第3次作业　正则语言的性质

4.1.2 证明下列语言都不是正则的

e) 由 0 和 1 构成的ww形式的串的集合，也就是某个串重复的串集合

证明：不妨称该语言为L。假设L是正则的，那么存在$n \in Z^{+}$，对$\forall s \in L(|s| \geq n)$满足泵引理。

取$s = 0^{n}1^{n}0^{n}1^{n}$，显然$s \in L$且$|0^{n}1^{n}0^{n}1^{n}| = 4n > n$，那么s可被分为$s = xyz$，满足$|xy| \leq n$且$y \neq \varepsilon$，因此y只能为$0^{m}(m > 0)$。那么$xy^{2}z = 0^{n+m}1^{n}0^{n}1^{n}(m > 0) \notin L$，而由泵引理知$xy^{2}z \in L$，矛盾。因此假设不成立，该语言不是正则的。

f) 由 0 和 1 构成的ww^{R}形式的串的集合，也就是由某个串后面跟着它的反转所构成的串的集合

证明：不妨称该语言为L。假设L是正则的，那么存在$n \in Z^{+}$，对$\forall s \in L(|s| \geq n)$满足泵引理。

取$s = 0^{n}1^{n}1^{n}0^{n}$，显然$s \in L$且$|0^{n}1^{n}1^{n}0^{n}| = 4n > n$，那么s可被分为$s = xyz$，满足$|xy| \leq n$且$y \neq \varepsilon$，因此y只能为$0^{m}(m > 0)$。那么$xy^{2}z = 0^{n+m}1^{n}1^{n}0^{n}(m > 0) \notin L$，而由泵引理知$xy^{2}z \in L$，矛盾。因此假设不成立，该语言不是正则的。

g) 由 0 和 1 构成的$w\bar{w}$形式的串的集合，其中$\bar{w}$是把w中所有的 0 都换成 1 同时把所有的 1 都换成 0 而得到的串，例如，$\overline{011} = 100$，因此 011100 是该语言中的一个串

证明：不妨称该语言为L。假设L是正则的，那么存在$n \in Z^{+}$，对$\forall s \in L(|s| \geq n)$满足泵引理。

取$s = 0^{n}1^{n}$，显然$s \in L$且$|0^{n}1^{n}| = 2n > n$，那么s可被分为$s = xyz$，满足$|xy| \leq n$且$y \neq \varepsilon$，因此y只能为$0^{m}(m > 0)$。那么$xy^{2}z = 0^{n+m}1^{n}(m > 0) \notin L$，而由泵引理知$xy^{2}z \in L$，矛盾。因此假设不成立，该语言不是正则的。

h) 所有由 0 和 1 构成的$w1^{n}$形式的串的集合，其中w是由 0 和 1 构成的长度为 n 的串

证明：不妨称该语言为L。假设L是正则的，那么存在$p \in Z^{+}$，对$\forall s \in L(|s| = 2n \geq p)$满足泵引理。取$s = 0^{n}1^{n}$，显然$s \in L$且$|0^{n}1^{n}| = 2n \geq p$，那么s可被分为$s = xyz$，满足$|xy| \leq p$且$y \neq \varepsilon$。由于$y \neq \varepsilon$，那么$|xy^{2}z|$必然大于$2n$，又L所包括的所有串的长度均为$2n$，故$xy^{2}z \notin L$。而由泵引理知$xy^{2}z \in L$，矛盾。因此假设不成立，该语言不是正则的。

4.1.3证明下列语言都不是正则的

a) 所有满足以下条件的串的集合：由 0 和 1 构成，开头的是 1，并且当我们把该串看作是一个整数时该整数是一个素数

证明：不妨称该语言为L。假设L是正则的，那么存在$n \in Z^+$，对$\forall w \in L(|w| \geqslant n)$满足泵引理。已知素数可以是任意大的，不妨取素数p，p对应的二进制串为$w(|w| \geqslant n)$，那么w可被分为$w = xyz$，满足$|xy| \leqslant n$且$y \neq \varepsilon$，不妨令x,y,z对应的十进制数为d_x, d_y, d_z。

将p表示为$p = d_x \cdot 2^{|y|+|z|} + d_y \cdot 2^{|z|} + d_z$，而由泵引理知$xy^pz$同样属于L

其对应的十进制数为$q = d_x \cdot 2^{p|y|+|z|} + d_y \cdot 2^{|z|}(1 + 2^{|y|} + 2^{2|y|} + \ldots + 2^{(p-1)|y|}) + d_z$

