

自然语言处理导论 作业二

李鹏, 10175501102@stu.ecnu.edu.cn

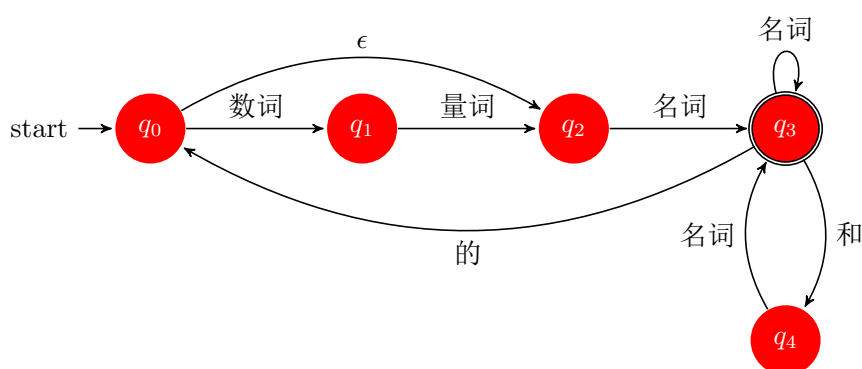
2019 年 11 月 28 日

1 有限状态自动机

构造识别名词词组的FSA

- 三本书
- 语法书
- 阿Q的书
- 阿Q的三本书
- 阿Q的三本语法书
- 阿Q的三本汉语语法书
- 阿Q的三本古代汉语语法书
- 阿Q和他的三本汉语语法书
-

答:



备注: 为了答案简单直观, 此处画出了有限状态自动机的图示, 并没有提供状态转移表;

2 上下文无关文法

对于语言 $L = \{ab, aabb, aaabbb, \dots, a^n b^n, \dots\}$, n 是自然数。请写出 L 的上下文无关文法, 要求产生式右部不能超过两个符号。

答:

一种可能的上下文无关文法: $G = (V_N, V_T, S, P)$, 其中 $V_N = \{S, A, B\}$, $V_T = \{a, b\}$,

P:

(1) $S \rightarrow a A$

(2) $A \rightarrow B b$

(3) $B \rightarrow a A$

(4) $A \rightarrow b$

备注：此处产生式右部可以先写出三个符号，然后利用Chomsky Normal Form将三支的规则转换成两个二分支的规则；

3 上下文无关文法与句法分析

(1) 写出可以产生汉语自然数表达式的CFG

(2) 用你写的CFG，画出下列数字的分析树：

一亿零三百万 三万六千五百八十一

答：(1)一个简单的CFG：

$G = (V_N, V_T, S, P)$, 其中 $V_N = \{S, A, B, HS, X, W\}$, $V_T = \{\text{零, 一, 二, 三, 四, 五, 六, 七, 八, 九, 十, 百, 千, 万, 亿}\}$, P :

1) $S \rightarrow A B \mid B$

2) $A \rightarrow A B \mid B$

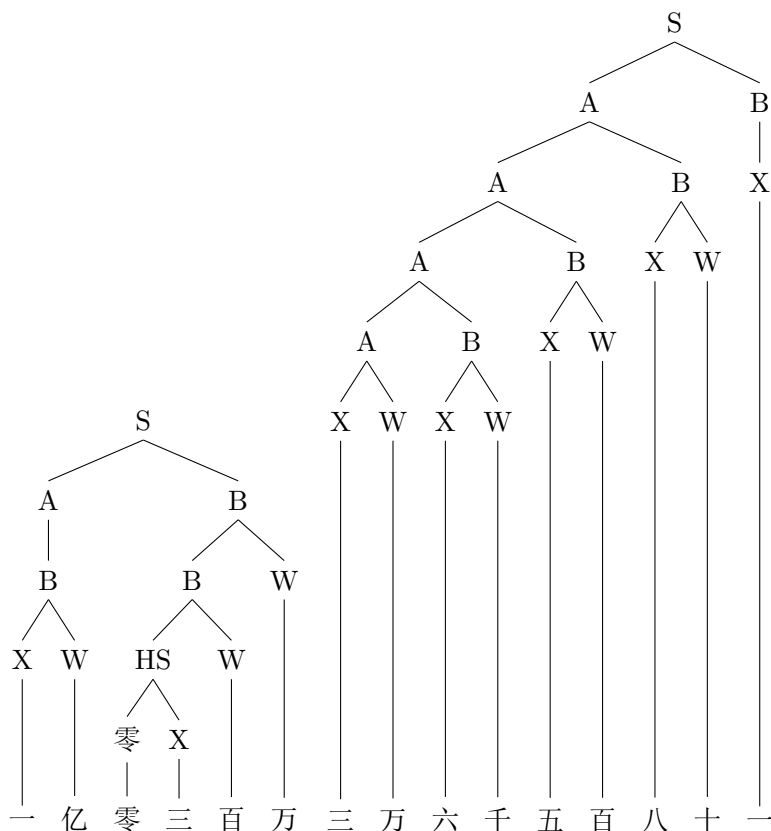
3) $B \rightarrow X \mid X W \mid HS W \mid B W \mid \text{十}$

