

第 12 章 语 义 分 析

中科院信息工程研究所第二研究室

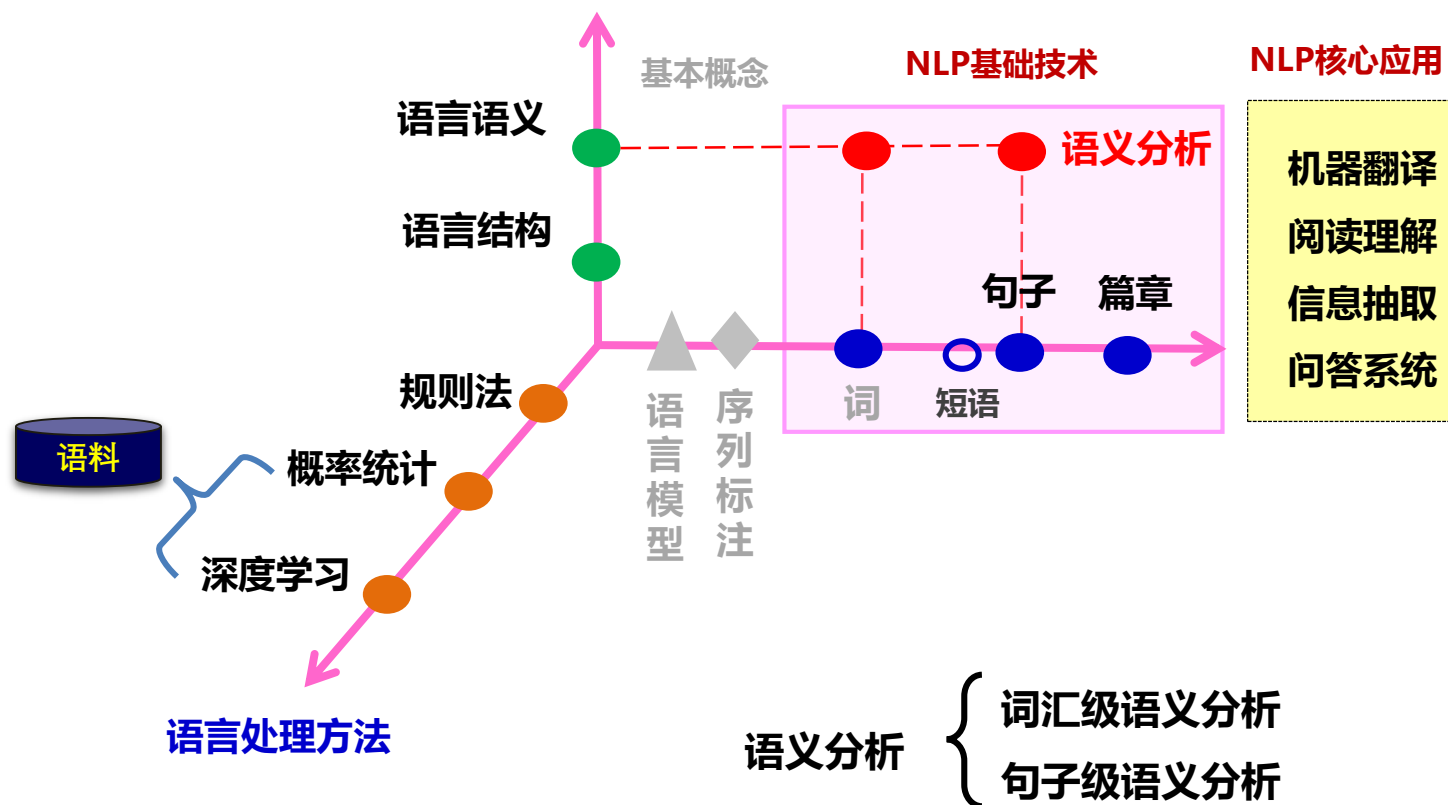
胡玥

huyue@iie.ac.cn

自然语言处理课程内容及安排

◇ 课程内容：

自然语言研究层面



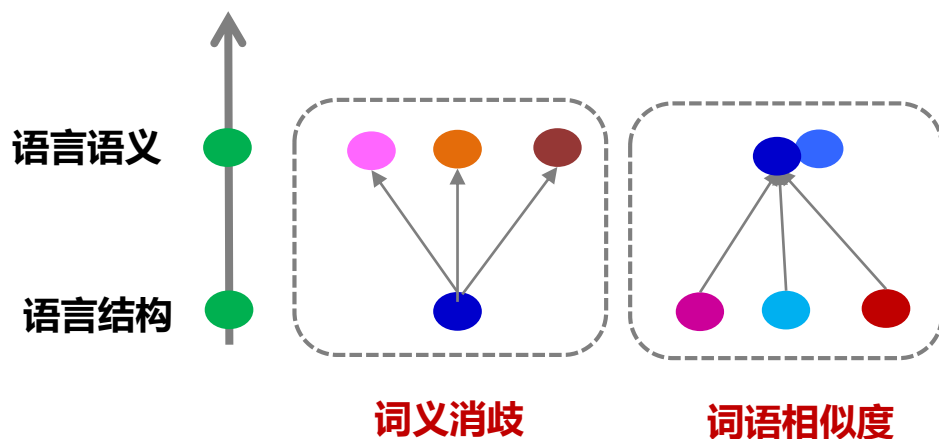
内 容 提 要

第一部分：词汇级语义分析

第二部分：句子级语义分析

词汇级语义分析概述

词汇级语义分析内容：



- **词义消歧**是自然语言处理中的基本问题之一在机器翻译、文本分类、信息检索、语音识别、语义网络构建等方面都具有重要意义。
- **词语语义相似度计算**在信息检索、信息抽取、词义排歧、机器翻译、句法分析等处理中有很重要的作用。

词汇级语义分析-内容提要

12. 1. 1 词义消歧

12. 1. 2 词语相似度

12.1.1 词义消歧

词义消歧(WordSense Disambiguation, WSD):

自然语言中一个词具有多种含义的现象非常普遍。如何自动获悉某个词的多种含义，或已知某个词有多种含义，如何根据上下文确认其含义，是词义消歧词研究的内容。

如：英文： bank: 银行/ 河岸
plant: 工厂/ 植物

汉语：

- | | |
|-------------------------|------------------------|
| (1) 他 打 鼓很在行。 | (9) 她会用毛线 打 毛衣。 |
| (2) 他会 打 家具。 | (10) 他用尺子 打 个格。 |
| (3) 他把碗 打 碎了。 | (11) 他 打 开了箱子盖。 |
| (4) 他在学校 打 架了。 | (12) 她 打 着伞走了。 |
| (5) 他很会与人 打 交道。 | (13) 他 打 来了电话。 |
| (6) 他用土 打 了一堵墙。 | (14) 他 打 了两瓶水。 |
| (7) 用面 打 浆糊贴对联。 | (15) 他想 打 车票回家。 |
| (8) 他 打 铺盖卷儿走人了。 | (16) 他以 打 鱼为生。 |

.....

12.1.1 词义消歧

Marrian-Webster袖珍词典		《现代汉语通用字典》	
词形	义项数	词形	义项数
go	63	打	26
fall	35	上	20
run	35	下	19
turn	31	干	19
way	31	子	18
work	31	着	18
do	30	生	18
draw	30	和	18
play	29	点	18
get	26	折	17

12.1.1 词义消歧

词义消歧的任务

输入：他善与外界 **打** 交道



- | | |
|-------------------------|------------------------|
| (1) 他 打 鼓很在行。 | (9) 她会用毛线 打 毛衣。 |
| (2) 他会 打 家具。 | (10) 他用尺子 打 个格。 |
| (3) 他把碗 打 碎了。 | (11) 他 打 开了箱子盖。 |
| (4) 他在学校 打 架了。 | (12) 她 打 着伞走了。 |
| (5) 他很会与人 打 交道。 | (13) 他 打 来了电话。 |
| (6) 他用土 打 了一堵墙。 | (14) 他 打 了两瓶水。 |
| (7) 用面 打 浆糊贴对联。 | (15) 他想 打 车票回家。 |
| (8) 他 打 铺盖卷儿走人了。 | (16) 他以 打 鱼为生。 |



方法：

根据词语的上
下文进行判断

输出： 确认输入句子中 “**打**” 字的义项（义项5）

12.1.1 词义消歧

词义消歧基本方法

语义消歧的方法大致分为四类：

- 基于背景知识的语义消歧 词典/规则方法
 - 监督的语义消歧方法
 - 半监督的学习方法
 - 无监督的学习方法
- 机器学习方法
-

12.1.1 词义消歧

- 基于背景知识的语义消歧

基本思想：词典中词条本身的定义作为判断其语义的条件。

如，cone 在词典中有两个定义：一个是指“松树的球果”，另一个是指“用于盛放其他东西的锥形物，比如，盛放冰激凌的锥形薄饼”。如果在文本中，“树(tree)”或者“冰(ice)”与cone出现在相同的上下文中，那么，cone的语义就可以确定了，tree对应cone的语义1，ice 对应cone的语义2。

12.1.1 词义消歧

● 监督的语义消歧方法

监督学习(Supervised Learning)的方法 就是数据的类别在学习之前已经知道。在语义消歧的问题上，就是说每个词所有可能的义项都是已知的。有监督的语义消歧方法通过一个已标注的语料库学习得到一个分类模型。

主要有：

- 基于贝叶斯分类器的词义消歧方法
- 基于最大熵的词义消歧方法
- 基于互信息的消歧方法

12.1.1 词义消歧

- 基于最大熵的消歧方法

基本思路: 每个词表达不同的含意时其上下文（语境）往往不同，即不同的词义对应不同的上下文，因此，可以将词的上下文作为特征信息利用最大熵模型对词的语义进行分类。

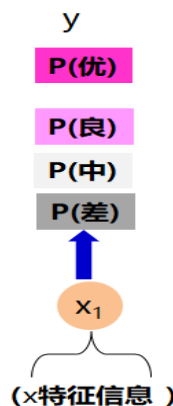
最大熵模型:

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。



关键问题: 特征选取问题

12.1.1 词义消歧

最大熵建模及训练

如： 他/P 很/D 会/V 与/C 人/N 打/V 交道/N 。 /PU

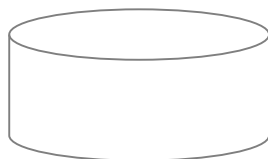


特征选取

上下文特征： ● 词形信息 ● 词性信息 ● 词形 + 词性信息

特征函数/信息： 选取适当的特征函数（如考虑：上下文窗口大小等）

训练语料



最大熵模型

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

12.1.1 词义消歧

最大熵语义消歧过：

打字的所有义项概率

- | | |
|-------------------------|------------------------|
| (1) 他 打 鼓很在行。 | (9) 她会用毛线 打 毛衣。 |
| (2) 他会 打 家具。 | (10) 他用尺子 打 个格。 |
| (3) 他把碗 打 碎了。 | (11) 他 打 开了箱子盖。 |
| (4) 他在学校 打 架了。 | (12) 她 打 着伞走了。 |
| (5) 他很会与人 打 交道。 | (13) 他 打 来了电话。 |
| (6) 他用土 打 了一堵墙。 | (14) 他 打 了两瓶水。 |
| (7) 用面 打 浆糊贴对联。 | (15) 他想 打 车票回家。 |
| (8) 他 打 铺盖卷儿走人了。 | (16) 他以 打 鱼为生。 |

输出：概率最大为结果

最大熵模型

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

输入：句子

他/P 对/D 打/V 鼓/N 很/D 在行/N 。/PU

词汇级语义分析-内容提要

12. 1. 1 语义消歧义

12. 1. 2 词语相似度

12.1.2 词语相似度

基本概念：

※ 词语相似度

词语相似度就是两个词语在不同的上下文中可以互相替换使用而不改变文本的句法语义结构的程度。

如果在不同的上下文中可以互相替换且不改变文本的句法语义结构的可能性越大，二者的相似度就越高，否则相似度就越低。

相似度以一个**数值表示**，一般取值范围在 $[0,1]$ 之间。一个词语与其本身的语义相似度为1。如果两个词语在任何上下文中都不可替换，那么其相似度为0。

相似度涉及到词语的词法、句法、语义甚至语用等方面的特点。其中，对词语相似度影响最大的应该是词的语义。

12.1.2 词语相似度

※ 词语距离

是度量两个词语关系的另一个重要指标；词语距离用一个 $[0, \infty)$ 之间的实数表示。

词语距离与词语相似度的对应关系：

- 1) 两个词语距离为0时，其相似度为1；（一个词语与其本身的距离为0）；
- 2) 两个词语距离为无穷大时，其相似度为0；
- 3) 两个词语的距离越大，其相似度越小（单调下降）

如两个词语 W_1 和 W_2 相似度记为 $\text{Sim}(W_1, W_2)$ ，其词语距离为 $\text{Dis}(W_1, W_2)$ 则有：

$$\text{Sim}(W_1, W_2) = \frac{\alpha}{\text{Dis}(W_1, W_2) + \alpha}$$

其中, α 是一个可调节的参数。 α 的含义是：当相似度为0.5时的词语距离值。

当直接计算词语的相似度比较困难时，可以先计算词语距离，再转换成词语的相似度。

12.1.2 词语相似度

※ 词语相关性

词语**相关性**反映的是**两个词语**互相关联的程度。可以用这两个词语在**同一个语境中共现**的可能性来衡量。

词语相关性和词语相似性是两个不同的概念。例如“医生”和“疾病”两个词语，其相似性非常低，而相关性却很高。**词语相似性**反映的是词语之间的聚合特点，而**词语相关性**反映的是词语之间的组合特点。

