作业 L1.13 - L1.18：泵引理与 DFA 极小化

2025 年 12 月 16 日

题目 1

给出下述语言的最小泵长度，并加以证明：

$$0001^*,\quad 0^*1^*,\quad 001 \cup 0^*1^*,\quad 01^*,\quad \epsilon,\quad 1^*01^*01^*,\quad 1011,\quad \Sigma^*.$$

Solution.

这里的“最小泵长度”按正则语言泵引理的定义取：对语言 L，最小的 $p \geq 1$ 使得对任意 $w \in L$ 且 $|w| \geq p$，存在分解 $w = xyz$ 满足 $|xy| \leq p$、$|y| \geq 1$ 且对所有 $i \geq 0$ 有 $xy^iz \in L$。

一般地，若一个 DFA 有 n 个状态，则 L 的一个泵长度可取 $p = n$。而“最小泵长度”往往等于最小 DFA 的状态数，但需逐个验证（这里给出可行的最小值与理由）。

(a) $L = 0001^*$**：最小泵长度** $p = 4$。

最小 DFA 需要区分前缀是否已经读到恰好三个 0 以及是否已经进入尾部 1^*，共至少 4 个状态；因此 $p \geq 4$。取 $p = 4$ 可泵：任取 $|w| \geq 4$ 的 $w \in L$，必形如 0001^k，取 $x = \epsilon, y = 0, z = 001^k$（或取 y 为第 4 位的 1 当 $k \geq 1$），可检查 $xy^iz \in L$。

(b) $L = 0^*1^*$**：最小泵长度** $p = 2$。

最小 DFA 有 3 个状态（处于 0 段、处于 1 段、陷阱），因此可取 $p = 3$；但最小泵长度可更小。取 $p = 2$：任取 $|w| \geq 2$ 且 $w \in 0^*1^*$，若 w 以 0 开头取 $y = 0$；否则以 1 开头取 $y = 1$，均不会破坏“先 0 后 1”的结构，故可泵。

(c) $L = 001 \cup 0^*1^*$**：最小泵长度** $p = 2$。

因为 $0^*1^* \subseteq L$，对长度 ≥ 2 的串，L 与 0^*1^* 的泵分解策略同样适用；特殊串 001 长度为 3，也可取分解 $x = \epsilon, y = 0, z = 01$，$xy^iz = 0^{i+1}1 \in 0^*1^* \subseteq L$。因此 $p = 2$ 可行，且 $p = 1$ 不行（例如 $w = 01$ 时 $|xy| \leq 1$ 迫使 $y = 0$，泵 $i = 2$ 得 $001 \notin 0^*1^*$ 且不等于 001，矛盾）。

(d) $L = 01^*$**：最小泵长度** $p = 2$。

$p = 1$ 不行（$w = 01$ 时 $|xy| \leq 1$ 只能在首位 0 上泵，得 $001 \notin L$）。取 $p = 2$：对任意 $w = 01^k$, $|w| \geq 2$，令 $x = 0, y = 1, z = 1^{k-1}$，则 $xy^iz = 01^{k-1+i} \in L$。

(e) $L = \{\epsilon\}$**：最小泵长度** $p = 1$。

因为不存在长度 ≥ 1 的串属于该语言，泵引理条件对所有 w 真 vacuously true；按定义最小 p 可取 1。

(f) $L = 1^*01^*01^*$**：最小泵长度** $p = 3$。

该语言包含至少两个 0，最短非空串为 00（长度 2），但泵要求对 $|w| \geq p$ 才适用。取 $p = 3$：任取 w，只要在前 3 位内找到一个 1 或 0 作为 y，泵不会改变“恰好两个 0 的相对顺序且允许任意多 1”的结构。$p = 2$ 不行可用反例（例如 $w = 010$ 的泵分解会在前两位强制泵到结构外）。

(g) $L = \{1011\}$**：最小泵长度** $p = 5$。

这是有限语言；对 $p \leq 4$，串 $w = 1011$ 满足 $|w| \geq p$，但任何 $|xy| \leq p$ 的分解都会在泵 $i = 0$ 或 $i = 2$ 时改变长度，得到不在语言中的串，故不满足。取 $p = 5$ 时条件 vacuously true（无 $|w| \geq 5$ 的串属于该语言）。

