

# 基于语义依存关系的句子理解模型

李涓子 王作英\*

清华大学计算机科学与技术系

\*清华大学电子工程系

ljz@keg.cs.tsinghua.edu.cn

**摘要:** 句子中词与词之间的远距离关系和句子意义的理解是建立语言模型中的两个重要问题。本文依据依存语法理论,用语义依存关系序列表示句子中词之间的句法语义关系,该方法可以减少人工标注句子语义结构时使用的句法分析器;在此表示方法下,本文提出一个基于语义依存关系的句子理解模型,模型可以解决词之间的远距离依存关系和句子意义理解问题。

**关键词:** 语言模型 依存语法 句法分析器 自然语言理解

## Sentence Understanding modeling based on semantic dependency Relationships

Li Juanzi Wang Zuoying\*

Depa. Of Computer Science and Technology, Tsinghua University

\*Depa. Of Electric Engineering and Technology, Tsinghua University

**ABSTRACT:** Long-distance dependency between words and sentence understanding are two important problems in language modeling. This paper, by using the theory of dependency grammar, uses a sequence of semantic or syntactic relations to denote the semantic structure of sentence so that it can speed the sentence tagging. Furthermore, an understanding model based on semantic dependency relations is proposed. The model can solve the problem of long-distance dependencies between words and the problem of sentence meaning understanding.

**Key words:** speech recognition, language model, dependency grammar, parser

### 一、引言

N 元语法模型是语音识别中被广泛使用的一种语言模型。但现有的语言模型存在两个问题:第一,不能解决 N 元语法模型中词之间的远距离搭配问题。第二,不能解决语音识别中句子的意义理解问题。在解决远距离搭配问题的语言模型的研究中,代表性的模型有 cache 模型[1]、Trigger 模型[2],可变长度和跨长度的 N 元语法模型(目前长度从 1 到 4)[3][4]。这些模型都试图在 N 元语法模型上找出句子中词之间的远距离关系,而这些关系并不完全是词之间真正的句法语义关系,因此也不可能达到对句子的理解。约翰霍普金斯大学(JHU)在宾州大学建立的句法分析树库(UPenn Treebank)上建立句法分析器,提出了语音识别的句法结构语言模型[5]。实现的模型在 SWICHBOARD 语料的实验表明,该方法比

传统的基于词的三元模型有改进。但是建立这样的模型需要有大规模的带句法结构的语料库，而汉语目前还不存在这样的大规模的带句法结构的语料库。另一方面，目前理解模型的研究还限定在特定领域，模型大都采用基于语义类和基于规则的语言模型。国外较成功的理解语言模型有 MIT 研制的 JUPITER 系统和 Philips 公司建立的理解模型[6][7]。这些理解模型存在以下问题。第一：需要手工编写语法规则，在向语法中添加新规则时需要保证语法的一致性，存在规则方法的不足。第二：在 Philips 的系统中使用人为赋予的概率规则，具有一定的局限性。第三：上面的语言理解模型并不是结构的语言理解模型。

本文利用依存语法理论，用一组语义依存关系表示句子的语义结构信息，提出了一个融解决词之间远距离搭配和意义理解问题为一体的句子理解机制。实验表明，理解模型具有较高的分析正确率。

