作业 L1.6 - L1.9：NFA 构造、正则表达式与正则变换

2025 年 12 月 17 日

说明（按题面顺序）

题面 `assets/problem_L01_06-L01_09.pdf` 中存在**重复题目**：

- 第 2 页的“题 1”与第 1 页的“题 1/题 2”部分重复

字母表约定

除非题目另行说明，本作业中所有 NFA/正则表达式的字母表均为 $\Sigma = \{0, 1\}$。

题目 1

（PDF 第 1 页·题 1）构造识别如下语言的 NFA，且符合规定的状态数：

(a) $\{w \mid w \text{ 以00 结束}\}$，3 个状态

(b) $\{w \mid w \text{ 含有偶数个0 或恰好 2 个1}\}$，6 个状态

Solution.

(a) 以 00 结束（3 状态）

取状态集合 $\{q_0, q_1, q_2\}$，初态 q_0，终态 $\{q_2\}$。定义转移：

$$\begin{aligned}
&\delta(q_0, 0) = \{q_0, q_1\}, && \delta(q_0, 1) = \{q_0\},\\
&\delta(q_1, 0) = \{q_2\}, && \delta(q_1, 1) = \emptyset,\\
&\delta(q_2, 0) = \{q_2\}, && \delta(q_2, 1) = \emptyset.
\end{aligned}$$

直观：q_0 在任意前缀上自环；当读到某个 0 时非确定性“猜测它是倒数第二个 0”，转入 q_1，再读到 0 到达接受态 q_2。

(b) 偶数个 0 或恰好 2 个 1（6 状态）

分别构造两个自动机并用 ϵ-并联实现并集。

分支 1：偶数个 0（2 状态 DFA 亦可视作 NFA）状态 $\{e, o\}$：e 表示已读到偶数个 0，o 表示奇数个 0；初态 e，终态 $\{e\}$。

$$\delta(e, 0) = o,\ \delta(o, 0) = e, \qquad \delta(e, 1) = e,\ \delta(o, 1) = o.$$

分支 2：恰好 2 个 1（3 状态 NFA）状态 $\{p_0, p_1, p_2\}$ 表示已读到的 1 的个数为 0/1/2；初态 p_0，终态 $\{p_2\}$，读到第 3 个 1 无路可走：

$$\begin{aligned} \delta(p_0,0) &= \{p_0\}, & \delta(p_0,1) &= \{p_1\}, \\ \delta(p_1,0) &= \{p_1\}, & \delta(p_1,1) &= \{p_2\}, \\ \delta(p_2,0) &= \{p_2\}, & \delta(p_2,1) &= \emptyset. \end{aligned}$$

并联合成（总 6 状态）新增总初态 s：

$$\delta(s,\epsilon) = \{e, p_0\},$$

总终态为 $\{e, p_2\}$。总状态数 $1+2+3=6$，满足题目要求。

题目 2

（PDF 第 1 页·题 2）构造识别如下语言的 NFA：

$$A = \{w \mid w \text{ 含有子串} 0101\}, \qquad B = \{w \mid w \text{ 不含子串} 110\}.$$

并构造识别 $A \cup B$、$A \circ B$（连接）与 B^* 的 NFA。

Solution.

(1) A：含有子串 0101（5 状态）

用“匹配到前缀长度”的 NFA（a_0 初态，a_4 终态）：

$$\begin{aligned} \delta(a_0,0) &= \{a_0, a_1\}, & \delta(a_0,1) &= \{a_0\}, \\ \delta(a_1,1) &= \{a_2\}, & \delta(a_1,0) &= \emptyset, \\ \delta(a_2,0) &= \{a_3\}, & \delta(a_2,1) &= \emptyset, \\ \delta(a_3,1) &= \{a_4\}, & \delta(a_3,0) &= \emptyset, \\ \delta(a_4,0) &= \{a_4\}, & \delta(a_4,1) &= \{a_4\}. \end{aligned}$$

其中 a_0 的自环实现“任意前缀”，并在读到某个 0 时非确定性开始匹配。

(2) B：不含子串 110（DFA 亦可视作 NFA）

状态 $\{b_0, b_1, b_2, \perp\}$，含义分别为“末尾不是 1”“末尾是 1”“末尾是 11”“已出现 110（陷阱）”。初态 b_0，终态 $\{b_0, b_1, b_2\}$：

$$\begin{aligned} \delta(b_0,0) &= b_0, & \delta(b_0,1) &= b_1, \\ \delta(b_1,0) &= b_0, & \delta(b_1,1) &= b_2, \\ \delta(b_2,1) &= b_2, & \delta(b_2,0) &= \perp, \\ \delta(\perp,0) &= \perp, & \delta(\perp,1) &= \perp. \end{aligned}$$

