

# 基于配价结构和语义依存关系的句法分析统计模型

袁里驰<sup>1,2</sup>

(1. 江西财经大学信息管理学院, 江西南昌 330013; 2. 江西财经大学数据与知识工程江西省高校重点实验室, 江西南昌 330013)

**摘要:** 目前主流的词汇化句法分析方法仅仅考虑词语之间的语义依存关系, 而没有引入语义搭配和语义类等语义信息。“配价”是词语的一个比较本质的特点, 一旦一个词语的配价结构确定下来, 它应该和怎样的词进行搭配也就比较清楚了, 从而也可以比较直接地导出句子的结构。本文结合中心词驱动句法分析模型, 提出了基于配价结构和语义依存关系的句法分析模型。模型在规则的分解及概率计算中引入丰富的语义信息, 既包括语义依存信息, 也包括配价结构等语义搭配信息。用改进的句法分析模型进行句法分析实验, 实验结果表明, 精确率和召回率分别为 88.76% 和 87.43%, 综合指标 F 值比 Collins 的中心词驱动句法分析模型提高了 6.65 个百分点。

**关键词:** 配价结构; 语义依存关系; 中心词驱动; 句法分析统计模型

**中图分类号:** TP391.1      **文献标识码:** A      **文章编号:** 0372-2112 (2013) 10-2029-06

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2013.10.025

## A Statistical Parsing Model Based on Valence Structure and Semantic Dependency

YUAN Li-chi<sup>1,2</sup>

(1. School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China;

2. Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China)

**Abstract:** The mainstream lexicalized syntactic parsing models currently only utilized semantic dependency between words, and did not incorporate other semantic information such as semantic collocation and semantic category. “valency” is an essential semantic feature of words, once the valency of word is determined, the collocation of the word is clear, and the sentence structure can be directly derived. Thus a syntactic parsing model combining valence structure with semantic dependency is purposed on the base of head-driven statistical syntactic parsing models. The parsing model incorporates rich semantic information including semantic dependency and semantic collocation in the decomposition and probability computation of the rules. Experiments are conducted for the refined statistical parser, it achieves 88.76% precision and 87.43% recall, F measure is improved 6.65% comparing with the head-driven parsing model introduced by Collins.

**Key words:** valence structure; semantic dependency; head-driven; statistical syntactic parsing model

## 1 引言

句法分析, 就是指根据给定的语法, 自动地识别出句子所包含的句法单位和这些句法单位之间的关系。句法分析在自然语言处理领域中具有十分重要的地位, 同时它也是公认的一个研究难题。汉语的理解一般分为以下步骤: 原文输入、句子词语切分及词语属性特征标注、语法及句法分析、语义及语用和语境分析、生成目标形式表示、句群及篇章理解等。句子分析上接篇章理解, 下联词汇分析, 起着承上启下的作用。词汇分析是基础, 句子分析是中心, 篇章理解是最终目的。那么, 一旦得到了

句子成分的计算机表示, 无论是应用于句群划分、篇章理解, 还是机器翻译、机器释义、人机对话或是情报检索等方面, 都有着实际意义。

句法分析的研究大体分为两种途径: 基于规则的方法和基于统计的方法<sup>[1-11]</sup>。为进行统计句法分析, 首先要遵循某一语法体系, 根据该体系的语法确定语法树的表示形式。目前, 在句法分析中使用比较广泛的有短语结构语法和依存语法。

当前短语结构句法分析普遍基于概率上下文无关文法 (Probabilistic Context Free Grammar, PCFG)。在早期研究工作中, 基于上下文无关文法的短语结构句法分析

方法直接从人工标注的树库中读取文法规则,并以相对频率作为规则的概率<sup>[12]</sup>.这类方法实现简单,但是先前的研究工作表明这种方法的性能并不理想.其主要原因在于上下文无关文法中的独立性假设,而这些独立性假设在实际情况中往往并不成立.

