

基于概念代数的事件表示与运算规则

张旭洁¹ 刘宗田² 刘念祖¹

(上海立信会计学院数学与信息学院 上海 201620)¹ (上海大学计算机工程与科学学院 上海 200072)²

摘要 “事件”作为大粒度的知识表示单元,是互联网上重要的信息载体,越来越受到学术界的重视。提出了一种基于概念代数的“事件”表示方法,其可以描述事件与其对象要素、环境要素、时间要素、动作要素之间的关系,表示事件之间的时间、逻辑与分类关系,表示事件的状态变化,记录各类事件要素的语言表现,并能进行“事件”概念的层次运算。最后通过实例说明该表示方法与运算规则的可行性和合理性。

关键词 事件,事件要素,概念代数,概念运算

中图法分类号 TP391 文献标识码 A

Concept Algebra-based Representation Model and Conceptual Calculation Rules for Event

ZHANG Xu-jie¹ LIU Zong-tian² LIU Nian-zu¹

(School of Mathematics and Information, Shanghai Lixin University of Commerce, Shanghai 201620, China)¹

(School of Computer Engineering & Science, Shanghai University, Shanghai 200072, China)²

Abstract As a large-grained unit of knowledge, “event” is an important information media on Web, which has attracted more and more attention and high regards from the academia. This paper proposed a novel representation model and conceptual calculation rules of events on the basis of Nilsson’s concept algebra. The representation model is able to express the relations between events and their elements. Furthermore, it can express dynamic procedure of event and relations between events. Case studies demonstrate the representation model and conceptual calculation rules are effective and widely applicable.

Keywords Event, Event elements, Concept algebra, Conceptual calculation

1 引言

同一般的概念相比,“事件”包含参与对象、时间、地点、状态变化等信息,被认知科学家认为是人类命题记忆的存储单元[1]。这样的“事件”存储单元不但存储组成事件的概念及其之间的关系,还包含了“事件”之间的联系[2]。如何有效、合理地运用知识表示方法对“事件”概念及其所包含概念之间的关系进行表示,并将“事件”概念运用到智能信息处理中,成为了近年自然语言处理、信息检索、舆情分析等信息处理领域所关注的内容。目前,有关“事件”的研究主要围绕事件的表示(事件本体)、事件与事件要素的信息抽取与推理等展开研究[3-9]。

本文在事件六要素表示模型[9]的基础上,从概念代数的角度给出“事件”的概念代数表达式。该表达式可以描述“事件”概念与事件各要素之间的关系,事件之间的关系,表示事件的状态变化,描述事件相关的语言表现等,并在此基础上对“事件”的概念代数表达式进行规范化。从概念层次的角度来看,“事件”的概念代数表达式可以同时用于表示“事件”、“事件类”。最后,在概念代数概念层次运算的支持下,给出“事件”的概念代数表达式运算规则,并以实例验证其有效性。

2 相关工作

什么是“事件”?这是一个很复杂的问题,目前还没有统一的定义。表 1 从哲学、认知科学、语言学、知识表示、信息和本体研究几个领域对“事件”的定义做出比较与分析[10]。

表 1 各领域对事件理解的比较与分析

领域	关注内容	优点	不足
哲学与认知科学领域	事件的本源、概念、定义	从哲学的角度对事件进行定义与剖析	忽略了事件、事件要素、概念之间的关系
语言学领域	事件的语法结构、词汇关系、修饰成分、语义	能部分描述实体间的关系,确定事件在句中与其它成分的关系	仅从语言的角度分析事件,忽略事件的本质涵义
知识表示领域	事件的结构、表示、推理	表示方法多样,可以描述事件的状态变化即动态性,表示事件之间的结构	受表示方法和表示能力的限制,有领域局限性
信息领域	事件的识别、获取、基于事件的应用	实现事件信息的浅层与初步应用	对事件的定义粒度过大,不能实现更细致的操作需求
本体研究领域	事件的表示、事件之间的关系、事件与概念之间的关系、应用等	完全或部分整合各领域对事件关注的内容和优点,应用前景光明	处于研究起步阶段,很多内容还需要具体实现并加以验证

本文受国家自然科学基金项目:事件本体形式化方法中的几个重要问题(61273328),国家自然科学基金项目:基于描述逻辑的事件推理关键问题研究(61305053)资助。

张旭洁(1980—),女,博士,讲师,主要研究方向为知识表示、数据挖掘,E-mail:zhangxujie@lixin.edu.cn;刘宗田(1946—),男,教授,博士生导师,主要研究方向为人工智能、事件本体;刘念祖(1955—),男,教授,主要研究方向为智能信息处理技术。

目前有关事件知识的表示方法主要包括:(1)基于传统本体概念层级结构的事件本体表示模型,如: Event Ontology^[3]、Linking Open Description of Event^[11]等;(2)基于逻辑方法的事件表示模型,如: Order-Sorted Logic Event Ontology^[12]、History Event Ontology^[13]等;(3)基于事件六要素的事件表示模型^[9]。