4) $HS \rightarrow \text{零} X \mid \text{十} X$

5) $X \rightarrow \text{零} \mid \text{一} \mid \text{二} \mid \text{三} \mid \text{四} \mid \text{五} \mid \text{六} \mid \text{七} \mid \text{八} \mid \text{九}$

6) $W \rightarrow \text{十} \mid \text{百} \mid \text{千} \mid \text{万} \mid \text{亿}$

(2)分析树如下：



备注：此CFG只是一个简单版本，汉语中还有很多情况没有考虑进去，更复杂的内容可以参考北京大学詹卫东教授《汉语自然数表达形式特征分析》的讲义；

4 句法分析

给定下面的CFG，

- 1) $S \rightarrow NP \ VP$
- 2) $NP \rightarrow N$
- 3) $NP \rightarrow Det \ N$
- 4) $VP \rightarrow V \ NP$
- 5) $VP \rightarrow VP \ PP$
- 6) $PP \rightarrow Prep \ NP$

分别用自顶向下、自底向上和Earley算法分析下面句子的结构。要求给出分析过程，格式可参考讲义。

Mike hits the dog with a rod

词性： N V Det N Prep Det N

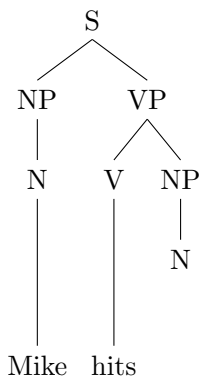
答：

(1)自顶向下方法：

使用的产生式规则顺序：

- 1) $S \rightarrow NP \ VP$
- 2) $NP \rightarrow N$
- 3) $VP \rightarrow V \ NP$
- 4) $NP \rightarrow N$

此时结果如下：

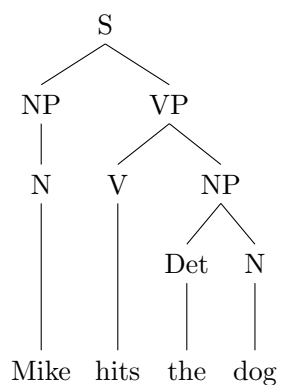


此时N与句子中的Det不再对应，回溯选择另一条路：

- 1) $S \rightarrow NP \ VP$
- 2) $NP \rightarrow N$
- 3) $VP \rightarrow V \ NP$