依存语法由法国语言学家 Lucien Tesnière 于 1959 年提出,依存语法是天然词汇化的,直接按照词语之间的依存关系工作.由于依存语法中词汇的依存本质是语义的,而不同语言间的语义层面是相通的,因此依存语法是一种跨越语言界限、客观揭示人类语言内在规律的句法理论.在汉语方面,依存句法分析的工作在最近几年开始受到重视.Zhou 是最早从事这方面的研究者之一,他采用分块的思想,应用一些制定的语法规则,先对句子进行分块处理,找出关系固定的语块,然后再对整个句子进行依存分析<sup>[13]</sup>.Lai 等人使用基于 span 的思想、Gao 等利用无指导的方法在汉语依存分析方面做了有价值的研究工作<sup>[14,15]</sup>.

句法结构是句法形式和语义内容的统一体.对句法结构不仅要作形式分析,例如句法层次分析、句法关系分析以及句型分析等,而且还要作种种语义分析.对句法结构的语义分析越全面、越深刻,就越有可能对句法形式上的各种现象给以科学合理的解释.目前的词汇化句法分析如中心词驱动句法分析模型<sup>[11]</sup>、依存语法仅仅考虑词语之间的语义依存关系,而没有引入更多的反映词语语义特点的信息,如语义类、语义搭配等语义信息,而这些语义信息对句法分析和语义计算是至关重要的.举例来说,在句子“Astronomers saw stars with telescopes”中词“telescopes”在语义搭配上既跟其直接的核心词“with”有关,也与整个句子的核心词“saw”有关,如果采用依存分析法,由于依存语法公理的制约,“telescopes”和“saw”之间无法建立依存关系,而这种关系对句法分析是至关重要的.

现有主流的句法分析理论并没有有效刻画出汉语的本质特性,导致目前汉语句法分析和语义计算的效果与英语相比相差较大.在汉语中,配价结构可以较好地刻画汉语句子的句法结构和语义构成关系,因此,我们有必要更系统广泛地考察和研究形式化语法理论,尤其是配价语法,并在此基础上建立句法分析模型.现有配价语法<sup>[16,17]</sup>的研究多集中于研究词语的配价特点,而没有考虑整个句子的配价结构.我们希望定义一种句子的配价结构,这种配价结构应该能反映出句子中所有词语之间的配价关系.

论文后续内容的安排如下:第一部分介绍配价语法和配价结构;第二部分介绍基于配价结构和短语结构树的语义分析;第三部分结合中心词驱动句法分析模型,提出基于配价结构和语义依存关系的句法分析

统计模型;第四部分给出模型的实验结果及分析.

## 2 配价语法和配价结构

配价语法与依存语法一样,同样被认为是来源于法国语言学家特斯尼耶尔的语言学思想.按照陆俭明先生在《现代汉语配价语法研究》(沈阳,郑定欧主编)序言中的说法,“价”(valency/valenz,亦称“配价”/“向”)这一术语借自化学,化学中“价”的概念用于说明在分子结构中各元素原子数目之间的比例关系,而特斯尼耶尔在语法学中引进“价”的概念,是为了说明一个动词能支配多少个名词词组.比如说,“吃”是一个二价动词,需要支配两个名词词组,分别说明“谁吃”和“吃什么”.而“给”是一个三价动词,需要支配三个名词词组,分别说明“谁给”、“给谁”、“给什么”.不难看出,配价语法和句子级的语义计算(特别是语义角色标注)有着紧密的联系.现在,配价的研究已经不仅仅局限于动词,形容词和名词的配价也有很多人在研究.比如说,形容词“年轻”和名词“姐姐”都是一价,分别需要支配一个名词词组,用于说明“谁年轻”和“谁的姐姐”.

### 2.1 动词的配价结构

配价语法强调动词是句子的中心,所以是动词中心论.动词的配价与句子的生成有极其密切的关系,要生成或理解一个句子,关键在动词.由动词为核心组成的动核结构(或称谓核结构)是生成句子的基底,任何句子都是运用语法手段让动核结构与一定的句法结构结合成句模并给以某种语用价值生成的.

