表述是指一段文字或话语表达了某个事件发生但未结束。

半事件实例采用 $\sim E(e, s, p, m, q, t_1, t_2)$ 表示。

正事件表述、负事件表述和半事件表述统称为事件表述。

2.2 对象要素匹配

事件中的对象是指事件的主要参与者,事件类描述的是一类事件的共同特征,因此事件类中的对象要素是这类事件中参与者的共同特征,在本文中称之为角色。按照事件的分工,一个事件类中的对象可能存在多个角色。例如,在食物中毒事件类中可能存在中毒者、食物和医生等角色。因此,在定义事件类过程中,需要详细给出角色的定义。一般情况下,角色可以认为是概念,所以对于角色的定义可以参考一般概念本体的定义方法。但不得不指出的是,在事件中人往往可以根据不同的情况继续进行分类,并且除了一般概念定义的方法以外,还可以通过特定事件中主要发生的动作对它进行约束,实现对角色的定义。例如,医生在描述逻辑中可定义为 $\text{Doctor} \sqsubseteq \text{People} \sqcap \text{treat. Patient}$,女医生则定义为 $\text{Doctor} \sqsubseteq \text{Female} \sqcap \text{People} \sqcap \text{treat. Patient}$ 。一个人在不同的事件下可能扮演不同的角色,例如一个女人,她的职业是医生,在家庭相关的事件中,她是孩子的母亲,是丈夫的妻子,在一些医疗事件中,她是一个医生。

具体事件中的参与者就用一个个体表示,如丹麦“世界最佳餐厅”食物中毒事件:

一名餐厅员工感染诺如病毒进入餐厅,导致一些顾客和员工出现呕吐和腹泻症状。

假设用 a 表示感染诺如病毒的员工,则可表示为 $\text{Staff}(a), \text{infect}(a, \text{norovirus})$,其中 Staff 和 infect 存在于对象要素子本体中, norovirus 是个体,表示诺如病毒。

定义 5 假设存在事件 e 和其候选事件类 E ,若 e 中存在一个参与者 a (a 表示对象实例), E 中存在一个角色 R (R 是一个概念), a 与 R 匹配当且仅当 e 中关于 a 的所有断言满足 R 的定义。

定义 6 对于任意关于个体 a 的断言集合 $\{C(a)\}$,存在一个匿名概念 AC (anonymous concept),使得 $AC(a) \sqsubseteq \{C(a)\}$ 。

例如,关于 a 的断言集合为 $\{A(a), B(a), C(a), r(a, b)\}$,其中 A, B, C 为概念, b 为概念 Object 的一个个体, r 为一个关系,则 $(A \sqcap B \sqcap C(\exists r. \text{Object}))(a)$ 与该断言集合的语义等价。

定理 1 对于匿名概念和角色概念的包含关系以及概念的可满足性, Tableau 算法是可判定的。

证明 事件类角色概念是 SHOQ(D)-概念,再结合定义 6 可知,匿名概念也是 SHOQ(D)-概念,对于 SHOQ(D)-概念的可满足性和包含关系推理,存在 Tableau 算法是可判定的^[1],因此,对于匿名概念和角色概念的包含关系以及概念的可满足性, Tableau 算法是可判定的。

定义 7 如果关于 a 的断言集合为对应匿名概念的实例 $AC(a)$,且 $AC \sqsubseteq R$, R 是事件类 E 中的角色,则称事件实例 e 的参与者 a 是匹配的。

算法 1 参与者—角色匹配

输入:事件实例 e , 候选事件类 E 。

输出:参与者与角色可匹配结果 $\text{Map}(\text{participant}, \text{List}(\text{role}))$ 。

- 1 获取候选事件类 E 中所有的角色,存放在数组 roles 中
- 2 获取事件实例 e 中所有的参与者,存放在数组 participants 中
- 3 根据定义 7 所述对 roles 与 participants 进行匹配,得到结果 Map

$(\text{participant}, \text{List}(\text{role}))$

4 返回 $\text{Map}(\text{participant}, \text{List}(\text{role}))$

在上述算法中,将参与者与事件类中的角色逐一匹配,找到与参与者可能匹配的所有事件类中的角色。算法 2 则将算法 1 得到的结果进行组合。通过使用算法 1 和 2 可以获取事件实例的参与者与事件类中角色匹配的所有情况。

