

Homework 2

Name: 方嘉聪 ID: 2200017849

Problem 1.(12 points). Give context-free grammars that generate the following languages. In all parts the alphabet Σ is $\{0, 1\}$.

- (a) $\{w \mid w \text{ starts and ends with the same symbol}\}$.
- (b) $\{w \mid w \text{ is a palindrome}\}$.

Answer. (a) $G_1 = (\{S, S_1\}, \Sigma, R, S)$, 其中 R 为:

$$S \rightarrow 0S_10 \mid 1S_11, \quad S_1 \rightarrow 0S_1 \mid 1S_1 \mid \varepsilon$$

- (b) $G_2 = (\{S_1\}, \Sigma, R, S_1)$, 其中 R 为:

$$S_1 \rightarrow 0S_10 \mid 1S_11 \mid 0 \mid 1 \mid \varepsilon$$

Problem 2.(12 points). For any language A with alphabet Σ , let

$$\text{PREFIX}(A) = \{w \mid wv \in A \text{ for some } v \in \Sigma^*\}$$

Show that the class of context-free languages is closed under the PREFIX operation.

Answer. 设存在一个满足 *Chomsky normal form* 的上下文无关文法 $G = (V, \Sigma, R, S)$ 使得 $L(G) = A$, 同时假设 G 中的每一个变量 A 都能够最终推导出字符串 w , 即 $\exists w, s.t. A \Rightarrow_R^* w$ (否则, 删去不满足这样条件的变量和与之相关的推导规则). 下面我们从 G 出发构造 $G' = (V', \Sigma, R', S')$ 使得 $L(G') = \text{PREFIX}(A)$: 初始状态 V', R', S' 均为空:

1. 首先对 $\forall A \in V$, 添加变量 A, A_0 到 V' 中.
2. 对于 R 中的每一个规则 $A \rightarrow BC$, 添加规则 $A \rightarrow BC, A_0 \rightarrow BC_0 \mid B_0$ 到 R' 中.
3. 对于 R 中的每一个规则 $A \rightarrow a$, 添加规则 $A \rightarrow a, A_0 \rightarrow a \mid \varepsilon$ 到 R' 中.

如果 R 中有规则 $S \rightarrow \varepsilon$, 则添加规则 $S \rightarrow \varepsilon, S_0 \rightarrow \varepsilon$ 到 R' 中. 同时令 $S' = S_0$. 注意到 G' 中任意一条规则的左侧如果是 G 中的变量则右侧的变量都是 G 中规则对应的变量, 故有:

$$X \Rightarrow_G^* w \text{ iff } X \Rightarrow_{G'}^* w \quad \forall w \in \Sigma^*, X \in V$$

下面我们证明 $L(G') = \text{PREFIX}(A)$, 即证明 $X_0 \Rightarrow_{G'}^* w$ iff $X \Rightarrow_G^* wv$ for some $v \in \Sigma^*$.

1. 首先证明对 $\forall X \in V, \forall w s.t. X_0 \Rightarrow_{G'}^* w$, 都 $\exists v \in \Sigma^* s.t. X \Rightarrow_G^* wv$. 我们对得到 w 的推导长度进行归纳:

- 当 $X_0 \Rightarrow_{G'} w$ 时, $w = \varepsilon$ 或 $w = c$ (其中 $c \in \Sigma$), 那么根据我们的构造, G 中存在规则 $X \rightarrow c$, 从而 $X \Rightarrow_G^* c$. 故 w 是 c 的一个前缀, 满足条件.
- 假设对于所有长度小于 n 的推导, 结论成立. 考虑长度为 n 的推导 $X_0 \Rightarrow_{G'}^* w$ 时, 有两种情况, $X_0 \Rightarrow_{G'} YZ_0$ 或 $X_0 \Rightarrow_{G'} Y_0$: (1) 对于前者, $\exists y, z$ s.t. $Y \Rightarrow_G^* y, Z_0 \Rightarrow_{G'}^* z$, i.e. $w = yz$. 由归纳假设知 $\exists z'$ s.t. $Z \Rightarrow_G^* z'$ 且 z 为 z' 的前缀. 从而有 $X \Rightarrow_G YZ \Rightarrow_G^* yz'$ 且 yz 为 yz' 的前缀, 命题成立. (2) 对于后者, 则有 $Y_0 \Rightarrow_{G'}^* w$, 由对那假设知 $\exists v \in \Sigma^*$, s.t. $Y \Rightarrow_G^* wv$, 进而有 $X \Rightarrow_G YZ \Rightarrow_R^* wvt$, 且 w 为 wv 的前缀, 命题成立.

综上, 这一命题成立.