动词根据它联系的动元(动词所联系的强制性语义成分)的数量来分类,即动词的“价”分类,可分为一价动词、二价动词和三价动词三类.袁毓林提出了配价层级<sup>[18]</sup>的思想,从而把单一的价的概念分化为联、项、位、元四个平面构成的配价层级.“联”是指一个动词在各种句子中所能关联的不同的语义角色的数量,“项”是指一个动词在一个句子中所能关联的名词性成分的数量(包括通过介词引导的名词性成分),“位”是指一个动词在一个句子中不借助介词所能关联的名词性成分的数量,“元”是指一个动词在一个简单的基础句中所能关联的名词性成分的数量.通过这种层级关系来充分反映动词在不同层面、不同句法框架中的组合和支配能力.

### 2.2 形容词的配价结构

张国宪等对形容词的配价进行了分类<sup>[19]</sup>.根据谓语句形容词所带补足语的数量,可分为单价、双价和三价;根据谓语句形容词对所带补足语的强制性程度,可分为必有价和可有价;根据补足语是否有标记介词,可分为有标记价和无标记价;根据谓语句形容词所带补足语的稳定性程度,可以分为静态价和动态价.

### 2.3 名词的配价结构

袁毓林受朱德熙对汉语动词的配价研究的直接影响,着手对汉语名词的配价研究<sup>[18]</sup>.从配价的角度看,现代汉语名词可分为无价名词(或零价名词和有价值名词两大类,这是根据名词有无配价要求分类的.有价值名词又分为两类:一类是从谓词派生出来的,另一类不是从谓词派生出来的,它们往往包含一个降级述谓结构.其中根据其支配能力又可以分为一价名词和二价名词两小类.

同样在英语中,大部分有价值名词为一价名词,如名词短语“pet food volume”中,词“food”、“volume”均为一价名词,词“food”修饰中心词“volume”,词“pet”修饰“food”而不是修饰中心词“volume”.文法规则 NPB→NN NN NN 的概率为:

$$P_l(L_i(l_i)|H, P, h, L_1(l_1)\cdots L_{i-1}(l_{i-1})) = P_l(L_i(l_i)|P, L_{i-1}(l_{i-1})) \quad (1)$$

$$P_l(R_i(r_i)|H, P, h, R_1(r_1)\cdots R_{i-1}(r_{i-1})) = P_l(R_i(r_i)|P, R_{i-1}(r_{i-1})) \quad (2)$$

部分有价值名词为二价名词,如名词短语“vanilla ice cream”中,词“cream”为二价名词,词“vanilla”、“ice”均修饰中心词“cream”.文法规则 NPB→NN NN NN 的概率为:

$$P_l(L_i(l_i)|H, P, h, L_1(l_1)\cdots L_{i-1}(l_{i-1})) = P_l(L_i(l_i)|H, P, h) \quad (3)$$

$$P_l(R_i(r_i)|H, P, h, R_1(r_1)\cdots R_{i-1}(r_{i-1})) = P_l(R_i(r_i)|H, P, h) \quad (4)$$

### 3 基于配价结构和短语结构树的语义分析

在配价语法中,领主属宾句指“王冕七岁上死了父亲”这种句子.与一般的句式相比,我们可看到这种句式有以下特点:(1)句中的主语与述语动词没有直接的语义关系,不是述语动词的必有语义成分,表现为主语类型的非典型性;(2)句中宾语多为述语动词的施事,表现为宾语类型的非典型性;(3)主语与宾语的联系不是靠动词而是靠两个成分之间在词汇语义上的“领有-隶属”关系,伴随这个特点的是述语动词(或形容词)为一价(或一向).