## 二、句子语义依存关系及其形式化表示

### 1、句子语义依存关系的形式化描述

句子的意义是由句子中词之间的句法语义关系体现出来的，而依存语法描述的正是具有直接句法语义联系的词之间的关系，因此我们在描述句子中词之间的语义关系时，采用了相对成熟的依存语法理论。设  $W$  是含有  $n$  个词的句子，表示为  $W=w_1, w_2, \dots, w_n$ ，由  $W$  中各词对应的语义类组成的语义类串为  $S=s_1, s_2, \dots, s_n$ 。则  $W$  的语义依存关系表可表示为  $SRL=\{SR(1), SR(2), \dots, SR(n)\}$ ，其中， $SR(i)=(j, R)$ ，它表示句子中第  $i$  个词是第  $j$  个词的修饰成分，他们之间的语义依存关系为  $R$ ，并且这个关系下的中心词为  $w_j$ ，中心词的语义类为  $s_j$ 。可以用  $SR(k)=(0, Top)$  表示句子的核心成分，其中  $w_k$  为句子的中心词。因此，一个含有  $n$  个词的句子就有  $n$  个关系。如句子  $W=$  叫 张三 老师 接 一下 电话。这个句子的语义依存关系表为： $SRL=\{(0, \text{核心成分})(3, \text{修饰})(1, \text{受事})(1, \text{使动})(4, \text{补语})(4, \text{对象})\}$ 。其中数字表示词在句子中的位置。

### 2、句子理解树与语义归并树

由  $n$  个词组成的句子理解树是一棵有  $n-1$  条边的二叉树，它直观反映了句子的意义结构。例句中基于词的句子理解树如图 1 所示。

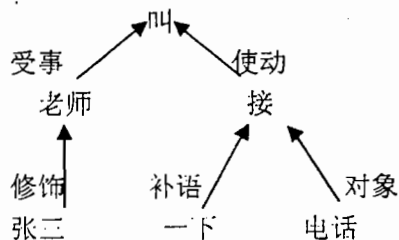


图 1: 例句的句子理解树

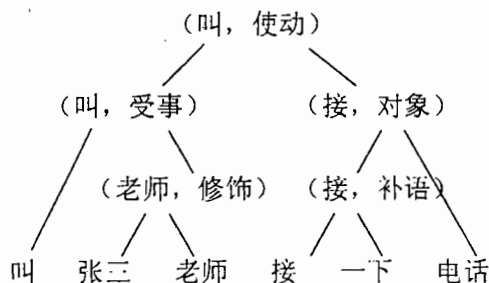


图 2: 语义分析树

可以把句子语义依存结构的分析过程看成是一个句子中两两词之间的归并过程，每次归并后产生的新结点是归并关系中的中心词以及归并时使用的语义依存关系。可以用一棵语义归并树表示词之间的归并过程。图 2 给出了上面例句的基于词的语义归并树。

本节给出了用语义依存关系表表示句子的语义结构的方法。这种表示方法使得在标注语料库时，对句中的每个词只需要指出这个词是哪个词的修饰词以及他们的语义依存关系。而后从句子的语义依存关系表可以直接得到这个句子的语义理解树和语义归并树的表示。

### 三、基于语义依存关系的理解模型

#### 1、理解模型框架

设  $W=w_1, w_2, \dots, w_n$  表示由  $n$  个词组成的句子， $W$  对应的语义类串表示为  $S=s_1, s_2, \dots, s_n$ ，它的语义依存关系表对应的基于语义类的语义归并树用  $ST$  表示。再设  $W_k=w_1, w_2, \dots, w_k$  ( $k < n$ ) 表示句子的前  $k$  个词组成的词串， $S_k$  为前  $k$  个词  $W_k$  组成的语义类串， $ST_k$  为  $W_k$  组成的语义类的语义归并树。设  $P(W, S, ST)$  表示句子  $W$  的语义类串为  $S$  并且对应的语义类的语义归并树为  $ST$  的概率，利用概率乘法公式，这个概率表示为：

$P(W, S, ST) = P(S, ST)P(W | S, ST)$ ，若假设  $s_k$  只与  $ST_{k-1}$  相关，则有公式：

$$P(S, ST) = \prod_{k=1}^n P(s_k | ST_{k-1}) * P(ST_k | ST_{k-1}, s_k)$$

再假设词  $w_k$  只与其所属语义类有关，最终概率  $P(W, S, ST)$  表示为：

$$P(W, S, ST) = \prod_{k=1}^n P(s_k | ST_{k-1}) * P(ST_k | ST_{k-1}, s_k) * P(w_k | s_k) \quad (1)$$