2. 接着证明对任意一个由 X 生成的字符串 w 的前缀都可以由 X_0 生成. 我们对字符串 w 的长度进行归纳.

- 当 $|w| = 1$ 时, 设 $w = c \in \Sigma$, 由于 G 是 *Chomsky normal form* 的, 那么有 $X \Rightarrow_G c$, 故 $X \rightarrow c$ 在 R 中. 根据我们的构造 $X_0 \rightarrow c \mid \varepsilon$ 在 R' 中, 故 c 的前缀 ε, c 都可以由 X_0 生成.
- 假设对任意长度小于 n ($n \geq 2$) 的 w 命题都成立. 考虑长度为 n 的 w , 由于 G 是 *Chomsky normal form* 的, 那么 w 是由这样形式的生成链 $X \Rightarrow_G YZ \Rightarrow_G^* w$ 生成的. 故 $\exists y, z$ s.t. $w = yz$ 且 $Y \Rightarrow_G^* y, Z \Rightarrow_G^* z$. 由归纳假设知 y, z 所有的前缀都可以分别由 Y_0, Z_0 生成. 而 $w = yz$ 的前缀要么是 y 的前缀, 要么是 yt (t 为 z 的前缀). 由我们的构造知 $X_0 \rightarrow YZ_0, X_0 \rightarrow Y_0$ 在 R' 中, 故有 $X_0 \Rightarrow_{G'} Y_0, X_0 \Rightarrow_{G'} YZ_0 \Rightarrow_{G'}^* yZ_0$, 即 X_0 可以生成 w 的所有前缀.

综上, 这一命题成立.

由上述证明知 $L(G') = \text{PREFIX}(A)$, 即上下文无关语言对于前缀操作是封闭的. ◁

Problem 3.(12 points).

- Prove that the language $L = \{0^{2^n} \mid n \geq 0\}$ is not context free.
- Let B be the language of all palindromes over $\{0, 1\}$ containing an equal number of 0s and 1s. Show that B is not context free.

Answer. (a) 假设 L 是上下文无关的, 设 p 为 *pumping length*. 考虑字符串 $w = 0^{2^k} \in L$, 其中 k 取充分大使得 $2^k > p$. 由 *pumping lemma* 知, 存在 $u = 0^a, v = 0^b, x = 0^c, y = 0^d, z = 0^e$ 使得 $w = uvxyz$, 即 $a + b + c + d + e = 2^k$ 且 $b + d > 0, b + c + d \leq q < 2^k$. 考虑字符串

$$w' = uv^2xy^2z = 0^{2^k + b + d}, \quad 2^k < |w'| = 2^k + b + d < 2^{k+1}$$

由 *pumping lemma* 知 $w' \in L$, 但由 L 的定义 $w' \notin L$. 矛盾, 故 L 不是上下文无关的.

- 假设 B 是上下文无关的, 设 p 为 *pumping length*. 考虑字符串 $w = 0^p 1^{2p} 0^p \in B$. 由 *pumping lemma* 知, 存在 u, v, x, y, z 使得 $w = uvxyz$, 且 $|vxy| \leq p, |vy| > 0$. 由于 $|vxy| \leq q$, 故 v 和 y 不能同时分别包含前 p 位和后 p 位的 0. 不妨设 uxy 落在前 $3p$ 位上, 下面分情况讨论:

- 如果 uxy 全都位于前 p 位, 考虑 $w' = uxz$, 那么前 p 位的 0 的个数为 $p - |vy| < p$, 与后 p 位的 0 的个数不相等, 故 $w' \notin B$.
- 如果 uxy 全都位于中间 $2p$ 位, 考虑 $w' = uxz$, 那么 w' 的 1 的个数为 $2p - |vy| < 2p$, 与 w' 的 0 的个数不相等, 故 $w' \notin B$.

- 如果 uxy 跨过 $0,1$ 分界, 考虑 $w' = uxz$, 那么前 p 位的 0 的个数减小, 后 p 位的 0 的个数不变, 故 $w' \notin B$.

综上, B 不是上下文无关的. \triangleleft

Problem 4.(16 points). Define a two-dimensional Turing machine to be a TM where each of its tapes is an infinite grid (and the read/write head can move not only Left and Right but also Up and Down). Show that for every $T : \mathbb{N} \rightarrow \mathbb{N}$ and every Boolean function f , if f can be computed in time $T(n)$ using a two-dimensional TM, then it can also be computed by a standard (one-dimensional) TM in time $O(T(n)^2)$. (T may not be time constructible) \blacktriangleleft

Answer. 用一个 5-tape 的一维图灵机来模拟二维图灵机的读写头移动. 我们先扩充二维图灵机 M_2 的符号集 Σ , 增加若干个特殊符号 (比如添加 \hat{a}) 用于标记 (可以使用 c 个格子来编码一个字符, 其中 c 是只与 $|\Sigma|$ 相关的常数), 记为 Σ' .