下图以“陈楠三十岁了儿子”给出了依存树、和设想中的配价结构.其中图2是我们设想的一种可能的配价结构形式,这种形式可能在我们的研究过程中还会发生变化和进步.可以看到,两个句子具有形式相同的依存树,却具有不同的配价结构,可见与传统的短语结构树和依存树相比,配价结构反映了更多的语义特点.同时,词语的配价信息比较稳定.最后,配价结构从形式上并非一棵树,而是一个有向图.因此,配价结构具有比短语语法和依存语法更强的表现能力,有潜力

获得更高的句法语义计算性能.因为在句子“陈楠三十岁了儿子”中,“陈楠”是零价的,“儿子”和“三十岁”都是一价的,“三十岁”并且作为时间副词修饰动词“生了”,动词“生了”是二价的.有了这些词语的配价信息,就可以比较准确地获得上述的配价结构.

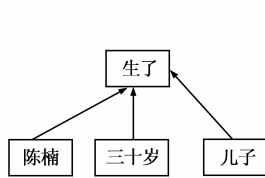


图1 句子“陈楠三十岁了儿子”依存树

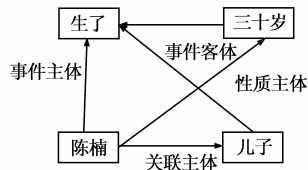


图2 句子“陈楠三十岁了儿子”的一种可能的配价结构

句法结构是句法形式和语义内容的统一体.对句法结构不仅要形式分析,例如句法层次分析、句法关系分析以及句型分析等,而且还要做种种语义分析.对句法结构的语义分析越全面、越深刻,就越有可能对句法形式上的各种现象给以科学合理的解释.我们的基本思想是:在句子短语结构或依存结构的基础上,利用基于配价理论开发的语义词典分析得到句子配价结构,反过来再利用句子配价结构对句中依存关系进行必要的修正.较详细的做法如下:

1 根据词语的配价信息和句子短语结构,可以得出句子配价结构的如下的一些推导规则(关于句式与词语配价关系的讨论可参考袁毓林的著作“汉语配价语法研究”):

1) 句子短语结构是词语配价的一个实现,词语的配价数必须在句子结构中得到满足;

2) 处于句子同一层级的词语,中心词可以支配其它词,同时除中心词外的词必须受某个词支配;

3) 处于句子同一层级的名词,后面的名词可以支配前面的名词.

2 在 Collins 的中心词驱动句法分析模型中计算文法规则的概率时,假定修饰成分间相互独立.而在我们的模型中,根据配价结构中提取的词语配价关系语义信息,有配价关系的修饰成分间不能相互独立.

在句法分析模型中引入丰富的语义信息,既包括由句法树或依存树确定的语义依存信息,也包括由句子分析树对应配价结构图确定的语义搭配信息.

下图以“Astronomers saw stars with telescopes”为例,给出了短语结构树和设想中的配价结构.其中图中(5)、(6)是两棵不同的句法树图,图中(7)、(8)为对应的在短语结构树基础上分析得到的可能配价结构形式,这种形式可能在我们的研究过程中还会发生变化和进步.

由配价结构图(7)、(8),结合句法树图(5)、(6),我们可以得到更多的语义知识:在句子分析树(5)中词

telescopes 与词 with 有语义依存关系,同时与词 saw 有语义搭配关系;而在句子分析树(6)中词 telescopes 与词 with 有语义依存关系,同时与词 stars 有语义搭配关系.

