

chart parsing

lecture11_句法分析1 PPT22页 例子的chart parsing的计算过程，结合了老师的课程板书来写。

思路如下：

(1) 从左往右扫描 $X(i,j)$ 填入agenda

(2) 将其视为CFG规则的右部填入边，结合点规则判断是活动边还是闭边。

(3) 如果是活动边，边保留在active里，消除此次的agenda（就是加入了chart，可以连线了），继续向右扫描；如果是完成边，将CFG规则左部加入agenda，消除此次的agenda和上一步留下的活动边。

对单词序列进行遍历，执行以上操作。

注意：边的括号注释表示当前扫描的位置。每一行表示每一次操作。

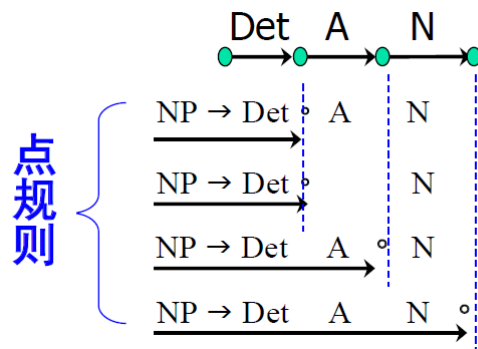
agenda	active	closed	act
Det(1,2)	NP->Det·N(1,2)		消除Det(1,2)
N(2,3)		NP->Det N·(1,3)	消除N(2,3) 和NP活动边
NP(1,3)	S->NP·VP(1,3)		消除NP(1,3)
V(3,4)	VP->V·NP(3,4)		消除V(3,4)
Det(4,5)	NP->Det·N(4,5)		消除Det(4,5)
N(5,6)		NP->Det N·(4,6)	消除N(5,6)和NP活动边
NP(4,6)		VP->V NP·(3,6)	消除NP(4,6)和VP活动边
VP(3,6)	VP->VP·PP(3,6)		消除VP(3,6)
PREP(6,7)	PP->PREP·NP(6,7)		消除PREP(6,7)
Det(7,8)	NP->Det·N(7,8)		消除Det(7,8)
N(8,9)		NP->Det N·(7,9)	消除N(8,9)和NP活动边
NP(7,9)		PP->PREP NP·(6,9)	消除NP(7,9)和PP活动边
PP(6,9)		VP->VP PP·(3,9)	消除PP(6,9)和VP活动边
VP(3,9)		S->NP VP·(1,9)	消除VP(3,9)和S活动边（也就是完成了）