- $P(s_k | ST_{k-1})$  表示前  $k-1$  个词的部分语义归并树为  $ST_{k-1}$  时第  $k$  个词语义类为  $s_k$  的概率。
- $P(ST_k | ST_{k-1}, s_k)$  表示根据前  $k-1$  词的语义归并树  $ST_{k-1}$ ，在第  $k$  个词的词义为  $s_k$  时，继续进行语义归并分析，产生的归并树结构为  $ST_k$  的概率。这个概率可进一步表示为：

$$P(ST_k | ST_{k-1}, s_k) = \prod_{i=1}^{N_k} P(r_i^k | ST_{k-1}, s_k, r_1^k \dots r_{i-1}^k) \quad (2)$$

公式 (2) 中  $r_i^k$  表示在  $ST_{k-1}$  和  $s_k$  下进行依存关系分析时进行的第  $i$  次归并操作， $r_i^k$  是进行第  $i$  次归并时使用的规则， $P(r_i^k | ST_{k-1}, s_k, r_1^k \dots r_{i-1}^k)$  为进行第  $i$  次归并操作时的概率， $N_k$  为在分析第  $k$  个词时使用规则的个数。

- $P(w_k | s_k)$ ：表示第  $k$  个词的语义类  $s_k$  时输出  $w_k$  的概率。

在我们实现的模型中使用了 2 元模型，即假设第  $k$  个词的语义类只与它紧邻的前  $k-1$  个词的语义归并树  $ST_{k-1}$  的中心词有关。上面的各种概率公式简化表示为下面两个公式，

$$P(s_k | ST_{k-1}) = P(s_k | s_{h0}) \text{ 和 } P(ST_k | ST_{k-1}, s_k) = \prod_{i=1}^{N_k} P(r_i^k | s_{h0}, s_k, r_1^k \dots r_{i-1}^k)。$$

其中， $s_{h0}$  为  $ST_{k-1}$  中最右边的部分语义归并树的中心词。

## 2、模型的特点

- (1) 模型在计算下一个可能的词的语义类时利用的是它前  $k-1$  个词的语义类组成的部分语义归并树的中心词，因此具有发现远距离语义依存关系的能力。如句子：叫教务科的张三老师接电话，在“接”之前的语义归并结构的中心词为“叫”，在计算下一个可能的词的语义类时利用“叫”的语义类，而不是“老师”的语义类。
- (2) 模型分析的结果可以得到句子的语义归并树，实现对句子意义的理解。设句子  $W$  的语义依存结构为  $ST'$ ，则：

$$ST' = \max_{S, ST} P(W, S, ST) \quad (3)$$

## 四、知识获取

为实现这个概率句子理解模型，首先要对语料库进行句子语义依存关系的标注。根据句子的语义依存关系表，导出句子的语义类的语义归并树，然后计算模型需要的各种参数。

在计算  $P(s_k | s_{h0})$  时，对语料库中的每个句子，根据句子中前  $k-1$  个词的语义依存关系，可以导出前  $k-1$  个词的基于语义类的语义归并树，找出该结构中的距  $s_k$  最近的一个结构的中心词，即为  $s_k$  在结构上接续的前一个词的语义类  $s_{h0}$ 。

$$P(s_k | s_{h0}) = \frac{C(s_{h0}, s_k)}{\sum_j C(s_{h0}, s_j)} \quad (4)$$

其中， $c(s_{h0}, s_j)$  表示在分析中  $s_k$  在语义类归并树中在结构上与  $s_j$  的同现次数。

在计算概率  $P(r | s_i, s_j)$  时，可以直接根据的句子依存关系序列，计算得到概率的依存规则。设  $SR(i) = (j, R)$  是语料库中的一个依存关系， $w_i$  和  $w_j$  对应的词的语义类分别是  $s_i$  和  $s_j$ ，则可将该关系表示为一个语义类的归并规则 Rule：

$$Rule = \begin{cases} s_j \rightarrow s_j + s_i & \text{if } i > j \\ s_j \rightarrow s_i + s_j & \text{if } i < j \end{cases}$$