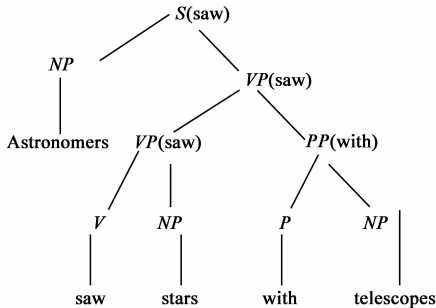


图3 句子“Astronomers saw stars with telescopes”分析树1

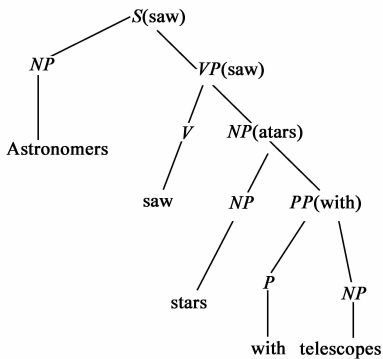


图4 句子“Astronomers saw stars with telescopes”分析树2

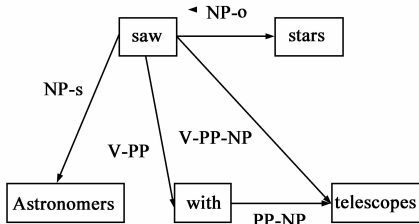


图5 句子分析树1对应的可能配价结构

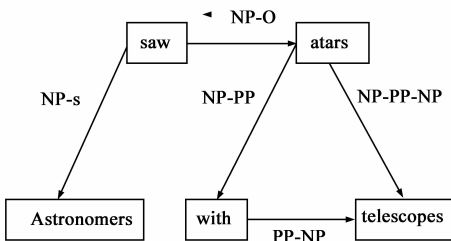


图6 句子分析树2对应的可能配价结构

## 4 基于配价结构和语义依存关系的句法分析统计模型

### 4.1 中心词驱动句法分析模型的基本原理

中心词驱动句法分析模型是最具有代表性的词汇化模型.为了发挥词汇信息的作用,中心词驱动模型为

文法规则中的每一个非终结符(none terminal)都引入核心词/词性信息.由于引入词汇信息,不可避免将出现严重的稀疏问题.为了缓解这个问题,中心词驱动模型把每一条文法规则的右侧分解为三大部分,分别为:一个中心成分;若干个在中心左边的修饰成分;若干个在中心右边的修饰成分.形式化地,可以写成如下形式:

$$P(ht, hw) - L_m(lt_m, lw_m) \cdots L_1(lt_1, lw_1) H(ht, hw) R_1(rt_1, rw_1) \cdots R_n(rt_n, rw_n) \quad (5)$$

其中,  $P$  为非终结符,  $H$  表示中心成分,  $L_1$  表示左边修饰成分,  $R_1$  表示右边修饰成分.  $Hw, lw, rw$  均是成分的核心词,  $ht, lt, rt$  分别是它们的词性.进一步假设,首先由  $P$  产生核心成分  $H$ , 然后以  $H$  为中心分别独立地产生左右两边的所有修饰成分.这样,形如式(1)的文法规则的概率为:

$$P_h(H | P(ht, hw)) \cdot \prod_{i=1}^{m+1} P_i(L_i(lt_i, lw_i) | H, P, h, \Delta_l(i-1)) \cdot \prod_{i=1}^{n+1} P_i(R_i(rt_i, rw_i) | H, P, h, \Delta_r(i-1)) \quad (6)$$

其中,  $L_{m+1}$  和  $R_{n+1}$  分别为左右两边的停止符号,  $\Delta_l(i-1)$  为距离函数, 补偿结构信息的缺失. 距离信息考虑了三种情况: 1 该成份前是否有成份; 2 该成份前是否出现动词; 3 该成份前是否出现有标点符号.

### 4.2 基于配价结构和语义依存关系并结合中心词驱动模型的句法分析模型

设  $P(h)$  表示句法树上当前核心词  $h$  所依赖的上层核心词, 其它符号的表示同上文一致. 在我们的句法分析模型中, 每一条文法规则写成如下形式:

$$P(ht, hw | P(h)) - L_m(lt_m, lw_m) \cdots L_1(lt_1, lw_1) \cdot H(ht, hw | P(h)) R_1(rt_1, rw_1) \cdots R_n(rt_n, rw_n) \quad (7)$$

形如式(7)的文法规则的概率为:

$$P_h(H | (ht, hw), P(h)) \cdot \prod_{i=1}^{m+1} P_i(L_i(lt_i, lw_i) | L_{i-1}(lt_{i-1}, lw_{i-1}), \cdots, L_1(lt_1, lw_1), (ht, hw), P(h)) \cdot \prod_{i=1}^{n+1} P_i(R_i(rt_i, rw_i) | R_{i-1}(rt_{i-1}, rw_{i-1}), \cdots, R_1(rt_1, rw_1), (ht, hw), P(h)) \quad (8)$$

