Lec 2 上下文无关文法和上下文无关语言

上下文无关文法(context-free grammars) $G=(V,T,P,S)$

V是非终结符集合，T是终结符集合，P是产生式的集合，S是开始符.

$V\cap T=\emptyset, S\in V$ 产生式形如$A\to\alpha,\, A\in V, \alpha\in(V\cup T)^*$

归约：由串向开始符归约

推导：由开始符向串推导 符号$\overset{*}{\Rightarrow}$

- 最左推导：每一步总是替换出现在最左边的非终结符 $\underset{lm}{\overset{*}{\Rightarrow}}$
- 最右推导

定义上下文无关文法G的语言为：$L(G)=\{w|w\in T^*\cap S\underset{G}{\overset{*}{\Rightarrow}}w\}$. 称为上下文无关语言。

例$\{w|w\in\{a,b\}^*, count(w,a)\neq count(w,b)\}$

$$\begin{aligned}S&\to AaA|BbB\\A&\to aAbA|bAaA|aA|Aa|\epsilon\\B&\to aBbB|bBaB|bB|Bb|\epsilon\end{aligned}$$

$\{w|w\in\{a,b\}^*, |count(w,a)-count(w,b)|=2\}$

$$\begin{aligned}S&\to CaCaC|CbCbC\\C&\to aCbC|bCaC|\epsilon\end{aligned}$$

证明给定语言L是某个文法G的语言:

- if $w\in L$ then $w\in L(G)$, 归纳于w的长度.
- if $w\in L(G)$ then $w\in L$, 归纳于推导w的步数.

Chomsky分类方法

- 0型文法 = 图灵机
- 1型文法 = 上下文有关文法
- 2型文法 = 上下文无关文法 = 下推自动机
- 3型文法 = 正规表达式 = DFA

消除二义性:

- 左结合
- 算符优先级连
- 最近嵌套匹配

Lec 3 正规表达式和正规语言

正规表达式的构造

例$\{xwx^R|x,w\in(a+b)^+\}$, x^R为x的反向(即反转).

$$a(a+b)^+a+b(a+b)^+b$$

代数定律具体化

设 E, F 为正规表达式，它们具有相同的变量集；采用同样的替换方式，得到对应于 E, F 的具体表达式分别为C,D. 则对 E, F 中的变量对应的所有语言，满足

$$L(E) = L(F) \iff L(C) = L(D)$$

Lec 4 有限状态自动机

DFA是一个五元组$A = (Q, \Sigma, \delta, q_0, F)$

- Q: 状态集
- Σ: 有限输入符号集
- δ: 转移函数, $Q \times \Sigma \to Q$
- q_0: 一个开始状态 $q_0 \in Q$
- F: 终状态集合，$F \subset Q$.

NFA与DFA的唯一不同之处, $\delta : Q \times \Sigma \to 2^Q$.

NFA构造DFA，子集构造法 最坏情况构造出的DFA可能有2^n个状态

ϵ-NFA, $\delta : Q \times (\Sigma \cup \{\epsilon\}) \to 2^Q$

ϵ-闭包，记为`ECLOSE(q)`，定义为从q经过所有ϵ路径能到达的状态，包括q自身.

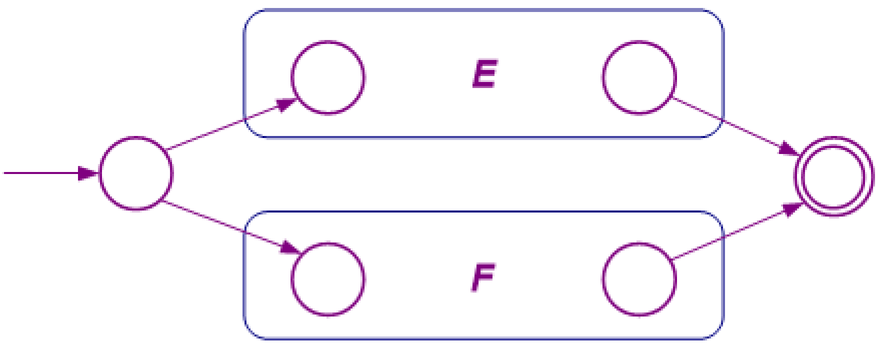
修改的子集构造法 最后一步取空闭包

最小化DFA 填表法

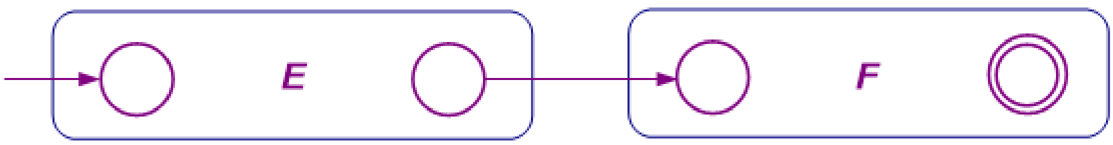
Lec 5 有限状态自动机 $\iff$ 正规表达式

正规表达式构造ϵ-NFA, Thompson构造法 (可靠程序)

