第十讲

✧ CFG 的简化及 Chomsky 范式

CFG 的简化及Chomsky 范式

- ✧ 消去无用符号
- ✧ 消去 ε 产生式
- ✧ 消去 Unit 产生式
- ✧ ***CFG* 的简化**
- ✧ ***Chomsky* 范式**

消去无用符号

✧ 有用符号和无用符号

– 有用符号（*useful symbol*）

对于 **CFG** $G = (V, T, P, S)$，称符号 $X \in V \cup T$ 是有用的，当且仅当 $S \overset{*}{\Rightarrow} \alpha X \beta \overset{*}{\Rightarrow} w$，其中 $w \in T^*$，$\alpha, \beta \in (V \cup T)^*$.

– 无用符号（*useless symbol*）

即非有用符号

消去无用符号

✧ 生成符号和可达符号

- 称符号 X 是生成符号（generating symbol），当且仅当存在 $w \in T^*$，满足 $X \stackrel{*}{\Rightarrow} w$.
- 称符号 X 是可达符号（reachable symbol），当且仅当存在 $\alpha, \beta \in (V \cup T)^*$，满足 $S \stackrel{*}{\Rightarrow} \alpha X \beta$.

✧ 与有用符号的关系

- 有用符号一定是生成符号和可达符号
- 反之，成立吗？

$$S \to AB \mid a$$
$$B \to b$$

消去无用符号

- 消去所有非生成符号
- 消去所有非可达符号
- 结果（定理7.2）
  - 剩余的符号都是有用符号
  - 新的文法与原来的文法是等价的

消去无用符号

✧ 消去非生成符号及不可达符号

设 CFG $G=(V, T, P, S)$，并假定$L(G)\neq\phi$（即 S 一定是一个生成符号），通过下列步骤（次序不能变）可以得到消去 G 中无用符号后的 CFG $G_1=(V_1, T_1, P_1, S)$:

（1）从G 中删除所有非生成符号以及所有包含这些符号的产生式，得到CFG $G_2=(V_2, T_2, P_2, S)$；

（2）从G_2 中删除所有不可达符号以及所有包含这些符号的产生式，得到CFG $G_1=(V_1, T_1, P_1, S)$；

✧ **结论** 通过上述步骤，G_1不包含无用符号，且$L(G_1)= L(G)$.

证明思路 一方面，G_1中不包含无用符号；另一方面，对任何 w ，$S \overset{*}{\underset{G_1}{\Rightarrow}} w$ iff $S \overset{*}{\underset{G}{\Rightarrow}} w$.

消去无用符号

✧ 步骤

- 计算生成符号集合
- 计算可达符号集合
- 消去非生成符号及不可达符号

次序是敏感的，例如：

$$S \to AB \mid a$$
$$B \to b$$

消去无用符号

✧ 计算生成符号集

– **步骤** 对于 CFG $G = (V, T, P, S)$，可通过下列归纳步骤计算生成符号集合:

基础 任何终结符 $a \in T$ 都是生成符号;

归纳 如果有产生式 $A \to \alpha$，其中 $\alpha \in (V \cup T)^*$ 的每一个符号都是生成符号，则 A 也是生成符号;

– **结论** 此步骤求出所有并只能求出 G 的生成符号

证明思路 一方面，所得到的符号的确是生成符号；另一方面，所有的生成符号都可由上述步骤得到.

消去无用符号

✧ 计算可达符号集

– 步骤 对于 CFG $G = (V, T, P, S)$，可通过下列归纳步骤计算可达符号集合：

基础 S 是可达符号；

归纳 如果 A 是可达符号，并且有产生式 $A \rightarrow \alpha$，其中 $\alpha \in (V \cup T)^*$，则 α 中的符号都是可达符号；

– 结论 上述步骤可以求出所有并只能求出 G 的可达符号.

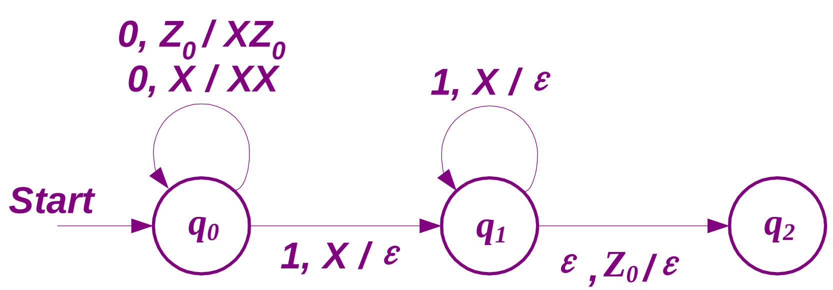
证明思路 一方面，所得到的符号的确是可达符号；另一方面所有的可达符号都可由上述步骤得到. （留作练习）

