

Chapter 3

讨论

0.1

请简要说明自动机的用途

自动机, 是一种 计算模型, 用于描述 有限状态 的 计算过程. 可以通过这种模型, 抽象化/规范化/形式化 的 描述, 识别, 分析 编程语言中的 词法 & 结构

0.2

为什么文法定义一个名字时, 产生式左边的文法符号只要能用 \rightarrow , 而不能用 $=$ 呢?

因为 $=$ 在 编程语言 中, 通常用于 赋值/等值比较, 为了避免混淆, 所以 文法 中使用 \rightarrow

实际上 \rightarrow 表示 右值由左值推出, $=$ 在 赋值 语义下表示 左值由右值得到 (更接近 \leftarrow), $=$ 在 比较 语义下表示 左值相当于右值 (由于不存在推出/得到 的逻辑关系, 故更接近于 \leftrightarrow)

0.3

已知正规式 r_1 和 r_2 , 请问如何证明 r_1 和 r_2 是否等价?

先用 Thompson 算法, 将两个 正规式 分别转化为 有限自动机 fa_1 & fa_2 , 再将两个有限自动机 确定化 为 dfa_1 & dfa_2 , 最后将 dfa_1 & dfa_2 化简, 如果可以得到 相同的 最小确定有限自动机, 那么两个正规式就是 完全等价的

作业

1.1

已知正规式:

$$r = b^*(a|bbb^*)b^*$$

给出与之等价的最小 DFA M

下午9:06 10月13日周五

大创

编译原理

$r = b^*(a|bbb^*)b^*$ (优先级 $*$ $>$ $|$)

	I	Ia	Ib
1	$\langle 1 \rangle$	$\langle \varnothing \rangle$	$\langle 1, 2 \rangle$
2	$\langle 1, 2 \rangle$	$\langle \varnothing \rangle$	$\langle 1, 2, 3 \rangle$
3	$\langle 1, 2, 3 \rangle$	$\langle \varnothing \rangle$	$\langle 1, 2, 3, \varnothing \rangle$
4	$\langle 1, 2, 3, \varnothing \rangle$	$\langle \varnothing \rangle$	$\langle 1, 2, 3, \varnothing \rangle$
5	$\langle \varnothing \rangle$	\varnothing	$\langle \varnothing \rangle$

已验证: 最小

1.2

已知正规式:

$$r = (a^*b)^*ba(b|a)^*$$

给出与之等价的最小 DFA M

1.2 错误答案

下午9:14 10月13日周五

大创

编译原理

$r = (a^*b)^*ba(b|a)^*$

(NFA)

(DFA)

→ DFAmin

I	Ia	Ib
0 <1>	1 <1>	2 <1, 2>
2 <1, 2>	1 <1>	3 <1, 2, 3>
3 <1, 2, 3>	4 <1, φ>	3 <1, 2, 3>
4 <1, φ>	4 <1, φ>	5 <1, 2, φ>
5 <1, 2, φ>	4 <1, φ>	5 <1, 2, φ>

$\langle 1, 2 | 3 \times \phi, 5 \rangle // \langle 1, 2, 3 \rangle a = \langle 1, \phi \rangle$

$\langle 1 | 2 \rangle \langle 3 \rangle \langle \phi, 5 \rangle // \langle 1, 2 \rangle a = \langle 1 \rangle$

$\langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle \phi, 5 \rangle // \langle 1, 2 \rangle b = \langle 2, 3 \rangle$

$\langle 1 \rangle \langle 2 \rangle \langle 3 \rangle \langle \phi, 5 \rangle$

1.2 正确答案

上午2:20 10月17日周二

大创

编译原理

修正: $r = (a^*b)^*ba(b|a)^*$

(NFA)

(DFA)

明显已为最小!