主要问题是 M_2 的垂直移动, 为了避免垂直移动的编号变化过大, 我们采用如下编号方式 (原点编号为 0 的平面直角坐标), 从起点 $(0,0)$ 出发先向右再沿对角线逆时针依次增加 1 (每次回到 y 负半轴就向右移一次):

34	19	8	17	30	47
20	9	2	7	16	29
10	3	0	1	6	15
22	11	4	5	14	27
38	23	12	13	26	43

这样可以将二维纸带的信息映射到一条一维纸带 T_0 上, 用 T_0 来模拟二维图灵机的读写头移动 (用特殊字符作为边界划分出各个区域内部有编号和应存储的对应字符).

同时使用两条纸带 T_1, T_2 来计算 x, y 两维的坐标信息:

- T_1 : 记录 $|x|$ 的信息 (用第一个位表示正负, 用从左到指针处的 $\$$ 的个数来表示绝对值大小), 用特殊字符标记边界
- T_2 : 记录 $|y|$ 的信息, 细节同 T_1 .

还有两个纸带用来记录 input 和 output 的字符串.

每次运算过程中, 先读取所在区域的内容, 按照 M_2 的状态转移, 修改内容并通过 T_0, T_1, T_2 的信息计算出下一步的编号 (读取当前 $|x|, |y|$ 的大小和符号, 需要用 $|x| + |y| + 2$ 步.), 移动到对应区域 (需要 $c\Delta$ 步, Δ 为编号变化, c 是一个常量).

此外采取边计算边扩展纸带的方式 (这样就不需要预先算出 $T(n)$ 并初始化所需的区域), 用特殊字符标记已经初始化区域的边界. 当进入未初始化的区域时, 边初始化边移动 (初始化指写入编号和内容, 如果是内容暂时为空的区域就先预留对应编码所需的最大空间).

设当前位置坐标为 (x, y) , $\delta = |x| + |y|$, 可以验证 M_2 每次朝上下左右移动的编号变化不会超过 $O(\delta)$, 而编号每变化 1 使用的时间是常数级别的, 注意到 $\delta < T(n)$, 故用一维图灵机模拟二维图灵机的一次移动的用时为 $O(T(n))$. 故整个过程的时间复杂度为 $O(T(n)^2)$. \triangleleft

Problem 5.(16 points). Let $T = \{\langle M \rangle \mid M \text{ is a TM that accepts } \alpha^R \text{ whenever it accepts } \alpha\}$. Show that T is undecidable. Note: α^R is the reverse string of α . You may assume the alphabet is $\{0, 1\}$. ◀

Answer. 我们考虑将 $A_{TM} = \{\langle M, \alpha \rangle \mid \text{TM } M \text{ accepts } \alpha\}$ 规约到 T 上. 给定 $\langle M, \alpha \rangle$, 我们希望构造一个 $\langle M' \rangle$ 使得 $\forall \langle M, \alpha \rangle \in A_{TM} \iff \langle M' \rangle \in T$. 我们如下构造 M' :

-
- 1: **Given** $\langle M, \alpha \rangle$, **on input** x , **Construct** M' **as follows:**
 - 2: **if** x is a string of the form $0^n 1^n$ **then**
 - 3: Run M on α
 - 4: **return** $M(\alpha)$
 - 5: **else if** x is a string of the form $1^n 0^n$ **then**
 - 6: **return** *Accept*
 - 7: **else**
 - 8: **return** *Reject*
-

注意到当 $\langle M, \alpha \rangle \in A_{TM} \implies L(M') = \{0^n 1^n, 1^n 0^n\} \implies \langle M' \rangle \in T$. 反之, 当 $\langle M, \alpha \rangle \notin A_{TM} \implies L(M') = \{1^n 0^n\} \implies \langle M' \rangle \notin T$. 故 $A_{TM} \leq_m T$, 由 A_{TM} 的不可判定性知 T 不可判定. ◀

Problem 6.(16 points). Define the language

$$C_{TM} = \{\langle M_1, M_2 \rangle \mid M_1, M_2 \text{ are two Turing machines such that } L(M_1) \subseteq L(M_2)\}$$

Show that C_{TM} is undecidable. ◀

Answer. 假设 C_{TM} 是可判定的, 设图灵机 S 可以判定 C_{TM} . 先构造两个图灵机 M_1, M_2 :

-
- 1: **On input** x , **Construct** M_1 **as follows:**
 - 2: **if** $x \neq w$ **then**
 - 3: **return** *Reject*
 - 4: **else**
 - 5: Run M on w . And **return** *Accept* if M *Reject*, **return** *Reject* if M *Accept*.
-