不妨令$S = 1 + 2^{|y|} + 2^{2|y|} + \ldots + 2^{(p-1)|y|} = \dfrac{2^{p|y|} - 1}{2^{|y|} - 1}$，则$q = d_x \cdot 2^{p|y|+|z|} + d_y \cdot 2^{|z|}S + d_z$。

又由费马小定理知$2^{p-1} \equiv 1(\bmod p)$，则有$2^{(p-1)|y|} \equiv 1(\bmod p)$，有$2^{p|y|} \equiv 2^{|y|}(\bmod p)$

所以$(2^{|y|} - 1)S = 2^{p|y|} - 1 \equiv 2^{|y|} - 1(\bmod p)$，因此$S \equiv 1(\bmod p)$。则有

$$
\begin{aligned}
q &= d_x \cdot 2^{p|y|+|z|} + d_y \cdot 2^{|z|}S + d_z \\
&\equiv d_x \cdot 2^{|y|+|z|} \cdot 2^{(p-1)|y|} + d_y \cdot 2^{|z|}S + d_z \\
&\equiv d_x \cdot 2^{|y|+|z|} + d_y \cdot 2^{|z|} + d_z \\
&\equiv p(\bmod p)
\end{aligned}
$$

因此$p|q$，q不是素数，这与泵引理矛盾。所以L不是正则的。

b) 所有满足以下条件的0^i1^j形式的串的集合：i 和 j 的最大公约数是 1

证明：不妨称该语言为L。假设L是正则的，那么存在$n \in Z^+$，对$\forall w \in L(|w| \geqslant n)$满足泵引理。取$w = 0^p1^{(p-1)!}$（p 是大于 n 的最小素数），显然$w \in L$且$|0^p1^{(p-1)!}| = p + (p-1)! > n$，那么 w 可被分为$w = xyz$，满足$|xy| \leqslant n$且$y \neq \varepsilon$，因此 y 只能为$0^m(m > 0)$。那么$xy^0z = 0^{p-m}1^{(p-1)!}(m > 0)$，因为$p - m > 1, \gcd((p-1)!, p-m) > 1$，所以$xy^0z \notin L$，而由泵引理知$xy^0z \in L$，矛盾。因此假设不成立，该语言不是正则的。

4.2.2

教材 4.2.2 如果 L 是一个语言，a 是一个符号，则 L/a(称作 L 和 a 的商)是所有满足如下条件的串 w 的集合：wa 属于 L。例如，如果 L={a, aab, baa}，则$L/a=\{\varepsilon, ba\}$，证明：如果 L 是正则的，那么 L/a 也是。提示：从 L 的 DFA 出发，考虑接受状态的集合

证明：因为 L 是正则的，因此可以构造出 L 的 DFA，现构造 L/a 的 DFA。其中 L/a 的状态集、字母表、转移函数、起始状态均与 L 相同，将接受状态定义为：若$\delta(q,a)$是 L 的接受态，则令所有

的q为 L/a 的接受态。则若 L 接受 wa，L/a 就接受 w。显然能做出 L/a 的 DFA，即 L/a 是正则的

4.2.7

教材 4.2.7 如果$w=a_1a_2...a_n$和$x=b_1b_2...b_n$是同样长度的串，定义 alt(w, x)是把 w 和 x 交叉起来且以 w 开头所得到的串，即$a_1b_1a_2b_2...a_nb_n$。如果 L 和 M 是语言，定义 alt(L, M)是所有形式为 alt(w, x)的串的集合，其中 w 是 L 中的任意串，而 x 是 M 中与 w 等长任意串。证明：如果 L 和 M 都是正则的，那么 alt(L, M)也是

证明：因为 L 和 M 都是正则的，不妨构造 L 和 M 的 DFA 为$L=(Q_1,\Sigma_1,\delta_1,q_{10},F_1)$，

$M=(Q_2,\Sigma_2,\delta_2,q_{20},F_2)$。则可以构造 alt(L, M)的 DFA 为$alt(L,M)=(Q_3,\Sigma_3,\delta_3,q_{30},F_3)$，

$Q_3=Q_1\times Q_2\times\{0,1\}$（其中×为笛卡尔积），即$Q_3$包含所有形如$(q_1,q_2,\alpha)$的三元组，其中

$q_1\in Q_1$、$q_2\in Q_2$、$\alpha\in\{0,1\}$；$\Sigma_3=\Sigma_1\cup\Sigma_2$；$q_{30}=(q_{10},q_{20},0)$；$F_3=F_1\times F_2\times\{0\}$（其中×为笛卡尔积），即$F_3$包括所有形如$(f_1,f_2,0)$的串，其中$f_1\in F_1$、$f_2\in F_2$；

