

第 5 章

正则语言的性质

在讨论正则语言的性质时,一般分两类进行。第一类是由于直接用有穷的形式表达无穷所表现出来的性质,这种性质用正则语言的泵引理表示。除了基本的泵引理外,还有一种形式可以称为扩充的泵引理,其证明与基本泵引理的证明具有相同的思想。除了证明泵引理的正确性外,更重要的是应用泵引理证明给定的语言不是正则语言。这涉及发现特殊串的问题。

正则语言性质的第二类是关于一系列运算的封闭性。包括对并、乘积、闭包、补、交等基本运算的封闭性和对正则代换、同态、逆同态等运算的封闭性。对应不同的运算,其封闭性证明将使用正则语言的不同描述方法。

本章讨论的第二部分内容是通过给定正则语言本身结构的固有特征寻求不同表示的一致性。这种一致性是通过该语言确定的右不变等价关系 R_L 表现出来的。它将相应字母表上的所有字符串分成一些等价类,而这一分法是由语言 L 决定的,所以说它表达的是 L 本身结构的固有特征。由于在一般情况下,等价类的直接划分比较困难,而且这种困难导致了对其进行自动化的困难,所以,人们将其转换为寻求接受该语言的、状态最少的、确定的有穷状态自动机——最小 DFA 的问题。而且在同构意义下,最小 DFA 是唯一的。这里给出的根据 DFA 去构造其相应最小 DFA 的方法是可以由计算机系统自动完成的。这一部分的内容可以分成 Myhill-Nerode 定理与 FA 的极小化。

本章的第三部分内容简要介绍正则语言相关的几个判定问题。包括空否、有穷否、两个 DFA 等价否以及成员关系等。

本章的内容比较多,其中最基本的有:正则语言的泵引理及其应用,正则语言的封闭性。这些是读者应该掌握的,尤其是相关证明的基本思想。所以,这些可以作为本章的重点内容。如果时间和精力允许,还可以将 Myhill-Nerode 定理的理解与应用作为重点。本章中比较难的内容是关于正则代换、同态、逆同态的封闭性证明。主要原因是相关的概念比较生,也比较复杂。另外,Myhill-Nerode 定理及其应用涉及 DFA 以及语言决定的等价关系对语言的结构特征的刻画问题,加上用等价关系进行等价分类也是比较难的问题,所以,这一部分也是比较难的。

5.1 正则语言的泵引理

1. 知识点

(1) 当 DFA 识别的语言 L 是无穷语言时, L 中必定存在一个足够长的句子, 使得 DFA 在识别该句子的过程中, 肯定要重复地经过某些状态。

(2) 泵引理给出的是正则语言的“必要条件”而不是“充分条件”, 它只能用来证明一个语言不是正则语言, 不能用来证明一个语言是正则语言。

2. 主要内容解读

引理 5-1(泵引理 pumping lemma) 设 L 为一个正则语言, 则存在仅依赖于 L 的正整数 N , 对于 $\forall z \in L$, 如果 $|z| \geq N$, 则存在 u, v, w , 满足:

- (1) $z = uvw$ 。
- (2) $|uv| \leq N$ 。
- (3) $|v| \geq 1$ 。
- (4) 对于任意整数 $i \geq 0$, $uv^i w \in L$ 。
- (5) N 不大于接受 L 的最小 DFA M 的状态数。

证明要点: DFA 在处理一个足够长的句子的过程中, 必定会重复地经过某些状态。换句话说, 在 DFA 的状态转移图中, 必定存在一条含有回路的从启动状态到某个终止状态的路。由于含有回路, 所以, DFA 可以根据实际需要沿着这个回路循环运行, 相当于这个回路中弧上的标记构成的非空子串可以重复任意次。

值得反复强调的一点是, 泵引理给出的是 RL 的必要条件, 所以, 该引理只能用来证明给定的语言不是 RL。条件 $|uv| \leq N$ 表明, 可以用来“泵进”和“泵出”的子串 v 在句子 z 的前 N 个字符中是存在的, 所以, 讨论可以集中在 z 的前 N 个字符上进行。然而, 如果在这 N 个字符中找不到引起矛盾的 v , 并不能说明给定的语言是 RL, 因为其他类型的某个具体语言也是满足这一要求的——泵引理给出的是必要而非充分条件。