算法 2 事件实例参与者与事件类对象匹配的所有结果 (matchingObjects)

输入:参与者与角色可匹配结果 $\text{Map}(\text{participant}, \text{List}(\text{role}))$ 。

输出:参与者—角色匹配的所有结果 $\text{List}(\text{Map}(\text{participant}, \text{role}))$ 。

1 当 $\text{map.size} = 0$ 时,返回 null ; 否则,在一个 map 中取出一个参与者 curParticipant 和它可匹配的角色 curRoles ;

2 在 map 中删除该参与者,并调用 matchingObjects 算法去匹配剩下的对象,得到结果记为 smallerList ;

3 curParticipant 与 curRoles 中一个角色进行匹配,并且加上 smallerList 中的一个匹配结果作为一个匹配结果,将所有匹配结果都加入到 $\text{List}(\text{Map}(\text{participant}, \text{role}))$;

4 返回 $\text{List}(\text{Map}(\text{participant}, \text{role}))$;

定义 8 如果事件某个参与者与事件类的一个角色是匹配的,则可以把事件类中关于该角色的断言替换成该参与者。把所有角色替换成参与者的过程称之为事件实例代换。

事件实例代换主要用于当确定或假设一个事件是某个事件类实例之后,用实例代换获得一般情况下,该类事件的事件要素之间普遍存在的事件关系,它其中一个重要的应用是事件缺省条件的补充。

2.3 语义补充

2.3.1 时间要素

人们在关注事件发生时往往是关注事件的某个时段。以此划分,有宏观时段和微观时段。宏观时段是要表述事件在一段时间内发生或经常发生这一事实,这是一般时态。微观时段是要表述事件发生到某一时间点的情况,通常为发生前时点、发生中时点和发生后时点。发生前时点是未来态,发生中时点是进行态,而发生后期点是完成态。用 TA 表示关注时段, N 表示当前时间。对于事件 e, t_1 为事件开始时间, t_2 为结束时间,则有:

事件的过去时: $(TA < N)$

事件的现在时: $(TA = N)$

事件的将来时: $(N < TA)$

事件的一般态: $(t_1, t_2 \in TA)$

事件的完成态: $(t_1 < t_2 < TA)$

事件的进行态: $(t_1 < TA < t_2)$

事件的将来态: $(TA < t_1 < t_2)$

三种时和三种态共组成七种。

规则 1 $E(e, s, p, m, q, t_1, t_2) \Rightarrow (s, < >) \text{ and } (p, < t_1) \text{ and } (m, t_1 < > t_2) \text{ and } (q, t_2 <)$ 其中 $(s, < >)$ 表示一直成立, $(p, < t_1)$ 表示断言 p 在 t_1 之前成立, $(m, t_1 < > t_2)$ 表示断言 m 在 t_1 到 t_2 之间成立, $(q, t_2 <)$ 表示断言 q 在 t_2 之后成立。

惯性原理 $(p, t <)$ 表示如果在 t 之后没有事件改变 p , p 在 t 之后一直成立; $(p, < t)$ 表示自 t 之前导致 p 的事件后直到 t 之前一直成立。

规则 2 $\neg E(e, s, p, m, q, t_1, t_2) \Rightarrow (\neg m, t_1 < > t_2) \text{ and } (\neg q, t_2 <)$

例如 昨天他没来,得

$\neg \text{come}(\text{he-come}, \text{able}(\text{he}, \text{move})), \neg \text{at}(\text{he}, \text{here}), \text{at}(\text{he},$

way-to-here), at(he, here), $t_1, t_2 \in TA = \text{yesterday} \Rightarrow t_1, t_2 \in TA = \text{yesterday}$ and (\neg at(he, way-to-here), $t_1 < > t_2$) and (\neg at(he, here, $t_2 <$)

规则 3 $\sim E(e, s, p, m, q, t_1, t_2) \Rightarrow (s, < >) \text{ and } (p, < t_1) \text{ and } (m, t_1 < > t_2) \text{ and } (\neg q, t_2 <)$

例如, 他现在还没来到 得

$\sim \text{come}(\text{he-come}, \text{able}(\text{he}, \text{move}), \neg \text{at}(\text{he}, \text{here}), \text{at}(\text{he}, \text{way-to-here}), \text{at}(\text{he}, \text{here}), t_1 < N < t_2) \Rightarrow t_2 < N \text{ and } (\text{able}(\text{he}, \text{move}), < >) \text{ and } (\neg \text{at}(\text{he}, \text{here}), < t_1) \text{ and } (\text{at}(\text{he}, \text{way-to-here}), t_1 < > N) \text{ and } (\neg \text{at}(\text{he}, \text{here}), N <)$

