

# 事件前提和后果常识知识分析方法研究

李闪闪<sup>1,2</sup> 曹存根<sup>2</sup>

(首都师范大学计算机科学联合研究院 北京 100037)<sup>1</sup>

(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)<sup>2</sup>

**摘要** 事件推理的研究需要事件的知识作支撑。事件知识作为一种常识性知识,其获取仍然是知识获取的瓶颈。提出一种事件知识分析方法,以从多个视角分析事件的前提和后果知识。在对事件知识分析中遵循基础分析、知识丰富、知识优化的流程。针对分析过程中所得知识不够全面和准确的问题,总结得出生理、心理、社会、物理世界4个常识知识角度,同时在知识丰富步骤中提出了角色细化的方法。实验表明,此方法能够得到较为全面和准确的事件知识。

**关键词** 事件,常识知识角度,多视角常识分析,角色细化

**中图分类号** TP391.1 **文献标识码** B

## Commonsense Knowledge Analysis Approach Based on Event Preconditions and Effects

LI Shan-shan<sup>1,2</sup> CAO Cun-gen<sup>2</sup>

(Joint Institute of Computer Science, Capital Normal University, Beijing 100037, China)<sup>1</sup>

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,

Chinese Academy of Sciences, Beijing 100190, China)<sup>2</sup>

**Abstract** Event reasoning requires support of commonsense event knowledge. We proposed a method of event knowledge analysis in order to analyze preconditions and effects of events from multiple perspectives. In the process of analysis we followed three phases: fundamental analysis, knowledge supplement, and knowledge optimization. We summarized four common sense knowledge dimensions to solve the problem of comprehensiveness and accuracy in analysis, which are physiological, psychological, social and physical. The method of role specification was proposed in the knowledge supplement phase. Experiments show that this method is promising for acquiring comprehensive and accurate commonsense event knowledge.

**Keywords** Event, Common sense knowledge dimension, Multi-perspective commonsense analysis, Role specification

## 1 引言

20世纪60年代John McCarthy提出了行动推理研究<sup>[1]</sup>,事件演算以其强大的表示和实现能力,在行动推理中取得了良好效果<sup>[2,3]</sup>。目前对于事件推理主要致力于事件的表示和推理方法<sup>[4]</sup>的研究,但这些方法的检验和应用需要有事件的前提和后果作支撑。智能系统中也必须有预先存储的事件的前提和后果知识才能进行推理。获取事件的前提和后果是进行事件推理的基础,但目前对于获取事件前提和后果常识的研究很少。

事件前提和后果的获取属于常识知识获取研究的范畴,相对国外来讲,国内在常识知识获取方面的研究起步较晚,但研究者们一直在不懈努力着。在常识知识库建立方面,中国科学院董振东教授开发了知网<sup>[5]</sup>。在常识语料抽取方面,朱耀对句子的常识获取的合适程度进行了分析研究<sup>[6]</sup>。在基于

概念的常识获取方面,曹亚男提出了一种将语言模型和统计模型相结合的方法以获取概念的属性类常识和对比类的常识<sup>[7,8]</sup>。在基于事件的常识获取方面,彭会良提出一种相关事件挖掘的方法<sup>[9]</sup>以帮助从文本中直接获取事件相关的常识知识,曹亚男对面向Web语料的因果知识获取进行了全面的研究<sup>[10]</sup>。总体来说常识知识获取任务是困难的,一方面常识定义模糊,对于已获取的常识知识难以评价;另一方面常识具有隐式的特点,即很多常识属于人们对现实世界的缺省假设,这些假设通常不在文本中显示表达。目前按获取的自动化程度可将常识知识获取分3类:手工获取(以Cyc<sup>[11]</sup>和OMCS<sup>[12]</sup>为代表)、半自动获取<sup>[13]</sup>和自动获取。手工获取方式能够得到隐式的常识,但效率低下而且难以保证知识的完备性。自动获取使得知识获取效率得到较大提高,但较难获取隐式的常识。半自动获取是对手工获取和自动获取的折中,需要有种子知识库的支持。总之,目前还没有一种公认的

到稿日期:2012-10-27 返修日期:2013-01-28 本文受国家自然科学基金(30973713,61173063,61035004,61203284,91224006),国家社科基金(10AYY003)资助。

李闪闪(1988—),女,硕士生,主要研究方向为大规模知识获取,E-mail:lss198810@163.com;曹存根(1964—),男,研究员,博士生导师,主要研究方向为知识获取与共享、文本挖掘。

较好的常识知识获取方法。

本文针对事件推理过程中缺乏事件前提和后果的现状给出一种获取事件前提和后果的方法(将事件的前提和后果作为事件的知识,将获取事件前提和后果的过程称为事件知识分析的过程。下文中的事件知识即指事件的前提和后果,事件知识分析即获取事件前提和后果的分析)。事件前提和后果的常识绝大多数属于隐式的常识,如对于事件“小明在喝水”有一条前提“小明是人”这类常识知识,人们不会在文本中明确写出,因此很难通过文本挖掘的方式自动获取。为了获取到这类隐式的常识,本文通过手工获取方式构建事件前提和后果的常识知识库。其目的一方面是获取到事件的前提和后果的常识知识库,另一方面是通过事件知识分析的过程总结得出获取事件前提和后果常识中的问题及解决方法,寻找自动或半自动的获取此类常识的线索,为将来实现半自动的获取事件常识知识做准备。本文以现代汉语基本句模为基础,通过对句模的修改和扩充得到事件体系,并对事件体系中的每一个事件进行知识分析。通过对交易、给予、索取3类中56种事件的考察,发现由于各人思维的主观性和局限性,对事件进行知识分析过程中容易产生遗漏、以偏概全、结果不准确的现象,为了得到全面准确的事件知识,需要解决以下难题:

- (1)总结分析事件知识的角度;
- (2)寻找扩充事件知识的方法;
- (3)提高所得知识的准确性。

针对上述问题,提出从生理、心理、社会、物理世界4个常识知识角度来思考事件的知识,并通过对角色的细化和相似事件的类比对事件的知识进行丰富,同时在角色细化的过程中通过添加条件状态提高知识的准确性,并将拥有海量开放信息的Web作为信息的来源,通过对相关事件的查询得到新的信息,以弥补个人思维的局限性。根据上述内容,我们将事件知识分析过程分为基础分析、知识丰富、知识优化3个阶段。分析过程中主要采用谓词的方式描述所得知识,为了便于以后的知识推理引入了缺省逻辑,为了减少冗余引入了继承理论<sup>[14]</sup>,分析中还对现有部分概念进行了扩充,如在“认知主体”概念的基础上提出了“强认知主体”的概念。实验表明,按此流程对事件进行知识分析可以得到较为全面和准确的事件知识。

本文第2节介绍事件的知识表示模型,第3节介绍了常识知识的角度;第4节结合实例介绍事件知识分析的过程;第5节通过实验来验证本方法的可行性;最后为总结和未来的工作。

## 2 事件的知识表示

事件作为知识表示的一种方法,其概念在诸如语言学、信息处理、动作推理、自然语言理解等不同领域均有出现。WordNet中“事件”定义为“在特定地点和时间发生的事情”<sup>[15]</sup>。自动文摘领域, Filatova将事件单元等同于一个句子<sup>[16]</sup>。刘宗田等将事件定义为〈动作、对象、时间、环境、断言、语言表现〉六元组<sup>[17]</sup>。鲁川将事件定义为以一个事元为中枢、一个或多个相关的事元为周边的语义组合所反映的客观现实<sup>[18]</sup>。我们将事件作为一种动作或状态的变化,认为事件是由特定参与者参与的、在满足一定条件的情况下发生的,并能对参与者或周边环境造成一定影响的行为。

知识分析的对象是事件,为了使所分析的事件之间具有系统性,减少事件选取的盲目性,我们将现代汉语基本句模中的每个句模作为一个事件。鲁川根据谓词的分类将现代汉语基本句模分为26类共100个句模<sup>[19]</sup>。其表示形式如图1所示。这些句模涵盖范围广泛且自成体系,但存在同一句模中谓词含义不一致等不利于事件知识分析的现象。为了保证分析的顺利进行,我们对句模做了适当的修改和扩充。首先,基本句模中同一句模的谓词含义并不完全相同,为使每个句模都表示一个确定的语义,需将同一句模中的谓词按照同义词进行分组,然后扩充为多个句模。其次,句模所描述的是一种非常规范的简单句式,是理想状态下的句子模式,与自然语言多样化的表达形式相距较远,为了处理实际语句,把表达同一语意的不同句式写到同一事件类中,同时句模中谓词数量有限,需要进行同义词的扩充。再次,句模中的介词均以“〈K〉”表示,不利于理解和处理,需要将其改为具体的介词类。最后,在用计算机进行自然语言处理的过程中,通常选用上下文无关文法对自然语言进行分析,文法中产生式的表示选用BNF范式。为了与这一习惯相统一,对句模的表示形式也进行了局部的修改,以使其表示为BNF范式的形式。图2是对图1的句模进行上述修改后得到的一个句模。经过上述修改,我们得到了事件的基本体系。

【施事】	〈K〉向事	【V】	【受事】	【捐,发,捐献,奉献,缴纳】
工作组	给灾民	发放	救济金。	(某人给某人无偿的“发放”某物)

图1 现代汉语基本句模的表示形式

〈捐献语句〉::=	〈施事〉〈给-对象介词词类 向-对象介词词类〉〈向事〉〈捐献词类〉[〈受事〉
	[〈施事〉〈捐献词类〉〈向事〉〈受事〉
	[〈施事〉〈捐献词类〉〈受事〉〈给-对象介词词类〉〈向事〉
	[〈施事〉〈把-对象介词词类〉〈受事〉〈捐献词类〉〈给-对象介词词类〉〈向事〉
〈捐献词类〉::=	捐 捐献 赠送 捐赠 捐钱 捐款 捐资

图2 修改后“捐献类”的表示形式

事件的表示模型指事件的组成元素及元素之间的关系。Nelson结合脚本的知识表示方法提出事件的GER模型<sup>[20]</sup>, Filatova将“元事件”表示为动词及3类相关的命名实体<sup>[6]</sup>,周文在此基础上提出了多元组表示模型<sup>[21]</sup>。王寅提出事件域认知模型(ECM)以研究事件元素之间的搭配关系<sup>[22]</sup>。ACE事件抽取任务中每类事件都有相应的事件模板。鲁川将事件表示成“中枢角色”及其相应的“周边角色”。

上述几种事件表示模型都无法应用于事件推理任务。这些事件模型的定义中所涉及到的事件元素都是在描述事件的语言中可见的,均通过语法和语意这些语言层次上的内容来揭示事件的组成元素及元素之间的关系,不涉及事件的前提、后果这些无法单纯从事件的描述语言中得到的信息。我们将这些事件表示为事件的语言表示模型,事件的语言表示模型能较好地应用于事件抽取任务,但无法应用于事件推理任务。

刘宗田等提出的事件六元组表示模型中<sup>[7]</sup>事件的断言包含了事件的前置条件、中间断言、后置条件,但此模型中对前置、后置条件表示较为粗略,不利于事件推理任务的进行。为使事件的表示模型能够应用于事件推理等任务,我们将前提和后果作为事件表示模型的一部分,提出了事件的知识表示模型。此模型保留了事件语言表示模型的内容,并在语言表示模型的基础上增加了前提、后果、角色细化的内容,定义如下:

定义1 事件的知识表示模型是一个六元组  $E=(\text{中枢角色,语言表示,谓词表示,前提,后果,角色细化})$ 。

中枢角色;也称谓词,表示事件的核心动作;

语言表示:指描述此事件的句式的集合,由现代汉语基本句模修改扩充而来;

谓词表示:由事件的语言表示改写成的一阶谓词逻辑表示形式;

前提:指事件的发生需要各事件参与者满足的约束条件;

后果:指事件发生后各事件参与者及其周边元素发生的状态变化;