该规则表示  $s_j$  和  $s_i$  可以归并，归并后的中心

词义为  $s_j$ 。则：
$$P(r | s_i, s_j) = \frac{C(s_i, s_j, r)}{\sum_{r_k} C(s_i, s_j, r_k)}$$
，其中  $C(s_i, s_j, r_k)$  表示语料中语义类  $s_i$

和  $s_j$  具有语义依存关系  $r_k$  的个数。

## 五、带概率的语义依存关系的理解算法

语义依存关系的理解算法采用由底向上的图表分析算法 (chart parser)。与已有算法不同的是在模型中加入了概率计算。算法的输入为一个待分析的句子，输出是通过分析得到的这个句子最优的语义归并树。过程首先利用公式 (1) 计算句子的所有的语义依存结构对应的概率值，然后利用公式 (3) 选择具有最大概率的语义归并树作为最终的输出结果。

### 1、数据结构

分析中用到两个数据结构：Agenda 堆栈和 Chart 表。Agenda 记录分析的进程，Chart 记录已经分析得到的部分分析结果以及已经部分匹配的规则。Chart 和 Agenda 都采用了下面 EedgeNode 的结构。结构中 Start\_vertex 和 End\_vertex 是当前成分在句子中的开始和结束位置，LHS\_Category、RHS\_Category 以及 Relation 是组成这个成分时使用的规则的左和右部和关系名，Structure 和 Lpro 是当前分析结果对应的语义依存结构和对应的概率值。

### 2、理解算法

理解过程对当前分析到的每条边，首先用这条边和 Chart 表中已经得到的部分的分析结果组成一个更大的结构，这个过程称为合并。其次还要检查当前边是否是句子中的尚未分析过的一个词，若是，则查找规则集，对这个边进行扩展，这个过程称为扩展。

#### 语义依存关系理解算法

- (1) 初始化 agenda 并且设 chart 表为空表；
- (2) 对句子中的每个词  $w_i$ ，做 (2) - (6)
  - 计算句子中第  $i$  个词的词义并放入 agenda 中；
  - 根据 chart 中记录的已分析的部分分析结果，查结构二元表，将它们的接续概率加入到已分析的结构中。
- (3) 在 agenda 表为空之前，做 (4) - (6)
- (4) 从 agenda 选一个成分 C 并放入 chart 表中；
- (5) 合并：根据新加入到 chart 表中的 C，检查该结点是否可以跟表中的一些结点合并，如过存在这样的边，则：

- 生成一个新的边;
  - 计算这个边对应的部分语义依存关系;
  - 将这条边放入 agenda 中;
- (6) 扩展: 如果成分 C 是一个具有完整结构的结点并且尚未扩展过, 则对该结点进行扩展。在规则集中查找规则, 使得规则的右部的第一个成分为该成分结点。如果存在这样的规则, 则生成一个新的边, 并放入 agenda 表中;
- (7) 从 chart 表中选择具有最大概率的  $P(W, S, ST)$  为分析结果。
- 

## 六、 实验及其结果分析

为了验证本文提出的理解模型, 我们在一个特定领域的较小规模的语料库上进行了实验, 语料库为电话语音转接这一特定领域中的句子, 共 566 句。根据语料定义了 39 个语义类和 19 种句法语义依存关系。之后, 对这 566 个句子逐一进行了标注和校对。训练语料包含 466 个句子, 测试语料包含 100 个句子。主要测试指标是召回率 recall 和正确率 precision,