另外, 从泵引理的证明不难看出, 可以在被判定语言的句子 z 的任何位置截取一个子串, 只要这个子串足够长, 就一定能在这个子串中找到可以用来“泵进”和“泵出”的子串 v 。事实上, 不妨取 $z = z_1 z_2 z_3 \in L$, $|z_2| = N$, $z_2 = a_1 a_2 \cdots a_N$, 将 DFA 开始扫描到 z_2 的首字符时所处的状态记为 q'_0 , 并且令

$$\delta(q'_0, a_1 a_2 \cdots a_h) = q_h$$

则在状态序列 $q'_0, q_1, q_2, \cdots, q_N$ 中就必定有重复的。例如, 不妨令最早相等的两个状态为 q_j 和 $q_k, k < j$, 则

$$\begin{aligned} & \delta(q_0, z_1 a_1 a_2 \cdots (a_{k+1} \cdots a_j)^i \cdots a_N z_3) \\ &= \delta(q'_0, a_1 a_2 \cdots (a_{k+1} \cdots a_j)^i \cdots a_N z_3) \\ &= \delta(q_k, (a_{k+1} \cdots a_j)^i \cdots a_N z_3) \\ &= \delta(q_j, (a_{k+1} \cdots a_j)^{i-1} \cdots a_N z_3) \\ &= \delta(q_j, a_{j+1} \cdots a_N z_3) \\ &= q_m \in F \end{aligned}$$

令 $u = a_1 a_2 \cdots a_k, v = a_{k+1} \cdots a_j, w = a_{j+1} \cdots a_N$, 上式表明, 对于 $i \geq 0, z_1 u v^i w z_3 \in L$ 恒成立。从而得到如下关于 RL 的扩展泵引理。

设 L 为一个 RL, 则存在仅依赖于 L 的正整数 N , 对于 $\forall z_1 z_2 z_3 \in L$, 如果 $|z_2| = N$, 则存在 u, v, w , 满足:

- (1) $z_2 = uvw$ 。
- (2) $|v| \geq 1$ 。
- (3) 对于任意整数 $i \geq 0, z_1 u v^i w z_3 \in L$ 。
- (4) N 不大于接受 L 的最小 DFA M 的状态数。

主教材总结的关于使用泵引理的 8 条注意事项对扩展的泵引理也是有意义的, 对这 8 条注意事项的体会需要在应用中加深。

用泵引理证明一个语言不是 RL 时, 最大的难点是寻找适当的 z , 其次是对 v 的各种取法进行论证。对部分读者来说, 可能存在如何给出良好的叙述, 这需要一个锻炼过程。因此, 建议读者完成适量的用泵引理证明一个语言不是正则语言的习题。

5.2 正则语言的封闭性

1. 知识点

- (1) 封闭性与有效封闭性。
- (2) 正则语言在并、乘积、闭包运算下的封闭性证明用正则表达式作为描述工具最方便。
- (3) 正则语言在补运算下的封闭性用 DFA 作为描述工具最方便。
- (4) 正则语言在交运算下的封闭性的证明也可以用 DFA 作为描述工具。
- (5) 代换与正则代换。
- (6) 正则语言在正则带换下是封闭的。
- (7) 同态映射。
- (8) 正则语言的同态像是正则语言。
- (9) 正则语言的同态原像是正则语言。
- (10) 商运算是非有效封闭的。

2. 主要内容解读

定义 5-1 如果任意的、属于同一语言类的语言在某一特定运算下所得的结果总是该类语言, 则称该语言类对此运算是**封闭**的, 并称该语言类对此运算具有**封闭性**。

给定一个语言类的若干个语言的描述, 如果存在一个算法, 它可以构造出这些语言在给定运算下所获得的运算结果的相应形式的描述, 则称此语言类对相应的运算是**有效封闭**的, 并称此语言类对相应的运算具有**有效封闭性**。