其中,  $L_{m+1}$  和  $R_{n+1}$  分别为左右两边的停止符号. 式(8)中的概率

$$P_i(R_i(rt_i, rw_i) | R_{i-1}(rt_{i-1}, rw_{i-1}), \cdots, R_1(rt_1, rw_1), (ht, hw), P(h)) \quad (9)$$

可分解为两个概率

$$P_i(rt_i | rt_{i-1}, rt_{i-2}, \cdots, rt_1, ht, rw_i) \quad (9)$$

$$P_i(rw_i | rw_{i-1}, rw_{i-2}, \cdots, rw_1, hw, P(h)) \quad (10)$$

的乘积, 记  $S(rw_i)$  表示词  $rw_{i-1}, rw_{i-2}, \cdots, rw_1, P(h)$  中与当前词  $rw_i$  有语义搭配关系的词(由句子分析树对应

配价结构图确定),则有:

$$P_i(rw_i | rw_{i-1}, rw_{i-2}, \dots, rw_1, hw, P(h)) = P_i(rw_i | hw, \Delta_r(i-1), S(rw_i)) \quad (11)$$

再假定  $hw, S(rw_i)$  关于  $rw_i$  条件独立有:

$$P_i(rw_i | hw, \Delta_r(i-1), S(rw_i)) = \frac{P_i(rw_i | hw, \Delta_r(i-1)) \cdot P_i(rw_i | S(rw_i))}{P_i(rw_i)} \quad (12)$$

式(12)中概率  $\frac{P_i(rw_i | \Phi(rw_i))}{P_i(rw_i)} = \frac{P_i(rw_i, S(rw_i))}{P_i(rw_i) \cdot P_i(S(rw_i))}$  即为  $rw_i, S(rw_i)$  间的互信息,因而整个式(12)概率意义十分明确,符合语言现象。

可以说,目前词汇化的上下文无关语法所做的独立性假设与语言现象不相符合,既不适合于英文,更加不适合于中文.在我们的句法分析模型中,用条件独立性假设取代了中心词驱动句法分析模型中的独立性假设.从统计学的角度来说,相对条件独立性假设,独立性假设是过强假设,与语言现象也不尽符合.因而,我们的句法分析模型更符合语言的实际物理过程.通过对 Collins 模型的规则进行分解和修改,基于配价结构并结合中心词驱动模型的词汇化句法分析模型能够更好地融入语义(既包括由句法树确定的语义依存信息,也包括由句子分析树对应的配价结构图确定的语义搭配信息)等语言方面知识,提高句法分析的准确率。

## 5 实验结果

试验数据取自宾州中文树库(CHTB)5.0版本,大部分取材于新华社新闻, Sinorama 新闻杂志以及香港新闻. CTB 是由语言数据联盟(LDC)公开发布的一个语料库,为汉语句法分析研究提供了一个公共的训练、测试平台.该树库包含了 507222 个词,824983 个汉字,18782 个句子,有 890 个数据文件.为了在训练集、开发集和测试集中平衡各种语料来源,我们将语料分割如下:我们将文件 301~320、611~630 作为调试集,将文件 271~300、631~660 作为测试集,其余文件作为训练集.本文的所有实验中,模型的参数都是从训练集中采用极大似然法估计出来的。

测试的结果采取了常用的 4 个评测指标,即准确率  $P$ 、召回率  $R$ 、综合指标  $F$  值和交叉括号 CB.其定义如下:

精确率(Precision)用来衡量句法分析系统所分析的所有成份中正确的成份的比例。

召回率(Recall)用来衡量句法分析系统分析出的所有正确成份在实际成份中的比例。

综合指标:  $F = (P \times R \times 2) / (P + R)$ 。

交叉括号 CB:给出了在一棵树中与其他树的成分边界交叉的成分数目的平均数。

实验中采用的句法分析 Baseline 系统是 Daniel M. Bikel 基于 Collins 模型实现的 DBParser.表 1 列出了 baseline 系统和改进模型的句法分析实验结果。