I	Ia	Ia	Ib	Ib
0 <0, 1>	<0>	<0, 1>	<2>	<0, 1>
1 <0, 2>	<0>	<0, 1>	<3>	<3>
2 <3>	<φ>	<φ>	φ	φ
3 <4>	<φ>	<φ>	<φ>	<φ>
4 φ	φ	φ	φ	φ

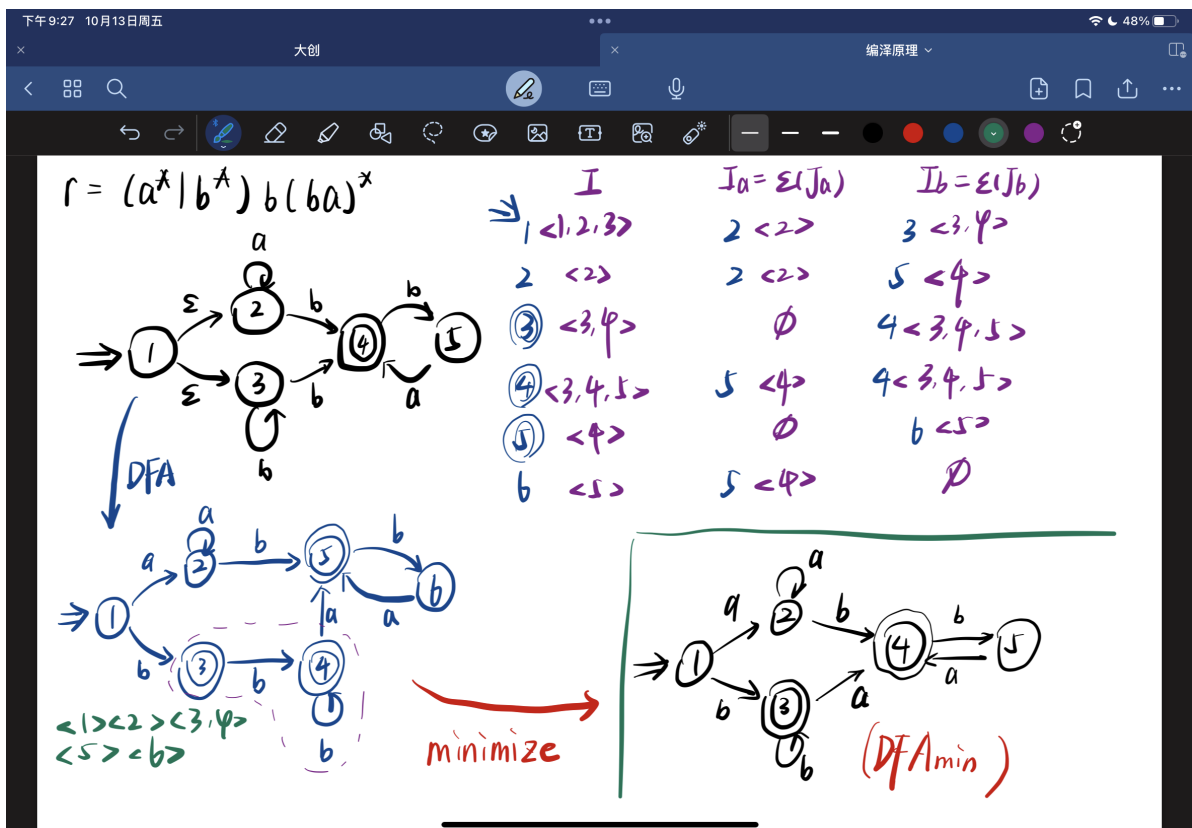
1.3

已知正规式:

$$r = (a^*|b^*)b(ba)^*$$

给出与之等价的最小 DFA M

1.3 错误答案



最开始的 NFA 的第一个节点有问题, 表示的是

$$(b|a)^*bba(b|a)^*$$

1.3 正确答案

上午 2:22 10月17日周二

大创

编译原理

$r = (a^* | b^*) b (ba)^*$

⇒

DFA

⇒

⇒ 不操作因为 $\delta(3, a) = \phi$ 也是一种状态

ϕ 作为错误处理

一定要画!!!