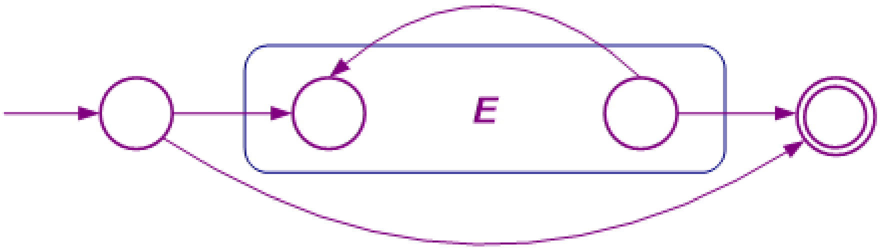
E+F:



EF;



E*:



DFA构造正规表达式 路径迭代法，状态消去法

自动机期末复习

Lec 6. 正规语言性质与运算

Pumping 引理是正规语言的**必要不充分**条件，用其否定 L 是正规语言:

1. 考虑任意的$n >= 1$.
2. 取串$w \in L, |w| \geq n$.
3. 任选满足$w = xyz, y \neq \epsilon, |xy| < n$ 的x,y,z
4. 找到一个 $k \geq 0$, 使 $xy^kz \notin L$.

判定问题:

- 是否为空，是否能接受某字符串，有限/无限，两个语言是否相等，是否相交...

正规语言的封闭运算:

- 并，星闭包，连接，差，同态
- 补 $(\Sigma^* - L)$: 调换 L DFA的终态和非终态
- 交: 构造 DFA $A = (Q_L \times Q_M, \Sigma, \delta, < q_L, q_M >, F_L \times F_M)$ 其中 $(< p, q >, a) =< \delta_L(p, a), \delta_M(q, a) >$.
- 反向: 由L的DFA构造反向语言的 $\epsilon - NFA$，逆向转移弧，以原初态为终态，加入以ϵ转移到原终态的新节点作为初态.
- 反同态：相应DFA 状态不变，转移函数 $r(q, a) = \delta'(q, h(a))$

可利用这些封闭性质证明某个语言不是正规语言。

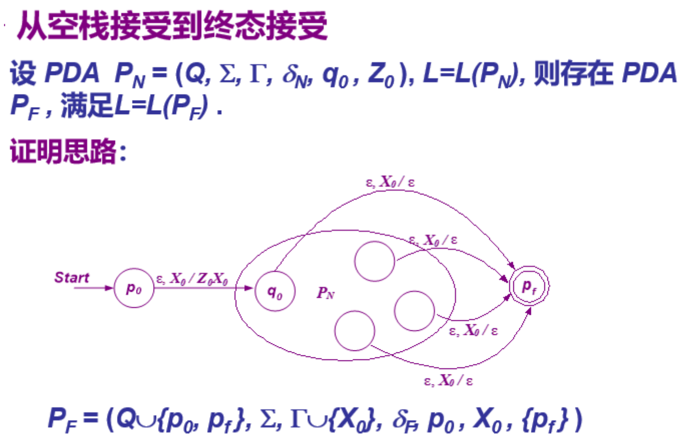
Lec 7. 下推自动机

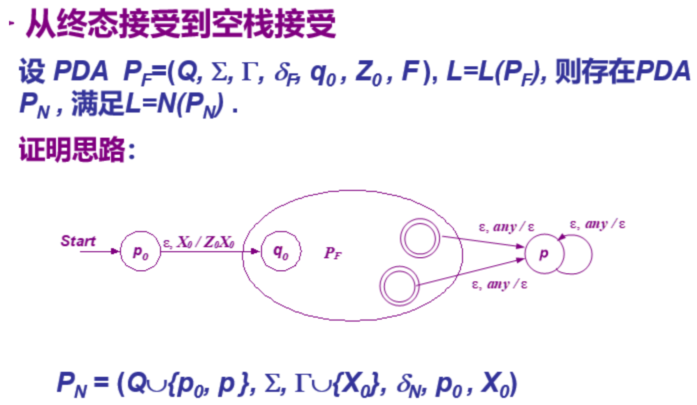
P = (状态集，字符集，栈符号集，转移函数，起始状态，起始栈符号[，终止态集])