词语相关性也以一个**数值表示**，一般取值范围在 $[0,1]$ 之间的实数。

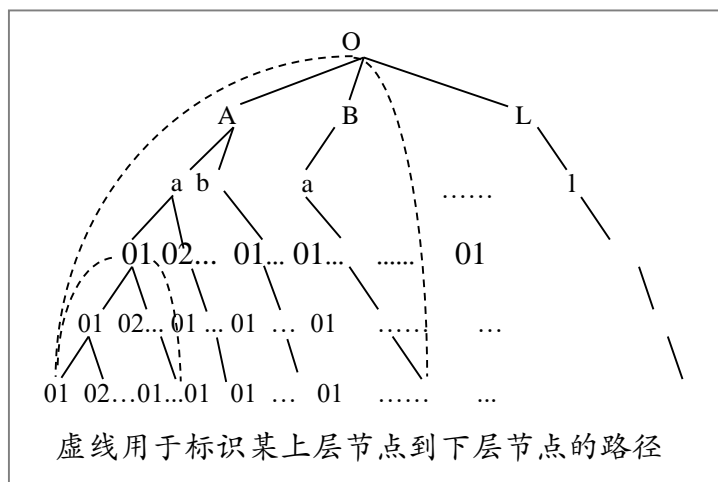
词语相关性和词语相似性又有着密切的联系。如果两个词语非常相似，那么这两个词语与其他词语的相关性也会非常接近。反之，如果两个词语与其他词语的相关性特点很接近，那么这两个词一般相似程度也很高。

12.1.2 词语相似度

词语相似度的计算方法

● 利用词典

利用一部同义词词典 (Thesaurus) 。一般同义词词典都是将所有的词组织在一棵或几棵树状的层次结构中。我们知道, 在一棵树形图中, 任何两个结点之间有且只有一条路径。于是, 这条路径的长度就可以作为这两个概念的语义距离的一种度量。



《同义词词林》语义分类树形图

除了结点间的路径长度外, 还可考虑其他一些因素, 如:

- 1) **概念层次树的深度**: 路径长度相同的两个结点, 如果位于概念层次的越底层, 其语义距离较大; 比如说: “动物”和“植物”、“哺乳动物”和“爬行动物”, 这两对概念间的路径长度都是2, 但前一对词处于语义树的较高层, 因此认为其语义距离较大, 后一对词处于语义树的较低层, 其语义距离更小;
- 2) **概念层次树的区域密度**: 路径长度相同的两个结点, 如果位于概念层次树中高密度区域, 其语义距离应大于位于低密度区域。

12.1.2 词语相似度

- 利用《知网》计算词语语义相似度

词语1	词语2	词语2的语义	语义相似度/方法1	语义相似度/方法2	语义相似度/方法3
男人	女人	人,家,女	1.000	0.668	0.833
男人	父亲	人,家,男	1.000	1.000	1.000
男人	母亲	人,家,女	1.000	0.668	0.833
男人	和尚	人,宗教,男	1.000	0.668	0.833
男人	经理	人,#职位,官,商	1.000	0.351	0.657
男人	高兴	属性值,境况,福,良	0.016	0.024	0.013
男人	收音机	机器,*传播	0.186	0.008	0.164
男人	鲤鱼	鱼	0.347	0.009	0.208
男人	苹果	水果	0.285	0.004	0.166
男人	工作	事务,\$担任	0.186	0.035	0.164
男人	责任	责任	0.016	0.005	0.010

方法1：仅使用《知网》语义表达式中第一独立义原来计算词语相似度

方法2：Li Sujian et al. (2002) 中使用的词语语义相似度计算方法

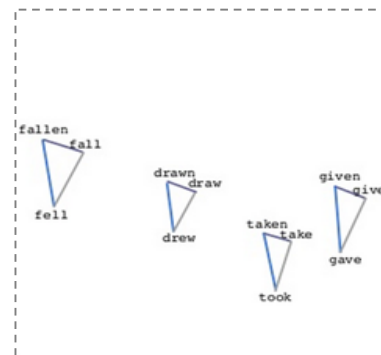
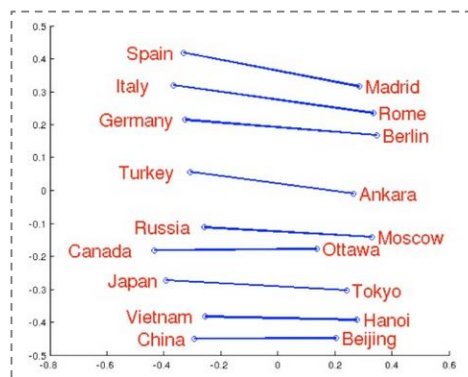
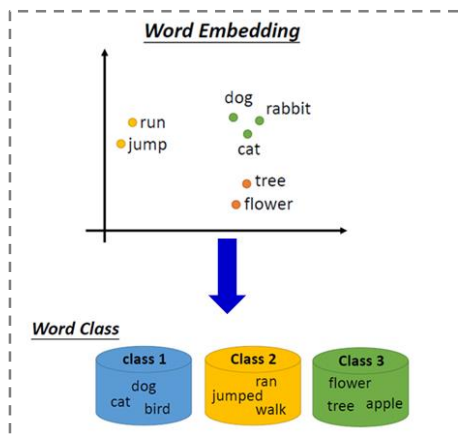
方法2：使用刘群基于《知网》的词汇语义相似度计算

12.1.2 词语相似度

● 神经网络-词向量

基于大规模的语料利用神经网络方法训练**词向量**，根据分布假说：语义相似的词其词向量空间距离更相近。训练出的词向量具有词义相近的词在向量空间聚簇的语言学特性

- 词义相近的词在向量空间聚簇
- 相似关系词对的词向量之差也相似



$$V(\text{king}) - V(\text{queen}) \approx V(\text{uncle}) - V(\text{aunt})$$

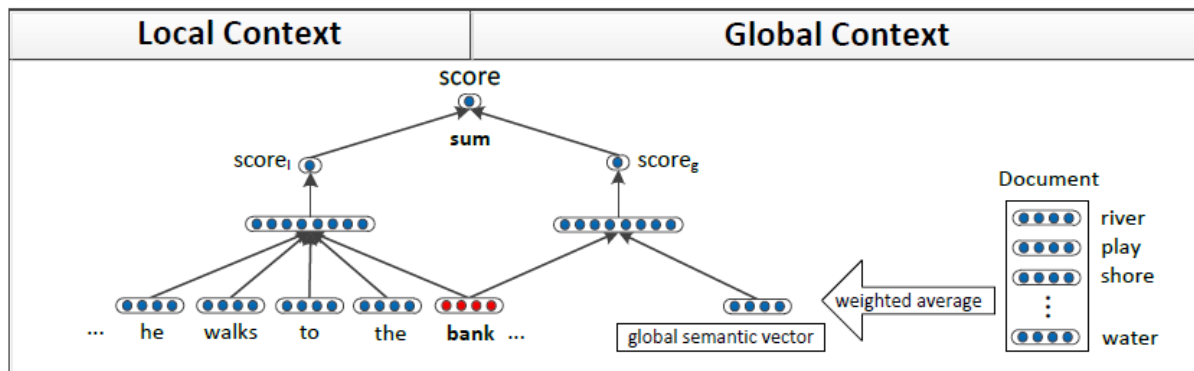
$$V(\text{hotter}) - V(\text{hot}) \approx V(\text{bigger}) - V(\text{big})$$

12.1.2 词语相似度

Huang语义强化模型

该工作在 C&W 的基础上改进而成。主要创新：1. 使用全文信息辅助已有的局部信息。2. 使用多个词向量来表示多义词。

■ 模型结构：

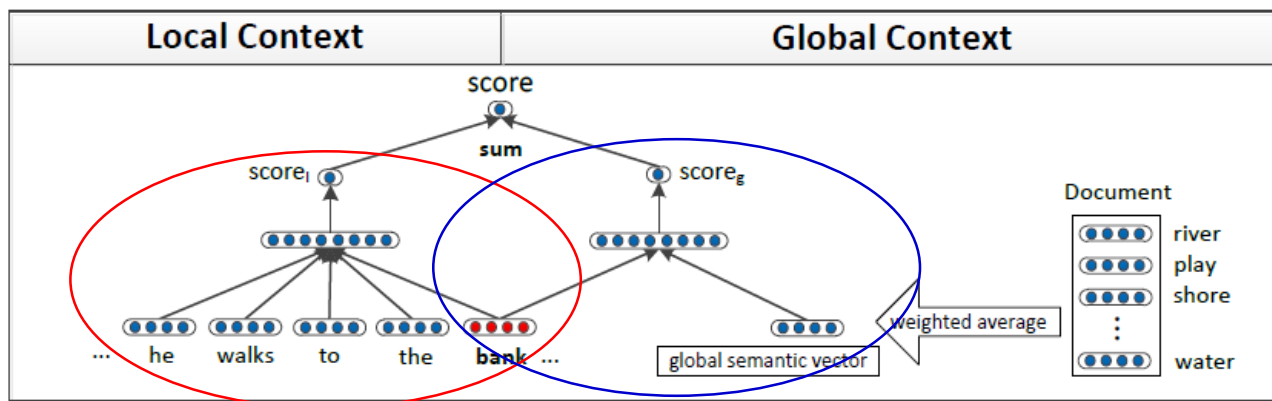


输入： 词序列S和出现S的文档d（s为局部上下文，d为全局上下文）

输出： 局部上下文和全局上下文联合打分

12.1.2 词语相似度

1. 全文信息辅助已有局部信息



Local Context

Local输入: 当前词及上文的m个词向量拼接

Local输出: 当前片段打分

$$a_1 = f(W_1[x_1; x_2; \dots; x_m] + b_1)$$

$$score_l = W_2 a_1 + b_2$$

局部score保留了文档中词的顺序以及句法信息

输出: $g(s, d) = score_l + score_g$

Global Context

Global输入: 当前词及全文信息的拼接

$$\text{全文信息 } c = \frac{\sum_{i=1}^k w(t_i) d_i}{\sum_{i=1}^k w(t_i)}$$

可有多
种形式

Global输出: 当前词和全文信息的打分

$$a_1^{(g)} = f(W_1^{(g)}[c; x_m] + b_1^{(g)})$$

$$score_g = W_2^{(g)} a_1^{(g)} + b_2^{(g)}$$

全局score保存了文档的语义和话题信息

12.1.2 词语相似度

■ 模型训练:

- **优化目标:** : 对于整个语料最小化:

$$C_{s,d} = \sum_{w \in V} \max(0, 1 - g(s, d) + g(s^w, d))$$

其中, s 是给定的词序列, d 是包含 s 的文档

- 正样本 $g(s,d)$ 来自语料
- 而负样本 $g(S^w,d)$ 则是将 S 最后一个词替换为其他词的序列 S^w

- **参数训练:**

采用pairwise的方式对文本片段进行优化, 即可得词向量

12.1.2 词语相似度

■ 单原型神经网络语言模型实验结果

Center Word	C&W	Our Model
markets	firms, industries, stores	market, firms, businesses
American	Australian, Indian, Italian	U.S., Canadian, African
illegal	alleged, overseas, banned	harmful, prohibited, convicted

Nearest neighbors of words based on cosine similarity. Huang model is less constrained by syntax and is more semantic.

12.1.2 词语相似度

2. 多原型神经网络语言模型 (multi-prototype neural language model)

该神经网络是为了解决一词多义的问题，使用不同的表示形式来描述某个词的多组意义。

在前面训练得到的词向量基础上：

- 针对每个词出现的位置设定一个固定大小的窗口（前后各5），得到一个短句，对窗口中的10 个词求其平均权重
- 对所有得到的上下文向量做 k-means 聚类，根据聚类结果给每个语料中的词打上标签（不同类中的同一个词，当作不同的词处理）
- 最后重新训练类别中词向量。

12.1.2 词语相似度

多原型神经网络语言模型中两个词之间的距离可以定义如下：

$$\text{AvgSimC}(w, w') = \frac{1}{K^2} \sum_{i=1}^k \sum_{j=1}^k p(c, w, i) p(c', w', j) d(\mu_i(w), \mu_j(w'))$$

其中 $p(c, w, i)$ 为词 w 在给定上下文 c 的情况下属于类别 i 的概率 $\mu_i(w)$ 表示第 i 个类别中心点 w ， $d(v, v')$ 为两个词之间距离计算函数。

■ 多原型神经网络语言模型试验结果

Center Word	Nearest Neighbors
bank_1	corporation, insurance, company
bank_2	shore, coast, direction
star_1	movie, film, radio
star_2	galaxy, planet, moon
cell_1	telephone, smart, phone
cell_2	pathology, molecular, physiology
left_1	close, leave, live
left_2	top, round, right

Nearest neighbors of word embeddings learned by Huang model using the multi-prototype approach based on cosine similarity. The clustering is able to find the different meanings, usages, and parts of speech of the words.