本文在事件六要素表示模型的基础上,给出了基于 Nilsson 概念代数的“事件”表达式。该表示方法的优点在于:(1)能灵活表示事件类与事件实例;(2)可以描述简化的事件动态过程;(3)运用该“事件”表达式的运算规则可以获得事件要素的属性取值信息;(4)方便描述事件要素与事件之间的关系、事件之间的分类与非分类关系、事件要素之间的关系;(5)动作要素 A 与语言表现要素 L 的描述可以实现事件网络的关联;(6)通过语言表现要素 L 的描述,可以关联相逆、相似与相同的事件。

“事件”表达式的事件表示方法与知识表示中的框架表示方法非常相似,但与传统的框架表示方法相比,事件框架表达式的优点在于:(1)可以描述事件的动态过程;(2)动作要素与语言表现要素可以构建事件间的关系网络,描述事件之间的分类与非分类关系;(3)概念代数的概念运算公理为事件类的层次结构表示以及事件要素的层次结构表示提供了理论支持,并为后期推理功能的实现提供服务。

3 基础知识

3.1 事件六要素表示模型

定义 1(事件, Event)^[9] 是指在某个特定时间和环境下发生的,由若干角色参与,表现出若干动作特征的一件事情。事件由六要素组成:动作要素 A,对象要素 O,时间要素 T,环境要素 V,断言要素 P,语言表现要素 L;形式化表示为: $e = \langle A, O, T, V, P, L \rangle$ 。

定义 2(事件类, Event Class)^[9] 指具有共同特征的事件的集合,用 EC 表示为 $EC = (E, C_1, C_2, \dots, C_6)$,其中 E 是事件的集合,称为事件类的外延; $C_i = \{c_{i1}, c_{i2}, \dots, c_{im}, \dots\}$ ($1 \leq i \leq 6, m \geq 0$) 称为事件类的内涵,是 E 中每个事件在第 i 个要素上具有的共同属性的集合, c_{im} 是事件类中每个事件在第 i 个要素上具有的一个共同属性。

有关事件六要素表示模型与事件本体的更多研究请参考文献[9,10,14,15]等。

3.2 概念代数

1994 年 Nilsson^[15-17] 教授提出用数学格来描述概念的思想。该思想的主要依据是:概念知识模型主要利用概念之间的包含关系(如:“is-a”、“subclass”等)来组织构建,类似于数学中的格结构。由于每个代数格都是一个偏序格^[18],也就是说,如果 (M, \leq) 是一个格,其中 M 表示集合, \leq 为偏序关系, \wedge 和 \vee 分别表示格内的交和并,那么: $\varphi \leq \Psi$ 当且仅当 $\varphi = \varphi \wedge \Psi$; $\varphi \leq \Psi$ 当且仅当 $\Psi = \varphi \vee \Psi$ ($\varphi \in M, \Psi \in M$)。

由于概念的包含关系也是一种偏序关系,因此可以建立代数格与概念包含之间的关系,即^[15-17]: $a \leq b$ 当且仅当 $a = a \times b$; $a \leq b$ 当且仅当 $b = a + b$ 。

概念代数中概念的继承、交、并分别用符号 \leq 、 \times 、 $+$ 表示,它们分别由代数格理论中的偏序、交、并约束。引入全概念(univ)和空概念(null)来定义格的界。

公理 1 设 a, b, c 为概念格的任意概念,对于分配有界概

念格,存在如下概念公理:

$$\text{幂等律: } a + a = a, a \times a = a \quad (1)$$

$$\text{交换律: } a + b = b + a, a \times b = b \times a \quad (2)$$

$$\text{结合律: } a + (b + c) = (a + b) + c \\ a \times (b \times c) = (a \times b) \times c \quad (3)$$

$$\text{吸收率: } a + (a \times b) = a, a \times (a + b) = a \quad (4)$$

$$\text{分配率: } a \times (b + c) = (a \times b) + (a \times c) \\ a + (b \times c) = (a + b) \times (a + c) \quad (5)$$

$$\text{有界律: } a + \text{null} = a, a \times \text{null} = \text{null} \\ a + \text{univ} = \text{univ}, a \times \text{univ} = a \quad (6)$$

概念之间除了简单的包含关系外,还存在其它复杂的关系,Nilsson^[15-17] 把概念之间的其它关系看作是概念的属性。属性函数(也称为属性项)表示为 $r(\varphi)$,其中 r 为属性名, φ 为属性 r 的值,用“:”运算符表示为 $r:\varphi$;属性函数通常作为概念项的一个因子出现,如“ $\Psi \times r(\varphi)$ ”表示概念项 Ψ 的子概念有属性 r,值为 φ 。下面给出添加了属性描述的概念项的一般表示形式:

$$c \times r_1(\varphi_1) \times r_2(\varphi_2) \times \dots \times r_m(\varphi_m) \quad (7)$$