将时间要素应用于事件实例语义补充的主要方式是根据事件实例的事件表述类型选择对应的规则进行推理, 其中, 惯性原理是推理需要遵守的原则。

2.3.2 环境要素

环境是事件的一个重要组成成分。在描述一个事件时, 往往会给出事件发生的地点, 如丹麦首都哥本哈根知名餐厅“诺马”。任何事件必然发生在一定的环境之中。在某些环境中, 对象之间存在一些特定的关系, 实例中的“餐厅”, 它既说明了事件发生的具体地点, 同时还隐含对象之间可能存在的关系, 如图 1 所示。

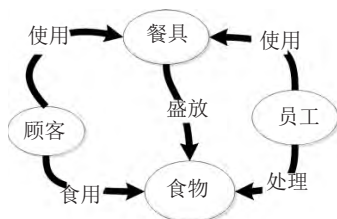


图 1 在餐厅的环境下部分对象之间的语义关系

因此本体中还需要对环境在本体中定义特定环境的一些特征, 如“餐厅”, 描述如下: $\text{Restaurant} \{ \text{use}(\text{Customer}, \text{Equipment}), \text{use}(\text{Staff}, \text{Equipment}), \text{eat}(\text{Customer}, \text{Food}), \text{deal_with}(\text{Staff}, \text{Food}), \text{put_on}(\text{Food}, \text{Equipment}) \}$ 。其中 Customer、Equipment、Staff、Food 都是在餐厅环境下不同的角色, use、eat、deal_with、put_on 存在于动作要素子本体之中。上述例子表达了在餐厅这个环境下, 顾客、餐厅的工作人员通常会使用桌子、碗碟等器具, 顾客是来吃食物的, 餐厅的工作人员烹饪食物, 食物是盛放在碗碟器具之上等事实。

如果事件实例中存在关于环境的断言并且该环境表示一个特定环境场景, 则用 2.2 节所述的方法对事件的参与者与环境中的角色进行映射, 然后用事件实例代换的方式生成关于参与者的断言添加到事件。例如, 一个事件的参与者有 People(Tom)、Staff(Jeff)、Food(food)、Equipment(bowl), 动作为 eat(Tom, food), 环境为 Restaurant(xuehai), 前置断言 deal(Jeff, food)、put_on(food, bowl)。首先识别 Tom、Jeff、food 在餐馆中的角色, 其中 Jeff 和 food、bowl 很容易判断出分别是 Restaurant 中的 Staff、Food 和 Equipment。Tom 处在环境 Restaurant 中, 且来吃东西, 满足 Restaurant 环境下顾客的定义, 则认为有可能是顾客, 然后进行实例代换, 把 $\{ \text{use}(\text{Tom}, \text{bowl}), \text{use}(\text{Jeff}, \text{bowl}) \}$ 断言加入到前置断言中。

2.3.3 动作要素

动作是对事件变化过程及其特征的描述, 也是体现事件动态性的关键, 包括对程度、方式、方法、工具等的描述。动作分为抽象动作和具体动作。抽象动作是对具体动作的抽象, 是一类动作总的概括; 具体动作则是抽象动作的一个实例。一个抽

象动作包含两种类型对象: a) 动作的主动者, 即主体; b) 动作的被动者, 即客体。一个动作是由形如 $\text{Action}(\text{subject}, \text{object})$ 表示。一个抽象动作采用动作发生的前置条件和动作产生的效果进行定义, 即 $\text{Action}(\text{subject}, \text{object}) = (\{ \text{Pre}, \text{Post} \})$ 。其中, Pre、Post 是关于 subject、object 的断言, 通过断言的形式表示动作的程度、方式、方法、工具等状况。

定义 9 动作实例代换是指动作 $\text{Action}(x_1, x_2)$ 中的所有变量用常量如 (a_1, a_2) 进行代换, 称 $\{a_1/x_1, a_2/x_2\}$ 是动作 Action 的一个实例代换。