内 容 提 要

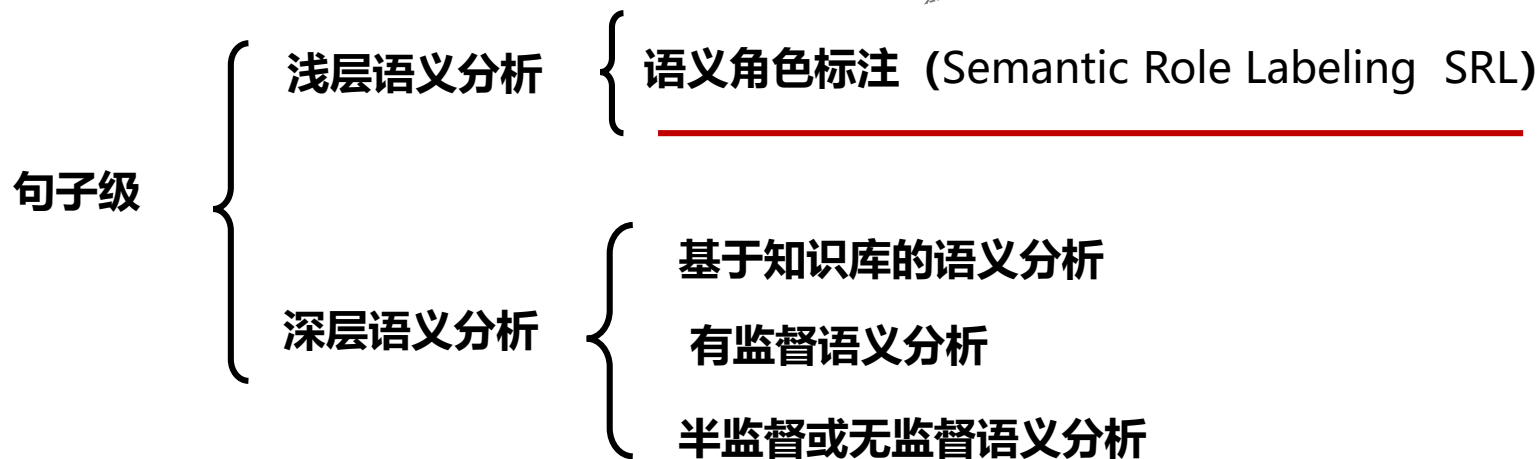
第一部分：词汇级语义分析

第二部分：句子级语义分析

句子级语义分析概述

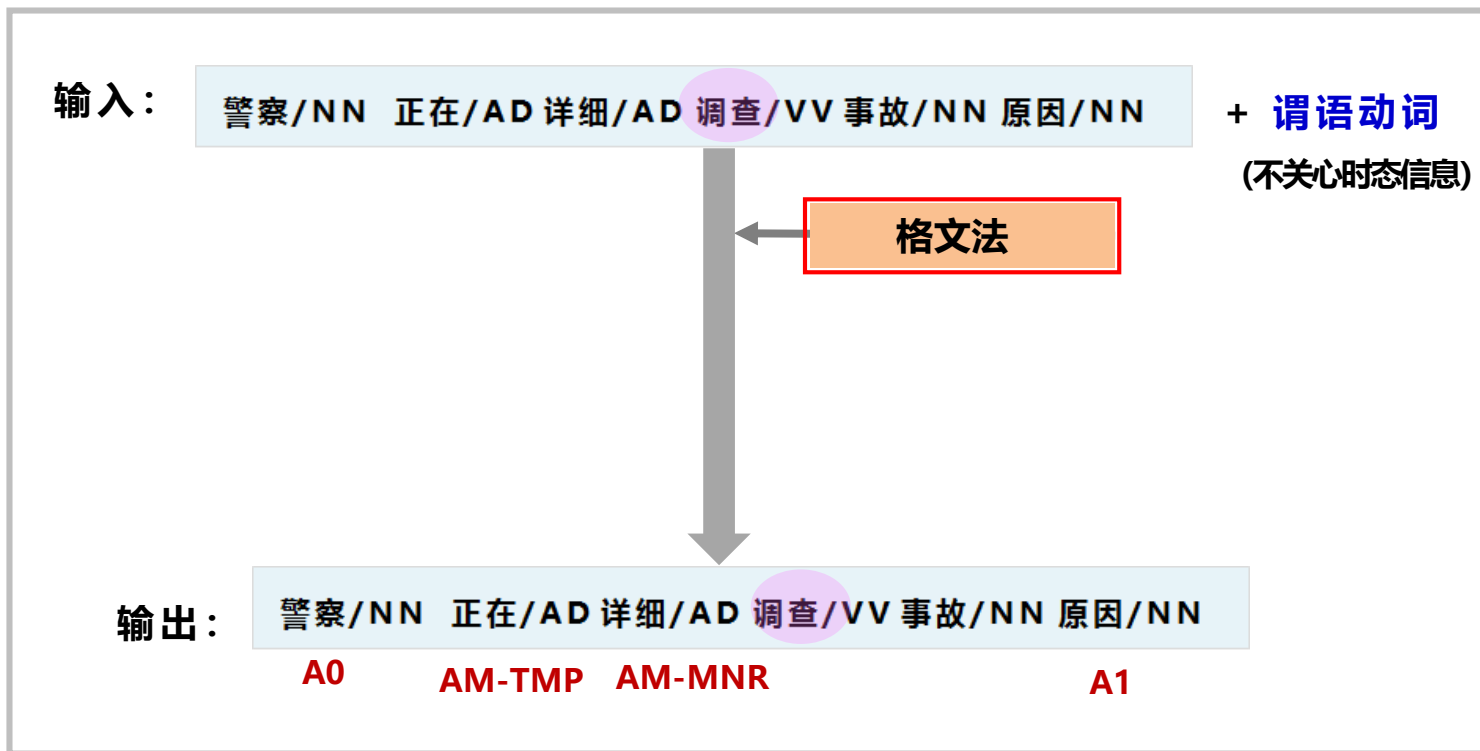
句子级语义分析内容：

主要围绕着句子中的**谓词**来分析各成分与其之间的结构关系，并用语义角色来描述这些结构关系



语义概角色标注

语义角色标注 (SRL) 任务



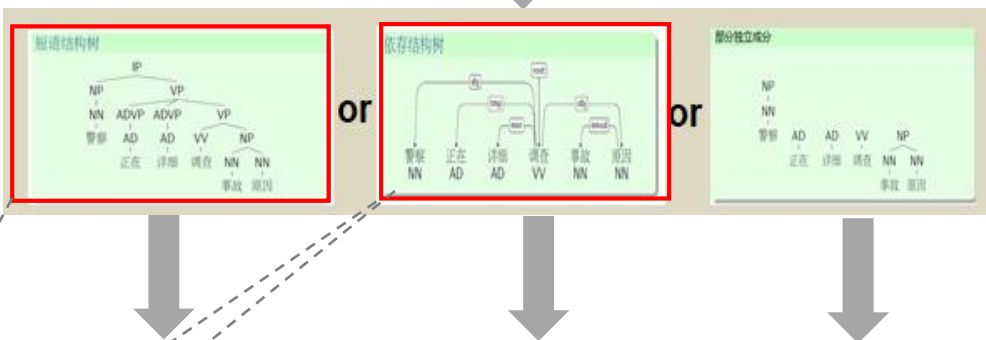
语义角色标注概述

语义角色标注 (SRL) 任务实现

输入：

警察/NN 正在/AD 详细/AD 调查/VV 事故/NN 原因/NN

+ 谓动词
(不关心时态信息)



输出：

警察/NN 正在/AD 详细/AD 调查/VV 事故/NN 原因/NN

A0

AM-TMP AM-MNR

A1

概率统计

深度学习

技术方法

句子级语义分析-内容提要

12. 2. 1 格文法

12. 2. 2 语义角色标注

12. 2. 3 语义角色标注评价

12.2.1 格文法

格语法(case grammar)



C. J. Fillmore

格语法(case grammar)是美国语言学家C. J. Fillmore于1966年提出的、从语义的角度出发,即从句子的深层结构来研究句子的结构。着重探讨句法结构与语义之间关系的文法理论。

12.2.1 格文法

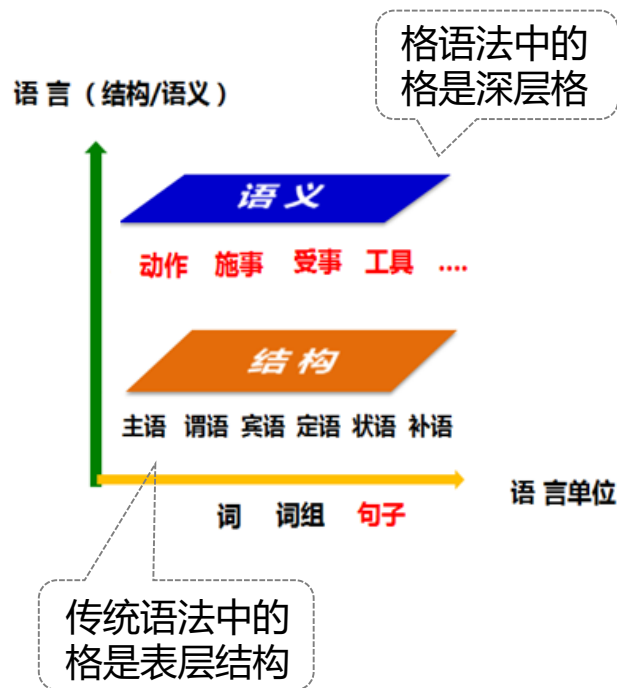
语言（结构/语义）



基本思想

Fillmore指出：诸如主语、宾语等语法关系实际上都是表层结构上的概念，在语言的底层，所需要的不是这些表层的语法关系，而是用施事、受事、工具、受益等概念所表示的句法语义关系。这些句法语义关系，经过各种变换之后才在表层结构中成为主语或宾语。

12.2.1 格文法

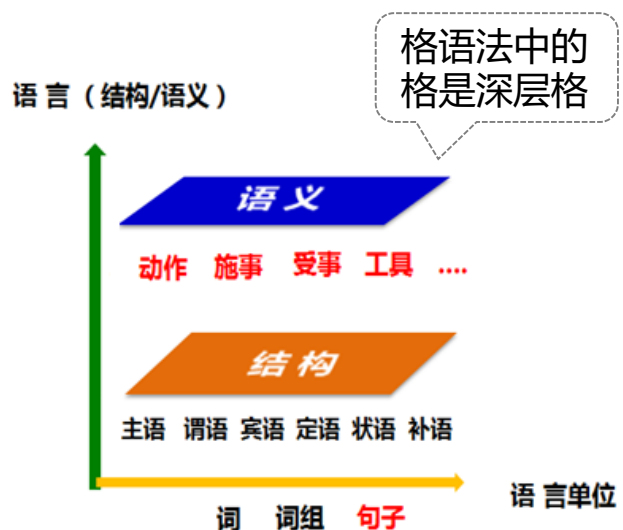


格的含义

传统语法中，“格”是指某些曲折语中特有的用于表示词语间语法关系的名词和代词的形态变化，如主格，宾格等。俄语有6个格，德语有4个格，（汉语名词和代词没有形态变化，所以没有格）。

格语法中的格是“深层格”是指句子中的体词（名词、代词等）和谓词（动词、形容词等）之间的及物性关系。如，动作和施事者的关系、动作和受事者的关系等，这些关系是语义的，是一切语言中普遍存在的。

12.2.1 格文法



这种格是在底层结构中依据**名词与动词之间**的**句法语义关系**来确定的，一经确定就固定不变，不管表层句法结构如何变化，底层格语法不变；底层的“格”与任何具体语言中的表层结构上的语法概念，如主语，宾语等，没有对应关系。

12.2.1 格语法

例如：

- a. John broke the window.
- b. The hammer broke the window.
- c. The window broke.
- d. John broke the window with the hammer.
- e. The window was broken by John.

broke : 主动词

John: 施事格

the window: 受事格 (客体格)

the hammer: 工具格

句子语义本质

表层结构

a. John broke the window.

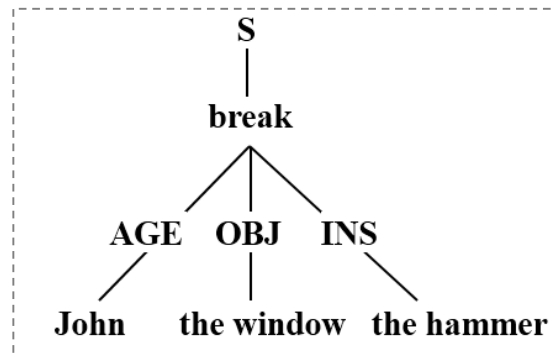
主 谓 宾

b. The hammer broke the window.

主 谓 宾

.....

底层语义



12.2.1 格语法

所有的语言都有表示深层结构的语义格，这些格是由NP 在句子中所起的作用来决定的：

Fillmore最初将命题中的格定义为6种：

- **施事格**(A=Agentive)：由动词确定的动作的有生命的动作发出者，
如：Mary cried.
- **工具格**(I=Instrumental)：对由动词确定的动作或状态而言作为某种因素而牵涉到的无生命的力量或客体，如：Mary cut the cake with a knife.
- **承受格**(D=Dative)：由动词确定的动作或状态所影响的有生物，
如：Mary is tall.

