第 4 次作业 – 正则语言的性质 2

2154312 郑博远

习题 1 利用 Myhill-Nerode 定理证明下列语言是否正则语言，如果是正则语言，请构造其 FA、RE 及 RG。

（1）$\{\, x \mid x = x^R,\ x \in \{0,1\}^+ \,\}$

（2）$\{\, x \mid x$ 中 0 的个数不少于 1 的个数, $x \in \{0,1\}^+ \,\}$

（3）$\{\, xx^R w \mid x, w \in \{0,1\}^+ \,\}$

解答：

1) 该语言（记为 L_1）不是正则语言，理由如下：

任取两字符串 $x, y \in \{0,1\}^+$，满足 $x \neq y$ 且 $|x| = |y|$。取 $z = x^R$，则 $xz = xx^R \in L_1$ 且 $yz = yx^R \notin L_1$。因此 $\{0,1\}^+$ 上所有长度相等的不同字符串均属于 R_{L_1} 中不同的等价类，R_{L_1} 具有无穷指数。根据 Myhill-Nerode 定理，该语言不是正则语言。

2) 该语言（记为 L_2）不是正则语言，理由如下：

分析该语言特征可得，可根据 $\{0,1\}^+$ 上字符串中 0 与 1 个数的差值来进行等价类的划分，例如：

[0] : 0 的个数与 1 的个数相同的字符串（包含 ε）所在的等价类；

[1] : 0 的个数比 1 的个数多 1 的字符串所在的等价类；

[2] : 0 的个数比 1 的个数多 2 的字符串所在的等价类；

[-1] : 0 的个数比 1 的个数少 1 的字符串所在的等价类；

[-2] : 0 的个数比 1 的个数少 2 的字符串所在的等价类；

……

由于 0 与 1 个数的差值有无穷多种，因此 R_{L_2} 具有无穷指数。根据 Myhill-Nerode 定理，该语言不是正则语言。

3) 该语言（记为 L_3）不是正则语言，理由如下：

对于 $x = 0^i 1^i$，字符串 $z = x^R 1 = 1^i 0^i 1$ 能够使得 $xz \in L_3$。对于 $y = 0^j 1^j$（$i < j$），$yz =$

$0^j1^j1^i0^i1 \notin L_3$。因此对于任意奇数 $i \neq j$, 有 $x_i = 0^i1^i$ 与 $x_j = 0^j1^j$ 属于不同的等价类，因此 R_{L_3} 具有无穷指数。根据 Myhill-Nerode 定理，该语言不是正则语言。

习题 2 判断下列命题，并证明你的结论。

（1） 正则语言的任意子集都是正则语言。

（2） 正则语言的补也是正则语言。

（3） 无穷多个正则语言的并不一定是正则语言。

解答：

1) 错误。$\{\, x |\, x = \{0,1\}^* \,\}$ 是正则语言，因为其能够被正则表达式 $0^* + 1^*$ 所描述。习题 1 中的（1）、（2）、（3）均是其子集，但都不是正则语言。

2) 正确。对于正则语言 L，由 Myhill-Nerode 定理可得，R_L 具有有穷指数。记 L 的补为 $\bar{L}$，则 $\forall x \in \Sigma^*, x \in L \Leftrightarrow x \notin \bar{L}$ 。因此对于满足等价关系 R_L 的字符串 $x, y \in \Sigma^*$，$x\, R_L\, y \Leftrightarrow (\forall z \in \Sigma^*,\ xz \in L \Leftrightarrow yz \in L) \Leftrightarrow (\forall z \in \Sigma^*,\ xz \notin \bar{L} \Leftrightarrow yz \notin \bar{L}) \Leftrightarrow (\forall z \in \Sigma^*,\ xz \in \bar{L} \Leftrightarrow yz \in \bar{L}) \Leftrightarrow x\, R_{\bar{L}}\, y$。由此可得，等价关系 R_L 与 $R_{\bar{L}}$ 相同。故 $R_{\bar{L}}$ 具有有穷指数，$\bar{L}$ 是正则语言（Myhill-Nerode 定理）。

3) 正确。取如下的无穷多个正则语言：

   $L_0 = \{\, \varepsilon \,\}$

   $L_1 = \{\, 01 \,\}$

   $L_2 = \{\, 0011 \,\}$

   $L_3 = \{\, 000111 \,\}$

   ……

   容易得到，语言 $L = \left\{\, 0^i1^i \,\middle|\, i \geq 0 \,\right\}$ 是上述无穷多个正则语言的并。假设该语言 L 是正则的，且 L 的泵长度为 p 。考察字符串 $s = 0^p1^p \in L$，由泵引理可知 $|xy| \leq p$，因此设 $y = 0^k$。$0^{p+k}1^p \notin L$，与假设矛盾，故该语言不是正则语言。由此可得，无穷多个正则语言的并不一定是正则语言。

习题 3 设 L 是正则语言，字母表是 Σ，定义 $L_{1/3} = \{w \in \Sigma^* \mid \exists\, x, y \in \Sigma^*,\ wxy \in L,\ |w| = |x| = |y|\}$。试证明 $L_{1/3}$ 是否正则语言吗？