或表示为:

$$c \begin{bmatrix} r_1:\varphi_1 \\ r_2:\varphi_2 \\ \dots \\ r_m:\varphi_m \end{bmatrix} \quad (8)$$

式(8)的表示形式称为概念框架表达式,需要说明的是属性值 φ_i ($i=1,2,\dots,m$) 为概念变量,可以定义为任意概念项。

公理 2 属性满足以下公理,其中 r_i ($i=1,2,\dots,m$):

$$r_i(x + y) = r_i(x) + r_i(y) \quad (9)$$

$$r_i(x \times y) = r_i(x) \times r_i(y) \quad (10)$$

$$r_i(\text{null}) = \text{null} \quad (11)$$

$$r_i(\text{univ}) = \text{univ} \quad (12)$$

定理 1 属性函数是单调的,即 $x \leq y \Rightarrow r_i(x) \leq r_i(y)$ 。文献[19]给出了证明。

定义 3 概念项由概念常量 \underline{C} 、概念变量 φ 、属性项 $r_i(\varphi_i)$ 、全概念(univ)和空概念(null),通过运算符“+”、“ \times ”、“:”连接而成,记概念项为 $\langle C, A \rangle$,其中 C 为概念标识符, A 为属性标识符;不含概念变量的概念项为基项,即概念常量 \underline{C} 。

概念项的定义类似于谓词公式,概念变量 φ 可以定义为任意概念项,这实现了概念项的递归定义。

定义 4 概念句是表示概念相等的代数等式,包括:

$$\text{一般等式: } \varphi = \Psi \quad (13)$$

$$\text{条件等式: } \varphi = \Psi \text{ if } \varphi = \Psi \quad (14)$$

例 1

$$\text{Hello Kitty} = \text{cat} \times \text{ISA}(\text{toy}) \times \text{material}(\text{VELBOA}) \times \text{color}(\text{white})$$

其中,cat 为概念常量,Hello Kitty 为 cat 的一个实例或子类,通过 ISA(toy),material(VELBOA)等属性描述,对 Hello Kitty 进行细化说明。

Nilsson 概念代数还可以通过逻辑中的 λ -表达式对概念代数的概念项和句子进行重构^[17]。文献[15]对概念代数五元组进行了扩充。

4 事件表示

4.1 基于概念代数的事件表示

为了能全面具体地表示事件,这里将事件六要素理解为

描述“事件”概念的 6 个属性,令 e_m 为某一“事件”概念基项; ${}^n e_m$ 为该“事件”概念的某一事件实例或子概念;再令 $R_A, R_O, R_V, R_T, R_P, R_L$ 为“事件”概念六要素的属性名, $\varphi_A, \varphi_O, \varphi_V, \varphi_T, \varphi_P, \varphi_L$ 为六要素的属性值,则“事件”的概念代数表达式可由式(15)表示如下:

$${}^n e_m = e_m \left[\begin{array}{l} \Pi R_A : \varphi_A \\ \Pi R_O : \varphi_O \\ \Pi R_V : \varphi_V \\ \Pi R_T : \varphi_T \\ \Pi R_P : \varphi_P \\ \Pi R_L : \varphi_L \end{array} \right] \quad (15)$$

其中,符号 Π 表示连续求概念交运算。以下将分别对事件六要素属性名和属性值进行详细说。

4.1.1 R_O 与 φ_O

对象要素 O ,包括参与事件的所有角色。为了能清晰地描述参与事件的对象在事件中所扮演的角色, R_O 对象属性名即为参与事件的角色名, φ_O 对象属性值即为参与事件的角色。表 2 给出了扮演事件中不同角色的角色名。

表 2 R_O 角色名

角色名 R_O	关系名	缩写
事件主体	Agent	AGT
事件客体	Patient	PNT
工具	Instrument	INS
参与者	Participator	POR
合作者	Collaborator	COR
涉及到的人或物	Involved	INV
其它

参与事件的角色 φ_O 包括人、物等概念。由定义 3 可知, φ_O 可由其它概念递归定义。为了进一步描述 φ_O ,需要一些通用的概念属性名,这里引入一些常用于描述 φ_O 的属性名 r_O ,详见表 3。

表 3 r_O 属性名

对象要素修饰说明属性名	关系名	缩写
个体整体关系	Is-a	ISA
子类关系	Subclass	SUB
部分整体关系	Part-of	PAT
成员组织关系	Member-of	MEMB
隶属	Belong to	BETO
性质	Property	PROP
颜色	Colour	COL
数量	Quantity	QUAN
质量	Quality	QUAL
单位	Unit	UNIT
性别	Gender	GEN
身份	Identity	IDEN
年龄	Age	AGE
其它

例 2 事件实例 He fell asleep. 表示如下:

$${}^i asleep = asleep \left[\begin{array}{l} AGT:he \\ \left[\begin{array}{l} ISA:human \\ GEN:male \end{array} \right] \end{array} \right]$$

