

67-70

# 基于概念图的自然语言语义解释

TP18

周继鹏 (西北大学计算机系 西安710069)

A

## 摘要

语法分析器应用文法规则生成语法树来表明一个句子的语法结构。本文讨论用概念图表示自然语言的语义,描述一个从语法树开始生成概念图来解释句子语义的方法,生成过程由语法树引导,把与每一个输入词相关的概念图连接成一个较大的概念图表示整个句子的语义。

关键词 概念图 自然语言

## 1. 引言

人们理解自然语言要利用大量的基础知识,包括词法知识、语法知识、语义知识和一些断言知识。词法知识和语法知识在语法分析时容易表示和处理,已经发展比较成熟,但是语法分析时不能解决语义问题和语义的二义性问题,因为自然语言理解中最困难的是语义知识的理解,它首先要有一个表示自然语言语义的方法,象语义网络,谓词逻辑,格语法(case grammar)等<sup>[1]</sup>。但现有的方法却表示复杂,操作困难或不能确切表示自然语言的语义,是自然语言语义解释的主要障碍。我们西北大学人工智能实验室经过几年的研究和系统模拟,认为 John Sowa 提出的概念图模型<sup>[2]</sup>综合了以前自然语言语义的多种表示,是一种适合于自然语言处理的知识表示方法,具有结构简单,表示范围广,易于操作,能确切地表示自然语言语义等多种特点,它将作为未来知识系统的核心知识表示模型。

在AI中,对自然语言语义处理的传统方法是遵循组合原则,即给出句子各部分的语义表示,然后把它们组合在一起得到整个句子的语义表示。一般情况下,从语法树开始,对每个词给出其对应的语义词(语义表示),由子节点的语义词的组合形成其父节点的语义表示,结果建立整个句子的语义表

示。例如 McCord 在处理语义时<sup>[3]</sup>,依逻辑中的 $\lambda$ 形式表达式表示每一个词对应的语义词,应用 $\lambda$ 操作把各个词的语义进行组合,最后得到整个句子的语义,如图1所示。

to Love,  $\lambda(x, y)[\text{Love}(x, y)]$   
 John,  $\lambda(\ )[\text{John}]$   
 girl,  $\lambda(x)[\text{girl}(x)]$   
 a,  $\lambda(p, x)[p(x)]$

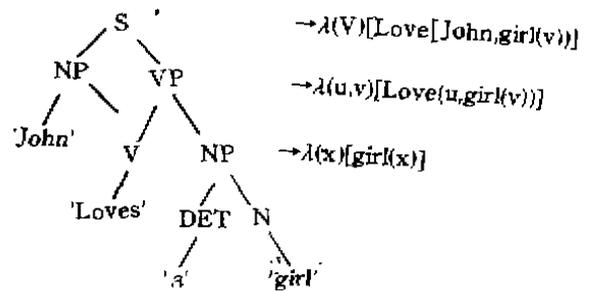


图1 基于逻辑的语义解释

本文描述如何应用文法规则对输入串进行分类(象名词、动词等)所产生的语法树,转换为概念图进行语义解释。

## 2. 概念图模型与自然语言语义

概念图知识表示方法包括概念图以及概念图上的操作在文[2]中已经给出,可知概念图是由概念节点和概念关系节点组成的二分图。在概念图中把有意义的图叫做正则图(Canonical graph)。概念图中的每一个概念由两部分组成,即概念类型和类型所指,如

[PERSON; LI MING], [COLOR; RED], [CITY; XIAN]。Sowa 所进行的关系类型, 象主格 (agentive), 宾格 (objective), 工具 (instrumented), 地方 (Location) 等实质上是语义格 (Case) 关系集。概念图的特点在于表达了词的语义限制, 例如, 正则图

$$[PERSON] \leftarrow (AGNT) \leftarrow [THINK] \rightarrow (OBJ) \rightarrow [PROPOSITION]$$

表示动词“to think”的主格 (agentive) 必须是人。

### 2.1 概念类型的分层

概念类型是有层次的, 例如 PERSON < ANIMAL, BUS < MOBILE-ENTITY, 狗 < 脊椎动物, EAT < ACT 等, 概念类型上的这种偏序关系形成格 (Lattice), 这个格的最顶层的概念类型是 UNIV (泛类型), 最底层的概念类型是 ABSURD (空类型)。

概念类型上的偏序关系 (<), 能扩展到有所指的概念, 例如 [PERSON; ZHOU HONG] < [PERSON], 由传递性有 [PERSON; ZHOU HONG] < [ANIMAL], 这时说 ANIMAL 或 PERSON 能限制到 [PERSON; ZHOU HONG]。两个类型相同, 所指不同的概念是语义上不可比较的。概念类型本身就是自然语言的语义词汇集, 概念类型上的这种分层实质上是自然语言词汇的语义分层, 格关系形成了一个语义网。