(h) $L = \Sigma^*$**：最小泵长度** $p = 1$。

任取 $|w| \geq 1$，令 $x = \epsilon, y$ 为首符号，z 为剩余部分，则对所有 i，$xy^iz \in \Sigma^*$ 恒成立。

题目 2

使用泵引理证明下述语言不是正则的：

$$A_1 = \{0^n 1^n 2^n \mid n \ge 0\}, \quad A_2 = \{www \mid w \in \{a,b\}^*\}, \quad A_3 = \{a^{2^n} \mid n \ge 0\}.$$

Solution.

(1) A_1 非正则

假设 A_1 正则，设泵长度为 p。取

$$w = 0^p 1^p 2^p \in A_1.$$

任意分解 $w = xyz$ 满足 $|xy| \le p$，则 y 只包含 0（即 $y = 0^t, t \ge 1$）。泵 $i = 0$ 得

$$xz = 0^{p-t} 1^p 2^p,$$

0 的个数与 1、2 的个数不相等，故 $xz \notin A_1$，矛盾。

(2) A_2 非正则

假设 A_2 正则，泵长度为 p。取 $w = a^p b^p a^p b^p a^p b^p$，令 $u = a^p b^p$，则 $w = uuu \in A_2$。任意分解 $w = xyz$ 且 $|xy| \le p$ 时，y 只含 a（在第一个 a^p 内）。泵 $i = 2$ 得到前段 a 个数改变，整串无法写成同一个串的三次重复（因为三个块必须长度相等且内容相同），与 A_2 定义矛盾。

(3) A_3 非正则

假设 A_3 正则，泵长度为 p。取 $w = a^{2^p} \in A_3$。任意分解 $w = xyz$ 且 $|xy| \le p$ 时，$y = a^t$，$1 \le t \le p$。泵 $i = 2$ 得

$$xy^2z = a^{2^p + t}.$$

但 $2^p < 2^p + t < 2^{p+1}$，因此 $2^p + t$ 不是 2 的幂，故 $xy^2z \notin A_3$，矛盾。

题目 3

考察语言

$$F = \{a^i b^j c^k \mid i, j, k \ge 0, \text{且若} i = 1, \text{则} j = k\}.$$

(a) 证明：F 不是正则的

(b) 证明：对于给定泵长度 p，F 满足泵引理的三个条件

(c) 解释为什么上述两个结论与泵引理不矛盾

Solution.

(a) F **非正则**
取子语言 $F' = F \cap ab^*c^*$（即限制 $i = 1$），则

$$F' = \{ab^jc^k \mid j = k\} = \{ab^nc^n \mid n \geq 0\},$$

这是经典非正则语言，因此若 F 正则则其与正则语言的交仍正则，矛盾。故 F 非正则。
(b) 对任意给定 p，F 满足泵引理条件
泵引理是"正则 $\Rightarrow$ 可泵"的必要条件，但"可泵"并不推出正则。这里要展示：对任意给定 p，对每个 $|w| \geq p$ 的 $w \in F$，确实存在某个分解满足泵条件。
设 $w = a^ib^jc^k \in F$ 且 $|w| \geq p$。只需给出一种分解方案：

- 若 $i \geq 1$，取 $x = \epsilon$，$y = a$，$z = a^{i-1}b^jc^k$。则对所有 $t \geq 0$，

$$xy^tz = a^ta^{i-1}b^jc^k = a^{t+i-1}b^jc^k.$$

  当 $t+i-1 = 1$ 时（即 $t = 2-i$，只可能在 $i = 1$ 且 $t = 1$），原串满足 $j = k$；其余情况下 a 的指数 $\neq 1$，语言对 j, k 无额外约束，因此 $xy^tz \in F$。

- 若 $i = 0$，则 $w = b^jc^k$ 自动满足 $i \neq 1$ 的情形。取 $x = \epsilon$，$y = b$，$z = b^{j-1}c^k$（若 $j = 0$ 则取 $y = c$），泵动不会引入 $i = 1$，故仍在 F。

因此对任意给定 p，F 满足泵引理的三个形式条件。
(c) 不矛盾的原因
泵引理是**正则语言的必要条件**而非充分条件：存在非正则语言也"满足"泵引理形式条件（即对某个 p，每个足够长的串都能找到某种可泵分解）。因此"F 非正则"与"对任意给定 p，F 仍可满足泵条件"并不矛盾。

题目 4

设 $L = \{1^q \mid q$ 为素数$\}$，用泵引理证明该语言不是正则的。

Solution.