所以以后一定要画出 ϕ 作为错误处理

$\phi \Rightarrow$ 显式标明

状态转移表:

	$I_a = \epsilon(Ia)$	$I_b = \epsilon(Ib)$
1	$\langle 1, 2, 3 \rangle$	$\langle 3, \phi \rangle$
2	$\langle 2 \rangle$	$\langle 5, \langle 4 \rangle \rangle$
3	$\langle 3, \phi \rangle$	$\langle 4, \langle 3, 4, 5 \rangle \rangle$
4	$\langle 3, 4, 5 \rangle$	$\langle 4, \langle 3, 4, 5 \rangle \rangle$
5	$\langle 4 \rangle$	$\langle \phi \rangle$
6	$\langle 5 \rangle$	$\langle \phi \rangle$
7	ϕ	ϕ

细节! 注意细节!

1.4

现在已知一个语言 L , 它具有以下的特征:

$$E = \{0, 1\}$$

L 由 E 构成, 是一个值能被 4 整除的, 开头不为 0 的, 二进制数字串

现在请列出:

1. L 的正规式
2. L 等价的 最小确定有限自动机

对于 复合问题, 一定要 分解成原子问题 分别考虑

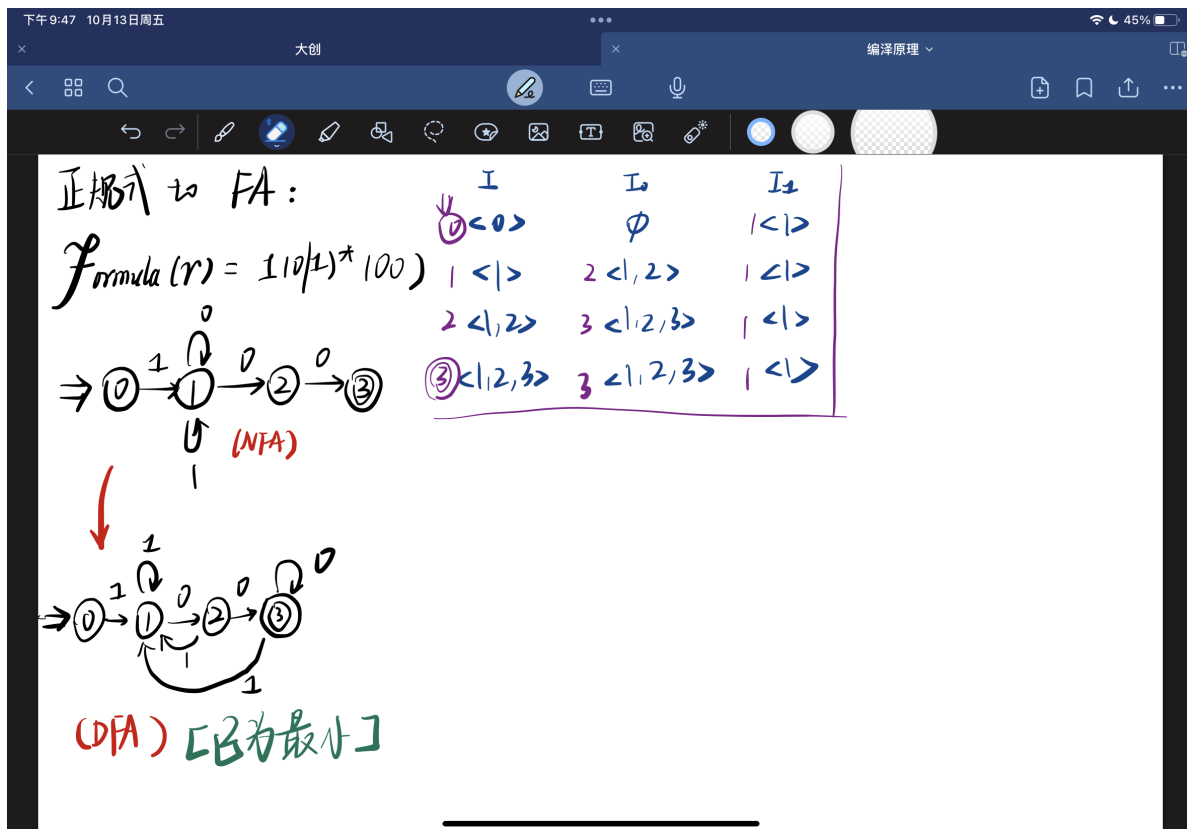
1. 值可以被 4 整除 :: 必须以 00 结尾, 证明略, 凡是结尾为 00 的二进制串, 值 ≥ 4 必然成立, 且一定都是 4 的倍数
2. 开头不为 0 的 二进制子串 :: $1(0|1)^*$