(3) $A \cup B$（并集）

新增初态 s，$\delta(s,\epsilon) = \{a_0, b_0\}$，终态取 $F_A \cup F_B$ 即可。

(4) $A \circ B$（连接）

把 A 的接受态 a_4 用 ϵ 边连到 B 的初态 b_0，并把整体接受态设为 B 的接受态 $\{b_0, b_1, b_2\}$。

(5) B^*（星号）

新增接受初态 s（用于接受 ϵ），$\delta(s,\epsilon)=\{b_0\}$，并对每个 B 的接受态 $f\in\{b_0,b_1,b_2\}$ 加回边 $\delta(f,\epsilon)\ni b_0$。

题目 3

（PDF 第 2 页·题 1）给出识别下述语言的 NFA，且符合规定的状态数（题面与前面有重复）：

(a) $\{w \mid w$ 以00 结束$\}$，3 个状态（**与第 1 页·题 1(a) 重复**）

(b) $0^*1^*0^+$，3 个状态（**新增**）

(c) $\{w \mid w$ 含有子串0101$\}$，5 个状态（**与第 1 页·题 2 的 A 重复**）

(d) $\{w \mid w$ 含有偶数个0 或恰好 2 个1$\}$，6 个状态（**与第 1 页·题 1(b) 重复**）

Solution.

(a) 重复说明： 同第 1 页·题 1(a)。
(c) 重复说明： 同第 1 页·题 2 中对 A 的 5 状态构造。
(d) 重复说明： 同第 1 页·题 1(b) 的 6 状态并联构造。
(b) $0^*1^*0^+$ **（3 状态）**

该语言表示：先若干个 0，再若干个 1，最后至少一个 0（且一旦进入 1 段后，只能在末尾进入 0 段并停留在 0 段）。给出 3 状态 DFA（也是 NFA）：

状态 $\{q_0,q_1,q_2\}$，初态 q_0，终态 $\{q_2\}$：

$$\begin{aligned}
\delta(q_0,0)&=q_0, & \delta(q_0,1)&=q_1,\\
\delta(q_1,1)&=q_1, & \delta(q_1,0)&=q_2,\\
\delta(q_2,0)&=q_2, & \delta(q_2,1)&=\emptyset.
\end{aligned}$$

其中 q_2 表示已经进入末尾的 0^+ 段；若再读到 1 则不可能回到 1^*，因此直接拒绝（无转移）。

题目 4

（PDF 第 3 页·题 2）给出识别下述语言**并集**的 NFA 状态图：

$$L=\{w \mid w\ \text{从1 开始且以0 结束}\}\cup\{w \mid w\ \text{含有至少 3 个1}\}.$$

Solution.

分别为两个语言构造 DFA/NFA，再用 ϵ-并联实现并集（与前面 $A\cup B$ 的做法相同）。

语言 1：从 1 开始且以 0 结束
状态 $\{s, \alpha, \beta, \perp\}$，$s$ 初态，β 终态：

$$\begin{aligned} \delta(s,1) &= \alpha, & \delta(s,0) &= \perp, \\ \delta(\alpha,1) &= \alpha, & \delta(\alpha,0) &= \beta, \\ \delta(\beta,0) &= \beta, & \delta(\beta,1) &= \alpha, \\ \delta(\perp,0) &= \perp, & \delta(\perp,1) &= \perp. \end{aligned}$$

语言 2：至少 3 个 1
状态 $\{c_0, c_1, c_2, c_3\}$，c_0 初态，c_3 终态：

$$\delta(c_i,0) = c_i, \quad \delta(c_0,1) = c_1,\ \delta(c_1,1) = c_2,\ \delta(c_2,1) = c_3,\ \delta(c_3,1) = c_3.$$

并联： 新增初态 s_*，$\delta(s_*, \epsilon) = \{s, c_0\}$，终态取 $\{\beta, c_3\}$。

题目 5

（PDF 第 3 页·题 3）给出识别下述语言**连接**的 NFA 状态图：

$$L = \{w \mid |w| \le 5\} \circ \{w \mid w \text{ 的奇数位置均为1}\}.$$

Solution.