在 NFA 的基础上多了一个堆栈，用栈记录历史。

$$(q, aw, X\beta) \vdash (p, w, \alpha\beta) \iff (p, \alpha) \in \sigma(q, a, X)$$

空栈接受与终态接受的相互转换





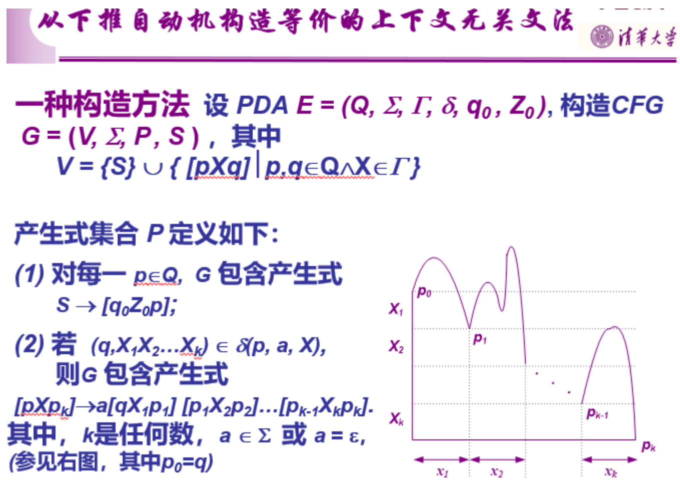
Lec 8. CFG 与 PDA

CFG -> PDA

设 CFG $G=(V,T,P,S)$, 构造空栈接受的相应PDA $E=(\{q\},T,V\cup T,\delta,q,S)$

- for $A\in V$, $\delta(q,\epsilon,A)=\{(q,\beta)|A\to\beta\in P\}$
- for $a\in T$, $\delta(q,a,a)=(q,\epsilon)$

PDA -> CFG



Lec 9. DPDA

一个PDA是 (终态接受) DPDA的条件

- $a\in\Sigma\vee a=\epsilon, x\in \mathrm{T}, \delta(q,a,x)$最多包含一个元素
- 若$\exists a\neq\epsilon$可以转移，则ε不可转移

前缀性质：不存在$x,y\in L, x\neq y$且 x 是 y 的前缀。

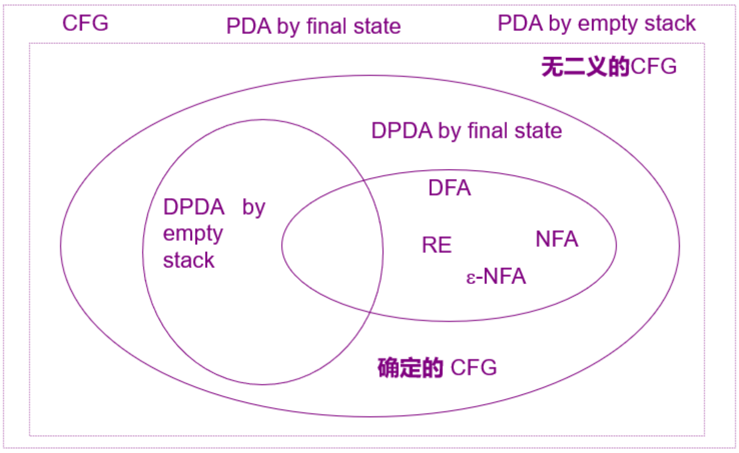
语言 L 是某个空栈接受DPDA P的语言，即$L=N(P)$, 当且仅当 L 满足前缀性质且是某个 DPDA 的语言。

- 若 $L=N(P)$，则 L 存在无二义文法。
- 可进一步推论若$L=L(P)$, 则 L 存在无二义文法。

DPDA不能表示所有上下文无关语言，包括所有固有二义的CFl和某些非固有二义的CFl.