- 1: **On input** x , **Construct** M_2 **as follows:**
 - 2: Whatever x is, **return** *Reject*.
-

下面我们构造一个图灵机 M' 来判定 A_{TM} :

-
- 1: **On input** $\langle M, w \rangle$, **Construct** M' **as follows:**
 - 2: Use the description of M, w to construct M_1, M_2 as above.
 - 3: Run S on $\langle M_1, M_2 \rangle$
 - 4: **Accept** if S *Accept*, **Reject** if S *Reject*.
-

注意到 $L(M_2) = \emptyset$, 注意到

$$\begin{aligned} M' \text{ accept } \langle M, w \rangle &\iff S \text{ accept } \langle M_1, M_2 \rangle \iff L(M_1) \subseteq L(M_2) = \emptyset \\ &\iff L(M_1) = \emptyset, \text{ i.e. } M \text{ accept } w. \end{aligned}$$

由此可知如果 S 可以判定 C_{TM} 则 M' 可以判定 A_{TM} . 但是 A_{TM} 是不可判定的, 这产生了矛盾. 故 C_{TM} 也是不可判定的. ◀

Problem 7.(16 points). Show that single-tape TMs that cannot write on the tape recognize only regular languages.

Hint: You may use the conclusion from Exercise 4 in HW1, which is a simplified version of Myhill-Nerode theorem. ◀

Answer. 设 $M = (Q, \Sigma, \Gamma, q_0, q_{\text{accept}}, q_{\text{reject}})$ 是一个单纸带的只读图灵机. 设 $k = |Q|$, 我们来证明 $L(M)$ 的等价类存在一个有限的上界 (大概是 $2^{\Omega(k)}$ 的量级).

定义等价关系 $\equiv_{L(M)}$ 为: $x \equiv_{L(M)} y \iff (\forall z \in \Sigma^*, xz \in L(M) \iff yz \in L(M))$. 注意到 M 的状态是有限的, 主要想法是建立一个从状态到等价类的映射, 从而证明等价类的个数是有限的.

记 $\#$ 为一个表示不停机状态的字符, 记 $Q_* = Q \cup \{\#\}$, 则 $|Q_*| = k + 1$. 定义 $f: \Sigma^* \rightarrow Q \times \{\#\}$ 为:

$f(w) = (q) \iff$ 从状态 q_0 , 指针从 w 第一个字符出发, 第一次指到了 w 后的字符时, 状态为 q

$f(w) = (\#) \iff$ 从状态 q_0 , 指针从 w 第一个字符出发, 在 w 内部无限循环

定义 $h_{q_i}(w): \Sigma^* \times Q \rightarrow Q \times \{\#\}$ 为:

$h_{q_i}(w) = (q) \iff$ 从状态 q_i , 指针在 w 最后一个字符出发, 第一次指到了 w 后的字符时, 状态为 q

$h_{q_i}(w) = (\#) \iff$ 从状态 q_i , 指针在 w 最后一个字符出发, 在 w 内部无限循环

由 f, g 出发考虑映射 $T: \Sigma^* \rightarrow Q_*^{k+1}$, 其中 Q_*^{k+1} 表示 $k+1$ 元组的笛卡尔积. 定义 T 为:

$$T(w) = (f(w), h_{q_1}(w), \dots, h_{q_k}(w))$$

那么我们可以推出 $T(w_1) = T(w_2) \implies w_1 \equiv_{L(M)} w_2$. 由此成功建立了从状态到等价类的映射, 故等价类的个数不超过 $|Q_*^{k+1}| = (k+1)^{k+1}$, 上界是有限的.

最后我们来简要证明一下 $T(w_1) = T(w_2) \implies w_1 \equiv_{L(M)} w_2$, 即

$$\forall v \in \Sigma^*, w_1v \in L(M) \iff w_2v \in L(M)$$

将 w_1v, w_2v 分别放在 M 的纸带上, 从状态 q_0 开始, 由于 $T(w_1) = T(w_2)$, 从而有 $f(w_1) = f(w_2)$, 如果 w_1 不停机那么 w_2 也不停机. 如果 w_1, w_2 会停机, 那么根据我们的定义可知, 当 M 把 w_1, w_2 全部读完时能够保持在同一状态下 (f 保证指针第一次离开 w_i 时状态相同, g 保证后续指针再次若干次进出 w_i 到最终读取完整个 w_i 的状态相同), 从而有 $w_1 \equiv_{L(M)} w_2$, 证毕.

综上所述, $L(M)$ 的等价类是有限的, 由 Myhill-Nerode 定理知 $L(M)$ 是正则的. ◁