其中,AGT(事件主体,见表 2)为事件角色名 R_O ,表示“he”是事件“asleep”的主体,而有关概念“he”的修饰说明则通过 ISA(个体整体关系,见表 3)和 GEN(性别,详见表 3)来表示,即表示“he”属于人类,性别为男。 ${}^i asleep$ 是“asleep”事件的一个实例。

4.1.2 R_V 与 φ_V

环境要素 V ,表示事件发生的场所及其环境特征。为了清晰地描述事件发生所处的环境, R_V 环境属性名即为事件发生地与事件的关系,详见表 4。

表 4 R_V 环境属性名

R_V	关系名	缩写
地点	Location of	LOC
出发地	Departure	DEP
到达地	Destination	DES
其它

φ_V 环境属性值即为事件发生的地点概念,包括地点、方位等。与 φ_O 采用相同的方法, φ_V 可由其它概念递归定义,这里引入属性名 r_V ,用于描述 φ_V 的各属性,详见表 5。

表 5 r_V 属性名表

环境要素修饰说明属性名	关系名	缩写
在某种介质内	In/Inside	IN
在某物上	On	ON
在某位置边	At	AT
接近某物	Near	NEAR
在某物下	Under	UNDER
在某物周围	Round/Around	ROUND
在某物之前	Among	AMONG
在什么中间	Between	BETWEEN
在什么后面	Behind	BEHIND
在对立的位置	Opposite	OPP
在什么旁边	Next to/by/Beside	BY
方向	Direction	DIRECT
其它

例 3 事件实例 Tom fell asleep under her bed. 表示如下:

$${}^i asleep = asleep \left[\begin{array}{l} AGT:Tom[ISA:cat] \\ INV:bed \left[\begin{array}{l} BETO:she \left[\begin{array}{l} ISA:human \\ GEN:female \end{array} \right] \\ \left[\begin{array}{l} UNDER:bed \\ ON:ground \end{array} \right] \end{array} \right] \end{array} \right]$$

其中,LOC(地点,详见表 4)为事件发生地与事件的关系名 R_V ,表示“place”是事件“asleep”的发生地点,但语句中未给出一个具体地点,而是对地点的属性进行了描述,通过地点属性名 UNDER(在某物下,见表 5)与 ON(在某物上,见表 5)表示,即“asleep”的地点“place”在床“bed”下的地板“ground”上;INV(事件涉及到的人或物,见表 2)为事件角色名 R_O ,表示“bed”为事件所涉及的物;“Tom”为事件主体;BETO(隶属,见表 3)修饰说明“bed”属于“she”。

4.1.3 R_T 与 φ_T

时间要素 T ,通常为一些关于时间的描述,这些时间描述用以说明事件在时间轴上的位置等有关时间的信息。 R_T 时间属性名即为事件发生时间与事件的关系,详见表 6。

表 6 R_T 时间属性名

R_T	关系名	缩写
时间描述	Temporal	TMP
其它

φ_T 时间属性值即为具体的时间描述。与 φ_O 采用相同的方法, φ_T 可由其它概念递归定义,这里引入属性名 r_T ,用于描述 φ_T 的各属性,详见表 7。

表7 r_T 属性名表

时间属性名	关系名	缩写
时态	Tense(bygone, now, future)	TENSE
事件持续时间长度	Time Span	TS
先后关系	Aspect	A
时间类型	Time Type	TT
其它

时间类型 (Time Type) 主要包括: 绝对时间 (Absolute time)、相对时间 (Relative time)、时间段 (Time duration)、开始时间 (Start time)、结束时间 (End time)、过程时间 (Duration time)、日期 (Date)、年 (Year)、月 (Month)、小时 (Hour)、分钟 (Minute)、秒 (Second)、周期时间 (Period)、频率时间 (Frequency) 等。

4.1.4 R_A 与 φ_A

动作要素 A , 是对动作的程度、方式、方法等的描述。为了能清晰地描述事件的程度、方式、方法以及该事件与其它事件(或动作)的时间关系、逻辑关系、分类关系, R_A 动作属性名即为该动作与事件的关系, 详见表 8。时间与逻辑动作关系属性名的说明请参考文献[21]。

表8 动作属性名 R_A

属性名类型	动作关系属性名	关系名	缩写
修饰	事件程度修饰	Degree	DEG
	事件方法修饰	Way	WAY
	事件方式修饰	Mode	MODE
时间	时间连接	Meets	MET
	时间跟随	Before	BEF
	时间重叠	Overlaps	OVL
	同一开始时间	Starts	STR
	同一结束时间	Ends	END
	时间包含	During	DUR
	时间等价	Equals	EQU
	逻辑	包含	Embedded
因果		Cause	CAU
选择		Selection	SEL
条件		Condition	CON
并列		Paralleling	PAR
循环		Loop	LOOP
分类	子事件	SubEvent	SUBE
	父事件	SuperEvent	SUPEE
其它

例4 事件实例 The magnitude 7.3 earthquake happened in Northeast of Japan on December 7th, 2012. 表示如下:

${}^k earthquake = earthquake$

$$\left[\begin{array}{l} \text{DEG:7.3 [UNIT:magnitude]} \\ \text{LOC:Japan [DIRECT:Northeast} \\ \quad \text{ISA:Country]} \\ \text{TMP:December 7th, 2012 [TENSE:bygone} \\ \quad \text{TT:Date} \times \text{Absolute-time]} \end{array} \right]$$

其中, TMP(时间描述, 见表 6) 为时间属性名 R_T , 表示“December 7th, 2012”是事件“earthquake”发生的时间; TENSE(时态, 见表 7)、TT(事件类型, 见表 7) 表示该时间的时态为“bygone”, 时间类型为日期型“Data”并且是绝对时间“Absolute time”; DEG(事件程度修饰, 见表 8) 为动作属性名 R_A , 表示程度值为“7.3”, 并用通过 UNIT(单位, 见表 3) 修饰说明单位为震级“magnitude”; LOC, DIRECT(方向, 见表 5) 的相关说明请参考 4.1.2 小节。

例5 事件实例 He sleeps snoring. 表示如下:

$${}^i sleep = sleep \left[\begin{array}{l} \text{PAR:snore [LOOP:snore]} \\ \text{AGT:He [ISA:human} \\ \quad \text{GEN:male]} \end{array} \right]$$

其中, PAR(逻辑并列, 见表 8) 为动作属性名 R_A , 表示事件“sleep”与打呼“snore”并列进行, 同时动作“snore”又是循环 LOOP 进行的。AGT 的相关说明请参考 4.1.1 小节。

4.1.5 R_L 与 φ_L

语言表现要素 L , 表示事件的语言规律, 包括核心词集、核心词表现、核心词搭配等。 R_L 语言表现属性名, 用于记录与该事件及其事件要素语义相关的其它关键词, 详见表 9。 φ_L 语言表现属性值即为对应的语义相关概念。

表9 R_L 语言表现属性名

类型	动作关系属性名	关系名	缩写
语义相关	语义相同	Synonym	SYNM
	语义相逆	Antonym	ANTM
	语义相似	Near-synonym	NSYNM
其它

例6 事件实例 Mother loves baby. 表示如下:

$${}^i love = love \left[\begin{array}{l} \text{AGT:mother [ISA:human} \\ \quad \text{GEN:female} \\ \quad \text{SYNM:Mum]} \\ \text{PNT:baby [ISA:human} \\ \quad \text{SYNM:honey]} \\ \text{SYNM:(like} \times \text{be-fond-of)} \end{array} \right]$$

其中, SYNM(语义相同, 见表 9) 为语言表现关系名 R_L , 表示“like”、“be fond of”为“love”的同义词。AGT、PNT 等详见 4.1.1 小节。

4.1.6 R_P 与 φ_P

断言要素 P , 分别描述事件的开始与结束状态, 用于表示事件的动态变化。要描述清楚事件的动态变化, 需要表示事件开始时刻与事件结束时刻 $\varphi_A, \varphi_O, \varphi_V$ 的状态, 因此将 R_P 断言属性名定义为时间标记, 即 $R_P = \{t_1, t_n\}$ 。 t_1 与 t_n 分别表示事件开始与结束时刻。为了简化表示内容, 仅对发生变化的 $\varphi_A, \varphi_O, \varphi_V$ 进行标记。

例7 事件实例 Exhaust gas from cars pollutes the air. 表示如下:

$${}^i pollute = pollute \left[\begin{array}{l} \text{AGT:gas} \\ \text{PNT:air [}(t_1 \times \text{QUAL):clearness} \\ \quad \text{(}t_n \times \text{QUAL):dirty]} \\ \text{WAY:exhaust [AGT:car]} \end{array} \right]$$

其中, 污染“pollute”事件的主体为“gas”, 被污染的客体是空气“air”, $(t_1 \times \text{QUAL})$ 与 $(t_n \times \text{QUAL})$ 表示事件“pollute”开始与结束时其事件客体“air”的状态变化, 由清洁“clearness”变为污浊“dirty”; WAY(事件方法修饰, 见表 8) 是事件“pollute”的动作属性名 R_A , 表示污染事件“pollute”的方法是排放“exhaust”。

4.2 事件的概念表达式规范化

通过 4.1 节的详细说明, 式(15) 可以进一步规范化为式(16)。

$${}^n e_m = e_m \left[\begin{array}{l} R_A : \varphi_A \left[\begin{array}{l} r_o : \varphi \rightarrow (t_1 \times r_o) : \varphi' \\ (t_n \times r_o) : \varphi'' \\ R_A : \varphi \rightarrow (t_1 \times R_A) : \varphi' \\ (t_n \times R_A) : \varphi'' \\ R_A : e \\ R_L : \varphi_A \end{array} \right] \\ R_O : \varphi_O \left[\begin{array}{l} r_o : \varphi \rightarrow (t_1 \times r_o) : \varphi' \\ (t_n \times r_o) : \varphi'' \\ R_L : \varphi_O \end{array} \right] \\ R_V : \varphi_V \left[\begin{array}{l} r_o : \varphi \rightarrow (t_1 \times r_o) : \varphi' \\ (t_n \times r_o) : \varphi'' \\ R_L : \varphi_V \end{array} \right] \\ R_T : \varphi_T \left[\begin{array}{l} r_i : \varphi \\ R_L : \varphi \end{array} \right] \\ R_L : e \end{array} \right] \quad (16)$$