$$recall = \frac{a}{b} \quad precision = \frac{a}{c}$$

其中, a 是分析器分析正确的句子个数, b 是分析器要识别句子的个数, c 是能被分析器分析的句子个数。

### 1、 基于词的语义依存关系理解模型的实验

这一理解模型在计算语义归并树时使用的是词而不是语义类信息。设  $W=w_1, w_2, \dots, w_n$  表示由 n 个词组成的句子, 它的语义依存关系表对应的基于词的语义归并树用  $WT$  表示。再设  $W_k=w_1, w_2, \dots, w_k$  ( $k < n$ ) 表示句子的前 k 个词组成的词串,  $WT_k$  为  $W_k$  组成的基于词的语义归并树。这个基于词的理解模型表示为:

$$P(W, WT) = P(w_1) \cdot \prod_{k=2}^n P(w_k | WT_{k-1}) \cdot P(WT_k | WT_{k-1}, w_k) \quad (5)$$

利用我们标注的语料和给出的参数计算方法, 得到基于词的规则 618 条, 词在语义归并树中结构同现个数 529 个。表 1 给出了基于词的理解模型的实验结果。

表 1: 基于词的理解模型实验结果表

	Recall	Precesion
测试结果	83. 5%	87. 0%

从表 1 看出, 基于词的语义依存结构分析模型抓住了词之间的精确的组合关系, 具有较高的分析正确率。但是由于训练语料的匮乏, 对于在语料中没有出现过的词的同现序列, 分析模型分析失败, 所以模型的召回率并不高。

### 2、 基于语义类的语义依存关系理解模型的实验

为了减轻人工标注语料的工作, 使在同数量的语料上建立起来的模型能够分析更多的

句子，我们实现了第二个模型，称为基于语义类的依存关系理解模型。模型采用公式(1)。训练得到的基于语义类的规则个数和语义类在结构上的同现个数分别为 267 个和 193 个。表 2 给出了基于语义类的理解模型的实验结果。

表 2: 基于语义类的理解模型的实验结果表

	Recall	Precesion
测试结果	91.9%	85.0%

从表 2 看出，这一模型的召回率比基于词的语义依存关系理解模型的召回率有很大的提高，但是模型分析的正确率却下降的不多。由此说明本文提出的基于语义类的语义依存关系的理解模型是合理的。

## 七、 结论

本文依据依存语法理论，提出一个基于语义依存关系的句子理解模型。模型具有以下特点：1) 用语义依存关系序列表示句子中词之间的句法语义关系，将句子的语义依存结构用一个线性关系序列表示，方便了句子的标注过程，减少了人工标注语料的时间。2) 在这个表示之上建立了一个概率的句子理解模型，从标注的句子中自动学习规则，并进行概率估值，从而将规则和概率方法有机融合在一起。3) 理解算法对句子的分析是由左向右完成的，因此可以将模型进一步应用于语音识别，实现结构的、理解的结构语言模型。

### 参考文献:

1. Iyer R, Rukmini M, Ostendorf M. Modeling long distance dependence in language: Topic mixtures versus dynamic cache models. IEEE Transactions on Speech and Audio Processing, 1999, v 7 (no 1): pp. 30-39
2. Tillmann C, Ney H. Selection criteria for word trigger pairs in language modeling. Lecture Notes in Artificial Intelligence, 1996. pp. 95
3. Jerome R Bellegarda. Speech recognition experiments using multi-span statistical language models. In: Proceedings of ICASSP, 1999. v2. pp 717-720
4. Manhung Siu, Mari Ostendorf, et al. Variable N-grams and extensions for conversational speech language modeling, IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 1, Journey, 2000. pp. 63-75
5. Chelba C., Exploiting syntactic structure for language modeling, A Dissertation of Johns Hopkins University, January 2000.
6. Victor Zue, Stephanie Seneff, et al. JUPITER: A telephone-based conversational interface for weather information, IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 1, January, 2000. pp. 85-96
7. Bernd Souvignier, Bernhard Rueber, et al. The thoughtful elephant strategies for spoken dialog systems, IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 1, Journey, 2000. pp. 51-62