从表 1 可以看出:由于在规则的分解及概率计算中,既利用了由句法树或依存树确定的语义依存信息,也利用了由句子分析树对应配价结构图确定的语义搭配信息.改进模型的准确率  $P$ 、召回率  $R$ 、综合指标  $F$  值、交叉括号比 Collins 的中心词驱动句法分析模型有了明显的提高.试验结果表明语言特征知识的应用对统计句法分析有很大的影响,这从一个侧面指出了汉语统计句法分析研究的一个方向:从语言学角度寻找更多的语法、语义、语用等特征知识。

表 1 句法分析实验结果

模型	准确率%	召回率%	$F\%$	交叉括号
Baseline	82.76	80.17	81.44	2.05
改进模型	86.13	85.21	85.66	1.83

数据稀疏问题是另一个严重影响句法分析系统性能的重要因素,改进模型采用了基于语义类和可变量模型的平滑技术,成功解决了数据稀疏问题,大大提高了系统性能.表 2 是采用平滑技术后的句法分析实验结果,这方面的详细讨论将另文发表。

表 2 句法分析实验结果

模型	准确率%	召回率%	$F\%$	交叉括号
Baseline	82.76	80.17	81.44	2.05
采用平滑技术				
后的改进模型	88.76	87.43	88.09	1.72

## 6 结论与讨论

1)在汉语中,配价结构可以较好地刻画汉语句子的句法结构和语义构成关系,因此,我们系统地考察和研究了配价语法相关理论,并在此基础上建立句法分析模型.基于配价结构和语义依存关系的句法分析模型在规则的分解及概率计算中,既利用了由句法树或依存树确定的语义依存信息,也利用了由句子分析树对应配价结构图确定的语义搭配信息,性能有了明显的提高。

2)现有配价语法的研究多集中于研究词语的配价特点,而没有考虑整个句子的配价结构.我们希望定义一种句子的配价结构,这种配价结构应该能反映出句子中所有词语之间的配价关系.我们所希望标注的配价关系不仅是涉及到动词与名词短语直接的关系,也涉及到名词与名词短语、形容词与名词短语、甚至副词与动词形容词短语之间的关系,也就是说,配价结构应该是一种完整的句法结构。