角色细化:指事件各参与者可能的不同类别,角色类别的不同会带来新的前提和后果,因此角色细化中同时包含与角色属性相关的事件知识。

中枢角色、语言表示、谓词表示都属于事件的语言层次的内容,图2中〈捐献语句〉后跟随的内容是“捐献类”的语言表示,〈捐献词类〉后的内容是“捐献类”的中枢角色。捐献类谓词表示为:捐献(〈施事〉,〈向事〉,〈受事〉)。前提、后果和角色细化(主要指跟角色细化相关的前提和后果)属于事件的知识层次的内容,是知识分析的重点。

### 3 事件的常识知识角度

事件知识分析就是把事件发生的前提和产生的后果深植于人脑而计算机却无从得知的知识形式化地表示出来。分析的目的在于得到全面准确的知识,但这些知识以不为人知的形式存储于人脑中,我们无从知晓这些知识的全貌,因此无法断言是否得到了事件的全部相关知识,我们所能做的只是尽可能多地获取这些知识。Cyc在进行知识录入过程中为了加快知识录入的速度和提高知识推理的有效性,提出用12个不同的维度来描述常识知识的上下文(Contexts,指常识知识成立的条件),其研究表明,这种表示方式确实可以在一定程度上提高知识录入速度和推理的有效性<sup>[23]</sup>。在对事件进行知识分析时,我们首先会根据个人的经验从常识的角度考察事件的前提和后果知识。类比Cyc的研究成果,我们认为将事件的前提和后果知识通过不同的维度来描述可以加快事件知识分析的效率,这就需要总结描述事件知识的维度。彭会良在获取人物相关事件的常识知识时,总结了人物相关事件的常识知识角度,并通过实验验证了所得角度的合理性<sup>[24]</sup>。为了使分析中的思考过程有章可循,减少遗漏,以彭会良总结得出的常识知识角度为基础,归纳总结出生理、心理、社会、物理世界4个常识的知识角度,如图3所示。其中生理、心理、社会方面主要是人类相关的,物理世界方面主要是物体相关的。

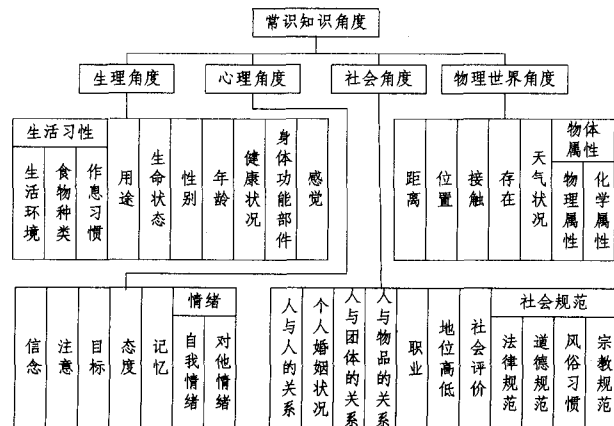


图3 常识知识角度结构图

### 3.1 生理角度

生理角度指事件参与者中有生命的那些角色的生理特点。当事件中参与者为有生命的事物时,事件的发生需要参与者具有某些生理特性,因此需要对这些事元进行生理方面的分析。生理角度包括生活习性、用途、生命状态、性别、年龄、身体功能部件、健康状况、感觉等。

生活习性指生活环境(水栖、陆栖、双栖)、食物种类(草食、肉食、混合)、作息习惯(昼伏夜出、白天活动)等。用途针对动物而言包括观赏、食用、劳作等。如耕牛是一种陆栖草食类动物,白天活动主要用于劳作。

生命状态包括健在和死亡两种状态,非特别指出时,均认为是活着的状态。性别指某些事件对角色性别的限制。年龄指某些事件对角色年龄的要求。比如说某人怀孕了,则这个人必须是活着的女性,而且到了可以怀孕的年龄。

健康状况指参与者是否健康,包括健壮、体弱、生病、受伤等。身体功能部件指参与者必须具有的身体的某种功能。感觉包括外部感觉和内部感觉。外部感觉指由于外界的刺激作用于感觉器官所感受到的,包括视觉、味觉、嗅觉、听觉、皮肤感觉。内部感觉指由于身体内部的刺激所引起的感受,包括饿、渴、恶心、疼痛等。如李明拿刀砍伤了王刚,则李明必须是有手的且王刚处于受伤状态会感受到疼痛。

### 3.2 心理角度

心理角度,指事件参与者中属于人类的那些角色的心理特征。在事件中,大部分是人类发出的动作,这些动作大多是意向性过程,因此具有一定的心理状态或认知状态。为此需要对施事、受事等进行心理角度的分析。心理方面包括信念、注意、意识、目标、态度、记忆、情绪等。

信念属于认知,指参与者内心相信为真的事。注意指参与者的注意力在哪里以及注意的方式,包括有意注意、无意注意、有意后注意。意识指参与者做某件事是有意识的还是无意识的,包括有意识、无意识、下意识、潜意识。如李明送给女朋友一束花,则李明在送花之前相信女朋友喜欢花,送花后李明的注意力在女朋友身上,送花是一种有意识的行为。

目标是参与者想要达到的状态或希望保持的状态。态度指参与者对其他的人或物的态度,包括喜欢、厌恶、同情、信任等。记忆指人对某些事记忆时间的长短,记忆的形式包括形象记忆、语词逻辑记忆、情绪记忆和运动记忆。情绪指由于外界的刺激产生的心情、心境等,包括自我情绪和对他情绪两种。自我情绪有高兴、悲伤、忧虑、害怕、吃惊等,对他情绪有感激、怨恨、敬佩等。

如李明借给王刚一万块钱帮助他度过难关,则李明的目标是帮助王刚,李明对王刚的态度是同情,王刚对李明的帮助行为会记忆深刻,这种记忆是一种情绪记忆,王刚对李明会产生感激的情绪,王刚自身会因为得到帮助而产生高兴情绪。