定理 5-1 RL 在并、乘积、闭包运算下是封闭的。

证明要点: 利用正则表达式的定义。

定理 5-2 RL 在补运算下是封闭的。

证明要点: 根据接受 L 的 DFA 构造接受 $\Sigma^* - L$ 的 DFA。

如果 DFA $M = (Q, \Sigma, \delta, q_0, F)$ 满足 $L(M) = L$, 则 DFA $M' = (Q, \Sigma, \delta, q_0, Q - F)$ 满足 $L(M') = \Sigma^* - L$ 。

定理 5-3 RL 在交运算下封闭。

证明要点:

方法 1: 由定理 5-1、定理 5-2 和 De Morgan 定理可得。

方法 2: 构造相应的 DFA。按照第 3 章习题 18 的解法进行。图 5-3 是相应的构造思想的来源描述,弄清楚了图,就掌握了相应的构造思想。

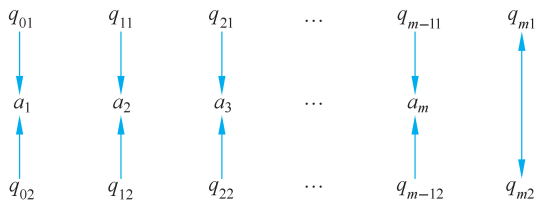


图 5-3 M_1 和 M_2 对应于同一个输入串的状态变换过程

定义 5-2 设 Σ, Δ 是两个字母表,映射

$$f: \Sigma \rightarrow 2^{\Delta^*}$$

被称为是从 Σ 到 Δ 的**代换**(substitution)。如果对于 $\forall a \in \Sigma, f(a)$ 是 Δ 上的 RL, 则称 f 为**正则代换**(regular substitution)。

先将 f 的定义域扩展到 Σ^* 上, $f: \Sigma^* \rightarrow 2^{\Delta^*}$

$$(1) f(\epsilon) = \{\epsilon\}.$$

$$(2) f(xa) = f(x)f(a).$$

再将 f 的定义域扩展到 2^{Σ^*} 上: $f: 2^{\Sigma^*} \rightarrow 2^{\Delta^*}$ 。对于 $\forall L \subseteq \Sigma^*$,

$$f(L) = \bigcup_{x \in L} f(x)$$

读者需要注意的是,代换是将一个字母映射成一个语言,对正则代换来说,这个字母所被映射成的语言是正则的。根据(正则)代换定义的扩展,(正则)代换也将一个语言映射成另一个语言。在这个映射过程中,一个句子将对应一个语言,所有句子对应的语言的并就是被代换语言的代换结果。

定义 5-3 设 Σ, Δ 是两个字母表,映射

$$f: \Sigma \rightarrow 2^{\Delta^*}$$

为正则代换,则

$$(1) f(\emptyset) = \emptyset.$$

$$(2) f(\epsilon) = \epsilon.$$

$$(3) \text{ 对于 } \forall a \in \Sigma, f(a) \text{ 是 } \Delta \text{ 上的正则表达式。}$$

$$(4) \text{ 如果 } r, s \text{ 是 } \Sigma \text{ 上的正则表达式,则}$$

$$f(r+s) = f(r) + f(s)$$

$$f(rs) = f(r)f(s)$$

$$f(r^*) = f(r)^*$$

是 Δ 上的正则表达式。

将正则代换的定义用在正则表达式上是为了叙述方便,也更利于理解。从上述定义可见,对一般的代换来说,由于 $f(a)$ 不一定是正则表达式,所以,就无法保证第(4)条成立了。有了这个定义之后,当证明下述定理时,对任意正则表达式 r 和正则映射 f ,只要证明 $f(r)$ 表示的是 $f(L(r))$,就可以说明 $f(L(r))$ 是正则语言了。

定理 5-4 设 L 是 Σ 上的一个 RL, $f: \Sigma \rightarrow 2^{\Delta^*}$ 是正则代换,则 $f(L)$ 也是 RL。