事件本体中动作要素用形如 $\text{Action}(x_1, x_2) = \{ \text{Pre}, \text{Post} \}$ 进行定义。一个具体的动作可以经过一组 $\{a_1/x_1, a_2/x_2\}$ 的动作实例代换获得, 并且在动作实例代换之后, Pre 和 Post 是关于个体 a_1, a_2 的断言。

事件关于动作要素的补充就是当事件中存在某个动作时, 用事件中具体动作的主体和客体代换抽象动作的主体和客体, 并把动作实例代换后的 Pre 和 Post 分别添加到事件的前置断言和后置断言之中。

2.4 事件实例检测流程

人类在描述一个事件时往往是通过简单、扼要的词句进行描述, 这些词句虽然少, 但是其他人能够理解事件发生过程, 这是因为每个人的大脑有一个共享的知识库(称为共享是因为对某个客观的事物, 每个人对它的认识基本上是一致的), 一个词句往往触发知识库里的其他信息进入人的大脑中, 使描述的事件具有更多的信息量, 进而听到或者看到某个事件描述时能够了解事件发生的全貌。

计算机认识事件也可以模仿人的思维方式。在本文的事件实例检测流程中, 先识别事件实例中参与者在事件类中的角色, 然后利用事件中存在的时间、动作和环境要素对事件实例进行语义补充, 再去判断事件实例是否满足事件类的定义。例如, 丹麦餐厅食物中毒事件先判断“丹麦餐厅食物中毒事件”这一事件是否是食物中毒事件类的一个实例时, 首先判断“丹麦餐厅食物中毒事件”中的参与者在事件类中充当什么角色, 若判断出现呕吐和腹泻情况的顾客和员工是食物中毒事件类中的中毒者, 在上述实例中虽然没有出现顾客和员工吃了坏掉东西的描述, 但是人们仍旧能了解他们是因为吃了带有病毒的食物才导致呕吐和腹泻, 这是“餐厅”和“感染”在人们的大脑中加工得到的信息。因此在事件实例检测中, 需要在事件本体中找到这些有隐含信息的要素, 然后将它们添加到事件实例中使事件更加完整, 最后根据事件实例的对象、时间、环境、动作等要素是否都满足事件本体库中关于“食物中毒”事件类中各要素的描述断言。当都满足时, 判定“丹麦餐厅食物中毒事件”属于“食物中毒”事件类的一个实例; 不满足时, 则认为“丹麦餐厅食物中毒事件”不是“食物中毒”事件类的实例。

3 实验和分析

3.1 实验数据集描述和实验设计

为了说明本文提出方法的有效性, 从网上搜取关于中毒事件相关的新闻报道和叙事文本共 144 篇, 并且以段落为事件的标注单元, 手工提取事件要素和识别事件指示词。以段落为事件的标注单元, 这是因为段落中的各个事件要素存在紧密的联系, 一个事件要素可能是多个事件的事件要素, 以句子为标注单元会使手工提取以后的事件缺少过多事件要素而影响实验

结果。此外,如果根据标注的人来判断事件的边界,则会引入太多的主观因素。本实验中抽取的事件均为文章中较为重要的事件。手工抽取的触发词、对象要素、动作要素和环境要素的具体情况如表 1 所示。其中在抽取的事件中 52.8% 的环境要素与事件指示词之间存在一定关系,这也验证了同种类型的事件往往发生在特定的环境之中这一观点。

表 1 实验中数据集事件要素统计

事件触发词	对象要素	动作要素	环境要素
363	668	497	216

在对实验语料进行预处理之后,事件检测分为三个步骤:

a) 通过匹配算法,找出所有事件中参与者和事件类中角色匹配的所有情况,取出其中一种情况对事件实例添加一些约束条件,例如,“丹麦世界最佳餐厅食物中毒事件”中呕吐的顾客是“食物中毒”的中毒者,本体中毒者的角色号为 role1,则对个体呕吐的顾客添加一条 $\text{hasRole}(a, \text{role1})$,其中 a 指的是呕吐的顾客;b) 先合理利用 2.3.1 节中三条规则对事件实例进行补充,再在本体中找到实例中的动作和环境对应的抽象动作和环境类,通过实例代换获得隐含信息,并将这些信息填充到事件实例中;c) 检测当前事件实例是否满足本体中事件类的定义,并且判断角色匹配可能情况是不是全部验证完,如果没有,重新回到步骤 a)。完整的事件实例检测流程如图 2 所示。

