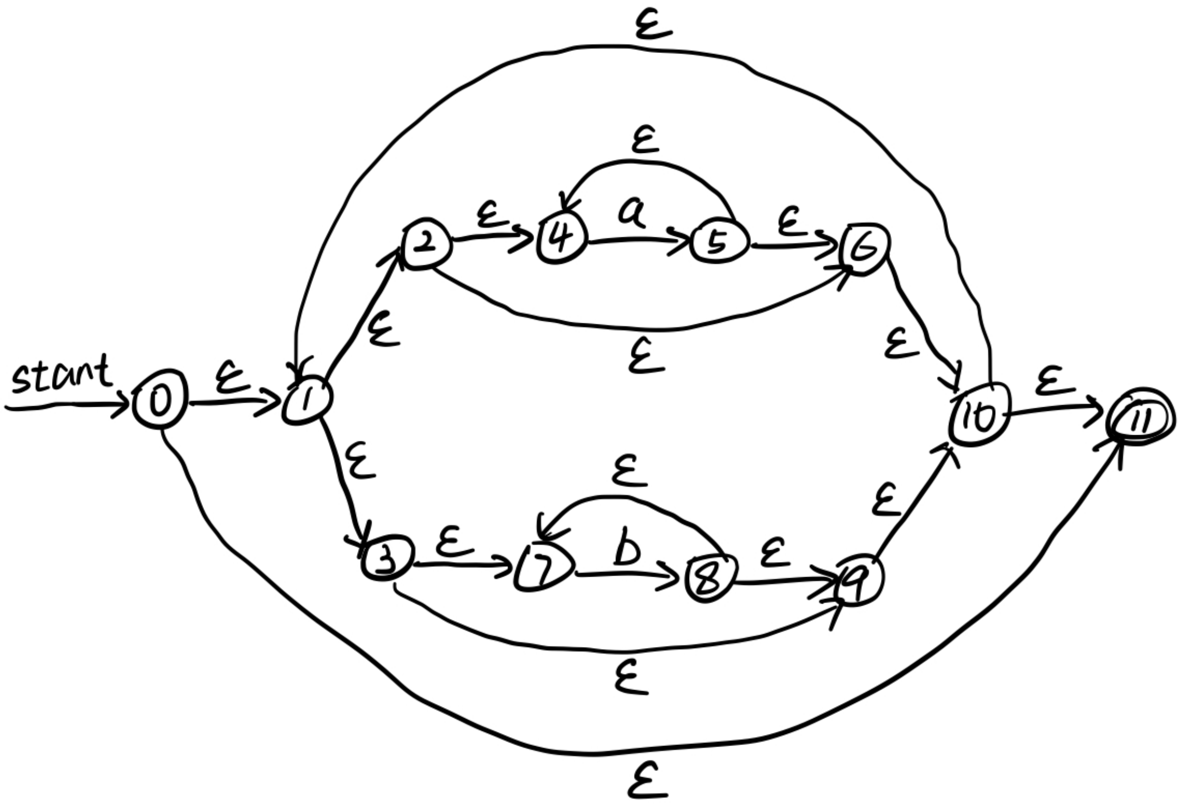
编译原理第二次作业

201300035 方盛俊

Ex. 3.7.3 (2)

将正则表达式 `(a*|b*)*` 转为 NFA 有:



通过子集构造法将 NFA 转换为 DFA 有:

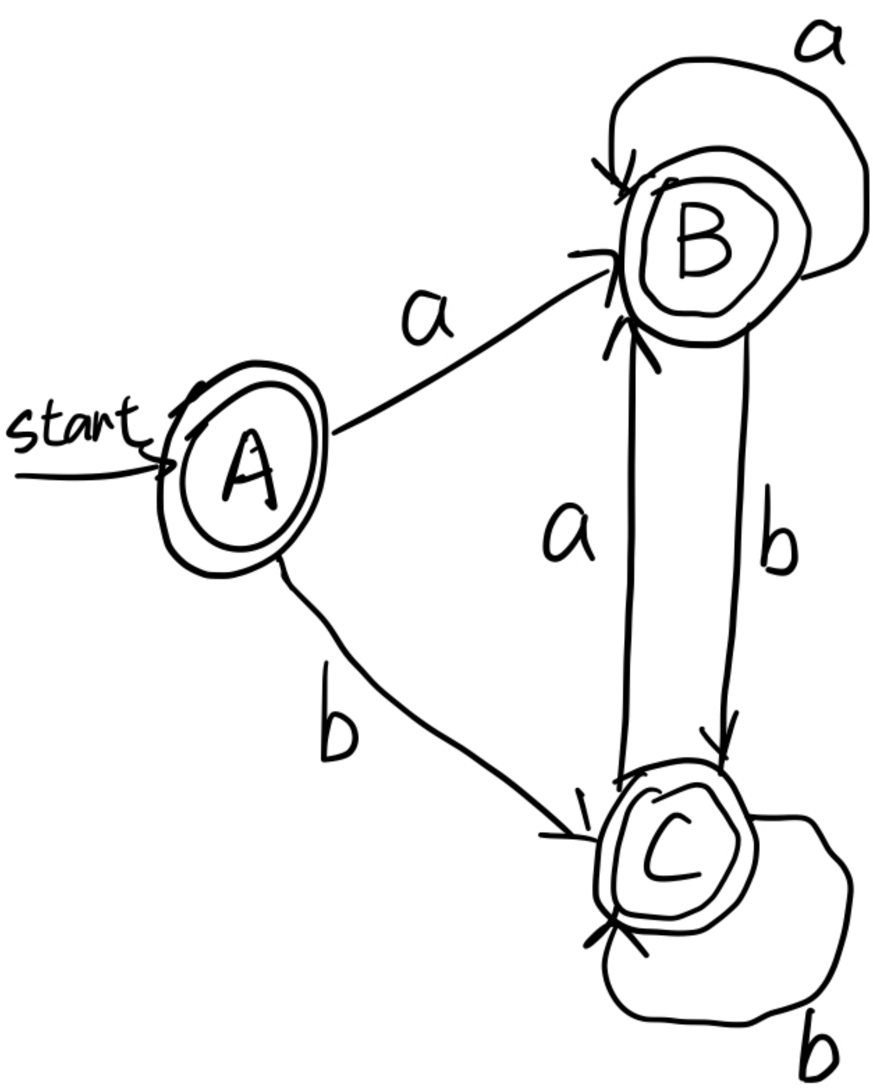
- A: ε-closure(0) = { 0, 1, 2, 3, 4, 6, 7, 9, 10, 11 }
- B: Dtran[A, a] = ε-closure(move(A, a)) = ε-closure({ 5 }) = { 1, 2, 3, 4, 5, 6, 7, 9, 10, 11 }
- C: Dtran[A, b] = ε-closure(move(A, b)) = ε-closure({ 8 }) = { 1, 2, 3, 4, 6, 7, 8, 9, 10, 11 }
- Dtran[B, a] = ε-closure({ 5 }) = B
- Dtran[B, b] = ε-closure({ 8 }) = C
- Dtran[C, a] = ε-closure({ 5 }) = B
- Dtran[C, b] = ε-closure({ 8 }) = C

则有 DFA:

- 开始状态: A
- 接受状态: A, B, C

NFA 状态	DFA 状态	a	b
{ 0, 1, 2, 3, 4, 6, 7, 9, 10, 11 }	A	B	C
{ 1, 2, 3, 4, 5, 6, 7, 9, 10, 11 }	B	B	C
{ 1, 2, 3, 4, 6, 7, 8, 9, 10, 11 }	C	B	C

作图即可得:



Ex. 4.2.1

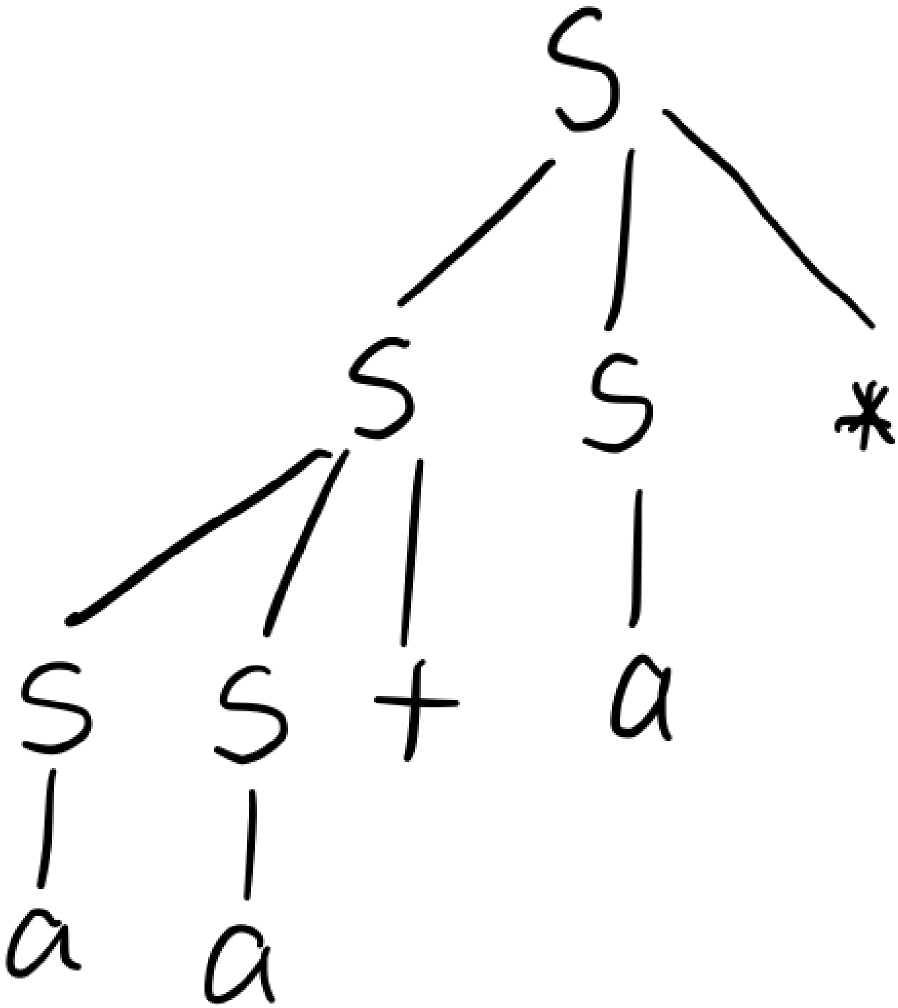
(1)