消去无用符号

✧ 消去非生成符号及不可达符号

– 举例 对于右下图的 PDA，构造CFG $G = (V,\{0,1\},P,S)$，其中 $V = \{S\} \cup \{ [pYq] \mid p,q \in \{q_0,q_1,q_2\} \wedge Y \in \{Z_0,X\} \}$

产生式集合 P 定义如下：



(1) $S \to [q_0Z_0q_0]$;
$S \to [q_0Z_0q_1]$;
$S \to [q_0Z_0q_2]$;

(2) $[q_0Z_0q_j] \to 0[q_0Xq_i]\,[q_iZ_0q_j]$, $i, j = 0,1,2$; $((q_0,XZ_0) \in \delta(q_0, 0, Z_0))$

(3) $[q_0Xq_j] \to 0[q_0Xq_i]\,[q_iXq_j]$, $i , j = 0,1,2$; $((q_0,XX) \in \delta(q_0, 0, X))$

(4) $[q_0Xq_1] \to 1$; $((q_1,\varepsilon) \in \delta(q_0, 1, X))$

(5) $[q_1Xq_1] \to 1$; $((q_1,\varepsilon) \in \delta(q_1, 1, X))$

(6) $[q_1Z_0q_2] \to \varepsilon$; $((q_2,\varepsilon) \in \delta(q_1, \varepsilon, Z_0))$

消去无用符号

✧ 消去非生成符号及不可达符号

– 举例（续前页）

$S \to [q_0Z_0q_0]$; $S \to [q_0Z_0q_1]$; $S \to q_0Z_0q_2]$;

$[q_0Z_0q_j] \to 0[q_0Xq_i]\,[q_iZ_0q_j]$, $i, j = 0,1,2$;

$[q_0Xq_j] \to 0[q_0Xq_i]\,[q_iXq_j]$, $i , j = 0,1,2$;

$[q_0Xq_1] \to 1$; $[q_1Xq_1] \to 1$; $[q_1Z_0q_2] \to \varepsilon$;

上述文法的生成符号包括:

终结符号: 0, 1

非终结符号: $[q_0Xq_1]$, $[q_1Xq_1]$, $[q_1Z_0q_2]$,
$[q_0Z_0q_2]$, S

消去无用符号

✧ 消去非生成符号及不可达符号

– 举例（续前页）

消去所有非生成符号，得到新文法：

$$S \to [q_0Z_0q_2]\,;$$
$$[q_0Z_0q_2] \to 0[q_0Xq_1]\,[q_1Z_0q_2]$$
$$[q_0Xq_1] \to 0[q_0Xq_1]\,[q_1Xq_1]$$
$$[q_0Xq_1] \to 1; \quad [q_1Xq_1] \to 1; \quad [q_1Z_0q_2] \to \varepsilon\,;$$

为简捷，记 $[q_0Z_0q_2]$ 为 A，$[q_0Xq_1]$ 为 B， $[q_1Xq_1]$ 为 C，$[q_1Z_0q_2]$ 为 D，上述文法的产生式改写如下：

$$S \to A; \quad A \to 0BD; \quad B \to 0BC;$$
$$B \to 1;\ C \to 1;\ D \to \varepsilon\,;$$

消去无用符号

✧ 消去非生成符号及不可达符号

– 举例（续前页）

以下产生式表示的文法中，*S*、*A*、*B*、*C*、*D*、*0*、*1* 都为可达符号，所以消除不可达符号后，结果没有变化：

$$S \rightarrow A;\quad A \rightarrow 0BD;\quad B \rightarrow 0BC;$$

$$B \rightarrow 1;\ C \rightarrow 1;\ D \rightarrow \varepsilon;$$

至此，该文法已经消去了所有无用符号.

消去 ε 产生式

✧ 目的 方便文法的设计, 利于文法规范化.

✧ 影响 消去 ε 产生式, 除文法不能产生串 ε 外, 不会影响到原文法相应的语言中其它字符串的产生.

✧ 可致空符号 (*nullable symbol*)

对于 CFG $G = (V, T, P, S)$, 称符号 $A \in V$ 是可致空的, 当且仅当 $A \overset{*}{\Rightarrow} \varepsilon$.

消去 ε 产生式及其影响, 需要计算可致空符号的集合.

消去ε产生式

✧ 计算可致空符号集

– 步骤 对于 CFG $G = (V, T, P, S)$，可通过下列归纳步骤计算可致空符号集合:

基础 对所有产生式 $A \to \varepsilon$，A 是一个可致空符号;

归纳 如果有产生式 $B \to C_1C_2\ldots C_k$，其中每一个 $C_i \in V$ 是可致空符号，则 B 是一个可致空符号;

– 结论 此步骤求出所有并只能求出G 的可致空符号。

证明思路 一方面，得到的符号的确是可致空符号;另一方面所有的可致空符号都可由上述步骤得到.