语言 1： $|w| \le 5$
用 6 个状态计长度：l_0（初态）到 l_5 全为终态；读入任意符号推进；超过 5 无路：

$$\delta(l_i,0) = l_{i+1},\ \delta(l_i,1) = l_{i+1} \quad (0 \le i \le 4), \qquad \delta(l_5,0) = \delta(l_5,1) = \emptyset.$$

语言 2：奇数位置均为 1
状态 $\{O, E, \perp\}$：O 表示“下一位是奇数位”，E 表示“下一位是偶数位”，$\perp$ 陷阱。初态 O，终态 $\{O, E\}$：

$$\delta(O,1) = E,\ \delta(O,0) = \perp, \qquad \delta(E,0) = O,\ \delta(E,1) = O, \qquad \delta(\perp,0) = \delta(\perp,1) = \perp.$$

连接： 把长度机的每个终态 l_i（$0 \le i \le 5$）用 ϵ 边连到 O，整体终态为 $\{O, E\}$。

题目 6

（PDF 第 3 页·题 4）给出识别下述语言**星号**的 NFA 状态图：

(a) $\{w \mid w \text{ 含有至少 3 个1}\}$

(b) $\{w \mid w$ 含有至少 2 个0 且至多含有 1 个1$\}$

Solution.

做法：先给出识别 L 的 NFA（或 DFA），再按 Kleene 星标准构造：新增接受初态 s，$\delta(s,\epsilon) = \{原初态\}$，并从原每个接受态加 ϵ 回到原初态。

(a) 至少 3 个 1 的语言 L

可直接复用上题"至少 3 个 1"的 DFA（c_0, c_1, c_2, c_3），将其按上述方式变为 L^*。

(b) 至少 2 个 0 且至多 1 个 1 的语言 K

用 DFA 记录 (z, o)：$z \in \{0,1,2\}$ 表示 0 的个数（2 表示 ≥ 2），$o \in \{0,1,2\}$ 表示 1 的个数（2 表示 ≥ 2，为陷阱）。初态 $(0,0)$，终态为 $\{(2,0),(2,1)\}$，转移（饱和计数）：

$$\delta((z,o),0) = (\min(z+1,2), o), \qquad \delta((z,o),1) = (z, \min(o+1,2)).$$

再按 Kleene 星构造得到 K^*。

题目 7

（PDF 第 4 页·题 1）写出表示下列语言的正则表达式：

(1) 字母表 $\{a,b,c\}$ 上包含至少一个 a 和至少一个 b 的串

(2) 倒数第 10 个符号是 1 的 0/1 串

(3) 至多只有一对连续 1 的 0/1 串

(4) 0 的个数被 5 整除的 0/1 串

(5) 不包括 101 作为子串的所有 0/1 串

(6) 0 的个数被 5 整除且 1 的个数是偶数的所有 0/1 串

Solution.

(1) 至少一个 a 且至少一个 b（$\Sigma = \{a,b,c\}$）

令 $T = (a+b+c)$，则：

$$T^*aT^*bT^* + T^*bT^*aT^*.$$

(2) 倒数第 10 个符号是 1（$\Sigma = \{0,1\}$）

$$(0+1)^*\,1\,(0+1)^9.$$

(3) 至多只有一对连续 1

先记"不含 11"的语言为 $N = (0+10)^*(\epsilon+1)$。"恰好一次出现 11"可写为：前缀必须以 0 或空结尾（避免形成 111），后缀必须以 0 或空开头（避免形成 111）：

$$(0+10)^*\,11\,(\epsilon + 0(0+10)^*(\epsilon+1)).$$

所以答案为

$$N+(0+10)^*\,11\,(\epsilon+0(0+10)^*(\epsilon+1)).$$

(4) 0 的个数被 5 整除

$$(1^*01^*01^*01^*01^*0)^*1^*.$$

(5) 不含子串 101
等价表述：所有的 1-块之间至少隔两个 0（否则会出现 $\cdots 1\,0\,1\cdots$）。因此：

$$0^*\big(\epsilon+1^+(00\,0^*\,1^+)^*0^*\big).$$

(6) 0 的个数被 5 整除且 1 的个数为偶数
可构造一个 DFA 以记录“0 的个数 mod 5”与“1 的个数 mod 2”的直积状态（共 10 状态），因此该语言正则。用状态消除法可得到一个等价正则表达式，例如：

$$\begin{aligned}R_6=\Big(&11+((01+10)(11)^*00+(00+(01+10)(11)^*(10+01))(11)^*(10+01))\\&\cdot(11+0(00(11)^*00+00(11)^*(10+01)(11)^*(10+01)))^*001\\&+((00+(01+10)(11)^*(10+01))(11)^*00\\&\quad+((01+10)(11)^*00+(00+(01+10)(11)^*(10+01))(11)^*(10+01))\\&\quad\cdot(11+0(00(11)^*00+00(11)^*(10+01)(11)^*(10+01)))^*\\&\quad\cdot(10+0(1+00(11)^*(10+01)(11)^*00))\Big)\\&\cdot\Big(1(1+00(11)^*(10+01)(11)^*00)\\&\quad+1(00(11)^*00+00(11)^*(10+01)(11)^*(10+01))\\&\quad\cdot(11+0(00(11)^*00+00(11)^*(10+01)(11)^*(10+01)))^*\\&\quad\cdot(10+0(1+00(11)^*(10+01)(11)^*00))\Big)^*\\&\cdot(0+101+1(00(11)^*00+00(11)^*(10+01)(11)^*(10+01))\\&\quad\cdot(11+0(00(11)^*00+00(11)^*(10+01)(11)^*(10+01)))^*001\Big)^*.\end{aligned}$$