复现完整树结构，只需要对完成边进行回溯。

关于chart parsing的几个问题：

1. 采用的CFG规则一定是位置连续的终结符/非终结符吗？

不一定，见下例Det 和N并不连续。（可是这样画出来的结构树会有交叉的，好奇怪）



2. chart parsing会出现歧义吗？

会的，是因为CFG规则。如果是PCFG，选择概率最大的一个。

分析的质量与规则的质量密切相关。

3. chart parsing的算法复杂度

每处理一个单词需要最多执行的最多操作次数为：

$$C + 1 + Sn^2 + 1 + Sn^2 + Sn^2 = 2 + C + 3Sn^2$$

由于算法对于长度为 n 的输入句子要执行 n 次循环，
因此，Chart 算法最大执行的操作次数为：

$$n \times (2 + C + 3Sn^2)$$

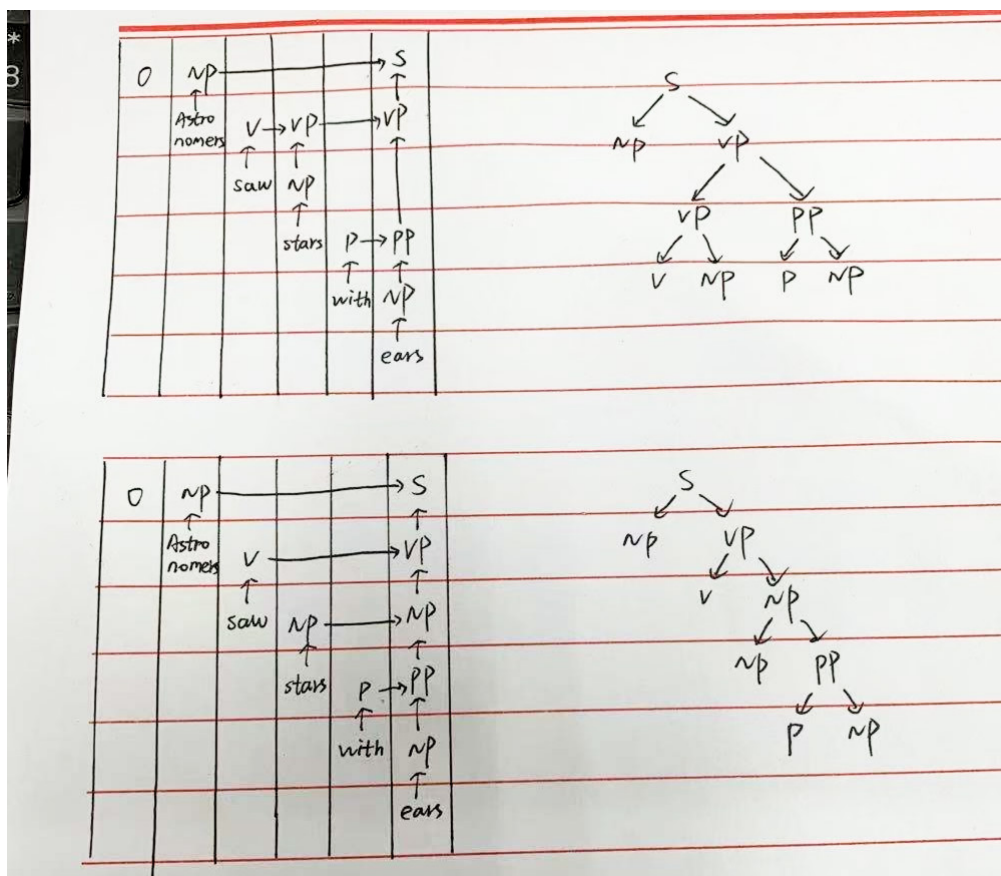
所以，Chart算法的时间复杂度为：

$$O(Kn^3) \quad (K \text{ 为一常数})$$

CYK算法

lecture11_句法分析1 PPT60页 例子

两种过程，歧义性的体现，考试时要注意



CYK算法的问题

1.语法规则

肖姆斯基规范化

$$\begin{aligned}
 A &\rightarrow w \quad \text{or} \quad A \rightarrow BC \\
 A, B, C &\in V_N \\
 w &\in V_T \\
 G &= (V_N, V_T, P, S)
 \end{aligned}$$

2.结构树还会有交叉情况吗?

不会。识别矩阵的行列位置 (i, j) 实际上就是终结符/非终结符的位置 (和chart parsing给的位置一样, 不过是从0开始), 而CYK的分析过程有这么一步:

- (3) 按平行于主对角线的方向, 一层一层地向上填写矩阵的各个元素 $t_{i,j}$, 其中, $i = 0, 1, \dots, n-d$, $j = d+i$, $d = 2, 3, \dots, n$ 。如果存在一个正整数 k , $i+1 \leq k \leq j-1$, 在文法 G 的规则集中有产生式 $A \rightarrow BC$, 并且, $B \in t_{i,k}$, $C \in t_{k,j}$, 那么, 将 A 写到矩阵 $t_{i,j}$ 位置上。

判断句子 x 由文法 G 所产生的充要条件是: $t_{0,n} = S$ 。

B 的列坐标总与 C 的行坐标相同, 也就是说 BC 总是相邻的。

3.CYK充要判据?

上图蓝色。