$$S \underset{lm}{\Longrightarrow} SS* \underset{lm}{\Longrightarrow} SS+S* \underset{lm}{\Longrightarrow} aS+S* \underset{lm}{\Longrightarrow} aa+S* \underset{lm}{\Longrightarrow} aa+a*$$

(2)

$$S \underset{rm}{\Longrightarrow} SS* \underset{rm}{\Longrightarrow} Sa* \underset{rm}{\Longrightarrow} SS+a* \underset{rm}{\Longrightarrow} Sa+a* \underset{rm}{\Longrightarrow} aa+a*$$

(3)



(4)

不是二义性的.

证明:

令 G 为文法 $S \rightarrow SS + |SS * |a$.

首先我们证明 $L(G)$ 是 p-表达式组成的集合.

我们称由 $\{a, +, *\}$ 组成的符号串 ω 为 p-表达式, 等且仅当:

- ω 不是空串;
- 对于任何 ω 的一个真后缀 φ 来说, 符号 a 的个数小于或等于操作符 $\{+, *\}$ 的个数, 不妨记作 $|\varphi|_a \leq |\varphi|_+$;

- ω 中符号 a 的个数恰好比操作符的个数多一个, 记作 $|\omega|_a - |\omega|_+ = 1$;
- 除 $\omega = a$ 的情况外, ω 以 aa 开头, 且以任意一个操作符结尾.

我们使用数学归纳法证明 $L(G)$ 是 p-表达式组成的集合.

归纳基础: 当 $\omega = a$ 或 $\omega = aa+$ 或 $\omega = aa*$ 时, 易见以上三个条件均满足, 此时 ω 是 p-表达式.

归纳假设: 当 $|\omega| < n$ 时, ω 是 p-表达式.

归纳步骤:

当 $|\omega| = n$ 时, 不失一般性, 我们认为 ω 是由 $S \to SS+$ 推导得到的, $S \to SS*$ 同理, 则 ω 有形式 $\omega_1\omega_2+$, 其中 $\omega_1, \omega_2 \in L(G)$.

由归纳假设可得, ω_1 和 ω_2 是 p-表达式.

对于第一个条件, 易见由于 $|\omega| = n$, 有 ω 不是空串.

对于第二个条件, 令 $\varphi+$ 是 $\omega = \omega_1\omega_2+$ 的一个真后缀:

- 若 $\varphi = \varepsilon$, 则 $\varphi+ = +$ 满足 $|\varphi+|_a = 0 \leq |\varphi+|_+ = 1$.
- 若 φ 是 ω_2 的真后缀, 则由于 ω_2 是 p-表达式, 我们可知 $|\varphi|_a \leq |\varphi|_+$, 则 $|\varphi+|_a = |\varphi|_a \leq |\varphi|_+ < |\varphi+|_+$ 同样满足 $|\varphi+|_a \leq |\varphi+|_+$.
- 若 $\varphi = \omega_2$, 则由于 ω_2 是 p-表达式, 我们可知 $|\omega_2|_a = |\omega_2|_+ + 1$, 则 $|\varphi+|_a = |\omega_2|_a = |\omega_2|_+ + 1 = |\omega_2+|_+ = |\varphi+|_+$, 满足 $|\varphi+|_a \leq |\varphi+|_+$.
- 若 $\varphi = \varphi_1\omega_2$, 我们可知 $\varphi_1\omega_2+$ 中 a 的个数与操作符个数的差值, 等于 φ_1 中 a 的个数与操作符个数的差值, 即 $|\varphi_1\omega_2+|_a - |\varphi_1\omega_2+|_+ = |\varphi_1|_a - |\varphi_1|_+$. 由于 φ_1 是 ω_1 的真后缀 (由 φ 是 ω 的真后缀可知 $\varphi_1 \neq \omega_1$), 且 ω_1 是 p-表达式, 因此 $|\varphi_1|_a - |\varphi_1|_+ \leq 0$, 则有 $|\varphi_1\omega_2+|_a - |\varphi_1\omega_2+|_+ \leq 0$ 满足 $|\varphi|_a \leq |\varphi|_+$.