其中, φ 表示用来对事件各要素进行修饰说明的其它概念集, 可以定义为概念代数中的任意概念项(概念常量或概念变量, 用以实现概念的嵌套定义); e 表示其它事件, 用来对当前事件 e_m 进行修饰或说明, e 可以通过 R_A 和 R_L 引入事件的概念表达式($e \in \varphi$)。事件的状态变化表示为: $\varphi \rightarrow \begin{matrix} (t_1 \times r_o) : \varphi' \\ (t_n \times r_o) : \varphi'' \end{matrix}$, 即如果事件过程中的概念 φ 发生改变, 其变化可以通过 φ 的开始状态 $(t_1 \times r_o) : \varphi'$ 和结束状态 $(t_n \times r_o) : \varphi''$ 来描述, 而没有发生变化的概念 φ 则无需对其状态变化进行表示, 最后, 由于不同事件要素的语言表现要素均不同, 因此式(16)用 $R_L : \varphi_A$ 和 $R_L : \varphi_O$ 等分别表示不同事件要素的语言表现。

5 事件概念运算规则

文献[20]对概念项进行“+”, “ \times ”运算在概念层次结构中的含义进行了解释, 其中 $c_1 + c_2$ 与 $c_1 \times c_2$ 的含义如图1所示, 图2为其实例表示。可以简要理解为: “+”运算求上层概念, “ \times ”运算求下层概念。

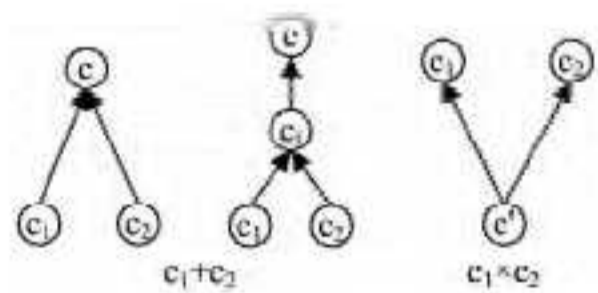


图1 概念运算与概念层次结构关系

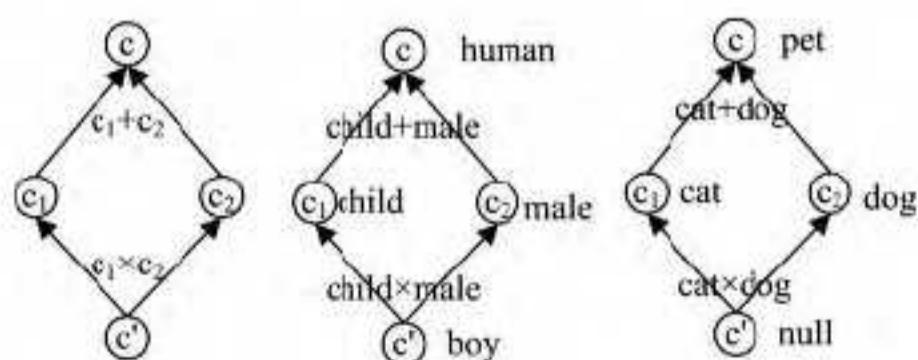


图2 概念运算实例

对于同一概念基项的不同概念项表达式 ${}^i c =$

$$c \begin{bmatrix} r_1 : \varphi_1 \\ r_2 : \varphi_2 \\ \dots \\ r_m : \varphi_m \end{bmatrix} \text{ 和 } {}^j c = c \begin{bmatrix} r_1 : \varphi_1 \\ r_2 : \varphi_2 \\ \dots \\ r_m : \varphi_m \end{bmatrix} \text{ 满足式(17)与式(18)的运算规则,}$$

$${}^i c + {}^j c = c \left(\begin{bmatrix} r_1 : \varphi_1 \\ r_2 : \varphi_2 \\ \dots \\ r_m : \varphi_m \end{bmatrix} + \begin{bmatrix} r_1 : \varphi_{j1} \\ r_2 : \varphi_{j2} \\ \dots \\ r_m : \varphi_{jm} \end{bmatrix} \right) = c \begin{bmatrix} r_1 : (\varphi_{i1} + \varphi_{j1}) \\ r_2 : (\varphi_{i2} + \varphi_{j2}) \\ \dots \\ r_m : (\varphi_{im} + \varphi_{jm}) \end{bmatrix} \quad (17)$$

$${}^i c \times {}^j c = c \left(\begin{bmatrix} r_1 : \varphi_1 \\ r_2 : \varphi_2 \\ \dots \\ r_m : \varphi_m \end{bmatrix} \times \begin{bmatrix} r_1 : \varphi_{j1} \\ r_2 : \varphi_{j2} \\ \dots \\ r_m : \varphi_{jm} \end{bmatrix} \right) = c \begin{bmatrix} r_1 : (\varphi_{i1} \times \varphi_{j1}) \\ r_2 : (\varphi_{i2} \times \varphi_{j2}) \\ \dots \\ r_m : (\varphi_{im} \times \varphi_{jm}) \end{bmatrix} \quad (18)$$