证明要点:

方法 1: 用 FA 作为 RL 的描述模型。在这里,应该假定接受 L 和 $f(a)$ 的 FA 都是 DFA,根据此,构造出接受 $f(L)$ 的 ϵ -NFA。

令 DFA $M = (Q, \Sigma, \delta, q_0, F)$ 识别 L 。DFA $M_a = (Q_a, \Delta, \delta_a, q_{0a}, F_a)$ 识别 $f(a)$ 。不妨假设:对 $a, b \in \Sigma$, 如果 $a \neq b$, 则 $Q_a \cap Q_b = \emptyset, Q_a \cap Q = \emptyset$ 。

取 FA

$$M_{RS} = (Q \bigcup_{a \in \Sigma} Q_a, \Delta, \delta_{RS}, q_0, F)$$

其中, δ_{RS} 定义为

① 对任意 $(q, a) \in Q \times \Sigma$, 如果 $\delta(q, a) = p$, 则定义 $q_{0a} \in \delta_{RS}(q, \epsilon)$, 并且对所有的 $r \in F_a, \delta_{RS}(r, \epsilon) = p$ 。

② 对任意 $(q, b) \in Q_a \times \Delta, \delta_{RS}(q, b) = \delta_a(q, b)$ 。

方法 2: 用正则表达式作为 RL 的描述模型。

对正则表达式 r 中运算符的个数 n 施以归纳,证明 $f(r)$ 是表示 $f(L)$ 的正则表达式。

关于定理 5-1 作为定理 5-4 的推论的问题,可以按照如下方式进行理解。

令 f 为正则代换, $L_1 = f(a), L_2 = f(b)$ 为 RL,

$L = \{a, b\}$ 是有穷集,所以是 RL, 由正则代换的封闭性, $f(\{a, b\})$ 是 RL, 而 $L_1 \cup L_2 = f(a) \cup f(b) = f(\{a, b\})$, 所以, $f(a) \cup f(b)$ 是 RL;

$L = \{ab\}$ 是有穷集,所以是 RL, 由正则代换的封闭性, $f(\{ab\})$ 是 RL, 而 $L_1 L_2 = f(a) f(b) = f(\{ab\})$, 所以, $f(a) f(b)$ 是 RL;

$L = a^*$ 是 RL, 令 $g(a) = L$, 是正则代换, 由正则代换的封闭性, $g(a^*)$ 是 RL, 而 $g(a^*) = (g(a))^* = L^*$, 所以, L^* 是 RL。

定义 5-4 设 Σ 和 Δ 是两个字母表, $f: \Sigma \rightarrow \Delta^*$ 为映射。如果对于 $\forall x, y \in \Sigma^*$,

$$f(xy) = f(x)f(y)$$

则称 f 为从 Σ 到 Δ^* 的**同态映射**(homomorphism)。

对于 $\forall L \subseteq \Sigma^*, L$ 的同态像

$$f(L) = \bigcup_{x \in L} f(x)$$

首先,与(正则)代换不同,同态映射是将一个字符映射成一个字符串。当将这个字符串看成是只将它自身作为唯一的元素的集合时,同态映射就成了(正则)代换的一种特例。因此,(正则)代换所具有的性质对代换都成立。

其次,同态映射是将一个字符映射成一个字符串,同时保持并置运算。一般地,对任意 $a_1, a_2, \dots, a_n \in \Sigma$,

$$f(a_1a_2\cdots a_n)=f(a_1)f(a_2)\cdots f(a_n)$$

最后,由于 $f(a_i)$ 都是唯一的,所以, $f(a_1)f(a_2)\cdots f(a_n)$ 也是唯一的。但是,对于 $f(a_1)f(a_2)\cdots f(a_n)$,却不一定只有 $a_1a_2\cdots a_n$ 与之对应。因此,