### 3.3 社会角度

社会角度,指事件参与者中属于人类的那些角色的社会状态。事件发生在社会中,事件的发生需要参与者具有一定的社会属性,因此需要对事件中属于人类的事元进行社会角度的分析。社会角度包括人与人的关系、人与团体的关系、个人婚姻状况、人与物品的关系、职业、地位高低、社会评价、社会规范等。

人与人的关系指人的社会角色关系,如父母子女关系、老师学生关系、领导职员关系、夫妻关系等。个人婚姻状况包括

未婚、已婚、离异、丧偶。社会规范指人们社会行为的规矩和准则,包括风俗习惯、法律规范、道德规范、宗教规范。如李明跟李红结婚了,则李明跟李红是夫妻关系,两人都处于已婚状态,相应地,就有了夫妻的责任以及应遵守的道德法律规范,即两人在离婚之前不能再婚、不能背叛对方等。

人与团体的关系指人是否属于某个团体,若属于某个团体,在团体中担任什么角色等。职业指人所从事的职业类别。地位高低包括人的家庭地位、社会地位、学术地位的高低。社会评价既包括长期以来对某人的印象,也包括由特定事件带来的评论。如王老师经常帮助学生解决学习中遇到的疑惑,则王老师属于某个学校或科研实体,王老师的职业是教师,老师的学术地位高于学生的学术地位,经常帮助学生会受到学生的尊敬,社会评价会是德高望重。

人与物品的关系包括拥有与失去等。如李明买了同学的一辆自行车,则李明现在拥有了这辆自行车,他的同学失去了自行车。

### 3.4 物理世界角度

物理世界角度,指可以用物理属性衡量的那些状态,包括距离、位置、接触、物体的属性、天气状况、存在等。

距离指不同参与者之间距离的大小。位置指不同参与者之间的位置关系。接触指不同参与者之间是否有物理上的接触。如总理亲切地跟他握手,则总理和他距离在一米范围以内,两人位于同一位置,握手过程中两个人的手有物理上的接触。

物体的属性主要指物体的形状、颜色、物质状态(固态、液态、气态)、物质结构属性、物理属性、化学属性等。存在指某些事物存在于某处,区别于人对事物的拥有。天气状况包括晴天、阴天、下雨、刮风等。如李明从河里钓了一条鱼,则河里的水没有结冰即水是液态的,且河里有鱼但鱼不属于河所有,因此表示成河里存在鱼,且天气不应该是处于下雨状态,因为从常识的角度考虑没有人会冒雨去钓鱼。

需要明确的一点是,并不是所有的事件都会包含常识4个角度的所有不同方面的知识,一个事件可能只具有某些方面的知识。

这些事件的常识知识角度为我们的知识分析提供了思考的方向,可以帮助我们有条理地总结事件的前提和后果知识,能有效减少信息的遗漏。

## 4 事件知识分析的过程

事件知识分析由基础分析、知识丰富、知识优化3个阶段构成,本节按照这3个阶段的流程以“捐献类”为例介绍事件知识分析的过程。事件知识分析的流程图如图4所示。

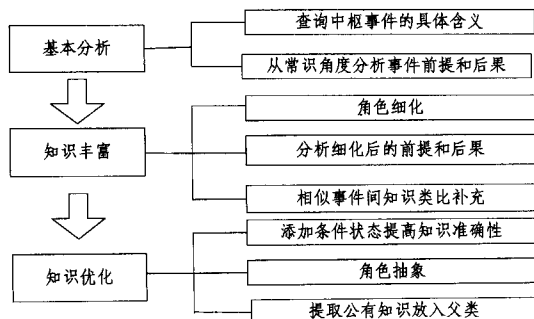


图4 事件知识分析的流程图

### 4.1 基础分析

基础分析是从常识的知识角度考察事件应该具有的前提和后果。我们将基础分析得到的知识称为基础知识。首先需要查询中枢角色的具体含义,以保证对事件理解的正确性,避免由于认识偏差造成分析不准确,对于某些事件还可以从具体含义中抽取出事件的前提和后果。然后再从常识的知识角度进行分析。

#### 4.1.1 查询中枢角色的具体含义

首先,我们需要通过查询词典了解待分析事件的意思、义项。针对“捐献”,我们查阅了《现代汉语大词典》<sup>[25]</sup>,可知“捐献”的含义:是一种无偿的奉献(将自己的财物,无偿地送给国家、集体、慈善机构、个人)。向国家、集体、慈善机构或个人献出自己所有的东西来帮助物质上需要帮助的人、集体或地区。

根据“捐献”的含义,可得到该事件的如下前提知识:

前提知识:

社会角度:

人与物品关系:

拥有(⟨施事⟩,⟨受事⟩)

~拥有(⟨向事⟩,⟨受事⟩)

心理角度:

信念:

相信(⟨施事⟩,需要⟨⟨向事⟩,⟨受事⟩⟩)

.....

#### 4.1.2 从常识知识角度分析事件的前提和后果

接下来从第3节中事件的常识知识角度来分析捐献类在各个角度的不同方面存在哪些常识知识。

前提方面,从生理角度看“捐献类”的施事应该是认知主体(认知主体指有认知能力的主体,包括人、机构等),但不能是婴幼儿、植物人、精神病人等无行为能力者,从心理角度讲施事应该相信向事会接受受事。

后果方面,心理角度施事是帮助了他人自身会感到高兴,社会角度捐献物品后会发生物品所有权的转移,物理世界角度所捐献的物品大多都是可以移动的,部分不可移动的物品主要是施事出资帮向事进行建造,且所建造的受事应该位于向事所在地区。

因此该步骤可以得到捐献类事件的如下前提和后果知识:

前提知识:

生理角度:

综合:

是一个(⟨施事⟩,强认知主体)

心理角度:

信念:

相信(⟨施事⟩,接受⟨⟨向事⟩,⟨受事⟩⟩)

.....

后果知识:

心理角度:

自我情绪:

高兴(⟨施事⟩)

社会角度:

人与物品关系:

失去(⟨施事⟩,⟨受事⟩)

得到(⟨向事⟩,⟨受事⟩)

物理世界角度:

位置:

not(可移动(⟨受事⟩))

—> 出资建造(⟨施事⟩, ⟨受事⟩)

∧ 同一地区(⟨向事⟩, ⟨受事⟩)

.....