- $L=\{ww^R|w\in\{0,1\}^*\}$不是任何DPDA的语言, 也非固有二义。

- $L=\{wcw^R|w=\{0,1\}^*\}$是DPDA的语言，但不是正规语言。



Lec 10. CFG 的简化及 Chomsky 范式

消去ϵ产生式

- 消去可致空符号，再在生成可致空符号的产生式中将可致空符号用空串替代

消去Unit产生式

消去无用符号：先消去非生成符号，再消去不可达符号。

- 有用符号, $\exists \alpha, \beta \in (V \cup T)^*, w \in T^*, S \Rightarrow^* \alpha X \beta \Rightarrow^* w.$
- 生成符号，$\exists w \in T^*, X \Rightarrow^* w$
- 可达符号，$\exists \alpha, \beta \in (V \cup T)^*$, $S \Rightarrow^* \alpha X \beta$

设CFG G的语言至少包含一个非 ε 字符串，经过以上简化得到G1，

$$L(G_1) = L(G) - \{\epsilon\}$$

Chomsky 范式 CNF: 只包含两类产生式的 CFG:

- $A \to BC, B, C \in V$
- $A \to a, a \in T$

对G1做变换，可以得到CNF G2

- 终结符替换 + 非终结符级联

Lec 11. 上下文无关语言的性质与运算

CFL也有必要不充分的pumping引理，否定 L 是 CFL 的证明思路:

1. 考虑任意的$n >= 1$.
2. 取串$w \in L$, $|w| \geq n$.
3. 任选满足$w = uvwxy, vx \neq \epsilon, |vwx| < n$ 的u, v, w, x, y
4. 找到一个 $k \geq 0$, 使 $uv^kwx^ky \notin L$.

判定问题:

- CFG是否为空：判定S是否是生成符号
- 是否能接受给定字符串：CYK算法

不可判定：CFG是否二义？CFL是否固有二义？两个CFL是否不相交，是否相等？

CFL的封闭运算

- 并，反向，闭包，连接，同态和反同态
- 与正规语言的交
- 替换

CFL之间的交，补，差不一定是CFL

Lec 12. 图灵机与递归可枚举语言

M = (状态集，字符集，带符号集，转移函数，开始态，空白带符，终态集)

图灵机可接受的语言是 **递归可枚举语言** (RE)

停机的图灵机可接受的语言是 **递归语言**

- 停机指图灵机不存在下一个移动，包括到达终态和无法接受两种情形

设计图灵机

- 带存储区的状态
- 多道图灵机：每个带符号表示为多元组
- 子例程

图灵机的扩展

语言接受能力等价于上述图灵机

- 多带图灵机: 可用2k多道模拟
- 非确定图灵机
- 半无穷带
- 多栈机，计数器机：两个堆栈/计数器可以接受RE语言
- 有无限地址的计算机

Lec 13. 计算理论初步

问题的判定：如果问题对应的语言是递归语言，问题是可判定的。

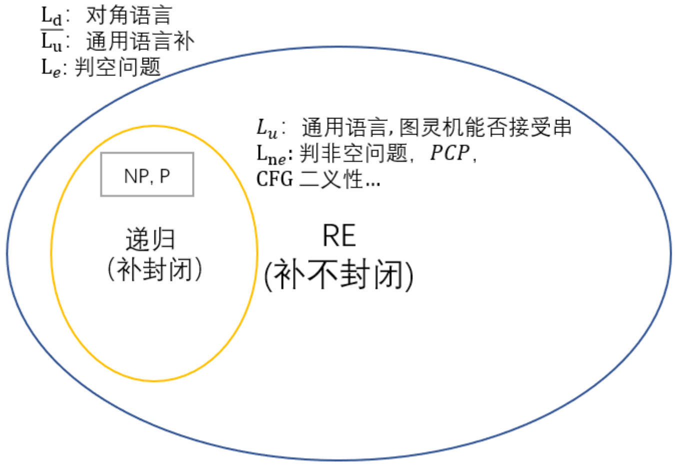
归约：若A能归约到B

- 如果B可判定，A也可判定。若A不可判定，B也不可判定。

Rice定理：有关RE语言的所有非平凡 Property 都是不可判定的。

设K是所有RE语言的集合，**Property P**定义为K的子集，若P不是空集或全集，则P是非平凡的。

- 由此推知，任给图灵机可以接受的语言（RE语言）L，以下问题都是不可判定的：
  - L是否是CFL，是否是正规语言？
  - 任给输入串w，w能否为图灵机接受？能否使图灵机停机？



P、NP、NP-完全、NP-难

- 在多项式时间内使图灵机/非确定图灵机停机。

对于**NP问题** P，若**任意NP问题都能在多项式时间内归约到P**，则P是NP-完全的。

对于问题 P，若**任意NP问题都能在多项式时间内归约到P**，则P是NP-难的。

- SAT是NP完全问题。