假设 L 正则，设泵长度为 p。取一个素数 $q > p$，并令 $w = 1^q \in L$。任意分解 $w = xyz$ 满足 $|xy| \leq p$，则 $y = 1^t$，其中 $1 \leq t \leq p$。泵 $i = q + 1$ 得

$$xy^{q+1}z = 1^{q+qt} = 1^{q(1+t)}.$$

指数 $q(1+t)$ 为合数（$q \geq 2$ 且 $1 + t \geq 2$），因此 $xy^{q+1}z \notin L$，与泵引理矛盾。故 L 非正则。

题目 5

极小化如下 DFA（两题，题面给出了状态图）。

Solution.

(1) 字母表 $\{a,b\}$ 的 DFA（状态 A,B,C,D,E，终态 E）
由题图读取转移：

	a	b
A	B	C
B	B	D
C	B	C
D	B	E
E	B	C

初态为 A，终态集合 $F=\{E\}$。

等价类划分（填表法同理）：初分割 $P_0=\{\{E\},\{A,B,C,D\}\}$。在 $\{A,B,C,D\}$ 内，D 在输入 b 下转入终态 E，其余不转入，故分裂为

$$P_1=\{\{E\},\{D\},\{A,B,C\}\}.$$

再看 $\{A,B,C\}$：在输入 b 下，$B\to D$，而 A,C 都留在 $\{A,B,C\}$，故再分裂为

$$P_2=\{\{E\},\{D\},\{B\},\{A,C\}\}.$$

最后检查 $\{A,C\}$：两者在 a 下都到 B，在 b 下都回到 $\{A,C\}$，因此等价。

所以最小 DFA 有 4 个状态：$[AC],[B],[D],[E]$，初态为 $[AC]$，终态为 $[E]$。其转移为：

	a	b
$[AC]$	$[B]$	$[AC]$
$[B]$	$[B]$	$[D]$
$[D]$	$[B]$	$[E]$
$[E]$	$[B]$	$[AC]$

(2) 字母表 $\{0,1\}$ 的 DFA（状态 $Q_0,\ldots,Q_5$，终态 Q_1,Q_2,Q_5）
由题图读取转移：

	0	1
Q_0	Q_1	Q_2
Q_1	Q_3	Q_4
Q_2	Q_4	Q_3
Q_3	Q_5	Q_5
Q_4	Q_5	Q_5
Q_5	Q_5	Q_5

终态集合 $F=\{Q_1,Q_2,Q_5\}$。

初分割：

$$P_0=\{\{Q_1,Q_2,Q_5\},\{Q_0,Q_3,Q_4\}\}.$$

在终态类中，Q_5 在 $0/1$ 下都回到自身，而 Q_1,Q_2 都会转入非终态类，故

$$P_1=\{\{Q_5\},\{Q_1,Q_2\},\{Q_0,Q_3,Q_4\}\}.$$

在 $\{Q_0,Q_3,Q_4\}$ 中，Q_3,Q_4 都在 $0/1$ 下转入 Q_5，而 Q_0 转入 $\{Q_1,Q_2\}$，因此

$$P_2=\{\{Q_5\},\{Q_1,Q_2\},\{Q_3,Q_4\},\{Q_0\}\}.$$

检查可知 $\{Q_1, Q_2\}$ 与 $\{Q_3, Q_4\}$ 都不再分裂。

故最小 DFA 的 4 个状态可记为 $[Q_0], [Q_{12}], [Q_{34}], [Q_5]$，初态 $[Q_0]$，终态为 $[Q_{12}]$ 与 $[Q_5]$。转移为:

	0	1
$[Q_0]$	$[Q_{12}]$	$[Q_{12}]$
$[Q_{12}]$	$[Q_{34}]$	$[Q_{34}]$
$[Q_{34}]$	$[Q_5]$	$[Q_5]$
$[Q_5]$	$[Q_5]$	$[Q_5]$

题目 6

利用迈希尔—尼罗德定理证明若干语言是否正则（题面列出 A_1, A_2, A_3 以及关于 A, B 的结论）。

Solution.