其中,“强认知主体”指不包括婴幼儿、植物人、精神病人等无行为能力者的“认知主体”,not()表示了 Negation As Failure 的缺省推理,可解释为没有证据表明。上例可解释为:在没有证据表明受事是可移动的条件下,可以推出受事是施事出资进行建造的,并且建好后与向事位于同一地区。

## 4.2 知识丰富

知识丰富是对基础分析的补充,以使分析的结果更加全面,减少遗漏。知识丰富可以由角色细化和相似事件间的知识类比补充得到。

### 4.2.1 角色细化

本工作中知识分析的对象是抽象事件,事件角色均由“施事”、“受事”等语义角色表示,没有具体的类别。分析抽象的事元角色得到的知识需要对所有满足此事件语义限制的角色都成立,其条件要求较强,得到的知识较少。角色细化的目的是将抽象的事元角色(施事、受事、向事等)具体为特定的角色(农民、小偷、富人等),以分析这些特定角色下事件应该具有的前提和后果,扩充事件的知识。

角色细化首先要获取可用的候选角色,这可以通过构造查询模式并查询 Web 语料再对其进行抽象概括得到。对于“捐献类”事件,查询过程中会发现有很多关于明星、公司、乞丐等的捐献事件的报道以及慈善机构接受捐款的事件,还有关于捐献器官的新闻等。总结得出“捐献类”的候选角色如下:

施事:穷人、富人、明星、富有企业、困难企业、非盈利机构、黑社会分子、企业家、乞丐、大学生、拾荒者.....

向事:灾民、贫困者、慈善机构、同性恋者、寺庙、学校、博物馆、病人.....

受事:钱、生活用品、食品、帐篷、药品、衣服、学习用品、文物、艺术品、器官、血液、遗体.....

得到候选角色后需要对角色进行筛选。为了使得细化后的角色能扩充事件的知识并且不存在冗余,我们认为筛选角色时应遵循如下原则:

(1)细化后得到的角色应具有新的知识。新的知识指未具体化前和已具体化的角色都没有的前提和后果。

(2)抽象程度较高原则。如果一些可用角色间具有上下位关系,并且它们所得出的知识相同,就选择上位角色作为具体化的角色。这一原则既有效减少了角色细化中角色的数量,也避免了事件知识的丢失。

(3)抽象事元角色细化时,应选择那些差异性较大、代表性较强的角色,即在选择角色时要尽量选取不同类别的角色。

根据上述原则,得到“捐献类”的角色细化如下:

施事:穷人、富人、明星、富有企业、困难企业、非盈利机构、黑社会分子

向事:贫困者、灾民、慈善机构、非盈利机构、博物馆、病人

受事:生活用品、学习用品、历史物品、人体组织、金钱

从上述细化结果中看出,角色筛选的同时还需要对某些角色进行进一步的抽象概括,如向事中将寺庙、学校概括为非盈利机构。

细化过程中发现有些情况下同一事件中某一角色确定后

会给另一角色带来限定,如“捐献类”中的受事细化为器官后,向事就应该是病人。为了在细化中表示角色之间的这种限定关系,需要在细化时对两个角色同时进行细化。同时发现对于某些事件,不同事元角色之间的相互作用也可以带来新的事件的前提和后果。因此角色细化时还需要考虑两个或两个以上角色间的相互作用。我们将涉及到多个事元角色的细化称为关联细化。相应地将细化某一个角色称为独立细化。

例如“赠送类”事件施事跟向事之间的关系会影响到事件的前提和后果。“父母给子女送东西”、“商人给贪官送东西”二者的前提和后果有很大差别。因此细化时需要按照两者的关系进行细化,如细化为“亲人关系”、“利益关系”等。

### 4.2.2 分析细化后的前提和后果

角色细化后同样是从常识的知识角度来分析事件在特定角色下所具有的前提和后果,只是这里需要考虑到角色相关的属性。与基础知识相对应,我们将角色细化中得到的知识称为衍生知识。如对于施事细化为明星后,考虑到明星喜欢通过炒作提高知名度,因此得到一条细化后的知识:

后果知识:

社会角度:

社会评价:

是一个(⟨施事⟩, 明星)

—> 有嫌疑(⟨施事⟩, 炒作)

角色细化后,还可以根据角色细化的结果查询相关事件的评论,根据评论可以使我们对知识的描述更加准确,同时也能激发联想,避免遗漏相关的知识。

如从评论“网民们在赞美、敬仰和歌颂老年乞丐的大义之举的同时,纷纷指责万科集团为富不仁、爱心缺失、社会责任感寡廉。捐款和爱心、善意、社会责任感”中可总结得出施事细化为穷人时有如下知识:

后果知识:

社会角度:

社会评价:

生活困难(⟨施事⟩)—> 品德高尚(⟨施事⟩)

∧ 社会责任感(⟨施事⟩, 强烈)

∧ 爱心(⟨施事⟩, 强)

大多数(X, 认知主体) (生活困难(⟨施事⟩))

∧ 不等(⟨施事⟩, X)—> 敬仰(X, ⟨施事⟩)

∧ 感动(⟨施事⟩, X)

### 4.2.3 相似事件间知识类比补充

类比可以清晰表达不同对象的相似性,允许在相似性的领域中进行知识的转换。类比的基本思想是:对于在某些方面具有相似性的两个对象在其他方面也有可能相似。类比的过程就是将一个对象的相关断言映射到另一个相似对象上的过程。将类比思想应用到事件知识分析上,提出以下假设:如果两个事件相似,那么它们的前提、后果也可能相似。

在我们的事件体系中,26类中的每一类称为一个一级类;对于一级类中的事件按照语义分成了不同的小组,这些小组称为二级类;每个二级类包含一种以上的事件,这些事件称为三级类。二级类中所包含的事件之间往往具有相似性。在这些二级类中的事件就可以进行知识的类比补充。