### 2.2 正则图的构成规则

正则图通过构成规则: 拷贝、限制、连接、化简得到新的正则图。

• 拷贝 (Copy): 正则图的准确拷贝仍是正则图。

• 限制 (Restrict): 是用子类型的标号替换概念的类型标号。例如, [GIRL] 是 [PERSON] 的子类型, 可用 [GIRL] 去替换 [PERSON]。也可用个体概念去替换一般概念。例如: [DOG] 是一般概念, 可以用个体概念 [DOG; Snoopy] 代换。

• 连接 (Join): 用两个图中完全相同的概念把两个概念图合并成一个概念图。

• 化简 (Simplify): 消除图中的冗余。

具体的构成规则见 [2], 可以看出正则图上的限制、连接实质上是语义的限制与组合。

### 2.3 概念图表示词的语义

正则图给出了概念模式和关系模式所表示的词的语义限制。例如, 正则图

$$[STUDENT] \leftarrow (AGNT) \leftarrow [READ] \rightarrow (OBJ) \rightarrow [BOOK]$$

表示 READ 的主格是学生, READ 的宾格是书。又如, 正则图

$$[MOBILE-ENTITY] \leftarrow (AGNT) \leftarrow [GO] \rightarrow (DEST) \rightarrow [PLACE]$$

表示 GO 的主格必须是能运动的实体 (MOBILE-ENTITY), GO 的目的地必须是一个地方 (PLACE)。

在概念图知识表示方法中, 引进了图式 (Schema) 的概念, 它是人们在推理中表示背景知识的基本结构。在自然语言理解中, 我们利用一个图式的集合, 称为图式串 (Schematic cluster), 表示自然语言中词的各种语义应用。例如, 词 “key” 对应的图式串为

$$(s_1) \quad [KEY] \leftarrow (INST) \leftarrow [OPEN] \rightarrow (OBJ) \rightarrow [DOOR]$$

$$(s_2) \quad [KEY] \leftarrow (PART) \leftarrow [KEYBOARD]$$

↓

(OBJ)

↓

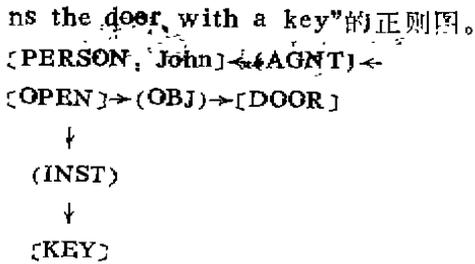
(PRESS) → (AGNT) → [PERSON]

若我们简单的假设, 概念类型和自然语言中的词汇有 1-1 对应关系, 那么在对句子进行语义解释时, 对句子中的一个词找出它的一个正则图或它的图式串中的一些图式与之对应。这一点在后面还要进行详细讨论。

例如, 设动词 “to open” 对应的概念类型 OPEN 的正则图为

$$(c1): [PERSON] \leftarrow (AGNT) \leftarrow [OPEN] \rightarrow (OBJ) \rightarrow [PHYSICAL-OBJECT]$$

那么把 “to open” 的正则图 (c1) 和 “key” 的图式串中的图式 (s1) 连接得到句子 “John ope-



### 3. 语法引导的概念图生成

对自然语言的语义解释从自然语言文法所生成的语法树开始，正确地决定每一个输入词对应的概念图（正则图或图式），然后把表示各部分语义的概念图通过连接组合，建立整个句子的概念图，用于表示整个句子的语义。为了说明这个过程，考虑句子“LI MING went to XIAN by bus”，图2给出了由PLNLP英语文法对于这个句子所生成的语法树，DECL表示这是一个陈述句，went是过去时态，John, XIAN, bus是单数形式，以及一些语法和词法特征。对这个句子的语义解释从这个语法生成树开始，首先找出这个树的每一个叶节点（词）的正则图或图式。例如，名词LI MING是PERSON的所指，

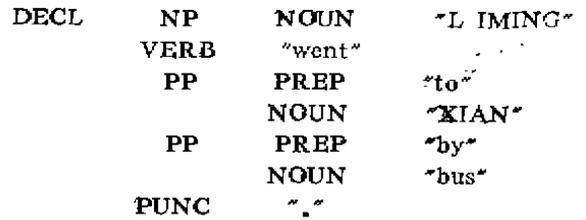
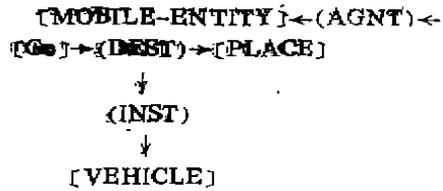
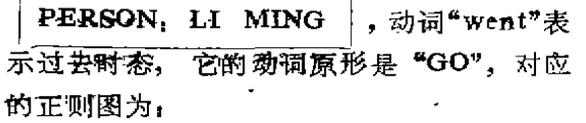
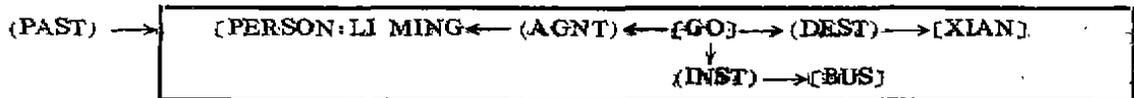