$\delta_3((q_1,q_2,\alpha),\sigma)=\begin{cases}(\delta_1(q_1,\sigma),q_2,1) & ,\alpha=0\\(q_1,\delta_2(q_2,\sigma),0) & ,\alpha=1\end{cases}$，因此构造出了 alt(L, M)的 DFA，则它也正则

4.2.8

教材 4.2.8 设 L 是一个语言，定义 half(L)是所有 L 中串的前一半构成的集合，即{w|对于某个满足|x|=|w|的 x，wx 属于 L}。例如，如果$L=\{\varepsilon, 0010, 011, 010110\}$，则$half(L)=\{\varepsilon, 00, 010\}$。注意，长度为奇数的串对于 half(L)没有贡献。证明：如果 L 是正则的，那么 half(L)也是

证明：假设L对应的自动机为$M=(Q,\Sigma,\delta,q_0,F)$。要证明$w\in half(L)$，就要找到$|x|=|w|$使得$wx\in L$。不妨令$\delta(q_0,w)=q_i$，也就是说要找到串x使得$\delta(q_i,x)\in F$

假设M从起始状态q_0开始已经输入了串$w(|w|=n)$，记S_n为M中能够通过输入长为n的串到达接受状态的状态的集合，以此类推，S_{n+1}为M中能够通过输入长为$n+1$的串到达接受状态的状态的集合。也就是说，如果$\delta(q_0,w)=q_i$，$q_i\in S_n$，则$w\in half(L)$。

下面我们便可以构建 DFA $M'=(Q\times 2^Q,\Sigma,\delta',(q_0,F),\{(q,S)|q\in Q,S\in 2^Q,q\in S\})$，其中$\delta'((q,S),c)=(\delta(q,c),\{p\in Q|\exists c\in\Sigma,q'\in S,\delta(p,c)=q'\})$。这样就构造出了half(L)对应的有穷自动机，因此，half(L)同样也是正则的

4.2.9

教材 4.2.9 我们把习题 4.2.8 推广到能够决定取走串中多大部分的一系列函数。如果 f 是一个整数函数，定义 f(L)为{w|对某个满足|x|=f(|w|)的 x，wx 属于 L}。例如，和运算 half 对应的

f 是恒等函数 f(n)=n，因为 half(L)的定义中有|x|=|w|。证明：如果 L 是正则的，那么对于以下的 f，f(L)也是正则的：

a） $f(n)=2n$（也就是取走串的前三分之一）

证明：假设L对应的自动机为$M=(Q,\Sigma,\delta,q_0,F)$。要证明$w\in f(L)$，就要找到$|x|=2|w|$使得$wx\in L$。

不妨令$\delta(q_0,w)=q_i$，也就是说要找到串x使得$\delta(q_i,x)\in F$

假设M从起始状态q_0开始已经输入了串$w(|w|=n)$，记S_n为M中能够通过输入长为2n的串达到接

受状态的状态的集合，以此类推，S_{n+1}为M中能够通过输入长为$2(n+1)$的串达到接受状态的状态的集合。也就是说，如果$\delta(q_0,w)=q_i$，$q_i\in S_n$，则$w\in f(L)$。

下面我们构建 DFA $M'=(Q\times 2^Q,\Sigma,\delta',(q_0,F),\{(q,S)|q\in Q,S\in 2^Q,q\in S\})$，其中$\delta'((q,S),c)=(\delta(q,c),\{p\in Q|\ \exists l\in\Sigma^2,q'\in S,\delta(p,l)=q'\})$。这样就构造出了f(L)对应的有穷自动机，因此，f(L)同样也是正则的

b） $f(n)=n^2$(也就是取走的长度是没取走部分长度的平方根)

证明：假设L对应的自动机为$M=(Q,\Sigma,\delta,q_0,F)$。要证明$w\in f(L)$，就要找到$|x|=|w|^2$使得$wx\in L$。

不妨令$\delta(q_0,w)=q_i$，也就是说要找到串x使得$\delta(q_i,x)\in F$

下面我们构建 DFA $M'=(Q\times 2^Q\times N,\Sigma,\delta',(q_0,F,1),\{(q,S,i)|q\in Q,S\in 2^Q,q\in S,i\in N\})$，其中N为小于等于$2n+1$的自然数集合，因为n有最大值，故M′状态集合是有穷的，而$\delta'((q,S,i),c)=(\delta(q,c),\{p\in Q|\ \exists l\in\Sigma^i,q'\in S,\delta(p,l)=q'\},i+2)$。这样就构造出了f(L)对应的有穷自动机，因此，f(L)同样也是正则的

c） $f(n)=2^n$(也就是取走的长度是剩下长度的对数)