（该表达式较长，但与题目语言等价，且满足“正则表达式”形式要求。）

题目 8

（PDF 第 5 页·题 2）给出下列正则表达式语言的自然语言描述：

(1) $(1+\epsilon)(00^*1)^*0^*$

(2) $(0^*1^*)^*000(0+1)^*$

(3) $(0+10)^*1^*$

Solution.

(1) $(1+\epsilon)(00^*1)^*0^*$

可选地以一个 1 开头（也可以不选），然后重复若干次“若干个 0（至少 1 个）再跟一个 1”的块，最后以若干个 0 结束。因此该语言中的串满足：除可能存在的首个 1 外，每个 1 的前面至少有一个 0；并且串以一段 0 结束。

(2) $(0^*1^*)^*000(0+1)^*$

因为 $(0^*1^*)^* = \Sigma^*$（任意 0/1 串都可分割成若干段“若干 0 后接若干 1”），所以该表达式等价于

$$\Sigma^*000\Sigma^*,$$

即：包含子串 000 的所有 0/1 串。

(3) $(0+10)^*1^*$

前半部分 $(0+10)^*$ 表示可由若干个 0 或 10 拼接得到（因此该部分若出现 1，则该 1 必紧跟一个 0），再接一个末尾的 1^*。等价描述：除末尾可能出现的一段连续 1 外，其余位置的每个 1 都必须紧跟一个 0。

题目 9

（PDF 第 6 页）把下列正则表达式转换成带 ϵ 转移的 NFA：

$$1)\ 01^*,\qquad 2)\ (0+1)01,\qquad 3)\ 00(0+1)^*.$$

Solution.

用 Thompson 构造法即可。下面用“状态 + 边”的方式给出 ϵ-NFA（也可直接画成状态图）。

(1) 01^*

状态 $\{s,u,v,f\}$，初态 s，终态 f：

$$s \xrightarrow{0} u,\quad u \xrightarrow{\epsilon} f,\quad u \xrightarrow{1} v,\quad v \xrightarrow{\epsilon} u.$$

其中 $u \to f$ 允许取 1^* 的 0 次，$u \xrightarrow{1} v \xrightarrow{\epsilon} u$ 实现循环。

(2) $(0+1)01$

先做并：$0+1$，再连接 01。状态 $\{s,a,b,c,d,f\}$，初态 s，终态 f：

$$\begin{gathered} s \xrightarrow{\epsilon} a,\ s \xrightarrow{\epsilon} b,\\ a \xrightarrow{0} c,\quad b \xrightarrow{1} c,\\ c \xrightarrow{0} d,\quad d \xrightarrow{1} f. \end{gathered}$$

(3) $00(0+1)^*$

先用两条边读入前缀 00，再对 $(0+1)^*$ 加星号。状态 $\{s,u,v,x,y,f\}$，初态 s，终态 f：

$$\begin{gathered} s \xrightarrow{0} u,\quad u \xrightarrow{0} v,\\ v \xrightarrow{\epsilon} f,\quad v \xrightarrow{\epsilon} x,\\ x \xrightarrow{0} y,\ x \xrightarrow{1} y,\quad y \xrightarrow{\epsilon} x,\ y \xrightarrow{\epsilon} f. \end{gathered}$$

其中 $v \to f$ 对应星号取 0 次；$x \xrightarrow{0/1} y$ 再回到 x 实现任意多次。

题目 10

（PDF 第 7 页）利用本节课讲授的方法，把题面给出的 DFA 转换成正则表达式（两小题，见图 (a)(b)）。

Solution.

(a) 两状态 DFA：a 自环，b 在两状态间切换，且仅状态 2 为终态

该 DFA 接受“b 的出现次数为奇数”的语言（a 不改变状态，b 翻转状态）。正则表达式可写为:

$$\boxed{a^*(ba^*ba^*)^*ba^*}.$$

(b) 三状态 DFA：初态 1 与状态 3 为终态（见题图）

设 $K = (a+b)(a+bb)^*ba$。用 Arden 引理/状态消除法可得从初态到终态的语言为:

$$\boxed{K^*\big(\epsilon + (a+b)(a+bb)^*b\big)}.$$