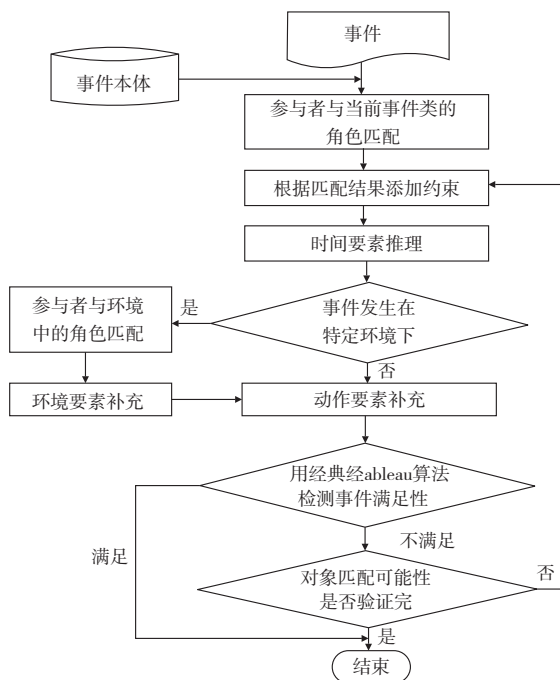


图 2 事件实例检测流程

为了进一步说明事件要素补充的必要性和事件要素补充对事件实例检测结果的影响,本文还设计了三组对比实验:仅使用了时间要素,仅仅使用了时间、动作要素,仅仅使用了时间、环境要素对事件实例进行检测。

3.2 评价标准

精确度(P)、召回率(R)和 F_1 值是广泛用于统计学分类领域的三个度量值,用来评价结果的质量。其中,精确度用来衡量系统的查准率,召回率用来衡量系统的召回率, F_1 值结合了准确率和召回率。本文实验的目的是了解事件类识别情况和事件要素补充情况,因此,需要重新给出精确度、召回率、 F_1 值的定义,新的计算公式如下:

$$P = \frac{\text{精确识别事件类的事件数}}{\text{待识别的事件数}} \times 100\% \quad (1)$$

$$R = \frac{\text{正确识别事件类的事件数}}{\text{待识别的事件数}} \times 100\% \quad (2)$$

$$F_1 = \frac{2PR}{P+R} \quad (3)$$

其中:精确定义的实验检测结果与人的事件识别结果一致,并且实例中事件参与者的角色匹配结果正确,以及不存在不正确的事件要素。当实验检测结果符合人的事件识别结果,则认为这次检测结果是正确的;当实验检测的结果与人判定结果不一致时,则认为是错误的。

3.3 实验结果分析

图 3 是 363 个事件进行实验的统计结果,横坐标表示 3.1 节所提及的四种方法,条形图的高度表示每种方法在 363 次检测中得到正确事件类的检测结果的事件数和精确确定语义关系的事件数。从图 3 可以直观地看出,对比其他方法,本文方法检测效果更好。

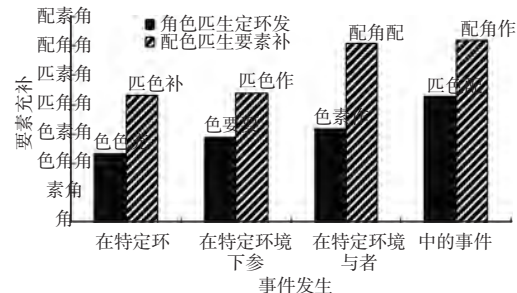


图 3 四种方法精确识别和正确识别事件数统计结果

精确率、召回率和 F_1 值是本次实验的三个重要指标。表 2 是根据式(1)~(3)计算获得的结果。从表 2 的实验结果可以发现,加入事件的环境要素和动作要素隐含的信息之后,检测结果的精确度和召回率都有所提高,并且同时补充这两个动作、环境要素隐含的消息比加入其中一个的效果来得更明显。但是在加入隐含信息的过程中还会出现两种情况:

a) 有些无关的信息会加入到事件实例中,这些断言虽然成立,但是与事件检测无关。不可否认,事件越复杂,事件检测效率越低。对于一个简单的事件,这个问题造成的影响不是太大;但是对于关系复杂的事件检测,这个问题是需要引起注意的。

b) 事件个别要素之间关系的识别精确度有待提高,如环境要素。当一个复杂事件中存在多个环境要素和动作要素环境,不能较准确地判断动作发生的环境。

表 2 四种方法的检测结果

实验组	precision / %	recall / %	F_1 measures / %
仅利用时间推理	32.23	59.50	41.81
仅时间、动作要素	39.67	60.33	47.87
仅时间、环境要素	43.80	83.47	57.45
本文方法	58.68	85.12	69.08