图2 语法树

因此，“LI MING”对应概念：



对于介词短语PP，它连接介词“to”的图和“XIAN”的图为 (DEST)→(XIAN)，因为XIAN是PLACE的所指，对于介词短语“by bus”，连接介词“by”的图和“bus”的图为 (INST)→(BUS)，因为BUS是INST的所指，结果得到整个句子的图如下：



从语法树引导生成概念图进行语义分析，并不都是确定性的，有时会出现二义性的问题，包括词法二义性和结构二义性。

• 词法二义性：就是一个词可以用多个概念图（正则图或图式）进行语义解释。例如，介词“to”可以表示目的地 (DEST)，接收者 (RCPT)，by可以表示工具 (INST)，地方 (LOC)，或主格 (AGNT)等。

• 结构二义性：就是语法分析树的生成有时并不是由语法规则唯一决定词或短语之间的关系，而是需要一些语义信息才能确定。例如，英语句子“John went to the chair by the window.”其中短语by the window是the chair的后修饰词，但是语法分析可能给出它是动词的后修饰词。

对于词法二义性，语义解释器必须考虑多个图的选择，通过语法分析和语义解释的过程进行图的连接操作，抛弃那些不能进行适当连接的图。在处理词法二义性时，参考Wilk的语义优先原则<sup>[4]</sup>，我们在概念图连接时，依照最大连接优先原则，即含有最少概念节点的图优先，这是因为最少概念节点的图拥有最多数目的匹配概念节点（即最大连接）。

对于结构二义性：首先通过语法分析生成一个语法树，然后通过正则图或图式在连接过程中决定词或短语之间的关系。对于“John went to the chair by the window”，在图连接过程中，语义解释器发现the window不是Go所接收的工具(INST)，

## 英文生成系统词汇结构与功能位的逻辑描述

卞世力 姚天顺 滕永林

(东北工学院计算机系 沈阳110006)

## 摘 要

This paper presents the structure and function of lexicon used to generate English from intermediate language in Chinese English Machine Translation (CETRAN). The transformation principles between each grammatical levels for the generation system are described. Some examples go through this paper to explain the method of English generation system. A grammar—Lexical Functional Position Grammar, which is applied to the generation system in the CETRAN, is advanced.

随着人工智能(AI)各领域的不断发展,越来越需要以自然语言形式进行人机对话,使其更加面向用户,特别是机器翻译的再度兴起,使得自然语言处理(NLP)面临新的挑战。在近几年,国内外一些语法理论陆续产生,如:管辖与约束理论(Government Binding Theory),词汇功能文法(Lexical Functional Grammar),功能合一文法(Functional Unification Grammar)和词汇语义驱动理论等,但是不管你是用哪一种方法,都离不开带有复杂特征集(Complex Feature Set—CFS)的词做为知识源。CETRAN系统生成器也是由CFS支持的,而且对于一个词X的复杂特征集有 $CFS(X) = (MNI(X), SFL(X), DFL(X))$ 。其中: MNI(X)—X主结点

信息; SFL(X)—静态属性表; DFL(X)—动态属性表。CFS(X)在这三个特征子集中又包含着许多特征信息,使得知识源十分丰富,系统的效率大大提高了。

CETRAN系统采用的是以中间语言媒介语做枢轴结构,力求这种枢轴结构能够成为独立于语种的通用性语言接口,目的是为此后汉英双向翻译,以及多语种翻译打下基础,整个系统不论是分析或生成系统都是基于词义的,利用CFS通过语义驱动自然语言的处理工作。

## 一、词汇结构与复杂特征集

CETRAN生成系统是分析系统的逆过程,是逐层次进行转换的,即从语义→句法→词法→输出目标语句。一个词X的CFS也

收到日期: 92-11-03.

因此去选择另外一个介词短语PP,修饰the chair,语义解释器发现the window是表明the chair的位置,因此,在连接时,by被指定为(LOC)关系。

## 4. 小结

概念图模型作为未来基于知识系统和自然语言处理系统的有效知识表示方法,它能够处理象时态、模态等问题,能灵活地表

示大量的语义词汇。几年来经过西北大学人工智能实验室全体同志和研究生认真的理论研究和“概念图处理器”工具系统的开发及英文的自然语言解释和汉语的自然语言生成的系统模拟<sup>[5,6,7,8]</sup>,证明了概念图模型是适合自然语言处理的知识表示方法。目前我们正在研究开发一个基于概念图的中英文翻译系统。(参考文献略)