证明：假设L对应的自动机为$M=(Q,\Sigma,\delta,q_0,F)$。要证明$w\in f(L)$，就要找到$|x|=2^{|w|}$使得$wx\in L$。

不妨令$\delta(q_0,w)=q_i$，也就是说要找到串x使得$\delta(q_i,x)\in F$

下面我们构建 DFA

$M'=(Q\times 2^Q\times N,\Sigma,\delta',(q_0,\{p\in Q|\ \exists l\in\Sigma,q'\in S,\delta(p,l)=q'\},1),\{(q,S,i)|q\in Q,S\in 2^Q,q\in S,i\in N\})$

其中N为小于等于2^n的自然数集合，因为n有最大值，故M′状态集合是有穷的，而

$\delta'((q,S,i),c)=(\delta(q,c),\{p\in Q|\ \exists l\in\Sigma^i,q'\in S,\delta(p,l)=q'\},i\times 2)$。这样就构造出了f(L)对应的有穷自动机，因此，f(L)同样也是正则的

补充1

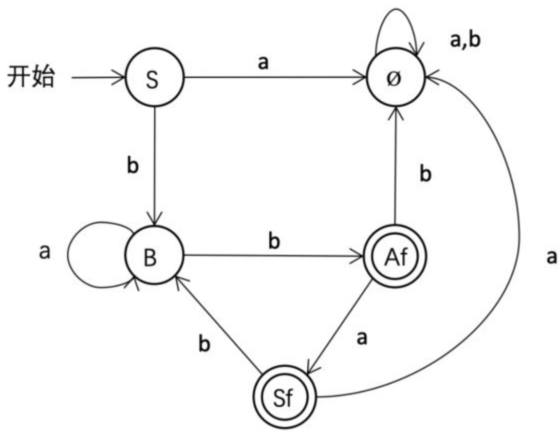
1) $G_1=(V,T,P_1,S)$

P_1=S→bB , B→aB | bA | b , A→a | aS

解：构造对应的有穷自动机 M=（ { S,A,B,f } , {a,b} , δ , S , {f} ），其中：

$\delta(S,b)=\{B\},\delta(B,a)=\{B\},\delta(B,b)=\{A\},\delta(B,b)=\{f\},\delta(A,a)=\{f\},\delta(A,a)=\{S\}$

对应 DFA 如下图：



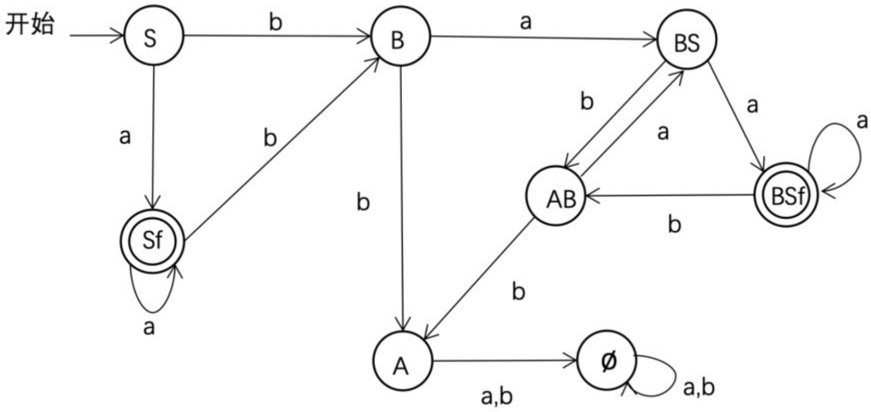
2) $G_2=(V,T,P_2,S)$

P_2=S→aS | bB | a , B→bA | aB | aS

解：构造对应的有穷自动机 M=（ { S,A,B,f } , {a,b} , δ', S , {f} ），其中：

$\delta'(S,a)=\{S\},\delta'(S,b)=\{B\},\delta'(S,a)=\{f\},\delta'(B,b)=\{A\},\delta'(B,a)=\{B\},\delta'(B,a)=\{$

对应 DFA 如下图：



补充2:

a)

解： 由 DFA M_1:

$A \to 0 \mid 0A \mid 1 \mid 1B$

$B \to 1 \mid 1B \mid 0C$

$C \to 0C \mid 1C$

或：

$A \to \varepsilon \mid 0A \mid 1B$

$B \to \varepsilon \mid 1B \mid 0C$

$C \to 0C \mid 1C$

b)

解：由 DFA M_2:

$A \to 0 \mid 0B \mid 1 \mid 1D$

$B \to 0 \mid 0A \mid 1 \mid 1D$

$C \to 1 \mid 1B \mid 0 \mid 0D$

$D \to 1 \mid 1B \mid 0C$

或：

$A \to \varepsilon \mid 0B \mid 1D$

$B \to \varepsilon \mid 0A \mid 1D$

$C \to 1B \mid 0D$

$D \to \varepsilon \mid 0C \mid 1B$

补充3:

令 L_1 的正则表达式为 $\mathbf{10}^*$,则 $L_1 = \{w : \exists n \geq 0, [w]_3 = 3^n\}$是正则语言.

对任意的 n,取 w 为 $D = 3^n$ 的二进制表示,则 $w \in L_2$,则 $D > 2^n$,所以 $|w| > n$.

如果可以把w拆成$w = xyz$,并且满足$y \neq \varepsilon$且$|xy| \leq n$. 设$|y| = l > 0$, $|z| = r$

令 $A = [x]_2, B = [y]_2 > 0, C = [z]_2$,则 $D = [xyz]_2 = A2^{l+r} + B2^r + C$

对任意的 $k > 0$:

$$
\begin{aligned}
E &= [xy^k z]_2 \\
&= A2^{r+kl} + B2^{r+(k-1)l} + B2^{r+(k-2)l} + \cdots + B2^{r+l} + B2^r + C \\
&= D + A2^{r+kl} - A2^{r+l} + B2^{r+(k-1)l} + B2^{r+(k-2)l} + \cdots + B2^{r+l} \\
&= D + 2^{r+l}(A(2^{(k-1)l} - 1) + B2^{(k-2)l} + B2^{r+(k-3)l} + \cdots + B) \\
&= D + 2^{r+l}(A(2^{(k-1)l} - 1) + B\frac{2^{(k-1)l} - 1}{2^l - 1}) \\
&= D + 2^{r+l}\frac{2^{(k-1)l} - 1}{2^l - 1}(A(2^l - 1) + B) \\
&= D + 2^{r+l}(1 + 2^l + \cdots + 2^{(k-2)l})(A(2^l - 1) + B)
\end{aligned}
$$

$2^0 \equiv 1(\bmod 3), 2^1 \equiv 2(\bmod 3)$

$\Longrightarrow 2^p \bmod 3 = 2^{p \bmod \varphi(3)} = 2^{p \bmod 2}$

所以 $3 \nmid 2^{r+l}$.

- 若 l 为偶数,则 $2^l \bmod 3 = 1$

  $(1+2^l+\cdots+2^{(k-2)l}) \bmod 3 = (1+1+\cdots+1) \bmod 3 = (k-1) \bmod 3$

  取 k 为 3 的倍数,则 $3 \nmid (1+2^l+\cdots+2^{(k-2)l})$

- 若 l 为奇数,则 $2^l \bmod 3 = 2$

  $(1+2^l+\cdots+2^{(k-2)l}) \bmod 3 = (1+2+1+2+1+2...) \bmod 3$

  取 k 为奇数,则上式等于$((1+2)(k-3)+1) \bmod 3 = 1$

  所以 $3 \nmid (1+2^l+\cdots+2^{(k-2)l})$

综上,总存在 $k>1$,使得 $3 \nmid (1+2^l+\cdots+2^{(k-2)l}) \Rightarrow 3 \nmid 2^{r+l}(1+2^l+\cdots+2^{(k-2)l})$

$2^l-1 < 2^{l+r}, 1 \le 2^r$,所以 $A(2^l-1)+B < A2^{l+r}+B2^r \le D = 3^n$,因此 $3^n = D \nmid (A(2^l-1)+B)$,即$(A(2^l-1)+B)$中素因子 3 的次数小于 n.

所以 $2^{r+l}(1+2^l+\cdots+2^{(k-2)l})(A(2^l-1)+B)$中素因子 3 的次数小于 n.

$\therefore D=3^n \nmid (D+2^{r+l}(1+2^l+\cdots+2^{(k-2)l})(A(2^l-1)+B))=E$

因为 $k>1 \Longrightarrow E>D$,所以 E 不可能是 3 的幂次,$xy^kz \notin L_2$.

根据 Pumping Lemma, L_2 不是正则语言.