本文方法的检测结果很大程度上依赖于事件本体,事件本体覆盖率越大,准确率越高,本文方法越有效。从实验中检测不正确的实例来看,检测出现错误的主要原因是:文档中已有的信息不能(能)满足本体中事件类中的定义,但是人却可以判断这个实例(不)成立。通过实验可以看出,把描述逻辑作为事件表示的工具,并利用关于事件的一些客观规律对一般事件进行信息补充,可以达到较准确的事件检测结果。

4 结束语

事件本体的提出为克服传统本体的缺陷提供了新的解决

思路,事件本体的推理和事件的表示同为事件本体的重要研究内容,事件的实例检测是事件推理中一个基础,同时也是重要的研究内容。扩展描述逻辑针对事件的特点扩展了描述逻辑的语法,并给出了它们的语义。本文在此基础上研究事件要素的内在联系,充分利用扩展描述逻辑的概念层次推理、概念满足性等可判定的推理服务功能,提出基于扩展描述逻辑的事件实例检测方法;然后通过大量文档实例对该方法进行测试,并且采用精确度、召回率、 F_1 测度三个评测标准对实验结果进行评价;最后对实验结果进行分析。

当然本文的方法还存在一些不足:a)本文方法虽然可以使用对象要素匹配识别方法,判断参与者在事件类中的角色,但是没有考虑事件实例中存在不同命名实体表示同一个人或物的情况,即没有将事件实例中同一个参与者进行归并;b)本文实验的数据采用以段落为标注单元,这种方法可以解决事件中对象要素、动作要素等重复利用的问题,但是不能解决所有问题。例如这个文档中如果存在多个事件和多个环境实体,不能较准确地判断事件实例和环境要素的关系。从实验的结果来看,这个问题虽然对检测事件是否是指定的事件类的实例未造成很大的影响,但是对识别事件实例中的事件要素的精确度有影响。如何完善事件本体,使事件本体具有较高的覆盖率和准确率,如何对事件实例参与者进行归并,如何更准确地判断事件要素(如环境要素、时间要素)是属于哪个具体事件,是今后工作的主要目标。

参考文献:

- [1] 梅靖,林作铨.从 ALC 到 SHOQ(D):描述逻辑及其 Tableau 算法[J]. 计算机科学,2005,32(3):1-11.
- [2] KANEIWA K, IWAZUME M, FUKUDA K. An upper ontology for event classifications and relations[C]//Advances in Artificial Intelligence. Berlin:Springer,2007:394-403.
- [3] NGUYEN P H P, KANEIWA K, CORBETT D R, et al. Representing event assertions in an upper event ontology[C]//Proc of International Conference on IEEE Knowledge and Systems Engineering. 2009:120-125.
- [4] KAWAZOE A, CHANLEKHA H, SHIGEMATSU M, et al. Structuring an event ontology for disease outbreak detection[J]. BMC Bioinformatics,2008,9(3):S8.
- [5] 郝秀兰,杨尔弘,舒鑫柱.基于 HowNet 的事件角色语义特征提取[J]. 中文信息学报,2001,15(5):26-32.
- [6] 刘宗田,黄美丽,周文,等.面向事件的本体研究[J]. 计算机科学,2009,36(11):189-192.
- [7] 仲兆满,刘宗田,李存华.事件本体模型及事件类排序[J]. 北京大学学报:自然科学版,2013,49(2):234-240.
- [8] 张旭洁,刘宗田,刘炜.事件与事件本体模型研究综述[J]. 计算机工程,2013,39(9):303-307.
- [9] LIU Wei, XU Wen-jie, FU Jian-feng, et al. An extended description logic for event ontology[C]//Advances in Grid and Pervasive Computing. Berlin:Springer,2010:471-481.
- [10] TSARKOV D, HORROCKS I. FaCT++ description logic reasoner: system description[C]//Lecture Notes in Computer Science. Berlin:Springer,2006:292-297.
- [11] HAARSLEV V, MÖLLER R. Racer: an owl reasoning agent for the semantic Web[C]//Proc of International Workshop on Applications, Products and Services of Web-based Support Systems. 2003:91-95.
- [12] SIRIN E, PARSIA B, GRAU B C, et al. Pellet: a practical OWL-DL reasoner[J]. Journal of Web Semantics: Science,2007,5(2):51-53.