12.2.1 格语法

- **使成格**(F=Factitive): 由动词确定的动作或状态所形成的客体或有生物, 或者理解为动词意义的一部分的客体或有生物,
如: John dreamed **a dream** about Mary.
- **处所格**(L=Locative): 由动词确定的动作或状态的处所或空间方向,
如: **Chicago** is windy.
- **客体格**(O=Objective): 由动词确定的动作或状态所影响的事物,
如: Mary cut **the cake**

统一确定有多少格较困难, 实际应用中需根据具体情况确定。

12.2.1 格文法

后来又增加了一些：

- **受益格**(B=Benefactive)：由动词所确定的动作为之服务的有生命的对象，如：
John sang a song for **Mary**.
- **源点格**(S=Source)：由动词确定的动作所作用到的事物的来源或发生位置变化过程中的起始位置，如：John bought a book from **Mary**.
- **终点格**(G=Goal)：由动词确定的动作所作用到的事物的终点或发生位置变化过程中的终端位置，如：John sold a book to **Mary**.
- **伴随格**(C=Comitative)：由动词确定的与施事者共同完成动作的伴随者，如：
John sang a song with **Mary**.

各种格关系的形式标志是介词或语序。

如. 我 去年 在北京 吃了 一顿烤鸭

施事 时间 处所 动作 受事

12.2.1 格文法

格语法组成

格语法有三部分组成：**基本规则**，**词汇部分** 和 **转换部分**

1. 基本规则

(1) $S \rightarrow M + P$

表示：句子S可以改写成情态(Modality)和命题(Proposition)两大部分，情态部分包括否定、时、式、体以及其它被理解为全句情态成分的状态语。命题涉及动词和依附于它的名词短语间的关系等。动词是句子的中心，名词短语按其特定的格属关系依附于该动词。

12.2.1 格文法

(2) $P \rightarrow V + C_1 + C_2 + \dots + C_n$

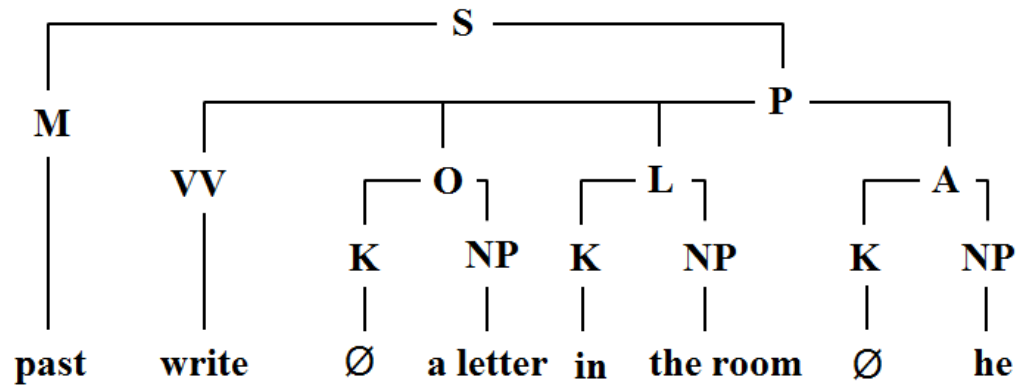
表示：命题P都可以改写成一个动词V和若干个格C。动词是广义上的动词，包括动词、形容词，甚至名词、副词和连词；格不一定要NP担任，某些内嵌小句也可以。

(3) $C \rightarrow K + NP$

表示：K为格标，是各种格范畴在底层结构中的标记。句子中的格范畴有的是独立地进入句子，有时要用介词等引进。这些介词即为格标，此外后缀词、零形式等也可以充当格标。通常，格范畴和它的格标被看作一个整体(即一个格)来看待的。

12.2.1 格文法

例：He wrote a letter in the room.



$S \rightarrow M+P$

$P \rightarrow V+C1+C2+\dots Cn$

$C \rightarrow K+NP$

12.2.1 格文法

2. 词汇部分

(1) 词库

在词库中除了要标明每一个词条在句法、语义和语音面的特征外，还需标明他们的底层格的特征。

(2) 词汇插入

格语法中词汇的插入问题主要是名词和动词的选择问题。

12.2.1 格文法

3. 转换部分

转换部分需要处理问题：底层结构中K成分的处理，即确定主语成分，解释词序以及特定成分的插入等。实际是同一个底层的K结构转换成表层结构上的显性表示，不同的语言采用不同的转换操作。

- Fillmore：每一个在深层结构中的格一般都与某个或几个介词相连接，经过一定的转型才形成表层结构。
- 在表层结构中，有的格保存了与之相连的介词，但有的却隐去了介词。
- 需注意的是：无论是哪一格，当其处于表层结构的主语或宾语位置时，与其相连的介词必须隐去。

12.2.1 格文法

格框架

分析的结果可用**格框架**表示。一个**格框架**由一个主要概念和一组辅助概念组成，这些辅助概念以一种适当定义的方式与主要概念相联系。在实际应用中，主要概念可以是动词，辅助概念为施事格、方位格、工具格等语义深层格。

[主动词

 [case-frame:

 [Agentive:

 Objective:

 Instrumental:

 Locative:

 ]

 [MODALs:

 time:

 voice:]]

把格框架中的格映射到输入句子中找到的短语上，识别一句话所表达的含义，即要弄清楚“**干了什么**”、“**谁干的**”、“**行为结果是什么**”以及发生行为的**时间**、**地点**和所用**工具**等。

12.2.1 格文法

使用格语法进行语义分析

例：In the room, John broke a window with a hammer.

[BREAK

[case-frame:

[Agentive: JOHN

Instrumental: HAMMER

Locative: ROOM]

[MODALs:

time: past

voice: active]]

12.2.1 格文法

评价

格文法理论自提出后，在语言界引起过强烈的反响。虽然70年代末以后该理论没有再继续发展，但它的思想已深入到其它语言理论中，成为计算机语言信息处理学科中的一种基本理论。

但格文法对描述汉语的流水句、无动句、连动、省略等结构存在一定的局限性。

句子级语义分析-内容提要

12.2.1 格文法

12.2.2 语义角色标注

12.2.2.1 语料资源

12.2.2.2 语义角色标注方法

1. 基于完全句法分析的统计SRL方法
2. 基于依存句法分析的统计SRL方法
3. 基于依存句法分析的神经网络SRL方法

12.2.3 语义角色标注评价

12.2.2.1 语料资源

用于语义角色标注 (SRL) 研究的语料资源:

- 英语命题库(PropBank)和英语名词命题库(NomBank)
- 中文 PropBank (CPB) 和 中文 NomBank 。
- 中文框架网(FrameNet)

12.2.2.1 语料资源

★ 命题库将宾州树库(Treebank) 中的《华尔街日报》语料和一部分布朗 语料(Brown Corpus) 进行了人工的语义角色标注。

语义角色	角色描述
ARG0	施事者
ARG1	受事者
ARG2	范围或程度
ARG3	动作起点
ARG4	动作结束点
ARGM-ADV	状语
ARGM-BNF	受益者
ARGM-CND	条件
ARGM-DIR	方向
ARGM-LOC	地点
ARGM-MNR	方式
ARM-PRP	目的
ARGM-TMP	时间
ARGM-TPC	主题
ARGM-PRD	次谓词

两类 语义角色:

1. 与 具体谓词直接相关的，这些角色用ARG0，ARG1，…，ARG5表示，如ARG0 通常表示动作的施事，ARG1 通常表示动作的影响等，ARG2-ARG5 对于不同的谓语动词会有不同的语义含义；

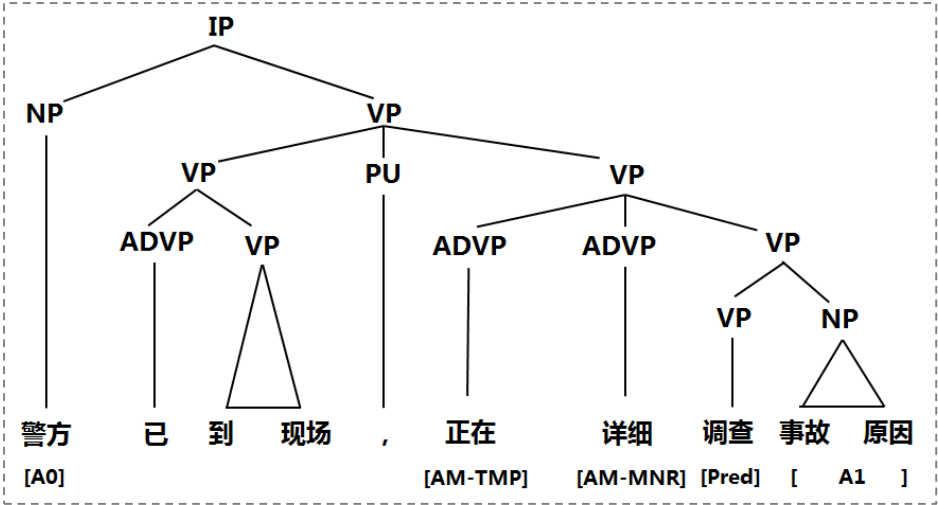
2. 起 修饰作用的辅助性角色，其角色标签都以ARGM 开头，常见的有表示时间的角色ARGM-TMP，表示地理位置的角色ARGM-LOC，表示一般性修饰成分的角色ARGM-ADV 等。

12.2.2.1 语料资源

★中文语义角色标注的语料库主要有Chinese PropBank (CPB) 和 Chinese NomBank。它们都是在中文树库（Chinese Treebank）的句法成分中加入了人工标注的语义角色信息。

语义角色	角色描述
ARG0	施事者
ARG1	受事者
ARG2	范围或程度
ARG3	动作起点
ARG4	动作结束点
ARGM-ADV	状语
ARGM-BNF	受益者
ARGM-CND	条件
ARGM-DIR	方向
ARGM-LOC	地点
ARGM-MNR	方式
ARM-PRP	目的
ARGM-TMP	时间
ARGM-TPC	主题
ARGM-PRD	次谓词

如，“警察已到现场，正在详细调查事故原因”



句子级语义分析-内容提要

12.2.1 格文法

12.2.2 语义角色标注

12.2.2.1 语料资源

12.2.2.2 语义角色标注方法

1. 基于完全句法分析的统计SRL方法
2. 基于依存句法分析的统计SRL方法
3. 基于依存句法分析的神经网络SRL方法

12.2.3 语义角色标注评价

12.2.2.2 语义角色标注方法

语义角色标注的基本方法

语义角色标注需要依赖句法分析的结果进行，因为句法分析包括完全句法分析、局部句法分析、依存关系分析，所以语义角色标注也分为：

- 基于完全句法分析的语义角色标注方法
- 基于局部句法分析的语义角色标注方法
- 基于依存句法分析的语义角色标注方法

但无论哪种方法，过程都是：

句法分析 --> 候选论元剪除--> 论元识别--> 论元标注--> 后处理（标注结果）

完全句法分析
局部句法分析
依存句法分析

尽量减少候
选项的数目

识别真正的论元
两类分类问题

标注过程：
多类分类问题

一些强制的一致性限制、删除重复论元等

- 与论元剪除合并
- 单独一步
- 与论元标注合并

12.2.2.2 语义角色标注方法

可以看出，语义角色标注的核心就是**对论元进行语义标签的分类**（方法：根据当前词与中心词之间的一系列特征,来判断当前词属于中心词的哪一类角色,不属于任何角色的词被赋予空类别）语义角色标注的不同方法，其实就是针对这个问题设计不同的更加高效的分类器。

句法分析是基础，在统计方法中后续步骤常常会构造的一些人工特征，这些特征往往也来自句法分析；在神经网络方法中可以用神经网络自动提取特征来实现分类（标注）任务。

可用最大熵、SVM等模型来完成分类（标注）任务

句子级语义分析-内容提要

12. 2. 1 格文法

12. 2. 2 语义角色标注

12. 2. 2. 1 语料资源

12. 2. 2. 2 语义角色标注方法

1. 基于完全句法分析的统计SRL方法

2. 基于依存句法分析的统计SRL方法

3. 基于依存句法分析的神经网络SRL方法

12. 2. 3 语义角色标注评价

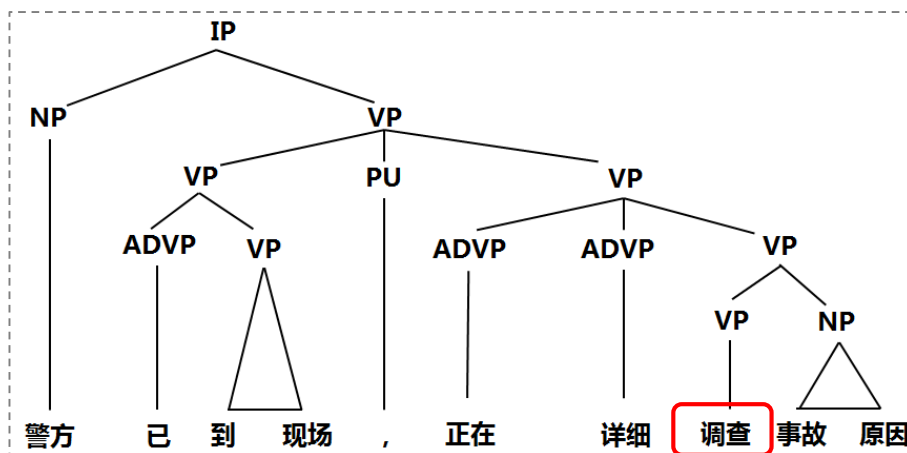
1.基于完全句法分析的统计SRL方法

★ 基于完全句法分析语义角色标注:

任务:

如，警察已到现场，正在详细调查事故原因。

输入:



输出:

A0

AM-TMP AM-MNR

A1

语义角色	角色描述
ARG0	施事者
ARG1	受事者
ARG2	范围或程度
ARG3	动作起点
ARG4	动作结束点
ARGM-ADV	状语
ARGM-BNF	受益者
ARGM-CND	条件
ARGM-DIR	方向
ARGM-LOC	地点
ARGM-MNR	方式
ARGM-PRP	目的
ARGM-TMP	时间
ARGM-TPC	主题
ARGM-PRD	次谓词

语义角色标注步骤:

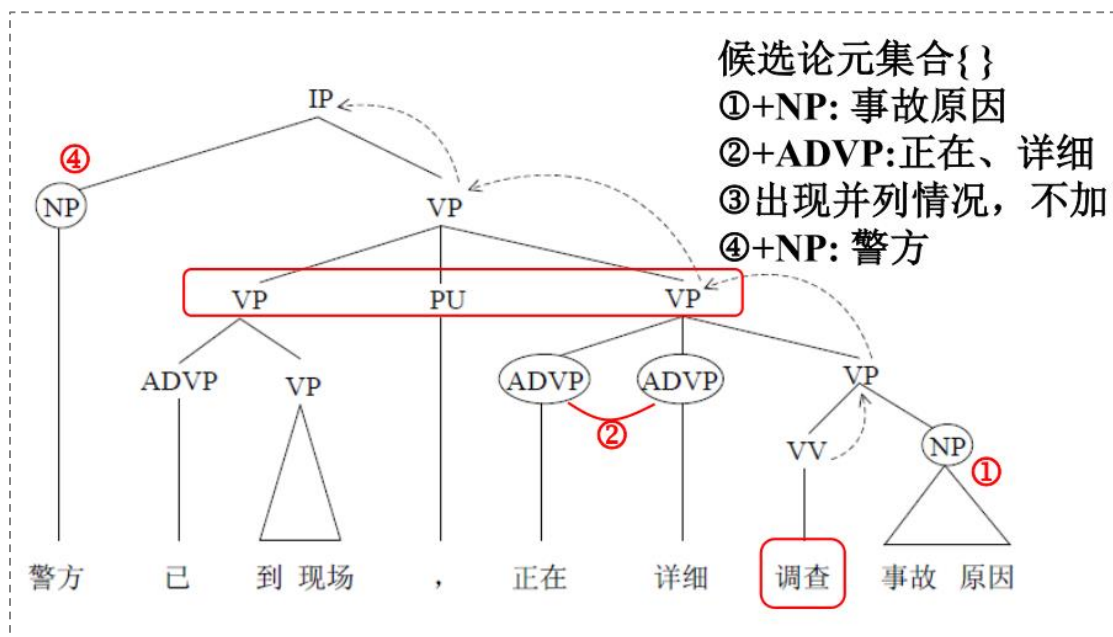


1.基于完全句法分析的统计SRL方法

□ 候选论元剪除（规则）：

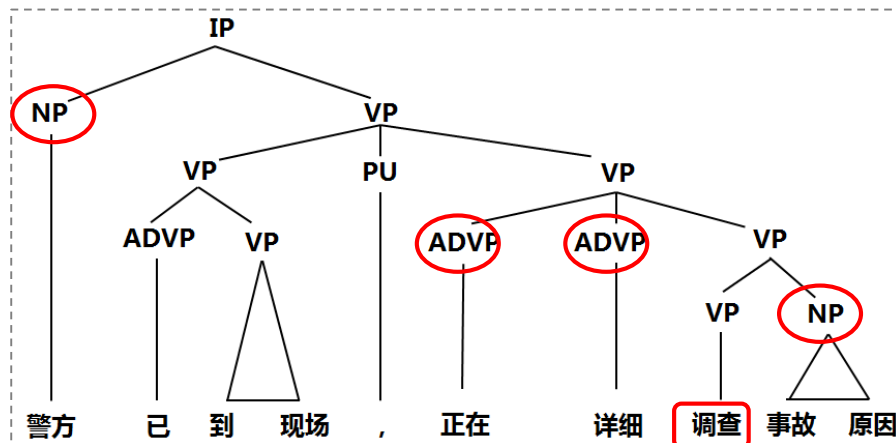
第1步：将谓词作为当前节点，依次考察它的兄弟节点：如果一个兄弟节点和当前节点在句法结构上不是并列的(coordinated)关系，则将它作为候选项。如果该兄弟节点的句法标签是PP，在将它的所有子节点也都作为候选项。

第2步：将当前节点的父节点设为当前节点，重复第1步的操作，直至当前节点是句法树的根节点。



1.基于完全句法分析的统计SRL方法

□ 论元识别： 可选其中一种：（1）与论元剪除合并 （2）**单独一步** （3）与论元标注合并



基本思路：

将论元识别作为一个序列标注问题来解决（二分类）

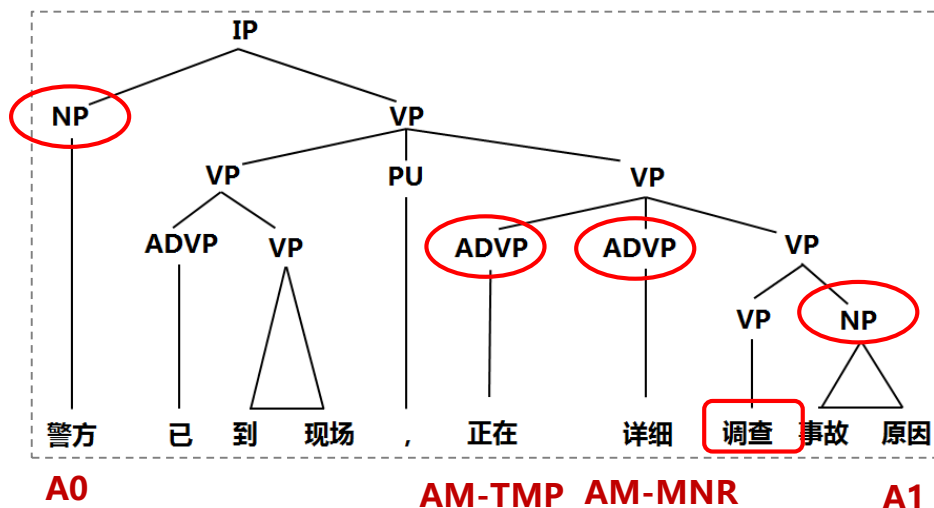
标签： 1：是论元；0：非论元

分类器：最大熵、CRF、SVM、感知机等

特征问题：见特征选取

1.基于完全句法分析的统计SRL方法

□ 论元标注:



基本思路:

将语义角色标注作为一个标注问题来解决

谓词论元关系可以表示为谓词与论元的中心词之间的关系。

分类器: 最大熵、CRF、SVM、感知机等

特征问题: 见特征选取

由于论元识别和论元标注是不同的问题, 可以用不同的特征

1.基于完全句法分析的统计SRL方法

特征选取:

常用的基本特征, 包括谓词、路径、短语类型、位置、语态、中心词、动词子类框架等能从不同的侧面反映待标注单元语义角色的信息. (对于论元识别和论元标注两个任务往往使用不同的特征)

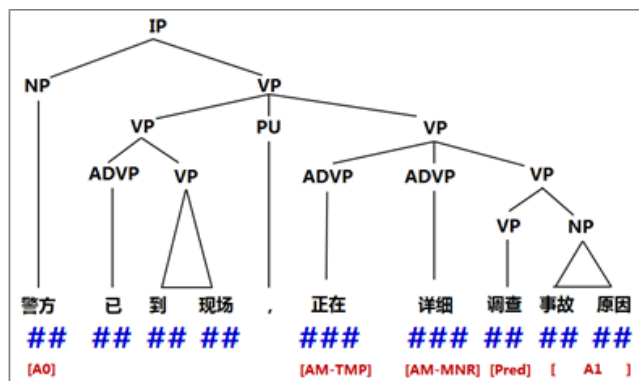
句法成分相关特征	
短语类型	
中心词词干	Collins 在其博士论文 ^[14] 中描述了识别一个成分中的中心词的规则.而对于介词短语,我们可以使用短语中的最后一个名词来代替传统的中心词.最后采用基于规则的方法 ^[19] 对一个成分中的中心词提取词干.
最后一个词的词干	一个成分中最后一个词的词干
最后一个词的词性	
命名实体	如果一个成分以命名实体结尾,则把命名实体的类型作为特征,类型包括 3 种:地点(LOC)、组织(ORG)、人物(PER)
中心词的词性	
句法成分前一个词	
句法成分后一个词	
句法成分前面第 2 个词	
句法成分后面第 2 个词	
谓词相关特征	
谓词原形	
语态	我们用规则识别谓词是主动语态还是被动语态,即谓词的词性为 VBN,并且前面有 AUX 则为被动语态.
子类框架	谓词所在 VP 的子类框架,如在图 2 中,子类框架为 VP->VBD, NP, PP
谓词的词性	
谓词的后缀	我们使用谓词的最后 3 个字母
谓词-句法成分关系特征	
路径	句法树中,从句法成分到谓词之间的句法路径.如在图 2 中,She 所在 NP 到谓词的路径为 NP↑S↓VP↓VBD.
位置	句法成分和谓词之间的位置关系,有两个取值:“前”和“后”.对于“覆盖”的情况,我们采用启发式规则直接将其忽略,因为这种情况下,句法成分不可能成为谓词的角色.
路径长度	一个成分和它的谓词之间的长度
部分路径	从一个成分到这个成分与它的谓词的最近共同祖先之间的路径
从句层级	在一个成分和它的谓词之间的路径上的从句的数量
谓词到句法成分的路径上 VP 个数	

1.基于完全句法分析的统计SRL方法

模型训练:

● 训练语料处理:

标注好语义角色的完全句法树训练语料



注: ##### 表示特征需要信息

● 模型训练:

分别用不同的特征训练2个模型: 论元识别模型 和 论元标注模型

模型1-论元识别

$$p_1(y|x) = \frac{1}{Z_1(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_1(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

模型2-论元标注

$$p_2(y|x) = \frac{1}{Z_2(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_2(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

模型: 选择最大熵模型

$$p_\lambda(y|x) = \frac{1}{Z_\lambda(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_\lambda(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

1.基于完全句法分析的统计SRL方法

语义角色标注:

例： 我在北京看电影

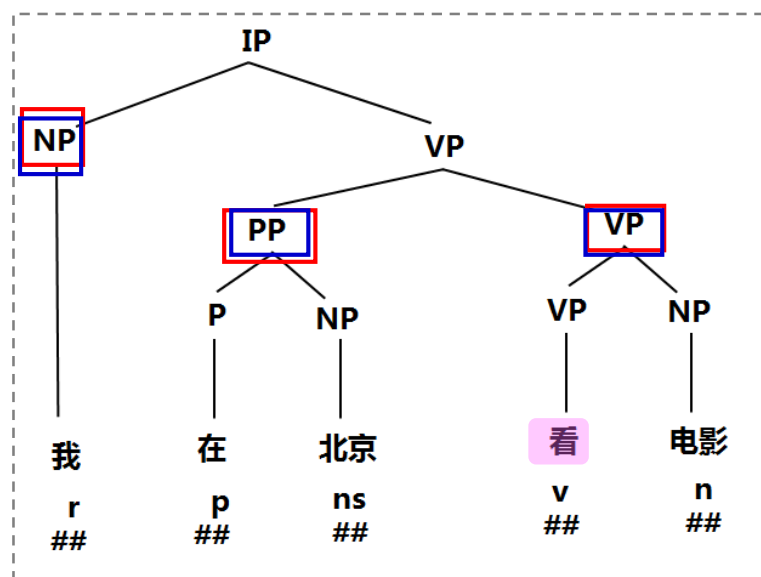
1.完全句法分析

2.候选论元剪除

3.论元识别

4. 论元标注

5. 后处理



B-A0 B-AM-LOC I-AM-LOC B-V B-A1

[A0] [LOC] [A1]

模型1-论元识别

$$p_2(y|x) = \frac{1}{Z_2(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_2(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

模型2-论元标注

$$p_2(y|x) = \frac{1}{Z_2(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_2(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