如“捐献类”和“赠送类”同属于“给予文法(一级类)”中的“自愿无偿给予类(二级类)”。在“捐献类”中容易得到以下知

识:

后果知识:

心理角度:

对他情绪:

急需(<向事>,<受事>)->感激(<向事>,<施事>)

分析可知这条知识在“赠送类”中也是成立的,如果“赠送类”中没有此条知识,就可以将这条知识添加到“赠送类”中。

对相似事件间的知识进行类比补充可以有效减少遗漏,使得知识更加全面丰富。

### 4.3 知识优化

上述知识分析过程中得到的知识存在不准确和冗余的现象。知识优化是为了使得到的知识更加准确,同时减少冗余使知识的结构更加简洁。

#### 4.3.1 添加条件状态提高知识准确性

对事件进行知识分析时,由于认识的偏差可能导致所得知识是不准确的,这种不准确是因为把部分条件下成立的知识错看成任何情况下都成立造成的,如果单纯地舍弃这类知识将使知识不完整,因此我们通过添加条件状态来表示部分情况下成立的知识。条件状态指在某个条件成立的情况下,事件的发生才能有的前提或者后果。

如开始分析时“捐献类”中得到以下知识:

前提知识:

心理角度:

态度:

同情(<施事>,<向事>)

在角色细化时发现,在器官的捐献中如果受事是单一器

官并且是整体器官的捐献,则施事处于脑死亡状态,即施事是没有意识的,因此上述知识是不准确的,应该将其修改为:

有意识(<施事>)->同情(<施事>,<向事>)

#### 4.3.2 角色抽象

角色抽象是角色细化的延伸,指考察角色细化中得到的知识,将可以脱离角色本身的属性而成立的衍生知识放入基础知识中。

角色细化是一种有效的激发联想的方式,不仅可以分析得到事件跟角色属性相连接所激发出的新知识,还可能得到由角色属性联想到的跟角色本身无关的知识。如在“捐献类”中向事细化为同性恋者时分析得到如下知识:有影响(<施事>)->关注(<社会中人>,<向事>),分析可知此条知识与向事角色无关,应该放在捐献类的基础知识中,即此条知识的所处位置不正确。角色抽象就是要把这类本该属于基础知识的衍生知识进行位置调整,重新放入事件的基础知识中。

#### 4.3.3 提取公共知识放入父类

由上文知事件体系中有一级、二级、三级类,其抽象程度依次降低。二级类继承一级类、三级类继承二级类。继承者为子事件,被继承者为父事件。由于一级类过于抽象,我们只对二级类和三级类进行知识分析。事件知识分析过程中先对父事件进行分析,再对子事件进行分析。父事件中已有的知识在子事件中不需要重复写出。并将分析的结果组织成继承树的结构,子事件继承父事件的知识,子事件所具有的知识是其所有父事件与其自身知识的总和。这种继承结构可以有效减少冗余。

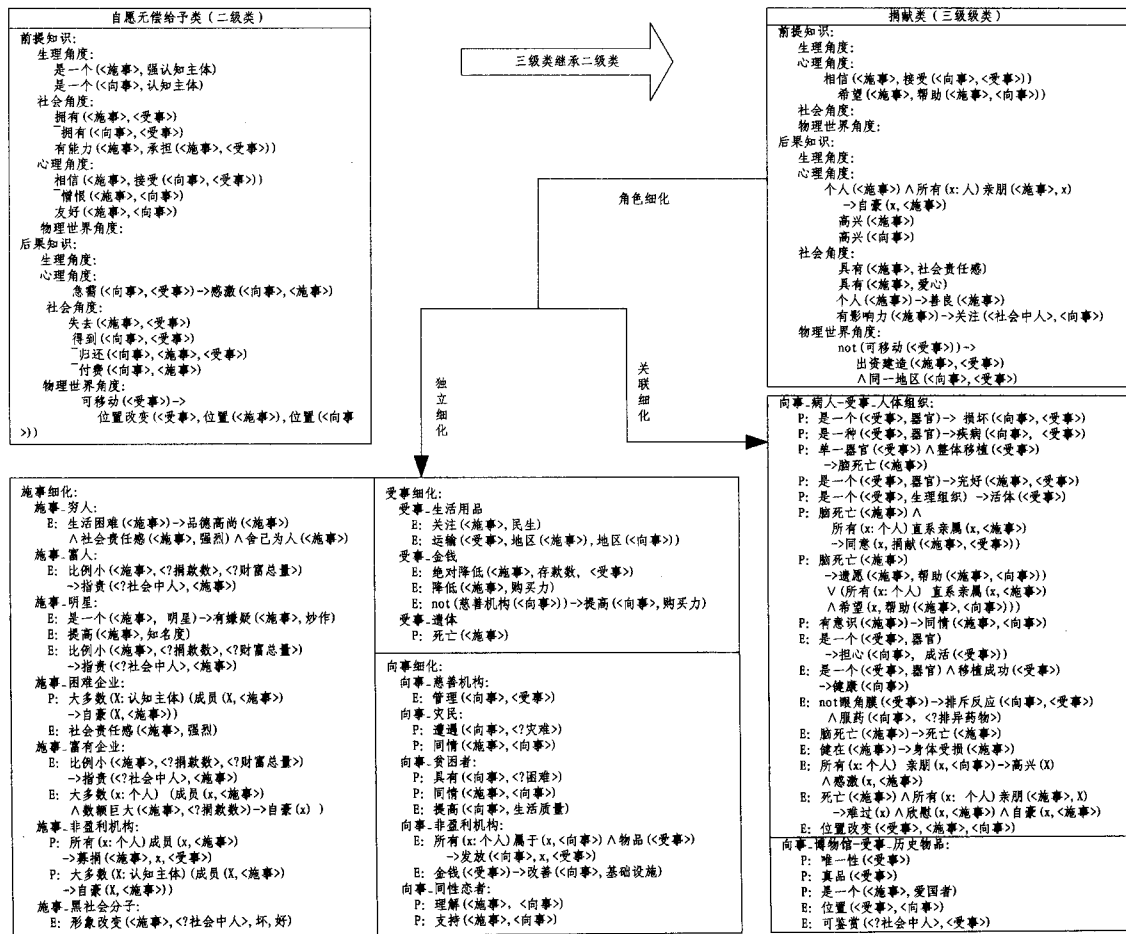


图5 “捐献类”及其父类知识分析结果

父事件抽象程度较高,分析中得到的知识有限,对子事件进行分析并进行相似事件间知识类比补充后可能出现同一父事件下的所有子事件都包含同一知识的现象。这种同时被所有子事件包含的知识即为公共知识。提取公共知识放入父类的过程就是把这种公共知识放入父事件中,同时从所有子事件中删除此条知识的过程。这样做可以减少冗余。