(上接第 3284 页)案 4 的比值为 1.16:1,方案 2、3 介于两者之间。

因此,在相同客运需求的条件上,实行高峰时段增加票价、平峰时段降低票价的定价策略将均衡高峰、平峰时段的客流量,同时能够增加企业收益,利于调动企业积极性,最终涉及多方利益的社会福利也将增大。并且实行高峰增加票价同时平峰降低票价的方案(即方案 4)比仅对高峰时段增加票价(即方案 2)或仅平峰时段降低票价(即方案 3)效果最为明显,平峰时段降低票价的弹性高于高峰时段增加票价的弹性(即方案 3 高峰/平峰公交客流比值较方案 2 小),这与高峰时段乘客出行中必须出行所占比重较大有关,即高峰时段票价增高对乘客改变出行时间或转变出行方式影响小于平峰时期降价的效果。

5 结束语

根据城市交通出行时间不均衡性的特点,提出了时间差别定价的概念,构建了上层为社会福利最优模型,下层通过非必要性出行系数描述多方式分时段城市交通网络随机弹性需求的公交时间差别定价模型。研究结果表明,时间差别定价乘客对高峰时段票价增加的敏感性低于对平峰时段票价折扣的敏感性,主要是因为高峰时段乘客大多为必要出行(如上班、上学等),从而有可能导致高峰时段公交客流并没有明显下降,而诱增了平峰时段的客流量。因此,时间差别定价可与目前我国很多城市已实施的错峰上下班政策协调实施,效果将更为显著。而模型中公交运营企业高峰/平峰时段的运营成本差异的表示、出行者高峰/平峰时段出行行为的描述及大规模城市复

合城市交通网络的算法求解可靠性还有待进一步研究。

参考文献:

- [1] CERVERO R. Time-of-day transit pricing: comparative US and international experiences [J]. Transportation Review, 1986, 6(4): 347-364.
- [2] CERVERO R. Transit pricing research: a review and synthesis [J]. Transportation, 1990, 17(2): 117-139.
- [3] BURRIS M W, PENDYALA R M. Discrete choice models of traveler participation in differential time of day pricing programs [J]. Transport Policy, 2002, 9(3): 241-251.
- [4] 王健,陈娟.基于差别定价的城市公共交通价格体系研究[J]. 科学技术与工程,2010,10(22):5467-5469.
- [5] 金键,杨德明.公交票制票价多样性及客流效应分析[J]. 交通运输系统工程与信息,2010,10(3):115-122.
- [6] 袁建,叶亮.公交服务水平对公交票价调整的影响分析[J]. 价格理论与实践,2007,4(5):23-27.
- [7] 姚丽亚,关宏志,严海.公交票价比对公交结构的影响及方式选择模型[J]. 北京工业大学学报,2007,3(8):42-46.
- [8] 刘海旭,蒲云.多类型弹性需求随机用户平衡分配模型[J]. 系统工程理论方法应用,2005,14(2):159-163.
- [9] 周晶,徐晏.弹性需求随机用户平衡分配模型及其应用[J]. 系统工程学报,2001,16(2):88-84.
- [10] 史若燃,李志纯.不同市场机制下多种交通方式定价模型研究[J]. 交通运输系统工程与信息,2010,10(5):91-97.
- [11] WILLIAMS H L. Formation of travel demand models and economic evaluation measures of user benefit [J]. Environment and Planning A, 1977, 9(3): 285-344.
- [12] EVANS A. A theoretical comparison of competition with other economic regimes for bus services [J]. Journal of Transport Economics and Policy, 1987, 21(1): 7-36.
- [13] 陈坚,杨飞,晏启鹏.多方式复合城市交通网络弹性需求随机用户平衡分配模型[J]. 计算机应用研究,2012,29(10):3693-3696.
- [14] 周晶,盛昭瀚,何建敏.弹性需求下公交网络系统票价结构的优化[J]. 自动化学报,2001,27(5):637-643.