题面（见 `assets/problem_L01_13-L01_18.pdf`）给出:

$$A_1 = \{0^n 1^n 2^n \mid n \geq 0\},$$
$$A_2 = \{www \mid w \in \{a, b\}^*\},$$
$$A_3 = \{a^{2^n} \mid n \geq 0\},$$
$$\Sigma = \{0, 1\}, \quad A = \{0^k u 0^k \mid k \geq 1,\ u \in \Sigma^*\}, \quad B = \{0^k 1 u 0^k \mid k \geq 1,\ u \in \Sigma^*\}.$$

用迈希尔—尼罗德定理：若语言正则，则其不可区分关系 $\equiv_L$ 只有有限多个等价类；反之若能构造无限多个两两可区分串，则语言非正则。

(1) A_1 不是正则语言

取集合 $X = \{0^n \mid n \geq 0\}$。对任意 $m \neq n$，令后缀 $z_n = 1^n 2^n$，则

$$0^n z_n = 0^n 1^n 2^n \in A_1, \qquad 0^m z_n = 0^m 1^n 2^n \notin A_1.$$

因此 0^m 与 0^n 可区分，$\equiv_{A_1}$ 有无限多个等价类，故 A_1 非正则。

(2) A_2 不是正则语言

取 $x_n = a^n$。令后缀

$$z_n = b\, a^n\, b\, a^n\, b.$$

则

$$x_n z_n = a^n b a^n b a^n b = (a^n b)(a^n b)(a^n b) \in A_2.$$

对任意 $m \neq n$,

$$x_m z_n = a^m b a^n b a^n b$$

若它属于 A_2，则必可写为 uuu。由于串中恰有 3 个字母 b，可推出 u 中恰有 1 个 b，且 u 必以 b 结尾，因此 u 只能形如 $a^t b$。于是 uuu 的三段 a 的长度必须相同，与本串三段 a 的长度为 m, n, n（且 $m \neq n$）矛盾。故 $x_m z_n \notin A_2$。因此 $\{a^n\}$ 两两可区分，A_2 非正则。

(3) A_3 不是正则语言

取 $x_n = a^{2^n}$。令后缀 $z_n = a^{2^n}$，则

$$x_n z_n = a^{2^n + 2^n} = a^{2^{n+1}} \in A_3.$$

若 $m \neq n$，则 $2^m + 2^n$ 不是 2 的幂，因此

$$x_m z_n = a^{2^m+2^n} \notin A_3.$$

所以 $\equiv_{A_3}$ 有无限多个等价类，A_3 非正则。

(4) A 是正则语言

注意到对任意 $w \in A$，w 至少以一个 0 开头且以一个 0 结尾；反之，任意以 0 开头且以 0 结尾的串都可取 $k = 1$ 写成 $0^k u 0^k$。因此

$$A = 0\Sigma^*0,$$

是正则语言（正则表达式即 $0(0+1)^*0$）。

(5) B 不是正则语言

取串集 $x_k = 0^k 1$ $(k \geq 1)$。对任意 $i < j$，取后缀 $z = 0^i$，则

$$x_i z = 0^i 10^i \in B \quad (\text{取} u = \epsilon),$$

但

$$x_j z = 0^j 10^i \notin B$$

因为其第一个 1 出现在 j 个 0 之后，若按 B 的形式必须以 $0^j 1$ 开头并以 0^j 结尾，但该串只以 0^i 结尾（$i < j$）。因此 $\{x_k\}$ 两两可区分，$\equiv_B$ 有无限多个等价类，故 B 非正则。