句子级语义分析-内容提要

12. 2. 1 格文法

12. 2. 2 语义角色标注

12. 2. 2. 1 语料资源

12. 2. 2. 2 语义角色标注方法

1. 基于完全句法分析的统计SRL方法

2. 基于依存句法分析的统计SRL方法

3. 基于依存句法分析的神经网络SRL方法

12. 2. 3 语义角色标注评价

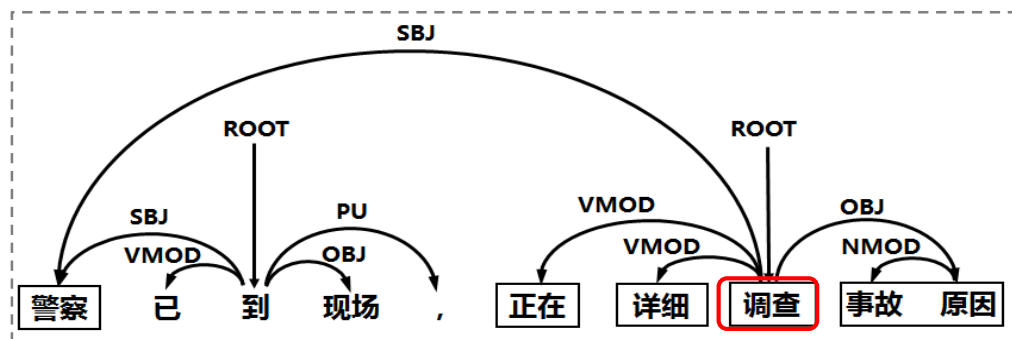
2.基于依存句法分析的统计SRL方法

★ 基于依存分析的统计法语义角色标注：

任务：

如，警察已到现场，正在详细调查事故原因。

输入：



输出：

A0

AM-TMP AM-MNR

A1

语义角色	角色描述
ARG0	施事者
ARG1	受事者
ARG2	范围或程度
ARG3	动作起点
ARG4	动作结束点
ARGM-ADV	状语
ARGM-BNF	受益者
ARGM-CND	条件
ARGM-DIR	方向
ARGM-LOC	地点
ARGM-MNR	方式
ARGM-PRP	目的
ARGM-TMP	时间
ARGM-TPC	主题
ARGM-PRD	次谓词

语义角色标注步骤：

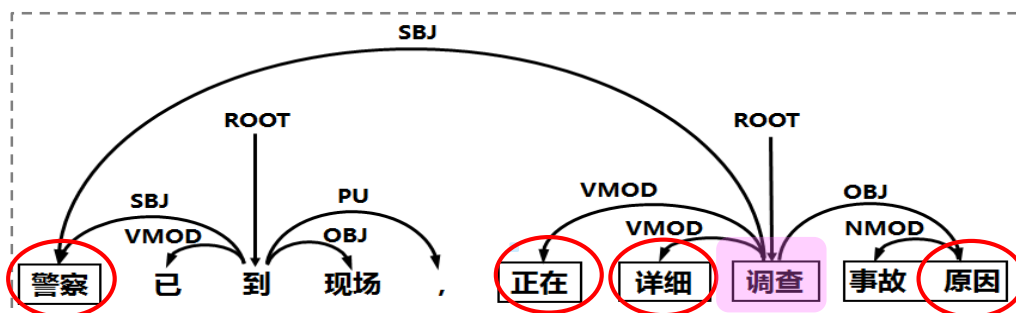


2.基于依存句法分析的统计SRL方法

□ 候选论元剪除（规则）：

第1步：将谓词作为当前节点，将它所有的孩子都作为候选项。

第2步：将当前节点设为它的父节点，重复第1步的操作，直到当前节点是依存句法树的根节点。

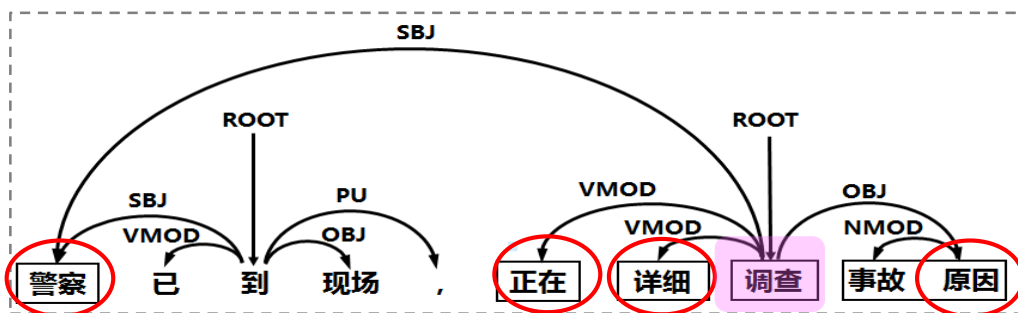


谓词 “调查” 的所有孩子{ 正在，详细，原因，警察 }

都加入到候选项中。这些孩子节点恰好是该谓词的所有论元（论元为词）

2.基于依存句法分析的统计SRL方法

□ 论元识别： 可选其中一种：（1）与论元剪除合并 （2）**单独一步** （3）与论元标注合并



基本思路：

将论元识别作为一个序列标注问题来解决（二分类）

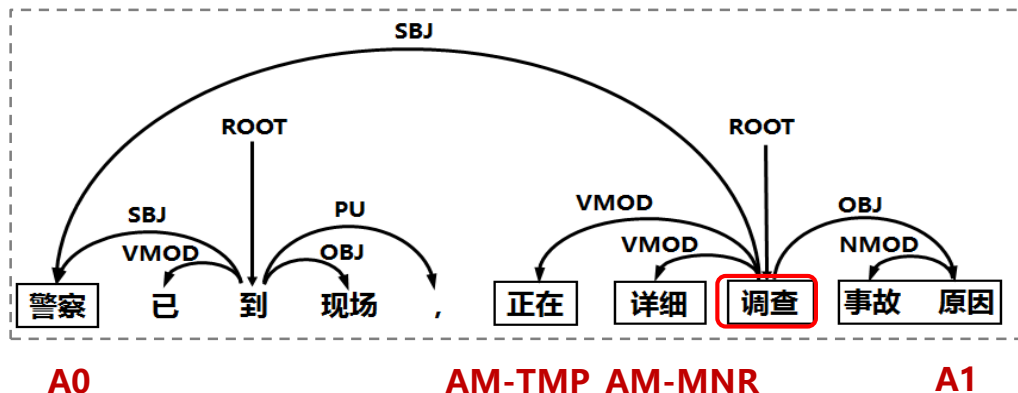
标签： 1：是论元；0：非论元

分类器：最大熵、CRF、SVM、感知机等

特征问题：见特征选取

2.基于依存句法分析的统计SRL方法

□ 论元标注:



基本思路:

将语义角色标注作为一个标注问题来解决

谓词论元关系可以表示为谓词与论元的中心词之间的关系。

分类器: 最大熵、CRF、SVM、感知机等

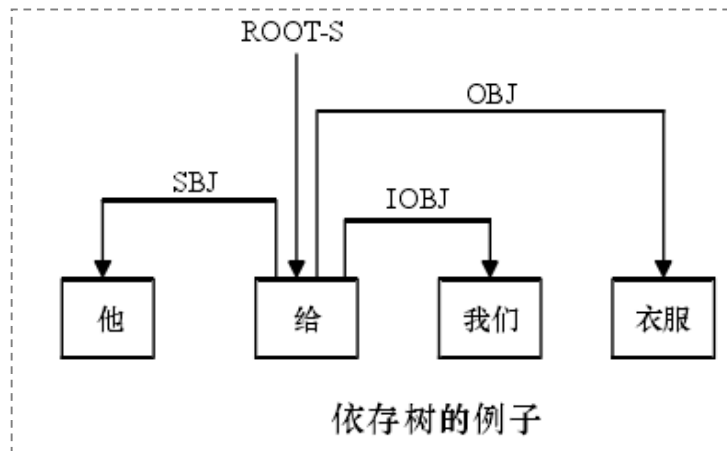
特征问题: 见特征选取

由于论元识别和论元标注是不同的问题，可以用不同的特征

2.基于依存句法分析的统计SRL方法

特征选取:

单一特征：



词: 词语本身

依存关系: 与父节点的依存关系

如“他”的依存关系为“SBJ” SBJ对于“他”的语义分类具有重要的指示作用。

首词及其词性:所有子孙节点中第一个词。

如，“给”的“首词”为“他”

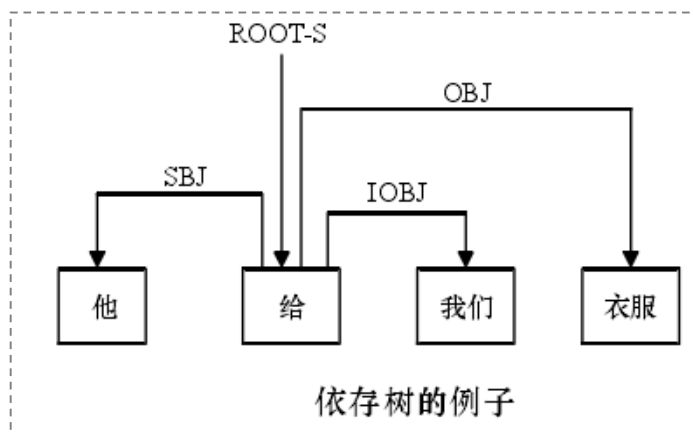
尾词及其词性:所有子孙节点中最后一个词。

如：“给”的“尾词”为“衣服”

词性模式: 在子孙节点中, 除去“首词”、“尾词”后, 剩余的词性标记构成一个“词性集合”, 集合中的元素没有顺序、不可重复。词性模式由 首词、词性集合与 尾词组成

2.基于依存句法分析的统计SRL方法

单一特征:



词性路径, 关系路径: 分别为从谓词到一个词的路径上的词性标记或依存关系构成的序列。方向有“上”、“下”、“左”“右”。相对于位置, 路径携带更多的信息, 如, 从“给”到“他”的关系路径为“SBJ左下”

路径长度:从起始节点到终止节点的路径长度。

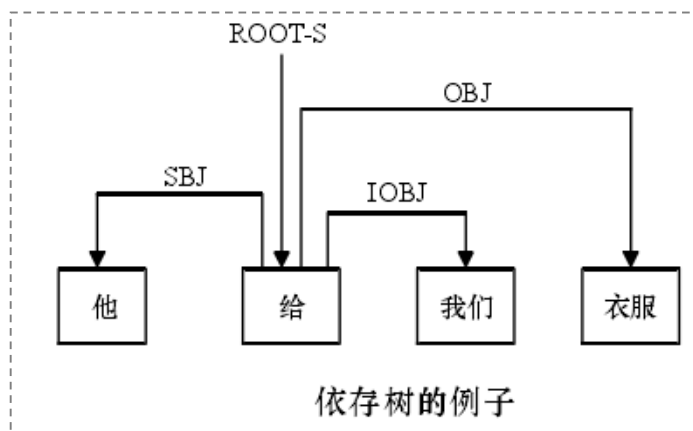
位置: 一个词相对于谓词的位置, 值为“左”或“右”。某一特定语义角色的节点通常出现在特定的位置, 如, 大部分附属角色出现在谓词前面。

谓词:动词在CPB中, 不同的动词具有不同的角色框架定义、不同的句法结构变化。

谓词子节点的词性序列:在测试集中出现相同谓词子节点的概率很小, 所以用词性代替词语, 提高系统的泛化能力。

2.基于依存句法分析的统计SRL方法

单一特征:



谓词子节点的概念首义原序列:由于词性泛化范围太大, 因此引入词语概念首义原来提高准确率。

谓词子节点的依存关系序列:即子类框架。如, “给” 的子类框架是SBJ+IOBJ+OBJ

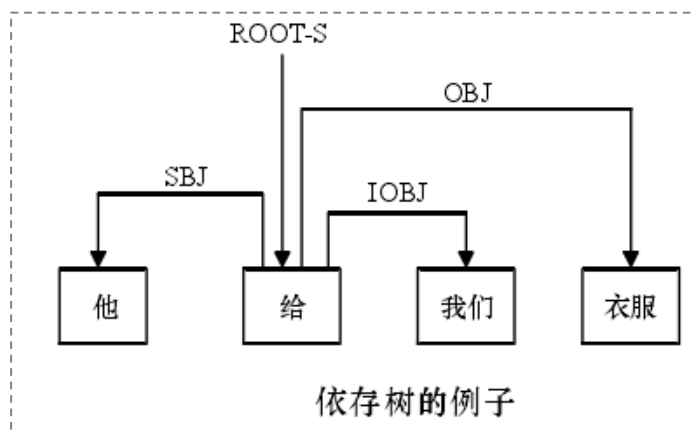
谓词兄弟节点的依存关系序列:有些句子有多个动词, 非根节点动词具有兄弟节点。

谓词的类别信息: 谓词的分类

并列动词的主语: 如果并列的动词只有一个主语, 则特征值为true, 反之为false。这个特征是为了解决并列动词的主语不确定性问题。如:在 “我唱歌, 他跳舞。” 中, 并列的动词共有2个主语;在 “我一边唱歌, 一边跳舞。” 中, 并列的动词共有1个主语。

2.基于依存句法分析的统计SRL方法

组合特征



由于最大熵分类器不能自动地对特征进行组合，因此，使用上述一些特征的组合来构造组合特征，经过实验验证后，去掉那些降低系统性能的，剩余的作为最终系统的特征。

- 词 + 依存关系 + 中心词
- 实体名词概念首义原 + 依存关系 + 中心词
- 词 + 谓词家族关系
(谓词家族关系: 一个词相对于谓词的关系。值为“孩子” “子孙” “双亲” “祖先” “自身” “兄弟” “无关系”)
- 词 + 关系路径
- 中心词 + 关系路径
- 中心词概念首义原 + 关系路径
- 中心词 + 上位关系路径
- 中心词 + 词性模式
- 谓词子节点的依存关系序列 + 谓词家族关系
- 词性路径 + 关系路径
- 谓词子节点的字性序列 + 谓词子节点的依存关系序列