解答：

设存在 DFA $M_A = (Q_A, \Sigma_A, \delta_A, q_{0A}, F_A)$，使得对于语言$L$，有 $L(M_A) = L$。下面构造 $M = (Q, \Sigma, \delta, q_0, F)$ 以识别 $L_{1/3}$。

其中：Q 所代表的 M 中的状态有如 (q, S) 的形式。$q \in Q_A$，用于记录在 M 中输入某个字符串后对应 M_A 所在的状态；S 表示在 M_A 中接受了当前读入字符串两倍长度的字符串后，能够到达接受状态的所有状态。对于已读入长度的记忆，需要通过转移函数来实现。

转移函数满足 $\delta\big((q, S), a\big) = (\delta_A(q, a), T)$，其中 T 表示接受**两个**字符后能够到达 S 中任意状态的 M_A 中的状态的集合。

此外，$q_0 = (q_{0A}, F_A)$，对应在 A 中的起始状态，且接受的字符串长为 0，即集合对应 M_A 中的 F_A。终止状态 F 对应所有满足 $q' \in S$ 的状态 (q', S)，即代表当前状态到接收状态的长度是当前读入字符串的长度的两倍。

易得上述 M 满足 $L(M) = L_{1/3}$。因此 $L_{1/3}$ 是正则语言。

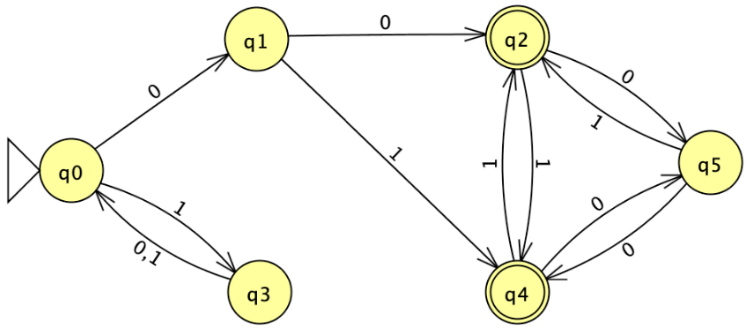
习题 4 用正则语言的**扩充泵引理**证明语言 $\{\, 0^n 1^m 0^m,\ n, m \geq 1 \,\}$ 不是正则的。

解答：

根据扩充泵引理，存在只依赖于 L 的正整数 k，对于任何串 x, y, z（这里 $xyz \in L$），只要 $|y| \geq k$，就可以将 y 写成 $y = uvw$（这里 $v \neq \varepsilon$, $|uv| \leq k$），使得对于任意 $i \geq 0$，都有 $xuv^i wz \in L$。

考察字符串 $0^n 1^k 0^k$，其中 $x = 0^n, y = 1^k, z = 0^k$。将 y 分解为 $y = uvw$，则 v 是由 1 构成的非空字符串。因此，$xuv^i wz = 0^n 1^{k+(i-1)|v|} 0^k$；当 $i \neq 1$ 时，$xuv^i wz \notin L$。由此可得，语言 $\{\, 0^n 1^m\, 0^m,\ n, m \geq 1 \,\}$ 不是正则的。

习题 5 对下图给出的 DFA，求出它的极小状态 DFA，**要求给出主要的求解步骤**。



解答：

运用极小化算法进行 DFA 化简，具体步骤如下：

（1） 为所有状态对 (p,q) $(p,q\in Q)$ 画一张表，开始时表中每个格子内均为空白（未做任何标记）；

（2） 对 $p\in F, q\notin F$ 的一切状态对 (p,q)，在相应的格子内做标记（例如画一个×）， 表示 (p,q) 是可以区分的。对接受状态和非接受状态的状态对的格子内做标记。如下图所示：

	0	1	2	3	4
1					
2	×	×			
3			×		
4	×	×		×	
5			×		×

（3） 重复下述过程，直到表中内容不再改变为止： 如果存在一个未被标记的状态对 (p,q)，且对于某个$a\in\Sigma$， 如果 $\left(r=\delta(p,a),\ s=\delta(q,a)\right)$ 已做了标记，则在 (p,q) 相应 的格子内做标记。

	0	1	2	3	4
1	×				
2	×	×			
3	×	×	×		
4	×	×		×	
5	×		×	×	×

- $(\delta(q_0,0) = q_1$，$\delta(q_1,0) = q_2)$已被标记，故标记(q_0,q_1)；
- $(\delta(q_0,0) = q_1$，$\delta(q_3,0) = q_0)$已被标记，故标记(q_0,q_3)；
- $(\delta(q_0,0) = q_1$，$\delta(q_5,0) = q_4)$已被标记，故标记(q_0,q_5)；
- $(\delta(q_1,0) = q_2$，$\delta(q_3,0) = q_0)$已被标记，故标记(q_1,q_3)；
- $(\delta(q_3,0) = q_0$，$\delta(q_5,0) = q_4)$已被标记，故标记(q_3,q_5)；
- 剩余状态对均不可被标记。

（4） 在完成（1），（2），（3）之后，所有未被标记的状态对 (p,q) 都是等价的，即 $p \equiv q$，状态 p 和状态 q 可以合并。合并 q_1、q_5 及 q_2、q_4后，DFA 如下：