因此, 我们可以得出:

- L 的 正规式:

$$Formal(r) : r = 1(0|1)^*00$$

- L 等价的 最小确定有限自动机:



1.5

对于 $\{0, 1\}$ 上包含子串 010 的所有数字串, 列出:

1. 它的正规式
2. 它的最小确定有限自动机

包含特定子串, 一般来说, 构造难度低于 不包含特定子串 (尤其是本题这种仅限定一种 模式 的情况)

对于

$$\Sigma = \{a_m, \dots, a_n\}$$

这样的字母表, 不限定模式, 能够构造出的 规模最大 的 正规式 的描述如下:

$$Formal(r) : r = (a_m | a_{m+1} | \dots | a_{n-1} | a_n)^*$$

按照这个结论, 可以得出 $(0|1)^*$ 是 $\{0, 1\}$ 不限定模式时, 产生的规模最大的正规式

限定单一子串, 只需要将指定子串两侧用规模最大的正规式包裹即可, 因此得出:

- 它的正规式:

$$Formal(r) : r = (0|1)^* 010 (0|1)^*$$

- 它的最小确定有限自动机:

下午 8:53 10月13日周五

大创

编译原理

复杂正规式: 包含子串

$\Sigma = \{0, 1\}$, 包含子串 '010'

NFA

DFA (min)

已为最简

I	I ₀	I ₁
$\Rightarrow 0 <0>$	$1 <0, 1>$	$0 <0>$
$1 <0, 1>$	$1 <0, 1>$	$2 <0, 2>$
$2 <0, 2>$	$3 <0, 3>$	$0 <0>$
$3 <0, 3>$	$3 <0, 3>$	$3 <0, 3>$
$<0, 1, 2> <3> <0, 1, 2>_0 = <1, 3>$		
$<0, 1, 2> <3> <0, 1, 2>_1 = <0, 2>$		
$<0> <2> <3> <0, 1>_0 = <0>$		
$<0> <2> <3> <0, 1>_1 = <0, 2>$		
$<0> <1> <2> <3>$		

1.6

给定了一个右线性文法 G:

$$G(S) : S \rightarrow 0S | 1S | 1A | 0B$$

$$A \rightarrow 1C | 1$$

$$B \rightarrow 0C | 0$$

$$C \rightarrow 0C | 1C | 0 | 1$$

求出与它等价的一个 左线性文法

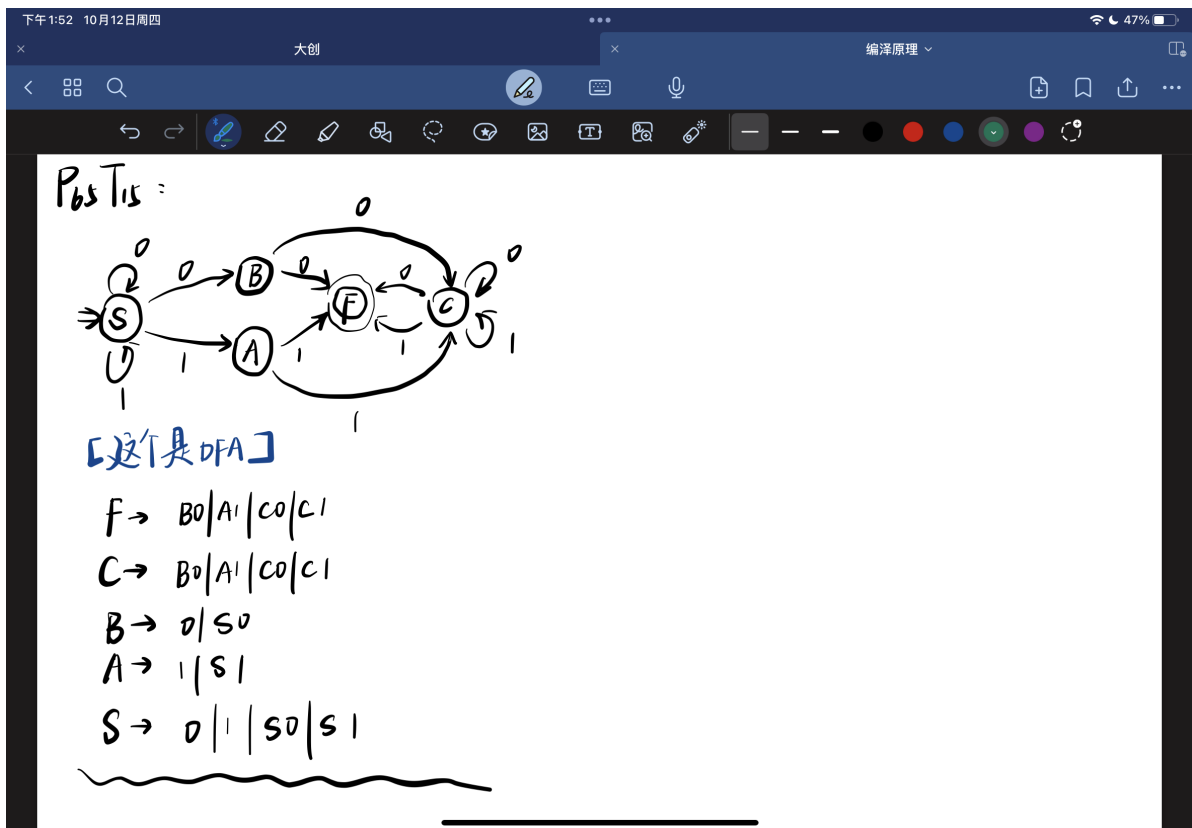
想要解决这个问题, 只要根据 G 构造 DFA, 再根据 DFA 构造 G_L (左线性文法) 即可

结论:

注意: 这里一定要写 $G_L(F)$ 就像前面写 $G(S)$ 一样, 标记文法的起始符号

$$\begin{aligned} F &\rightarrow B0|A1|C0|C1 \\ C &\rightarrow B0|A1|C0|C1 \\ G(S) &\Leftrightarrow G_L(F) : \begin{aligned} B &\rightarrow 0|S0 \\ A &\rightarrow 1|S1 \\ S &\rightarrow 0|1|S0|S1 \end{aligned} \end{aligned}$$

细节:



拓展

2.1

对于 $\{0, 1\}$ 上 不包含 子串 010 的所有数字串, 列出:

1. 它的正规式
2. 它的最小确定有限自动机

不包含 子串问题, 考虑的出发点 (初态), 一定是 字母表 中 不同于 给定子串 开头字母 的 元素

格外注意, 不包含 子串, 隐含 长度不限, 因此完全有可能允许: 多个 单字符 结尾的情况

分析本题:

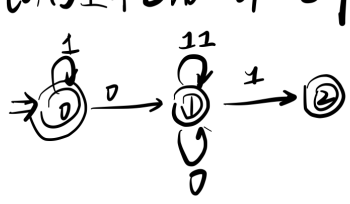
2.1 不规范分析

下午10:01 10月13日周五

大创

编译原理

0,1上不含子串010的所有数字串


$$r = r_1 | r_2 | r_3 = 1^* | 1^* 0 (11^* 0)^* | 1^* 0 (11^* 0)^* 1$$
$$r_1 = 1^*$$
$$r_2 = 1^* 0 (11^* 0)^*$$
$$r_3 = r_2 1$$

2.1 规范分析

上午 2:23 10月17日周二

大创

编译原理

Pos Tis :

```

graph LR
    S((S)) -- 0 --> S
    S -- 1 --> A((A))
    A -- 1 --> F(((F)))
    F -- 0 --> B((B))
    B -- 0 --> F
    F -- 0 --> C((C))
    C -- 1 --> F
    C -- 0 --> C
  
```

[这个是 DFA]

$GF = F \rightarrow B0|A1|C0|C1$

$C \rightarrow B0|A1|C0|C1$

$B \rightarrow 0|S0$

$A \rightarrow 1|S1$

$S \rightarrow 0|1|S0|S1$

7/26