4) $NP \rightarrow Det\ N$

此时结果如下：



此时句法树扩展完毕，但是句子没有完全匹配成功，再次回溯选择另一条路：

1) $S \rightarrow NP\ VP$

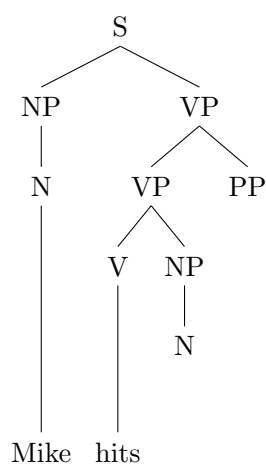
2) $NP \rightarrow N$

3) $VP \rightarrow VP\ PP$

4) $VP \rightarrow V\ NP$

5) $NP \rightarrow N$

此时结果如下：



此时 N 与 Det 仍然不匹配，再回溯选择另一条路：

1) $S \rightarrow NP\ VP$

2) $NP \rightarrow N$

3) $VP \rightarrow VP\ PP$

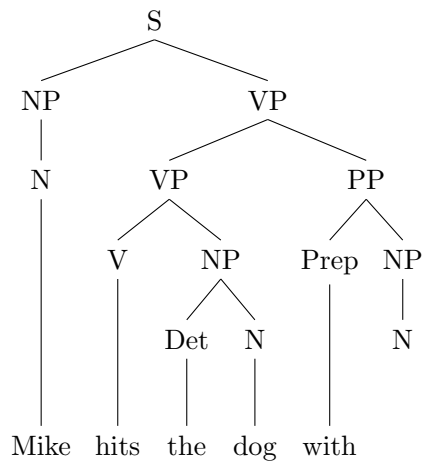
4) $VP \rightarrow V\ NP$

5) $NP \rightarrow Det\ N$

6) $PP \rightarrow Prep\ NP$

7) $NP \rightarrow N$

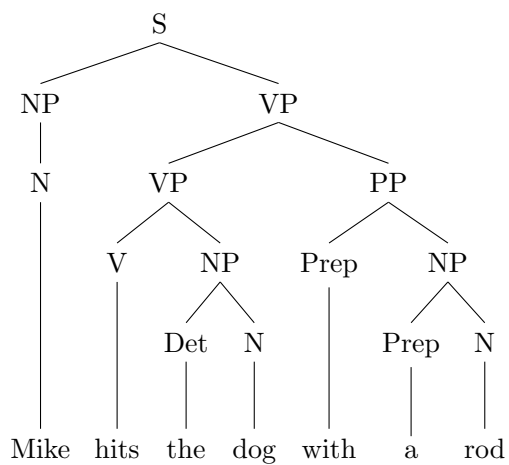
此时结果如下：



此时 N 与 Det 仍然不匹配，再回溯选择另一条路：

- 1) $S \rightarrow NP \quad VP$
- 2) $NP \rightarrow N$
- 3) $VP \rightarrow VP \quad PP$
- 4) $VP \rightarrow V \quad NP$
- 5) $NP \rightarrow Det \quad N$
- 6) $PP \rightarrow Prep \quad NP$
- 7) $NP \rightarrow Det \quad N$

最终实现了句子成分匹配：

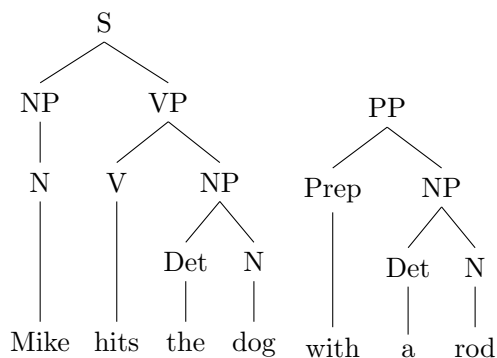


(2)自底向上方法：

使用的产生式规则顺序（这里采用逐层规约）：

- 1) $NP \rightarrow N$
- 2) $NP \rightarrow Det \quad N$
- 3) $NP \rightarrow Det \quad N$
- 4) $VP \rightarrow V \quad NP$
- 5) $PP \rightarrow Prep \quad NP$
- 6) $S \rightarrow NP \quad VP$

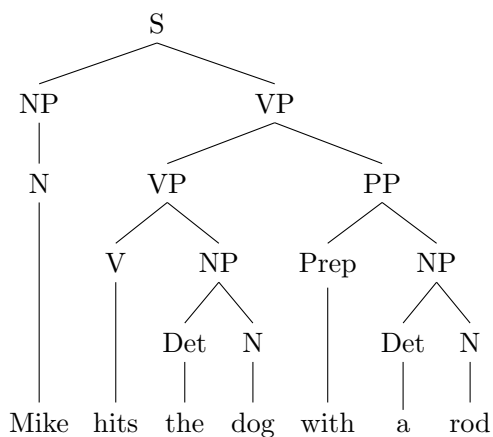
此时结果如下：



无规则可用，回溯：使用的产生式规则顺序（这里采用逐层规约）：

- 1) $NP \rightarrow N$
- 2) $NP \rightarrow Det \ N$
- 3) $NP \rightarrow Det \ N$
- 4) $VP \rightarrow V \ NP$
- 5) $PP \rightarrow Prep \ NP$
- 6) $VP \rightarrow VP \ PP$
- 7) $S \rightarrow VP \ NP$