例如“捐献类”是“自愿无偿给予类”的子类,经过对“自愿无偿给予类”中的相似事件进行知识类比补充,得到“捐献类”中的知识:急需(<向事>,<受事>)->感激(<向事>,<施事>),其对所有兄弟事件均适用,因此需要将此条知识放入“自愿无偿给予类”中,而“捐献类”及其所有兄弟事件均不需要再包含此知识。

由上述分析步骤得到的捐献类及其父类的知识分析的结果如图 5 所示。对于二级类由于其比较抽象,不进行角色细化。图 5 中,白色空心箭头表明“捐献类”继承自“自愿无偿给予类”,小箭头表示分析得到的知识。前提与后果中分别列出从生理、心理、社会、物理世界 4 个角度分析得到的内容,施事细化指对施事进行角色细化后得到的知识,每个细化类别由两部分组成,第一部分表示细化的具体类别,第二部分表示细化得到的知识。由图 5 中“捐献类”及其父类的知识分析结果可以看出,应用本文的方法可以得到较全面的事件前提和后果知识。

## 5 实验和分析

为了进一步验证本方法的有效性,做了如下实验。从事件体系中选取了“购买类”、“抢劫类”、“审讯类”、“死亡类”4 个三级类事件,这些事件来自不同的一级类。首先进行了对比实验,即在不使用和使用本文方法的情况下分别对上述 4 类事件进行知识分析,所得结果如表 1 所列。

表 1 对比实验结果

	类别	购买类	抢劫类	审讯类	死亡类
直接分析	前提知识	5	6	4	2
	后果知识	6	9	8	4
使用本文 防范分析	前提知识	9	15	11	4
	后果知识	15	15	8	11
	角色细化知识	23	22	2	6

表 1 分别列出了不使用本文方法和本文方法进行事件知识分析时得到的前提和后果知识的数目,其中角色细化知识是对角色细化后得到的跟角色属性相关的知识的数目。

由表 1 可以看出,使用本文方法可以明显增加前提和后果的数目,同时角色细化也能得到角色相关的知识,在表中还可看出审讯类角色细化后得到的知识比较少(只有两条),这主要是由于参与审讯类的角色本身有较多限定,可细化的角色类别较少。

为了进一步验证分析结果的准确性,选取了 10 个不同的人对由使用本文方法所得知识在准确性方面进行评分。这 10 个人均为在读硕士研究生,年龄在 21~24 岁之间,在阅历、文化背景、教育程度、专业知识方面有能力强弱,事件知识的准确性且未参与到事件知识分析的过程中。

准确性指所得到的事件前提和后果是否是合适的,有没

有例外情况。评分标准如下:

- 0: 完全不是前提或后果;
- 1: 个别情况下是前提或后果;
- 2: 小部分情况下是一个前提或后果;
- 3: 大部分情况下是一个前提或后果,能想到反例;
- 4: 感觉存在不成立的情况,但暂时想不到反例;
- 5: 是一个前提或后果,想不出反例。

通过评分,以每条知识的平均得分作为知识的最终得分,知识的得分分布情况如图 6 所示。

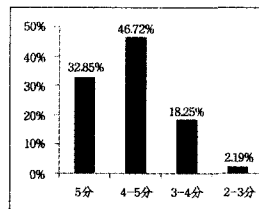


图 6 知识得分分布图

事件的知识属于常识知识。由常识知识模糊性的特点,可以认为得分在 3 分以上的知识就是一条正确的常识知识。由图 6 可以看出,有 97.81% 的知识得分达到了 3 分以上,由此可认为本文方法得到的知识在准确性方面达到了很好的效果。

对于得分较低的情况,主要是由于个人思维的有限性使得在对事件进行知识分析的过程中考虑得不够全面,导致将部分条件下成立的知识误认为所有条件下成立的知识。这类知识并不是错误的,只要添加适当的条件状态就可以使其成为准确的知识。

**结束语** 本文将事件作为知识表示的基本单位对其进行知识分析,给出了事件的知识表示模型,阐述了常识知识的角度及知识分析的步骤。本工作着重研究得到事件前提、后果的方法,这是一项十分基础的工作,对于动作推理、常识知识获取都有重要的意义。利用手工整理的方式获取事件的前提和后果的优势在于可以发掘出文本中没有的隐性知识,且能保证所得知识是有用的。但这种方式工作量大且繁琐,需要知识工程师有极大的耐心与细心,同时也不能保证所得知识的完备性。下一步工作包括进一步完善事件知识分析的方法,提高知识分析的效率;将分析得到的结果应用于动作推理和规划等领域,以使其具有实用的价值。

## 参考文献

- [1] McCarthy J, Hayes P. Some philosophical problems from the standpoint of artificial intelligence[J]. Machine Intelligence, 1969, 4: 463-502
- [2] Shanahan M. The event calculus explained: Artificial Intelligence Today[C] // Recent Trends and Developments. Berlin: Woodrige M J, Veloso M M, 1999: 409-430
- [3] Shanahan M. An abductive event calculus planner[J]. The Journal of Logic Programming, 2000, 44(1-3): 207-240
- [4] 朱景阳. 基于贝叶斯方法的事件推理[J]. 科技信息, 2010(33): 219-225
- [5] 董振东, 董强. 知网[EB/OL]. <http://www.keenage.com/>