消去 ε 产生式

✧ 消去 ε 产生式及所有可致空符号的影响

设 CFG $G = (V, T, P, S)$，通过下列步骤可以得到消去 G 中 ε 产生式及其影响，由此得到 CFG $G_1 = (V, T, P_1, S)$:

（1）计算 G 的可致空符号集合;

（2）对每一产生式 $A \to A_1A_2\ldots A_k$ ，在 G_1 中对应有一组产生式，每一个可致空符号都可能出现或不出现；若包含 $m < k$ 个可致空符号，则该产生式能够对应G_1 中的 2^m 个产生式；若包含 k 个可致空符号，则该产生式能够对应G_1 中的 2^k-1 个产生式;

（3）G_1 中不包含 G 的所有 ε 产生式：$A \to ε$.

消去 ε 产生式

✧ 举例

以下产生式表示的文法中，D 为可致空符号：

$S \to A$; $A \to 0BD$; $B \to 0BC$;

$B \to 1$; $C \to 1$; $D \to \varepsilon$.

消去 ε 产生式，得到如下产生式集合：

$S \to A$; $A \to 0BD$; $A \to 0B$; $B \to 0BC$;

$B \to 1$; $C \to 1$.

✧ 结论 通过上述步骤从 G 构造 G_1，满足 $L(G_1)= L(G) - \{\varepsilon\}$.

证明思路 欲证对任何 w，$S \overset{*}{\underset{G_1}{\Rightarrow}} w$ iff $(S \overset{*}{\underset{G}{\Rightarrow}} w \wedge w \neq \varepsilon)$.

消去 Unit 产生式

✧ Unit 产生式（*unit productions*） 形如 $A \to B$ 的产生式，其中*A*、*B* 为非终结符.

✧ 消去 Unit 产生式的目的 可简化某些证明，减少推导步数, 利于文法规范化.

✧ Unit 偶对（*unit pairs*）

对于 CFG $G = (V, T, P, S)$，A，$B \in V$ ，称 (A, B) 是 Unit 偶对，当且仅当 $A \overset{*}{\Rightarrow} B$ ，且该推导过程仅使用过 Unit 产生式.

消去 Unit 产生式时，需计算所有 Unit 偶对的集合

消去 Unit 产生式

✧ 计算 Unit 偶对的集合

– **步骤** 对于 CFG $G = (V, T, P, S)$，可通过下列归纳步骤计算所有 Unit 偶对的集合：

基础 对于任何 $A \in V$，(A,A) 是一个 Unit 偶对；

归纳 如果 (A,B) 是一个 Unit 偶对，及 $B \rightarrow C$ 是产生式（$C \in V$），则 (A,C) 是一个 Unit 偶对.

– **结论** 上述步骤恰好能求出 G 的所有 Unit 偶对：

证明思路 一方面，得到的偶对的确是 Unit 偶对；另一方面所有的 Unit 偶对都可由上述步骤得到.

消去 Unit 产生式

✧ 步骤

设 CFG $G = (V, T, P, S)$，通过下列步骤消去 G 中的 Unit 产生式，由此得到 CFG $G_1 = (V, T, P_1, S)$:

（1）计算 G 的 Unit 偶对集合;

（2）对每个 Unit 偶对 (A, B)，在 G_1 中加入产生式 $A \rightarrow \alpha$，其中 $B \rightarrow \alpha$ 为一个非 Unit 产生式;

（3）G_1 中包含 G 的所有非 Unit 产生式.

✧ 结论 通过上述步骤从 G 构造 G_1，$L(G_1) = L(G)$.

证明思路 欲证对任何 w，$S \overset{*}{\underset{G_1}{\Rightarrow}} w$ iff $S \overset{*}{\underset{G}{\Rightarrow}} w$.

消去 Unit 产生式

✧ 举例

以下产生式表示的文法中，Unit 偶对包括 (S，S)，(A，A)，(B，B)，(C，C)，(D，D) 以及 (S，A)：

$S \to A$; $A \to 0BD$; $A \to 0B$; $B \to 0BC$;

$B \to 1$; $C \to 1$.

通过上述步骤消去 Unit 产生式，得到产生式集合：

$S \to 0BD$; $S \to 0B$; $A \to 0BD$; $A \to 0B$;
$B \to 0BC$; $B \to 1$; $C \to 1$.

CFG 的简化

✧ **小结**　设 CFG $G = (V, T, P, S)$，可以通过下列步骤对 G 进行简化：

（1）先消除 ε 产生式；

（2）再 Unit 产生式；

（3）最后消除无用符号.

✧ **注意** 以上简化步骤的次序.