最终实现了句子成分匹配：



(3)Earley算法：

- 1) 初始： $S \rightarrow \cdot NP \ VP \ [0, 0]$
- 2) 预测： $NP \rightarrow \cdot N \ [0, 0]$
- 3) 预测： $NP \rightarrow \cdot Det \ N \ [0, 0]$
- 4) 扫描： $NP \rightarrow N \cdot \ [0, 1]$
- 5) 规约： $S \rightarrow NP \cdot VP \ [0, 1]$
- 6) 预测： $VP \rightarrow \cdot V \ NP \ [1, 1]$
- 7) 预测： $VP \rightarrow \cdot VP \ PP \ [1, 1]$
- 8) 扫描： $VP \rightarrow V \cdot NP \ [1, 2]$

- 9) 规约: $S \rightarrow NP VP \cdot$ [1, 2] (失败)
- 10) 规约: $VP \rightarrow \cdot VP PP$ [1, 2]
- 11) 预测: $PP \rightarrow \cdot Prep NP$ [2, 2]
- 12) 预测: $NP \rightarrow \cdot Det N$ [2, 2]
- 13) 预测: $NP \rightarrow \cdot N$ [2, 2]
- 14) 扫描: $\rightarrow Det \cdot N$ [2, 3]
- 15) 规约: $VP \rightarrow \cdot VP PP$ [2, 3]
- 16) 预测: $NP \rightarrow Det \cdot N$ [3, 3]
- 17) 扫描: $NP \rightarrow Det N \cdot$ [3, 4]
- 18) 规约: $VP \rightarrow VP \cdot PP$ [3, 4]
- 19) 预测: $PP \rightarrow \cdot Prep NP$ [4, 5]
- 20) 扫描: $PP \rightarrow Prep \cdot NP$ [4, 5]
- 21) 归约: $VP \rightarrow VP \cdot PP$ [4, 5]
- 22) 预测: $NP \rightarrow \cdot Det N$ [5, 5]
- 23) 预测: $NP \rightarrow N \cdot$ [5, 5]
- 24) 扫描: $NP \rightarrow Det \cdot N$ [5, 6]
- 25) 归约: $NP \rightarrow Det \cdot N$ [5, 6]
- 26) 归约: $PP \rightarrow Prep \cdot NP$ [5, 6]
- 27) 预测: $NP \rightarrow Det \cdot N$ [6, 6]
- 28) 扫描: $PP \rightarrow Prep \cdot NP$ [6, 7]
- 29) 归约: $PP \rightarrow \cdot Prep NP$ [6, 7]

(emmm...并没有完全掌握这个算法)

5 合一运算

给出下列特征结构合一运算的结果

- 1) $[NUMBER: SG] \cup [NUMBER: SG]$
- 2) $[NUMBER: SG] \cup [NUMBER: PL]$
- 3) $[NUMBER: SG] \cup$
- 4) $[NUMBER: SG] \cup [PERSON: 3]$
- 5) $[AGREE: [NUMBER: SG]] \cup \left[\begin{array}{c} CAT: NP \\ AGREE: [PERSON: 3] \end{array} \right]$
- 6) $[NUMBER: RL] \cup [ARGEE: [NUMBER: PL]]$

答: 各式合一结果如下

1) [*NUMBER* : *SG*]

2) \emptyset

3) [*NUMBER* : *SG*]

4) $\left[\begin{array}{c} \textit{NUMBER} : \textit{SG} \\ \textit{PERSON} : 3 \end{array} \right]$

5) $\left[\begin{array}{c} \textit{CAT} : \textit{NP} \\ \textit{AGREE} : \left[\begin{array}{c} \textit{NUMBER} : \textit{SG} \\ \textit{PERSON} : 3 \end{array} \right] \end{array} \right]$

6) $\left[\begin{array}{c} \textit{NUMBER} : \textit{RL} \\ \textit{ARGEE} : [\textit{NUMBER} : \textit{PL}] \end{array} \right]$