作业 L1.10 - L1.12：正则表达式与 Arden 引理

2025 年 12 月 16 日

题目 1

用分配律化简如下正则表达式，得到两个不同但更简单的等价表达式：

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

Solution.

设 $X=(0+1)$，原式为 X^*1XX+X^*1X。

化简 1：右因子提取

$$X^*1XX+X^*1X=X^*1X(X+\epsilon).$$

其中 $(X+\epsilon)$ 表示“可选的一个 X”，因此该式更简洁地表达为“在某个 1 之后跟 1 个或 2 个符号”。

化简 2：左因子提取

$$X^*1XX+X^*1X=X^*1(XX+X)=X^*1X(X+\epsilon).$$

两种写法分别是提取右因子/左因子，均比原式更紧凑，且等价。

题目 2

证明：$(L+M)^*=(L^*M^*)^*$。

Solution.

证明两边语言互相包含。

(1) $(L+M)^*\subseteq(L^*M^*)^*$

任取 $w\in(L+M)^*$，则 w 可写为有限个因子连接：

$$w=x_1x_2\cdots x_n,\quad x_i\in L \text{ 或 } x_i\in M.$$

把相邻的 L-因子合并成一个 L^*-块，相邻的 M-因子合并成一个 M^*-块，可得

$$w=y_1y_2\cdots y_k,$$

其中每个 y_j 属于 L^* 或 M^*，并且它们在序列中交替出现（可能从 L^* 或 M^* 开始）。将相邻的一对 $(L^*)(M^*)$ 视为一个块（允许其中一个为空串 ϵ），则每个块属于 L^*M^*，因此 $w \in (L^*M^*)^*$。

(2) $(L^*M^*)^* \subseteq (L+M)^*$

任取 $w \in (L^*M^*)^*$，则

$$w = z_1 z_2 \cdots z_t, \quad z_i \in L^*M^*.$$

每个 z_i 可写为 $z_i = uv$，其中 $u \in L^*$，$v \in M^*$。而 u 是若干个 L 中串的连接，v 是若干个 M 中串的连接，因此 z_i 也是由若干个属于 L 或 M 的串连接得到，所以 $z_i \in (L+M)^*$。于是 w 作为若干个 z_i 的连接仍属于 $(L+M)^*$。

综上，两边相等：$(L+M)^* = (L^*M^*)^*$。

L1.12：利用 Arden 引理求正则表达式

题目 3

利用 Arden 引理将如下有穷自动机转换成正则表达式。

(1) 三状态自动机：起始态 A，接受态 C。转移为

$$A \xrightarrow{a} A, \quad A \xrightarrow{a} B, \quad B \xrightarrow{b} B, \quad B \xrightarrow{b} A, \quad B \xrightarrow{a} C, \quad C \xrightarrow{b} B.$$

(2) 三状态自动机：起始态 q_1，接受态 q_3。转移为

$$q_1 \xrightarrow{a} q_1, \quad q_1 \xrightarrow{b} q_2, \quad q_2 \xrightarrow{b} q_2, \quad q_2 \xrightarrow{a} q_3, \quad q_2 \xrightarrow{a} q_1, \quad q_3 \xrightarrow{a} q_2.$$

Solution.

对每个状态 X，令 R_X 表示“从起始态出发到达 X 的所有串”所构成的正则表达式（语言）。

(1) 自动机 (A,B,C)

由转移关系写方程：

$$\begin{aligned} R_A &= \epsilon + R_A a + R_B b, \\ R_B &= R_A a + R_B b + R_C b, \\ R_C &= R_B a. \end{aligned}$$

先由第三式代入第二式：

$$R_B = R_A a + R_B b + (R_B a)b = R_A a + R_B(b + ab).$$

由 Arden 引理（若 $X = XA + Y$ 且 $\epsilon \notin A$，则 $X = YA^*$）得

$$R_B = R_A a(b + ab)^*.$$

代回第一式：

$$R_A = \epsilon + R_A a + \big(R_A a(b+ab)^*\big)b = \epsilon + R_A\big(a + a(b+ab)^* b\big).$$

再次用 Arden 引理：

$$R_A = \epsilon\big(a + a(b+ab)^* b\big)^*.$$

因此
$$R_C = R_B a = R_A a(b+ab)^* a = \big(a + a(b+ab)^* b\big)^* a(b+ab)^* a.$$
接受态为 C，所以所求正则表达式可取:
$$\boxed{\big(a + a(b+ab)^* b\big)^* a(b+ab)^* a}.$$

(2) 自动机 (q_1, q_2, q_3)
写方程:
$$R_1 = \epsilon + R_1 a + R_2 a,$$
$$R_2 = R_1 b + R_2 b + R_3 a,$$
$$R_3 = R_2 a.$$
代入 $R_3 = R_2 a$ 到第二式:
$$R_2 = R_1 b + R_2 b + (R_2 a)a = R_1 b + R_2(b + aa).$$
由 Arden 引理:
$$R_2 = R_1 b(b+aa)^*.$$
代回第一式:
$$R_1 = \epsilon + R_1 a + \big(R_1 b(b+aa)^*\big)a = \epsilon + R_1\big(a + b(b+aa)^* a\big),$$
因此
$$R_1 = \big(a + b(b+aa)^* a\big)^*.$$
接受态为 q_3，故
$$R_3 = R_2 a = R_1 b(b+aa)^* a = \big(a + b(b+aa)^* a\big)^* b(b+aa)^* a.$$
所以可取正则表达式:
$$\boxed{\big(a + b(b+aa)^* a\big)^* b(b+aa)^* a}.$$