## 参考文献

- [1] Manning C D, Schutze H. Foundations of Statistical Natural Language Processing[M]. London: the MIT Press, 1999: 184 - 197.
- [2] Jesus Vilares, Miguel A. Alonso, Manuel Vilares. Extraction of complex index terms in non-English IR: A shallow parsing based approach[J]. Information Processing and Management, 2008, 44(4): 1517 - 1537.
- [3] 刘水, 李生, 赵铁军等. 头驱动句法分析中的直接插值平滑算法[J]. 软件学报, 2009, 20(11): 2915 - 2924.  
LIU Shui, LI Sheng, ZHAO Tie-Jun, et al. Directly Smooth Interpolation Algorithm in Head-Driven Parsing[J]. Journal of Software, 2009, 20(11): 2915 - 2924. (in Chinese)
- [4] 代印唐, 吴承荣, 马胜祥, 钟亦平. 层级分类概率句法分析[J]. 软件学报, 2011, 22(2): 245 - 257.  
DAI Yin-Tang, WU Cheng-Rong, MA Sheng-Xiang, ZHONG Yi-Ping. Hierarchically Classified Probabilistic Grammar Parsing[J]. Journal of Software, 2011, 22(2): 245 - 257. (in Chinese)
- [5] Aviran S, Siegel P H, Wolf J K. Optimal Parsing Trees for Run-length Coding of Biased Data[J]. IEEE Transaction on information Theory. 2008, 54(2): 841 - 849.
- [6] ZHOU De-yu, HE Yu-lan. Discriminative Training of the Hidden Vectors State Model for Semantic Parsing[J]. IEEE Transaction on Knowledge and Data Engineering. 2009, 21(1): 66 - 77.
- [7] 袁里驰. 基于词聚类的依存句法分析[J]. 中南大学学报: 自然科学版, 2011, 42(7): 2023 - 2027.  
YUAN Li-chi. Dependency language parsing model based on Word Clustering[J]. Journal of Central South University: Natural Science, 2011, 42(7): 2023 - 2027. (in Chinese)
- [8] 孙昂, 江铭虎, 贺一帆, 陈林, 袁保宗. 基于句法分析和答案分类的中文问答系统[J]. 电子学报, 2008, 36(5): 833 - 839.  
SUN Ang, JIANG Ming-hu, HE Yi-fan, CHEN Lin, YUAN Bao-zong. Chinese Question Answering Based on Syntax Analysis and Answer Classification[J]. Acta Electronica Sinica, 2008, 36(5): 833 - 839. (in Chinese)
- [9] 陈毅恒, 秦兵, 宋凡, 刘挺, 李生. 基于 ontology 抽取优化初始选择的检索结果聚类[J]. 电子学报, 2008, 36(12A): 166 - 171.  
CHEN Yi-heng, QIN Bing, SONG Fan, LIUTing, Search Result Clustering Based on Centroid Optimization by Ontology Extraction[J]. Acta Electronica Sinica, 2008, 36(12A): 166 - 171. (in Chinese)
- [10] 袁里驰. 融合语言知识的统计句法分析[J]. 中南大学学报: 自然科学版, 2012, 43(3): 986 - 991.  
YUAN Li-chi. Statistical Parsing with Linguistic Features [J]. Journal of Central South University: Natural Science, 2012, 43(3): 986 - 991. (in Chinese)
- [11] Collins M. Head-Driven Statistical Models for Natural Language Parsing[J]. Computational Linguistics, 2003, 29(4): 589 - 637.
- [12] Daniel Jurafsky, James H. Martin. Speech and language processing[M]. New Jersey, Prentice Hall, 2009: 210 - 265.
- [13] M Zhou. A block-based dependency parser for unrestricted Chinese text. [C]//Proceedings of the 2nd Chinese Language Processing Workshop, Hong Kong, 2000: 78 - 84.
- [14] J F Gao, H Suzuki. Unsupervised Learning of Dependency Structure for Language Modeling. [C]//Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, 2003: 521 - 528.
- [15] T B Y Lai, C N Huang, M Zhou, J B Miao, and K C Siu. Span-based Statistical Dependency Parsing of Chinese[A]. Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS2001)[C], Tokyo, Japan, 2001: 677 - 684.
- [16] 沈家焯. 句式和配价[J]. 中国语文, 2000, (4): 291 - 297.  
SHEN Jia-xuan, Valency and sentence patterns [J]. ZHONGGUO YUWen, 2000, (4): 291 - 297. (in Chinese)
- [17] 聂鸿英. 汉语“配价”语法研究综述[J]. 延边大学学报 (社会科学版), 2011, 44(2): 39 - 42.  
NIE Hong-ying. Review of “Coordination Valence” in Chinese Grammar [J]. Journal of Yanbian University (Social Science), 2011, 44(2): 39 - 42. (in Chinese)
- [18] 袁毓林. 汉语配价语法研究 [M]. 北京, 商务印书馆, 2010: 55 - 170.  
YUAN Yu-lin. The study of Chinese Valence grammars [M]. Beijing, Commercial press, 2010: 55 - 170. (in Chinese)
- [19] 周国光. 现代汉语配价语法研究 [M]. 北京, 高等教育出版社, 2011: 21 - 82.  
ZHOU Guo-guang. The study of modern Chinese Valence grammars [M]. Beijing, Higher education press, 2011: 21 - 82. (in Chinese)

## 作者简介



袁里驰 男, 1973 年 5 月出生于湖南邵阳, 江西财经大学信息管理学院副教授, 硕士生导师. 研究方向为自然语言处理.

E-mail: yuanlichang@sohu.com