6 事件表示与运算实例

以不同地震实例为例, 进行事件概念运算。

例8 事件实例 The magnitude 5 earthquake. 表示如下:

$${}^i \text{earthquake} = \text{earthquake}[\text{DEG}; 5[\text{UNIT}; \text{magnitude}]]$$

例9 事件实例 The magnitude 7.8 earthquake happened in Sichuan. 表示如下:

$${}^j \text{earthquake} = \text{earthquake} \left[\begin{array}{l} \text{DEG}; 7.8[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{Sichuan}[\text{ISA}; \text{province}] \end{array} \right]$$

运用式(17)与式(18)对例8、例9、例4的地震事件进行概念运算, 首先采用 Nilsson 概念代数公理对这些概念项进行变形, 表示如下:

例8的概念项变形为

$$\begin{aligned} {}^i \text{earthquake} &= \text{earthquake}[\text{DEG}; 5[\text{UNIT}; \text{magnitude}]] \\ &= \text{earthquake} \left[\begin{array}{l} \text{DEG}; 5[\text{UNIT}; \text{magnitude}] \\ \text{univ} \end{array} \right] \end{aligned} \quad (\text{依据式(6)一(8)})$$

$$= \text{earthquake} \left[\begin{array}{l} \text{DEG}; 5[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{univ} \end{array} \right] \quad (\text{依据式(12)})$$

$$= \text{earthquake} \left[\begin{array}{l} \text{DEG}; 5[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{univ} \\ \text{TMP}; \text{univ} \end{array} \right]$$

例9的概念项变形为

$${}^j \text{earthquake} = \text{earthquake} \left[\begin{array}{l} \text{DEG}; 7.8[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{Sichuan}[\text{ISA}; \text{province}] \end{array} \right]$$

$$= \text{earthquake} \left[\begin{array}{l} \text{DEG}; 7.8[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{Sichuan}[\text{ISA}; \text{province}] \\ \text{TMP}; \text{univ} \end{array} \right]$$

• 概念“+”运算

$${}^i \text{earthquake} + {}^j \text{earthquake} + {}^k \text{earthquake} = {}^l \text{earthquake}$$

$$\begin{aligned} &= \text{earthquake} \left[\begin{array}{l} \text{DEG}; 5[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{univ} \\ \text{TMP}; \text{univ} \end{array} \right] + \\ &\text{earthquake} \left[\begin{array}{l} \text{DEG}; 7.8[\text{UNIT}; \text{magnitude}] \\ \text{LOC}; \text{Sichuan}[\text{ISA}; \text{province}] \\ \text{TMP}; \text{univ} \end{array} \right] + \end{aligned}$$

项,而在寻找“看了多长时间”的答案时,把₂作为侯选项;如果都标注为 time,就多了一个干扰选项。

比较_{3a}和_{3b},其中“人”在_{3a}中标记为“goal”,这个语义角色是相对于前面的述语动词“让”来说的,是“让”的客体,它并不表示“人”相对于后面的“看”的语义角色;_{3b}在以“看”为目标词标注时,“人”表示为 perc(感知者)。可见, Sinica Treebank 以“显性”语法关系为依据,只有具有直接句法关系的成分之间才表示出语义角色,而汉语框架语义标注体系中,表示出了一部分隐性的语法关系。

比较₄和₅中跟空间有关的语义成分的标注;_b组将表示 location 的“从水里”和“从车窗”通过是否自主感知加以区分,而_a组则视为一致。₄强调通过水看见了自己,感知的是事件发生的外部条件,就是说所看到的事物(自己的影子)所在的场所;₅强调的是“看”,“从车窗”只是感知者“国强”所在的位置,“车窗”不该看做感知对象。因此,比较而言,_b组的分词结果更恰当。设想一个问答系统,在回答人或物体在什么位置这样的问题时,显然_b组的标注会使得系统输出正确答案的可能性更大。

结束语 本文以框架语义理论为基础,构建汉语语义分析的范畴知识体系,其语义角色细化,语义信息丰富,使得计算机语义理解更加精确化,始终采用基于语料库的方法,避免“从意义到意义”的理性思维,使研究结果可靠、实用;用基于规则的方法解决语义标注歧义问题,避免了基于统计的方法面临的语义资源匮乏、数据稀疏的问题,通过实验证明本研究所归纳的规则对于解决语义角色标注问题十分有效。