- [6] Zhu Yao, Zang Liang-jun, Cao Ya-nan, et al. A Manual Experiment On Commonsense Knowledge Acquisition From Web Corpora[C]// International Conference on Machine Learning and Cybernetics, Kunming, China, 2008
- [7] Cao Ya-nan, Cao Cun-gen, Cao Liang-jun, et al. Acquiring Commonsense Knowledge about Properties of Concepts from Text [C]// 5th International Conference on Fuzzy Systems and Knowledge Discovery, 2008
- [8] Cao Ya-nan, Cao Cun-gen, Cao Liang-jun, et al. Extracting Comparative Commonsense from the Web [C]// 6th International Conference on Intelligent Information Processing, 2010
- [9] 彭会良, 曹存根. 相关事件挖掘与角色联系发现的研究[J]. 计算机科学, 2010, 37(12): 149-155
- [10] 曹亚男. 面向 Web 语料的因果知识获取研究[D]. 北京: 中国科学院研究生院, 2012
- [11] Lenat D. CYC: A Large-scale Investment in knowledge Infrastructure[J]. Communications of the ACM, 1995, 38(11): 33-38
- [12] Singh P. The Public Acquisition of Commonsense Knowledge [C]// AAAI Spring Symposium on Acquiring Linguistic Knowledge for Information Access, 2002
- [13] Chklovski T A. Using Analogy to Acquire Commonsense Knowledge from Human Contributors[D]. Boston; Massachusetts Institute of Technology, 2003
- [14] 刘东立, 姚天顺. 自然语言处理中继承理论的分类和应用[J]. 东  
北大学学报, 1997, 18(5): 486-489
- [15] WordNet. A lexical database for the English language [EB/OL]. <http://wordnet.princeton.edu/>
- [16] Filatova E, Hatzivassiloglou V. Domain-independent detection, extraction, and labeling of atomic events [C]// Proceeding of RANLP, 2003: 145-152
- [17] 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
- [18] 鲁川. 知识工程语言学[M]. 北京: 清华大学出版社, 2010: 245-248
- [19] 鲁川, 黎瑞隆, 董丽萍. 现代汉语基本句模[J]. 世界汉语教学, 2000(4): 11-24
- [20] Nelson K, Gruendel J. Event Knowledge; Structure and Function in Development[M]. Hillsdale; Erlbaum, 1986
- [21] 周文. 基于概念的若干知识表示模型及相关方法研究[D]. 上海: 上海大学, 2007
- [22] 王寅. 事件域认知模型及其解释力[J]. 现代汉语, 2005, 28(1): 17-26
- [23] Lenat D. The Dimensions of Context-Space [EB/OL]. <http://www.casbah.org/resources/cycContextSpace.shtml>
- [24] 彭会良. 人物相关事件的常识知识获取方法研究[D]. 北京: 首都师范大学, 2010
- [25] 阮智富, 郭志新, 乐嘉民, 等. 现代汉语大辞典[M]. 上海: 汉语大词典出版社, 2000: 860-860

(上接第 176 页)

**结束语** 本文在标签集约束路径查询问题的基础上, 提出标签集约束路径集合查询问题, 给出求解的生成树压缩存储结构与倒排索引技术。同时, 理论证明倒排索引存在良好的压缩效益。给出两个优化算法解决树结构上无源点信息的集合查询问题。实验证明, 在大规模图上, 生成树方法与在线的 DFS 和预存储标签路径传递闭包相比, 在时间和空间开销上都有优势。下一步, 在本文的基础上, 继续研究如何改进压缩路径标签的存储结构以及查询集合包含关系的算法, 以进一步减少时间和空间开销。

### 参 考 文 献

- [1] Jin Ruo-ming, Hong Hui, Wang Hai-xun, et al. Computing Label-Constraint Reachability in Graph Database [C]// SIGMOD'10, 2010: 123-134
- [2] Zou Lei, Xu Kun, Yu J X. Answering Label-Constraint Reachability in Large Graphs[R]. TR-DB-ICST-PKU-2011-002. Institute of Computer Science and Technology
- [3] Fang Wei. TEDI: Efficient Shortest Path Query Answering on Graphs [C]// SIGMOD '10, 2010: 99-110
- [4] Jin Ruo-ming, Xiang Yang, Ruan Ning, et al. 3-HOP: A High-Compression Indexing Scheme for Reachability Query [C]// SIGMOD '09, 2009: 813-826
- [5] Wang Hai-xun, He Hao, Yang Jun, et al. Dual labeling: Answering graph reachability queries in constant time [C]// ICDE '06, 2006: 75
- [6] Yan Y, Wang C, Zhou A, et al. Efficiently querying RDF data in triple stores [R]// Tech report, 2008
- [7] Gou Gang, Chirkova R. Efficiently querying large xml data repositories; A survey [J]. IEEE Trans. Knowl. Data Eng., 2007, 19(10): 1381-1403
- [8] Jagadish H V. A compression technique to materialize transitive closure [J]. ACM Trans. Database Syst., 1990, 15(4): 558-598
- [9] Cohen E, Halperin E, Kaplan H, et al. Reachability and distance queries via 2-hop labels [C]// Proc of the 13th annual ACM-SIAM Symp on Discrete Algorithms, 2002: 937-946
- [10] Chomsky, Noam. Three Models for the Description of Language [J]. IRE Transactions on Information Theory, 1956, 2(3): 113-124
- [11] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512
- [12] Abiteboul S, Vianu V. Regular path queries with constraints [C]// PODS, 1997: 122-133
- [13] Ramasamy K, Patel J M, Naughton J F. Set Containment Joins: The Good, The Bad and The Ugly [C]// Proc of the 26th VLDB Conf. Cairo, Egypt, 2000
- [14] Mendelzon A O, Wood P T. Finding regular simple paths in graph databases [J]. SIAM J. Comput., 1995, 24(6): 1235-1258
- [15] Leskovec J, Singh A, Kleinberg J M. Patterns of influence in a recommendation network [C]// PAKDD'06, 2006: 380-389