对于 $\forall w \subseteq \Delta^*$, w 的同态原像是一个集合:

$$f^{-1}(w)=\{x \mid f(x)=w \text{ 且 } x \in \Sigma^*\}$$

对于 $\forall L \subseteq \Delta^*$, L 的同态原像是一个集合:

$$f^{-1}(L)=\{x \mid f(x) \in L\}$$

一般地, $f(f^{-1}(L)) \neq L$ 。但是, $f(f^{-1}(L)) \subseteq L$ 。

推论 5-1 RL 的同态像是 RL。

定理 5-5 RL 的同态原像是 RL。

证明要点:

(1) 接受 RL 的同态原像的 FA 的构造思想。

① 使用 DFA 作为 RL 的描述。

② 让新构造出的 FA M' 用一个移动去模拟 M 处理 $f(a)$ 所用的一系列移动。

③ 对于 Σ 中的任意字符 a ,如果 M 从状态 q 开始处理 $f(a)$,并且当它处理完 $f(a)$ 时到达状态 p ,则让 M' 在状态 q 读入 a 时,将状态变成 p 。

④ M' 具有与 M 相同的状态,并且,在 M' 对应的状态转移图中,从状态 q 到状态 p 有一条标记为 a 的弧当且仅当在 M 的状态转移图中,有一条从状态 q 到状态 p 的标记为 $f(a)$ 的路。

(2) 接受 RL 的同态原像的 FA 的形式描述。

设 DFA $M=(Q,\Delta,\delta,q_0,F)$, $L(M)=L$

取 DFA $M'=(Q,\Sigma,\delta',q_0,F)$,

其中,对 $\forall (q,a) \in Q \times \Sigma$,

$$\delta'(q,a)=\delta(q,f(a))$$

(3) 等价证明。

① 施归纳于 $|x|$,证明对于 $\forall x \in \Sigma^*$,

$$\delta'(q_0,x)=\delta(q_0,f(x))$$

② 对于 $\forall x \in \Sigma^*$,

$$\delta'(q_0,x) \in F \Leftrightarrow \delta(q_0,f(x)) \in F$$

定义 5-5 设 $L_1, L_2 \subseteq \Sigma^*$, L_2 除以 L_1 的商(quotient)定义为

$$L_1/L_2=\{x \mid \exists y \in L_2 \text{ 使得 } xy \in L_1\}$$

计算语言的商主要是考虑语言句子的后缀。只有当 L_1 的句子的后缀在 L_2 中时,其相应的前缀才属于 L_1/L_2 。所以,当 $\epsilon \in L_2$ 时,

$$L_1 \subseteq L_1/L_2$$

成立。读者应该注意的是,此时并不一定有 $L_1=L_1/L_2$ 成立,也就是说,有可能是

$$L_1 \subset L_1/L_2$$

实际上,取 $L_1=\{000\}$, $L_2=\{\epsilon\}$, $L_3=\{\epsilon,0\}$, $L_4=\{\epsilon,0,00\}$, $L_5=\{\epsilon,0,00,000\}$,则有如下很值得注意的结果:

$$L_1/L_2 = \{000\} = L_1$$

$$L_1/L_3 = \{000, 00\}$$

$$L_1/L_4 = \{000, 00, 0\}$$

$$L_1/L_5 = \{000, 00, 0, \varepsilon\}$$

可见,与一般的求商运算不同,这里,在充当“被除数”的集合不变的情况下,充当“除数”的集合越“大”,所得的“商”可能越“大”。这种变化趋势在充当“除数”的集合依次使包含关系成立时呈现。

另外,按照“依次使包含关系成立”的要求,随着充当除数的集合变成 $\{\varepsilon\}$,商变成充当“被除数”的集合。

定理 5-6 设 $L_1, L_2 \subseteq \Sigma^*$, 如果 L_1 是 RL, 则 L_1/L_2 也是 RL。

证明要点:

设 DFA $M = (Q, \Sigma, \delta, q_0, F)$, $L(M) = L_1$

取 DFA $M' = (Q, \Sigma, \delta, q_0, F')$,

其中, $F' = \{q \mid \exists y \in L_2, \delta(q, y) \in F\}$ 。

由于 L_2 可以是各种语言,所以, F' 的取法并不是一个有效的取法,从而导致这种封闭性并不是有效封闭性。

5.3 Myhill-Nerode 定理与 DFA 的极小化

5.3.1 Myhill-Nerode 定理

1. 知识点

(1) DFA M 确定的 Σ^* 上的右不变等价关系 R_M 。

(2) L 确定的 Σ^* 上的右不变等价关系 R_L 。

(3) $|\Sigma^*/R|$ 是 R 关于 Σ^* 的指数。

(4) L 是 RL 的充要条件是 R_L 具有有穷指数,因此,可以通过判定 R_L 是否将 Σ^* 分成有穷多个等价类来判定 L 是否为 RL。

(5) 如果 L 是 Σ^* 上的某一个具有有穷指数的右不变等价关系 R 的某些等价类的并,则 L 是 RL,并且 R 一定是 R_L 的加细。

(6) 任意 DFA $M = (Q, \Sigma, \delta, q_0, F)$, 则 $|\Sigma^*/R_{L(M)}| \leq |Q|$ 。

(7) 在同构意义下,接受 L 的最小 DFA 是唯一的。

2. 主要内容解读

定义 5-6 设 DFA $M = (Q, \Sigma, \delta, q_0, F)$, M 所确定的 Σ^* 上的关系 R_M 定义为: 对于 $\forall x, y \in \Sigma^*$,

$$x R_M y \Leftrightarrow \delta(q_0, x) = \delta(q_0, y)$$

R_M 为 M 所确定的 Σ^* 上的关系。

这个定义可以简化成: DFA $M = (Q, \Sigma, \delta, q_0, F)$ 确定的关系 R_M 为, 对于 $\forall x, y \in \Sigma^*$, $x R_M y \Leftrightarrow \delta(q_0, x) = \delta(q_0, y)$ 。

读者应该注意,这个等价关系与第3章的定义3-6所定义的集合 $set(q)$ 紧密相关。即

$$x R_M y \Leftrightarrow \exists q \in Q, x, y \in set(q)$$

事实上, $\{set(q) | q \in Q\}$ 是 R_M 关于 Σ^* 的等价划分。显然,对于任意的 DFA $M=(Q, \Sigma, \delta, q_0, F)$, $|\{set(q) | q \in Q\}| \leq |Q|$, 而且等号仅在 Q 中不存在不可达状态时成立。

例5-8给出了依据 R_M 划分等价类的一个例子,由于在第3章中对相应的概念已经有所接触,该例子则是在给出了确切定义之后,让读者对这种等价关系有进一步的认识。

满足关系 R_M 的字符串组成的集合的特征体现了语言的结构特征。这一特征,可以与相应的 DFA 的状态转移图对应起来进行理解。再与例5-9的讨论联系起来,就会得到更深入的理解。

定义 5-7 设 $L \subseteq \Sigma^*$, L 确定的 Σ^* 上的关系 R_L 定义为:对于 $\forall x, y \in \Sigma^*$,

$$x R_L y \Leftrightarrow (\text{对 } \forall z \in \Sigma^*, xz \in L \Leftrightarrow yz \in L)$$

R_L 为 L 确定的 Σ^* 上的关系。

可以证明,对于任意的 RL L , DFA M 接受 L , 则 R_M 是 R_L 的加细。也就是说,对于 $\forall q \in Q, x, y \in set(q)$, 必有 $x R_L y$ 成立,即 $x R_{L(M)} y$ 。事实上,因为

$$x, y \in set(q)$$

所以

$$\delta(q_0, x) = \delta(q_0, y) = q$$

从而使得对于 $\forall z \in \Sigma^*$,

$$\begin{aligned} \delta(q_0, xz) &= \delta(\delta(q_0, x), z) \\ &= \delta(q, z) \\ &= \delta(\delta(q_0, y), z) \\ &= \delta(q_0, yz) \end{aligned}$$