语义标注因汉语框架语义范畴的知识体系而有众多更加细化的语义角色,其标注方法体系针对汉语特殊句法形式给出了具体的处理方案。因此,与其格语法体系表达模式相比,框架语义标注模式识别出了更多的语义内容,语义描述也更加准确,能够为计算机应用系统提供更加丰富而实用的信息。

参考文献

(上接第 88 页)

- [7] 林萍,黄卫东. 基于 LDA 模型的网络舆情事件话题演化分析[J]. 情报杂志, 2013, 32(12): 26-30
- [8] 王晓峰, 张大鹏, 王绯, 等. 基于语义轨迹的视频事件探测[J]. 计算机学报, 2010, 33(10): 1845-1858
- [9] 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, 199
- [10] 张旭洁, 刘宗田, 刘炜, 等. 事件与事件本体模型研究综述[J]. 计算机工程, 2013, 39(9): 303-307
- [11] Shaw R, Troncy R, Hardman L. Lode: Linking open descriptions of events [J]. The Semantic Web, 2009, 31(8): 153-167
- [12] Kaneiwa K, Iwazume M, Fukuda K. An upper ontology for event classifications and relations [J]. AI, Advances in Artificial Intelligence, 2007, 11(8): 394-403
- [13] Corda I, Bennett B, Dimitrova V. A Logical Model of an Event Ontology for Exploring Connections in Historical Domains [C]// Proceedings of the 10th International Semantic Web Conference (ISWC). Bonn, Germany, 2011: 1107-1114
- [14] Liu Wei, Xu Wen-jie, Fu Jian-feng, et al. An extended description logic for event ontology [C]// Proceedings of the 5th International Conference on Grid and Pervasive Computing(LNCS). Springer, 2010: 471-451
- [15] Nilsson J F. A concept object algebra CA [C]// Proceedings of

- [1] Gildea D, Jurafsky D. Automatic Labeling of Semantic Roles [J]. Computational Linguistics, 2002, 28(3): 245-288
- [2] Hellen, Lars, Beermann D. Classification of Prepositions Senses for Deep Grammar Applications [C]// Proceedings of the Second ACLSINSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications. Colchester, 2005: 74-83
- [3] Charles F. Frames and the Semantics of Understanding [C]// Quaderni di Semantica. 1985: 222-254
- [4] Baldwin. Prepositions in Applications: A Survey and Introduction to the Special Issue [J]. Computational Linguistics, 2009, 35(2): 119-149
- [5] Patrick S-D. Syntactic and Semantic Frames in PrepNet [C]// Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJ-CNLP-08). Hyderabad, 2008: 763-768
- [6] Jorgensen F, Lønning J T. A Minimal Recursion Semantic Analysis of Locatives [J]. Computational Linguistics, 2009, 35(2): 229-270
- [7] 陈群秀. 一个现代汉语语义知识库的研究和实现 [C]// 中文信息处理前沿进展, 中国中文信息学会成立二十五周年学术会议论文集, 2006: 172-181
- [8] 冯志伟. 从格语法到框架网络 [J]. 解放军外国语学院学报, 2006, 29(3): 1-9
- [9] 贾君枝, 毛海飞. 汉语框架网络问答系统问句处理研究 [J]. 图书情报工作, 2008, 52(10): 55-57
- [10] 由丽萍, 范开泰, 刘开瑛. 汉语语义分析模型研究述评 [J]. 中文信息学报, 2005(6): 59-65
- [11] 由丽萍. 构建现代汉语框架语义知识库技术研究 [D]. 上海: 上海师范大学, 2006

the 2nd European-Japanese Seminar on Information Modeling and Knowledge Bases, V: principles and formal techniques. 1993: 121-132

- [16] Nilsson J F. A logico-algebraic framework for ontologies [C]// Proceedings of the 1st International OntoQuery Workshop. 2001: 189-110
- [17] Nilsson J F. An algebraic logic for concept structures [C]// Preprint of the 3rd European-Japanese Seminar on Information Modeling and Knowledge Bases, V: principles and formal techniques. 1993: 75-81
- [18] Wille R. Formal concept analysis: mathematical foundations [M]. Springer Verlag, 1999: 117
- [19] 时百胜, 余泓. 概念知识表示和推理 [J]. 小型微型计算机系统, 2006, 27(9): 1618-1622
- [20] 车敦仁, 周立柱. 概念代数——新一代数据库系统的理论 [J]. 计算机研究与发展, 1996, 33(1): 32-38
- [21] Zhang Xu-jie, Liu Zong-tian, Liu Wei, et al. Research on Event-Based Semantic Annotation of Chinese [C]// 2012 IEEE International Conference on Computer Science and Network Technology (ICCSNT). 2012: 1883-1888
- [22] 付剑锋, 刘宗田, 刘炜, 等. 基于层叠条件随机场的事件因果关系抽取 [J]. 模式识别与人工智能, 2011, 24(4): 567-573