对于第三个条件, $|\omega|_a - |\omega|_+ = |\omega_1\omega_2+|_a - |\omega_1\omega_2+|_+ = (|\omega_1|_a - |\omega_1|_+) + (|\omega_2|_a - |\omega_2|_+) - 1 = 1 + 1 - 1 = 1$ 满足条件.

对于第四个条件, 若 $\omega_1 = a$, 则 $\omega_1\omega_2+ = a\omega_2+$ 必然以 aa 开头, 以 $+$ 结尾. 若 $\omega_1 \neq a$, 则 $|\omega_1| \geq 3$, 且由 ω_1 为 p-表达式, 则 ω_1 也是以 aa 开头的, 则 $\omega_1\omega_2+ = a\omega_2+$ 必然以 aa 开头, 以 $+$ 结尾.

由以上归纳证明可知, $L(G)$ 是 p-表达式组成的集合.

接着我们使用数学归纳法证明 G 是无二义性的, 其中 $\omega \in L(G)$.

归纳基础: 当 $\omega = a$ 时, 易见 $S \to a$ 是 ω 唯一的推导, ω 有着唯一的语法树.

归纳假设: 当 $|\omega| < n$ 时, ω 有着唯一的语法树.

归纳步骤:

当 $|\omega| = n$ 时, 不失一般性, 我们认为 ω 是由 $S \to SS+$ 推导得到的, $S \to SS*$ 同理, 则 ω 有形式 $\omega_1\omega_2+$, 其中 $\omega_1, \omega_2 \in L(G)$.

如果 ω 没有唯一的语法树, 则我们会有多个符合条件的 $\omega_1\omega_2$ 对, 以 ω_2 为例, 不同的 ω_2 有着不同的长度 $|\omega_2|$, 我们要证明的就是 $\omega_1\omega_2$ 对是唯一的, 即 $|\omega_2|$ 是唯一的.

从右到左, 我们逐个扫描符号, 获取当前的真后缀 $\varphi+$, 我们始终追踪 $|\varphi+|_a - |\varphi+|_+$ 的值:

- 若 $\varphi = \varepsilon$, 则 $|\varphi+|_a - |\varphi+|_+ = -1$
- 若 φ 是 ω_2 的真后缀, 则 $|\varphi+|_a - |\varphi+|_+ = |\varphi|_a - |\varphi|_+ - 1 \leq -1$.
- 若 $\varphi = \omega_2$, 则 $|\varphi+|_a - |\varphi+|_+ = |\omega|_a - |\omega|_+ - 1 = 1 - 1 = 0$.
- 若 $\varphi = \varphi_1\omega_2$,
  - 若 $\omega_1 = a$, 则已经扫描了所有真后缀 $\varphi+$ 了, 已经终止了.
  - 若 $\omega_1 \neq a$, 则 ω_1 以 aa 开头,
    - 若 φ_1 是 ω_1 的真后缀且不包括 ω_1 开头的 aa 中的 a, 则 $|\varphi+|_a - |\varphi+|_+ = |\varphi_1|_a - |\varphi_1|_+ < 0$.
    - 若 φ_1 是 ω_1 的真后缀且包括 ω_1 开头的 aa 中的一个 a, 则 $|\varphi+|_a - |\varphi+|_+ = |\varphi_1|_a - |\varphi_1|_+ = 0$.

可以看出, 扫描真后缀 $\varphi+$ 时, 最多只有两次 $|\varphi+|_a - |\varphi+|_+ = 0$, 且只有第一次 $|\varphi+|_a - |\varphi+|_+ = 0$ 时, 我们有 $\varphi = \omega_2$.

因此我们证明了 $\omega_1\omega_2$ 对是唯一的, 说明 $|\omega_2|$ 是唯一的, 即 ω_2 是唯一的, 因此 ω_1 也是唯一的, ω 有着唯一的语法树.

由以上归纳证明可知, $S \to SS + |SS * |a$ 是无二义性的.

(5)

包含加法 "+", 乘法 "*" 和操作数 "a" 的后缀表达式.

Ex. 4.2.3 (1)

该文法应该对应正则表达式 `1*(01+)*`, 则文法为

- `S -> 1S | A`
- `A -> 0BA | ε`
- `B -> 1B | 1`