这就是说,

$$\delta(q_0, xz) \in F \Leftrightarrow \delta(q_0, yz) \in F$$

即对于 $\forall z \in \Sigma^*$,

$$xz \in L \Leftrightarrow yz \in L$$

表明

$$x R_L y$$

也就是

$$x R_{L(M)} y$$

成立。

定义 5-8 设 R 是 Σ^* 上的等价关系,对于 $\forall x, y \in \Sigma^*$,如果 $x R y$,则必有 $xz R yz$ 对于 $\forall z \in \Sigma^*$ 成立,则称 R 是**右不变的**(right invariant)等价关系。

定义 5-9 设 R 是 Σ^* 上的等价关系,则称 $|\Sigma^*/R|$ 是 R 关于 Σ^* 的**指数**(index),简称为 R 的指数。 Σ^* 的关于 R 的一个等价类,也就是 Σ^*/R 的任意一个元素,简称为 R 的一个等价类。

命题 5-1 对于任意 DFA $M=(Q, \Sigma, \delta, q_0, F)$, M 所确定的 Σ^* 上的关系 R_M 为右不变的等价关系。

证明要点:

(1) 证明 R_M 是等价关系。

通过证明 R_M 具有三歧性来完成。即此关系具有自反性、对称性、传递性。

(2) R_M 是右不变的。

证明 R_M 是右不变的。

命题 5-2 对于任意 $L \subseteq \Sigma^*$, L 所确定的 Σ^* 上的关系 R_L 为右不变的等价关系。

(1) 证明 R_L 是等价关系。

通过证明 R_L 具有三歧性来完成。

(2) R_L 是右不变的。

证明 R_L 是右不变的。

上述 4 个定义和 2 个命题都是为证明定理 5-7 服务的。甚至可以看作将 Myhill-Nerode 定理的证明拆分成 4 个定义、2 个命题和 1 个定理来完成。配上相关例题的分析,会使读者理解起来更加容易。其主要目的是为了将难点划分开来,使得读者容易建立比较清晰的思路。

定理 5-7 (Myhill-Nerode 定理) 如下 3 个命题等价:

(1) $L \subseteq \Sigma^*$ 是 RL。

(2) L 是 Σ^* 上的某一个具有有穷指数的右不变等价关系 R 的某些等价类的并。

(3) R_L 具有有穷指数。

证明要点: 与一般的多命题等价的证明方法一样,通过循环等价的证明来完成它们相互等价的证明。实际上,主教材所给的只是这种循环证明顺序的一种,在掌握了主教材给出的证明之后,读者很容易给出按照其他顺序进行的证明。

在由(1)推出(2)时,用到的是 DFA $M=(Q, \Sigma, \delta, q_0, F)$ 所确定的等价关系 R_M 将 Σ^* 分成有穷多个等价类

$$\{set(q) \mid q \in Q \text{ 并且 } q \text{ 是可达的}\}$$

由 Q 的有穷性,得到 $\{set(q) \mid q \in Q \text{ 并且 } q \text{ 是可达的}\}$ 的有穷性,也就是 R_M 指数的有穷性。而且

$$L = \bigcup_{q \in F} set(q)$$

由(2)推出(3)时,是基于当 L 是 Σ^* 上的具有有穷指数的右不变等价关系 R 的某些等价类的并时, R 一定是 R_L 的加细这一事实。有 3 点需要注意:一是 R 必须是 Σ^* 上的右不变等价关系,而不是其他集合上的;二是要求 R 具有有穷指数;三是 L 必须是某些等价类的并,否则也不能成立。

由(3)推出(1)时,实际上就是根据 L 所确定的关系 R_L 构造一个 DFA。这个 DFA 实际上就是接受 L 的状态个数最少的 DFA——最小 DFA。难点是对该 DFA 状态的理解。在所给的证明中,状态是用等价类命名的。而且直接地用这种命名表达出 DFA 的状态转移函数

$$\delta'([x], a) = [xa]$$

这可能会使读者感到比较“虚”。实际上,在前面曾经有意识地将等价类与状态对应起来,而且在第 3 章中,特意给出了一些例子,在给出的例子中,没有用类似 q 或 p_i 这样的符号去命名状态,而是用了 $q[00], q[01], q[a_1 a_2 \cdots a_i], A, Z$ 进行过渡,在本书第 3 章的典型习题解析中,更用了像 $[\varepsilon], [\text{奇 } 0], [\text{偶 } 1]$ 等形式的状态名。