✧ **结论** 设 CFG G 的语言至少包含一个非 ε 的字符串，通过上述步骤从 G 构造 G_1，则有

$L(G_1)= L(G) - \{\varepsilon\}$.

Chomsky 范式

✧ **Chomsky 范式 CNF**（*Chomsky Normal Form*）

任何不含 ε 的非空 **CFL**都存在一个 **CFG G**，其产生式具有如下两种简单形式之一:

1. $A \rightarrow BC$，其中 A，B，C 都是非终结符;

2. $A \rightarrow a$，其中 A 是非终结符，a 是终结符;

并且 **G** 中不包含无用符号.

这样的文法形式称为**Chomsky 范式**.

Chomsky 范式

✧ 如何获得 *Chomsky* 范式

任何不含 ε 的非空 *CFL*（上下文无关语言）都存在一个 *CFG G*，首先通过下列步骤对 *G* 进行简化，得到不含 ε 产生式、不含 Unit 产生式以及不含无用符号的文法 G_2：

（*1*）消除 ε 产生式;

（*2*）消除 Unit 产生式;

（*3*）消除无用符号.

由前所述，$L(G_2)= L(G)$.

Chomsky 范式

✧ 如何获得 **Chomsky** 范式（续前页）

举例　以下产生式表示的文法中，已经不存在 ε 产生式和 Unit 产生式:

$$S \rightarrow 0BD;\ S \rightarrow 0B;\ A \rightarrow 0BD;\ A \rightarrow 0B;$$
$$B \rightarrow 0BC;\ B \rightarrow 1;\ C \rightarrow 1.$$

进一步消去无用符号 *D*，*A* 后，得到如下产生式集合:

$$S \rightarrow 0B;\quad B \rightarrow 0BC;\quad B \rightarrow 1;\quad C \rightarrow 1.$$

Chomsky 范式

✧ 如何获得 **Chomsky** 范式（续前页）

由于 **CFG** G_2 不含 ε 产生式，因此每个产生式的右部的长度至少为 *1*；

又由于 **CFG** G_2 不含 Unit 产生式，因此一个产生式的右部如果长度为 *1*，则只可能形如 $A \rightarrow a$（其中A 是非终结符，a 是终结符），已经符合 **CNF** 的要求.

举例　前面步骤后得到的产生式集合:

$S \rightarrow 0B$;　$B \rightarrow 0BC$;　$B \rightarrow 1$;　$C \rightarrow 1$.

Chomsky 范式

✧ 如何获得 Chomsky 范式（续前页）

下一步，将 G_2 做如下变换，得到CFG G_1：

（1）如果某一终结符 *a* 出现于某些右部长度大于 1 的产生式中，则引入一个新的非终结符，如 *A*，将这些产生式中的 *a* 替换为 *A*，并增加新的产生式 $A \rightarrow a$；至此，右部长度大于 1 的产生式中只包含有非终结符

Chomsky 范式

- 如何获得 Chomsky 范式（续前页）

举例　前面步骤后得到的产生式集合:

$S \to 0B$;　$B \to 0BC$;　$B \to 1$;　$C \to 1$.

引入新的非终结符 A，用 A 替换 0，并增加新的产生式 $A \to 0$，得到如下产生式集合:

$S \to AB$; $B \to ABC$; $A \to 0$; $B \to 1$; $C \to 1$.

✧ 如何获得 Chomsky 范式（续前页）

（2）将右部长度大于2的产生式采用级连（cascade）的方法转变为只包含两个非终结符；如对于产生式

$$A \to B_1B_2\ldots B_k\,,\quad k>2,$$

引入k-2个新的非终结符C_1，C_2，…，C_{k-2}，则将原产生式替换为以下一组产生式:

$$A \to B_1C_1\,,\ C_1 \to B_2C_2\,,\ \ldots,\ C_{k-3} \to B_{k-2}C_{k-2}\,,$$
$$C_{k-2} \to B_{k-1}B_k\,.$$

Chomsky 范式

◇ 如何获得 Chomsky 范式（续前页）

举例　前面步骤后得到的产生式集合:

$S \to AB;\ \ B \to ABC;\ \ A \to 0;\ \ B \to 1;\ \ C \to 1.$

进一步引入新的非终结符 D，将其变换为:

$S \to AB;\quad B \to AD;\quad D \to BC;$

$A \to 0;\quad B \to 1;\quad C \to 1.$

◇ 结论　设 CFG G 的语言至少包含一个非 ε 的字符串，通过上述步骤从 G 构造 G_1，则 G_1 符合 CNF 的要求，且满足 $L(G_1)= L(G) - \{\varepsilon\}$.

课后练习

✧ 必做题:

- *Ex.7.1.3*
- *Ex.7.1.9(b)*

✧ 思考题:

- **!Ex.7.1.10*

That's all for today.

Thank You