2.基于依存句法分析的统计SRL方法

由于测试集与训练集中出现相同的参数中心词的情况不多，因此大部分系统引入词性来解决这个问题。但在很多情况下，中心词词性、与谓词的关系相同，但语义角色不同。

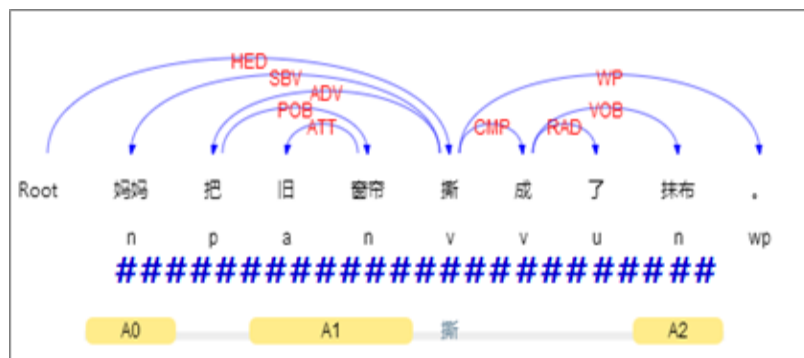
如:吃食堂，吃米饭，“食堂”与“米饭”都为实体名词，这2个短语都为动宾关系，通过词性差异无法区分;为此引入词语的概念首义原来提高系统的召回率。例如，“米饭”的概念首义原为“食品”，“食堂”的概念首义原为“场所”，当“吃”的宾语首义原为“食品”时，语义关系为Arg I，当宾语首义原为“场所”时，语义关系为 ArgM-LOC 。

2.基于依存句法分析的统计SRL方法

模型训练:

● 训练语料处理:

标注好语义角色的依存句法树训练语料



注: ##### 表示特征需要信息

模型: 选择最大熵模型

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

● 模型训练:

用不同的特征分别训练**论元识别模型** 和 **论元标注模型**

模型1-论元识别

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

模型2-论元标注

$$p_{\lambda}(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中:

$$Z_{\lambda}(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

与基于句法成分的语义角色标注系统相比较, 基于依存关系的系统对词汇的依赖性较弱, 受小语料库的影响较小。

2.基于依存句法分析的统计SRL方法

语义角色标注：

例： 我在北京看电影

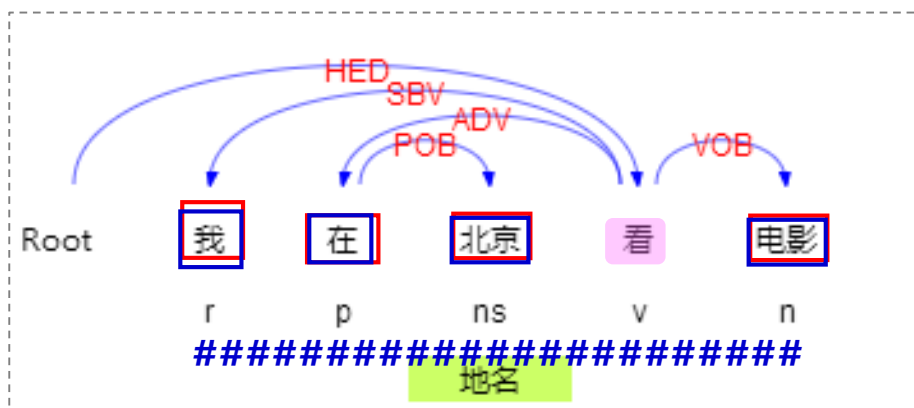
1.依存句法分析

2.候选论元剪除

3.论元识别

4. 论元标注

5. 后处理



B-A0 B-AM-LOC I-AM-LOC B-V B-A1

[A0] [LOC] [A1]

模型1-论元识别

$$p_1(y|x) = \frac{1}{Z_1(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

$$Z_1(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

模型2-论元标注

$$p_2(y|x) = \frac{1}{Z_2(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

其中：

$$Z_2(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

称为归一化因子。

2.基于依存句法分析的统计SRL方法

现有语义标注方法存在的主要问题：

- 对句法分析器性能的严重依赖性
- 领域适应能力差

基本性能：

英语、汉语：F1值大约为：70%左右(68%~76%)。

句子级语义分析-内容提要

12.2.1 格文法

12.2.2 语义角色标注

12.2.2.1 语料资源

12.2.2.2 语义角色标注方法

1. 基于完全句法分析的统计SRL方法

2. 基于依存句法分析的统计SRL方法

3. 基于依存句法分析的神经网络SRL方法

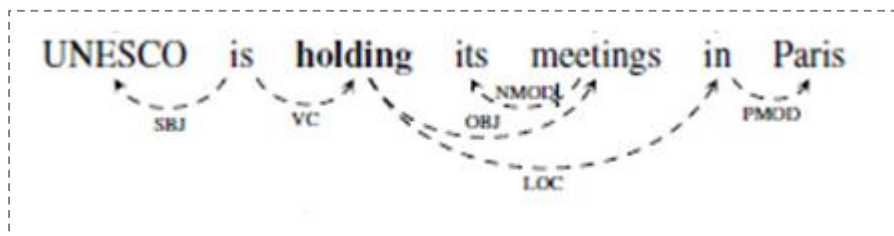
12.2.3 语义角色标注评价

3.基于依存句法分析的神经网络SRL方法

★ 基于依存分析的神经网络语义角色标注:

已知:

依存树
谓词语

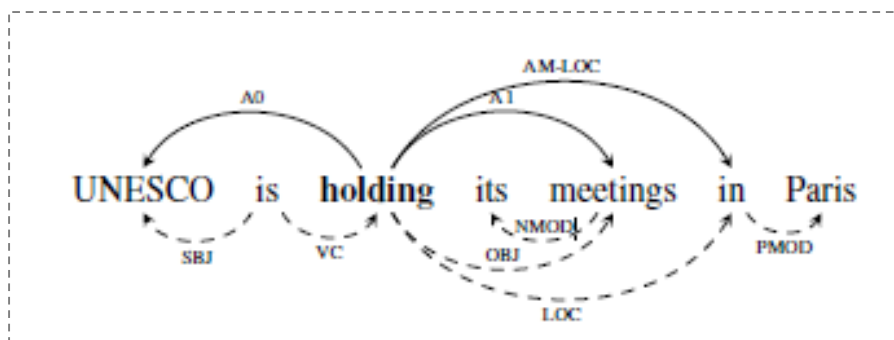


输入:

词特征、上下文、句法路径

目标:

语义角
色标注



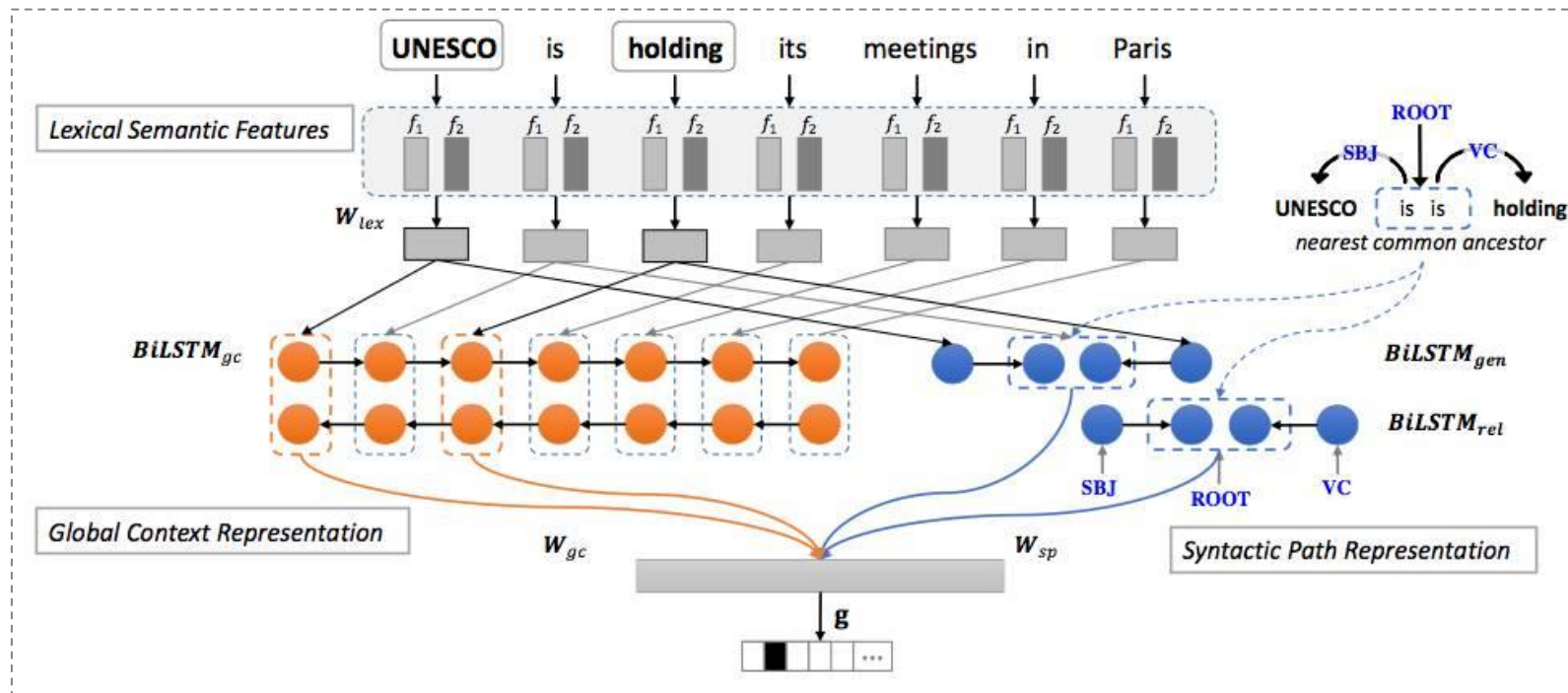
输出:

语义角色标签

= 语义角色标签总数+非语义角色类

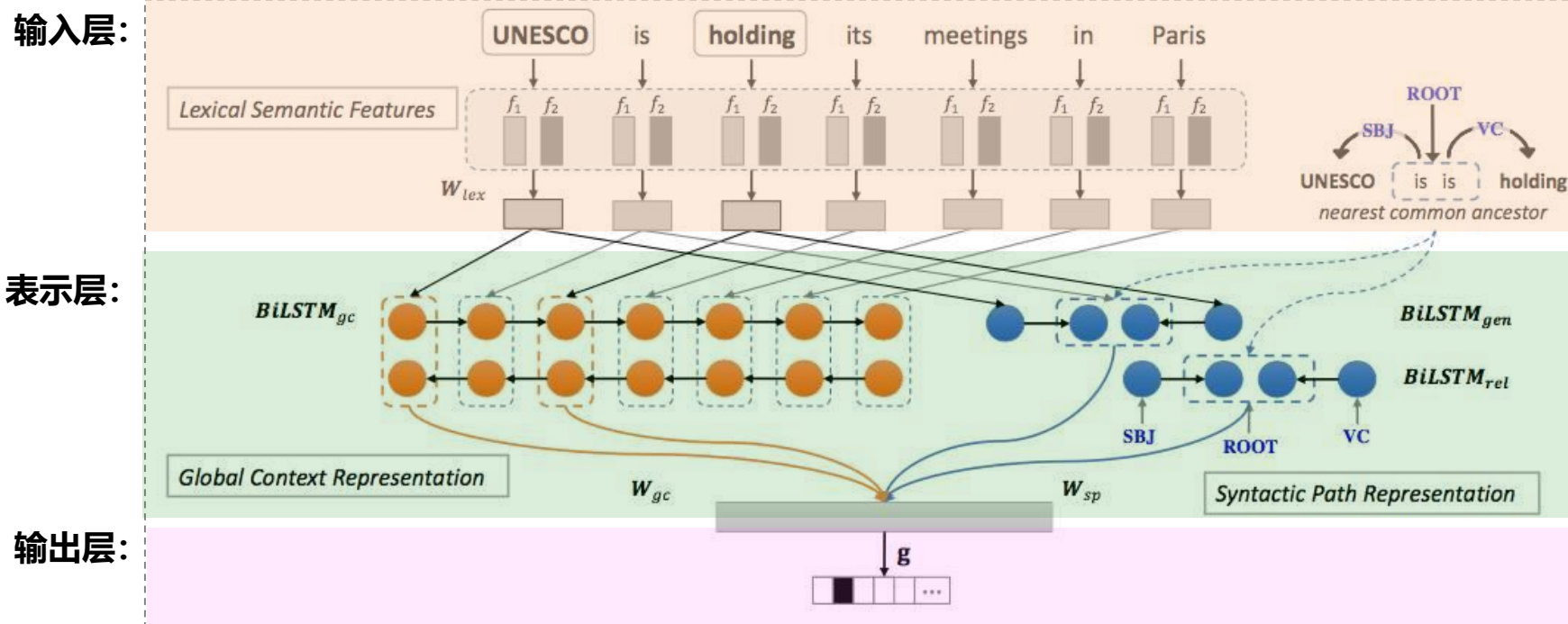
3. 基于依存句法分析的神经网络SRL方法

语义角色标注模型架构



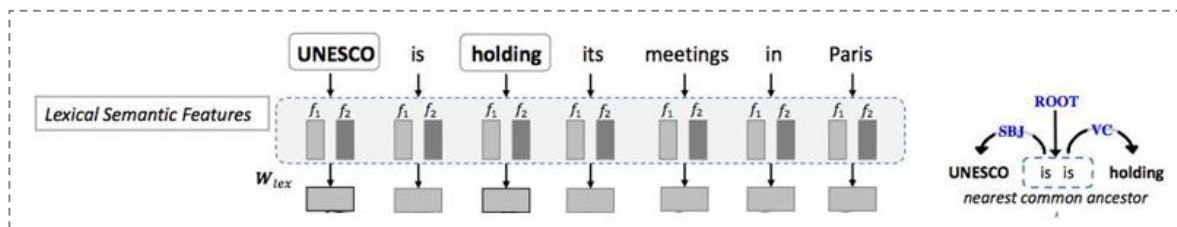
3. 基于依存句法分析的神经网络SRL方法

语义角色标注模型架构



3. 基于依存句法分析的神经网络SRL方法

输入层:



- 句子:

词汇语义特征表示(Lexical semantic Feature)

$$\mathbf{x}_i = \text{ReLU}(\mathbf{W}_{lex} \Phi_i + \mathbf{b}_{lex}), \text{ where } \Phi_i = [\mathbf{w}_i; \mathbf{p}_i]$$

其中, \mathbf{w}_i 为词 \mathbf{p}_i 为词性

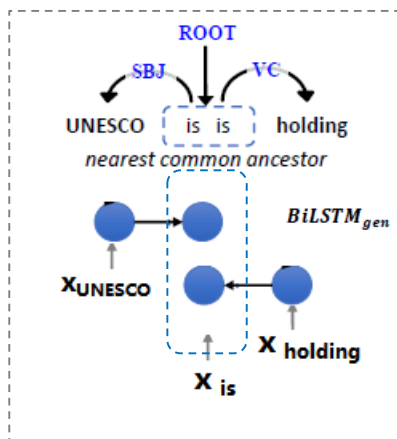
- 依存树:

句法路径表示 (Syntactic Path Representation)

定义 : $\text{nca}(e_1, e_2)$ 为 e_1 和 e_2 的最近共同祖先

$$e_1 \text{ 到 } e_2 \text{ 的路径表示为 } \mathbf{R}_{(e_1, e_2)} = [\vec{\mathbf{h}}_{\text{nca}(e_1, e_2)}; \overleftarrow{\mathbf{h}}_{\text{nca}(e_1, e_2)}]$$

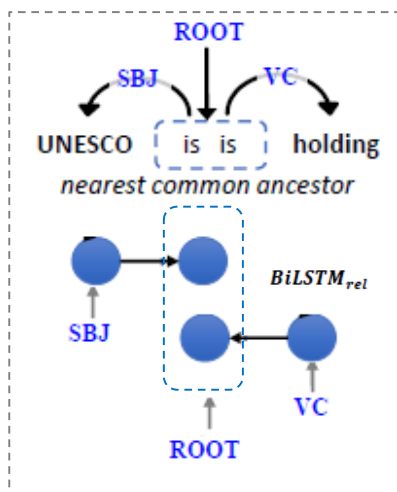
3. 基于依存句法分析的神经网络SRL方法



句法路径:

1. 通用路径：以词汇语义特征表示 x_i 为输入的句法路径

$$R_{(e_1, e_2)}^{gen} = [\vec{h}_{nca(e_1, e_2)}^{gen}; \overleftarrow{h}_{nca(e_1, e_2)}^{gen}]$$

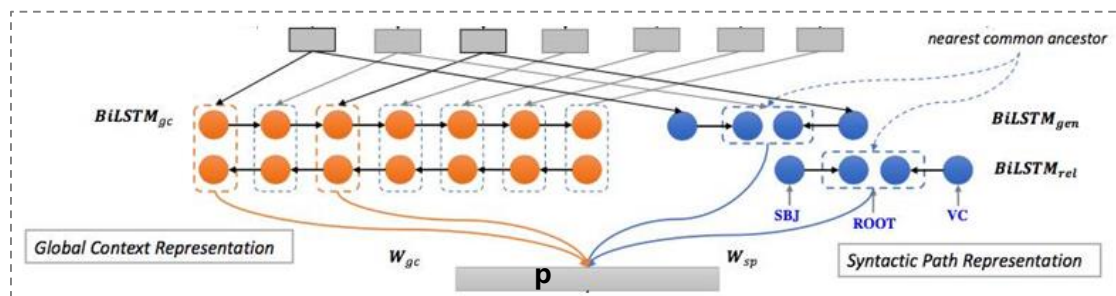


2. 关系路径：以依存关系为输入的句法路径

$$R_{(e_1, e_2)}^{rel} = [\vec{h}_{nca(e_1, e_2)}^{rel}; \overleftarrow{h}_{nca(e_1, e_2)}^{rel}]$$

3. 基于依存句法分析的神经网络SRL方法

表示层:



- 句子:

全局上下文表示 (Global Context Representation)

$$\mathbf{R}_{e_1}^{gc} = [\vec{\mathbf{h}}_{e_1}; \overleftarrow{\mathbf{h}}_{e_1}]; \quad \mathbf{R}_{e_2}^{gc} = [\vec{\mathbf{h}}_{e_2}; \overleftarrow{\mathbf{h}}_{e_2}]$$

- 依存树:

句法语义路径表示:

1. 通用路径表示

$$\mathbf{R}_{(e_1, e_2)}^{gen} = [\vec{\mathbf{h}}_{nca(e_1, e_2)}^{gen}; \overleftarrow{\mathbf{h}}_{nca(e_1, e_2)}^{gen}]$$

2. 关系路径表示

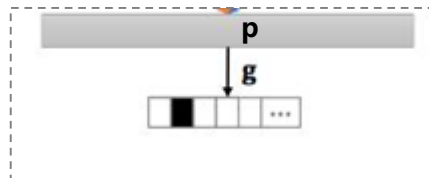
$$\mathbf{R}_{(e_1, e_2)}^{rel} = [\vec{\mathbf{h}}_{nca(e_1, e_2)}^{rel}; \overleftarrow{\mathbf{h}}_{nca(e_1, e_2)}^{rel}]$$

- 会聚:

$$\mathbf{p} = \text{ReLU}(\underbrace{\mathbf{W}_{gc} [\mathbf{R}_{e_1}^{gc}; \mathbf{R}_{e_2}^{gc}]}_{\text{Global Context}} + \underbrace{\mathbf{W}_{sp} [\mathbf{R}_{(e_1, e_2)}^{gen}; \mathbf{R}_{(e_1, e_2)}^{rel}]}_{\text{Syntactic Path}} + \mathbf{b})$$

3. 基于依存句法分析的神经网络SRL方法

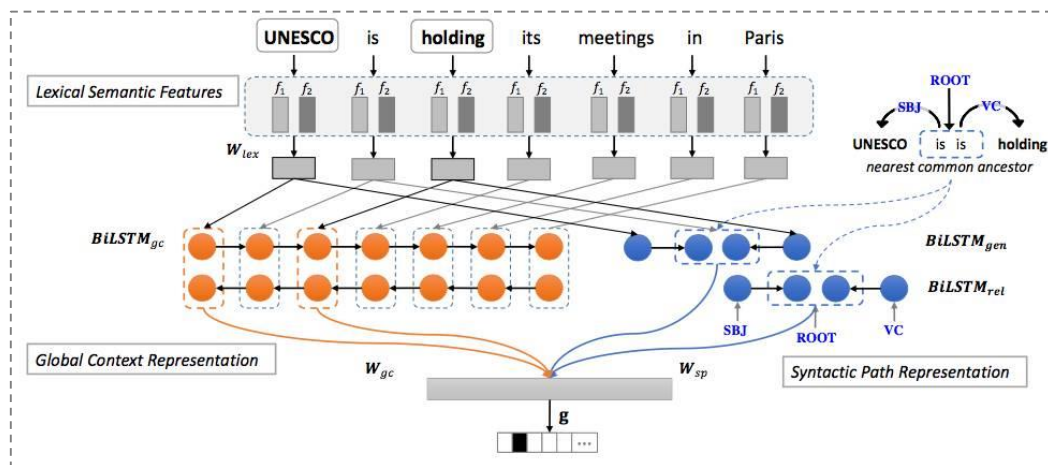
输出层:



$$p(c|p) = \text{softmax}(g_c^T p + q_c)$$

其中,
 g_c 为类别标签表示(embedding)
 q_c 为类别偏置

■ 模型结构



输入: 句子 (词特征、上下文)
依存树 (2种句法路径)

输出: 语义角色标签 (分布)

参数: $\theta = \{ \text{BiLSTM}_{gc}, \text{BiLSTM}_{gen}, \text{BiLSTM}_{rel}, W_{lex}, W_{gc}, W_{sp}, q_c \}$

3.基于依存句法分析的神经网络SRL方法

■ 模型训练

- **优化目标：** 最大化 $\mathcal{L}(\theta) = -\sum_{i=0}^N \log p(c_i|\mathbf{p}_i)$ N为训练实例数

$$Y=f(x,\theta)$$

$$p(c|\mathbf{p}) = \text{softmax}(\mathbf{g}_c^\top \mathbf{p} + \mathbf{q}_c)$$

$$\mathbf{p} = \text{ReLU}(\underbrace{\mathbf{W}_{gc} [\mathbf{R}_{e_1}^{gc}; \mathbf{R}_{e_2}^{gc}]}_{\text{Global Context}} + \underbrace{\mathbf{W}_{sp} [\mathbf{R}_{(e_1,e_2)}^{gen}; \mathbf{R}_{(e_1,e_2)}^{rel}]}_{\text{Syntactic Path}} + \mathbf{b})$$

$$\mathbf{R}_{(e_1,e_2)}^{gen} = [\vec{\mathbf{h}}_{nca(e_1,e_2)}^{gen}; \overleftarrow{\mathbf{h}}_{nca(e_1,e_2)}^{gen}]$$

$$\mathbf{R}_{(e_1,e_2)}^{rel} = [\vec{\mathbf{h}}_{nca(e_1,e_2)}^{rel}; \overleftarrow{\mathbf{h}}_{nca(e_1,e_2)}^{rel}]$$

$$\mathbf{R}_{e_1}^{gc} = [\vec{\mathbf{h}}_{e_1}; \overleftarrow{\mathbf{h}}_{e_1}]; \quad \mathbf{R}_{e_2}^{gc} = [\vec{\mathbf{h}}_{e_2}; \overleftarrow{\mathbf{h}}_{e_2}]$$

$$\mathbf{x}_i = \text{ReLU}(\mathbf{W}_{lex} \Phi_i + \mathbf{b}_{lex})$$

- **模型训练：** 使用随机梯度下降（SGD）方法优化目标函数得到参数所有 θ

3.基于依存句法分析的神经网络SRL方法

根据语义角色标注的限制条件及后处理:

- 1.每个词只能标为一种标签 (包括NULL)
- 2.分布概率小的标注(如, 小于0.3) 不标注(NULL除外).
- 3.一种角色(NULL除外)通常在句子中只出现一次, 因此在角色标注序列中,应无重复角色出现。

句子级语义分析-内容提要

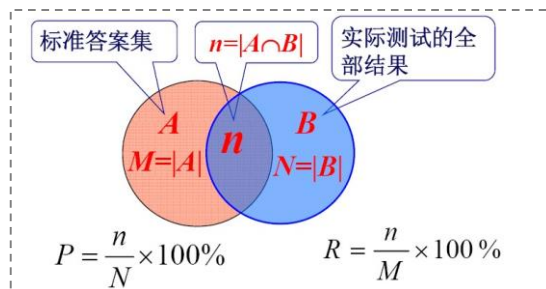
12. 2. 1 格文法

12. 2. 2 语义角色标注

12. 2. 3 语义角色标注评价

12.2.3 语义角色标注评价

评价指标:



正确率 (precision, **P**) : 测试结果中正确标注的个数占系统所有输出结果的比例

$$P = \frac{n}{N} \times 100\%$$

召回率 (Recall ratio, **R**): 测试结果中正确结果的个数占标准答案总数的比例

$$R = \frac{n}{M} \times 100\%$$

F-度量值: 正确率和召回率的综合值

$$F - measure = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \times 100\%$$

一般地, 取 $\beta = 1$, 即

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%$$

参考文献：

刘群，李素建.基于《知网》的词汇语义相似度的计算

宗成庆，统计自然语言处理（第2版）课件

安强强，张蕾，基于依存树的中文语义角色标注

刘挺，车万翔等，基于最大熵分类器的语义角色标注

王鑫,孙薇薇等，基于浅层句法分析的中文语义角色标注研究

Jiang Guo, Wanxiang Che , A Unified Architecture for Semantic Role Labeling and Relation Classification, Proceedings of COLING 2016

Eric H. Huang, Richard Socher, Improving Word Representations via Global Context and Multiple Word Prototypes

在此表示感谢！

谢谢各位！