另一处不易理解的是关于 δ' 定义相容性的证明。主要是这一步证明的必要性。那么为什么要证明其相容性呢? 主要是因为我们用 $[x]$ 代表 x 所在的等价类,一般地讲,该等价类中可

能含有除 x 以外的其他字符串,如 y 。这里要证明的是,无论将这个等价类记作 $[x]$ 还是记作 $[y]$,都是一样的,不会产生矛盾。这就是说,如果 $[x]=[y]$,对任意的 $a \in \Sigma$,由于 $\delta'([x],a)=[xa]$, $\delta'([y],a)=[ya]$,所以,必须有 $[xa]=[ya]$,否则 δ' 的定义是无效的。

由于此定理给出的是 RL 的充分必要条件(3 个命题等价),所以,它既可以被用来证明一个语言是 RL,也可以被用来证明一个语言不是 RL。通常是通过证明 R_L 的指数的有穷性来完成相应证明的。如果 R_L 有穷,则 L 为 RL;如果 R_L 无穷,则 L 不是 RL。这样一来,证明 L 是否是 RL 就转换成了根据 R_L 对 Σ^* 划分等价类的问题。例 5-10 是一个既便于理解,又能说明问题的例子,建议读者仔细地阅读。特别要注意的是,在该例子中怎样根据语言的结构特征来完成等价类的划分,从中找到一点启发。

另外,当 L 不是 RL 时,如果是用该定理进行判断,只看 R_L 是否将 Σ^* 分成了无穷多个等价类,如果是,则 L 不是 RL。这就是说,如果能找到无穷多个等价类,即可说明 L 不是 RL,一般此时都不用找出所有的等价类来,除非是问题要求计算出所有的等价类。下面的例子就说明了这一点。

例 5-10 用定理 5-7 证明 $\{0^n 1^n \mid n \geq 0\}$ 不是 RL。

该例的关键是, L 的句子具有以下两个主要特点:

- (1) 句子中所含的字符 0 的个数与所含的字符 1 的个数相同。
- (2) 所有的 0 都在所有的 1 的前面。

根据这两个特点,发现 0 和 00 不在同一个等价类中。进而由此受到启发,得出:如果 $h \neq j$,则 0^h 与 0^j 不在同一个等价类中,从而得到 R_L 的无穷多个等价类。

$[\epsilon]$ —— ϵ 所在的等价类
 $[1]$ ——0 所在的等价类
 $[2]$ ——00 所在的等价类
 $[3]$ ——000 所在的等价类
 $\vdots$
 $[n]$ —— 0^n 所在的等价类
 $\vdots$

所以, R_L 的指数是无穷的。因此, L 不是 RL。从上述还可以看出,因为只需要说明有无穷多个等价类,所以,并没有给出每个等价类所含的全部具体元素。例如,只指出 $[3]$ 代表的是 000 所在的等价类。

Myhill-Nerode 定理的如下两个推论说明接受 RL L 的最小 DFA 的存在性和同构意义下的唯一性。

推论 5-2 对于任意 RL L , 如果 DFA $M = (Q, \Sigma, \delta, q_0, F)$ 满足 $L(M) = L$, 则 $|\Sigma^* / R_{L(M)}| \leq |Q|$ 。

推论 5-3 对于任意 RL L , 在同构意义下, 接受 L 的最小 DFA 是唯一的。

证明要点:

(1) 先证明接受 L 的最小 DFA $M = (Q, \Sigma, \delta, q_0, F)$ 的状态数与 R_L 的指数相同, 也就是说, 这个最小 DFA 的状态数与 Myhill-Nerode 定理证明中构造的 $M' = (\Sigma^* / R_L, \Sigma, \delta', [\epsilon], \{[x] \mid x \in L\})$ 的状态数是相同的。