

理论计算机科学基础

期中整理

郭嘉睿

ntguojiarui@pku.edu.cn

2022 年 1 月 16 日

0 预备知识

在本文档中, 字母表 Σ 定义为任意非空有穷集合, 字符串由字母表中若干字母组成, 串长度定义为串所包含的字母数, 用 $|x|$ 表示, 子串定义为串中连续长度的一段,

$$\Sigma^* = \{x|x \text{ 为 } \Sigma \text{ 上的有穷长度的串}\},$$

$$\Sigma^+ = \{x|x \text{ 为 } \Sigma \text{ 上的正有穷长度的串}\},$$

$$\Sigma^\infty = \{x|x \text{ 为 } \Sigma \text{ 上的无穷长度的串}\},$$

语言定义为串的集合 $A \subseteq \Sigma^*$, 空语言定义为空集 \emptyset .

1 正则语言

定义 1.1 (DFA). 有穷自动机是一个五元组 $M = (Q, \Sigma, \delta, q_0, F)$, 其中, Q 为有穷状态集, Σ 为输入字母表, $\delta: Q \times \Sigma \rightarrow Q$ 为转移函数, $q_0 \in Q$ 为初始状态, $F \subseteq Q$ 为接收状态集.

定义 1.2 (正则语言). 有穷自动机 M 接受的语言称为正则语言, 用 $L = L(M)$ 表示.

定义 1.3 (正则运算). 设 A, B 是两个语言, 正则运算是指

1. 并: $A \cup B = \{x|x \in A \vee x \in B\}$;
2. 连接: $AB = \{xy|x \in A \wedge y \in B\}$;
3. 星号: $A^* = \bigcup_{i=0}^{\infty} A^i = \{x_1 \cdots x_k | k \geq 0, x_i \in A\}$.

定义 1.4 (NFA). 非确定性有穷自动机是一个五元组 $N = (Q, \Sigma, \delta, q_0, F)$, 其中, Q 为有穷状态集, Σ 为输入字母表, $\Sigma_\epsilon = \Sigma \cup \{\epsilon\}$, $\delta: Q \times \Sigma_\epsilon \rightarrow P(Q)$ 为转移函数, $q_0 \in Q$ 为初始状态, $F \subseteq Q$ 为接收状态集.

对于非确定性有穷自动机, N 接受 w 当且仅当存在接受计算。

定理 1.5 (NFA 和 DFA 的等价性). 每个 NFA 都有等价 DFA.

证明. 思路: 构造等价 DFA, 对于 NFA 的 k 个状态, 用 DFA 的 2^k 个状态去模拟。 □

定义 1.6 (REX). R 是正则表达式, 当且仅当 R 是 (递归定义)

1. $a, a \in \Sigma$;

2. ε ;
3. \emptyset ;
4. $R_1 \cup R_2$, R_1, R_2 都是正则表达式;
5. $R_1 R_2$, R_1, R_2 都是正则表达式;
6. R_1^* , R_1 是正则表达式;

这里, 运算优先级规定为 $* > \cdot > \cup$.

定理 1.7 (REX 与正则语言的等价性). 一个语言是正则的当且仅当可用正则表达式描述该语言.

定理 1.8 (泵引理). 设 A 是正则语言, 则存在常数 p , s.t. 若 $s \in A$ 且 $|s| \geq p$, 则 $s = xyz$, 且满足以下条件:

1. $\forall i \geq 0, xy^i z \in A$;
2. $|y| > 0$;
3. $|xy| \leq p$.

2 上下文无关语言

定义 2.1 (CFG). 上下文无关文法是一个四元组 $G = (V, \Sigma, R, S)$, 其中, V 为有穷变元集, Σ 为有穷终结符集, R 为有穷规则集 (规则形如 $A \rightarrow w, w \in (V \cup \Sigma)^*$), $S \in V$ 是一个初始变元.

定义 2.2 (CFL). 上下文无关文法生成的语言称为上下文无关语言, 用 $L = L(G)$ 表示.

定理 2.3. 正则语言都是 CFL.

定义 2.4 (CNF). 称一个 CFG 为 Chomsky 范式, 若它的每一条规则都具有如下形式:

1. $S \rightarrow \varepsilon$;
2. $A \rightarrow BC$;
3. $A \rightarrow a$.

这里, A, B, C 是任意变元, B, C 不是初始变元, a 是任意终结符.

定理 2.5. 任何 CFG 都有等价 CNF.

定义 2.6 (PDA). 下推自动机是一个六元组 $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$, 其中, Q 为有穷状态集, Σ 为输入字母表, $\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$, Γ 为栈字母表, $\Gamma_\varepsilon = \Gamma \cup \{\varepsilon\}$, $\delta: Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \rightarrow P(Q \times \Gamma_\varepsilon)$ 为转移函数, $q_0 \in Q$ 为初始状态, $F \subseteq Q$ 为接收状态集.

定理 2.7. 一个语言是 CFL 当且仅当存在 PDA 识别它.

定理 2.8 (泵引理). 设 A 是上下文无关语言, 则存在常数 p , s.t. 若 $s \in A$ 且 $|s| \geq p$, 则 $s = uvxyz$, 且满足以下条件:

1. $\forall i \geq 0, uv^i xy^i z \in A$;
2. $|vy| > 0$;
3. $|vxy| \leq p$.

3 图灵机

定义 3.1 (TM). 单带图灵机是一个七元组 $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{acc}, q_{rej})$, 其中, Q 为有穷状态集, Σ 为输入字母表, 空格符 $B \notin \Sigma$, Γ 为带字母表, $\Sigma \cup B \subseteq \Gamma$, $\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{L, R\}$ 为转移函数, $q_0 \in Q$ 为初始状态, q_{acc} 为停机接受状态, q_{rej} 为停机拒绝状态, $q_{acc} \neq q_{rej}$.

对于图灵机, 它的计算结果包括停机接受、停机拒绝和不停机.

	可判定 (可计算)	可识别 (半可计算)	补可识别 (补半可计算)
$x \in A$	停机接受	停机接受	停机接受/不停机
$x \notin A$	停机拒绝	停机拒绝/不停机	停机接受

定理 3.2. 图灵可识别等价于图灵可枚举.

证明. \Rightarrow : 枚举 Σ^* 逐个识别. 利用楔形, 定义 $A_i = \{w | w \text{ 运行 } i \text{ 步后接受}\}$, 则 $A = \bigcup_{i=1}^{\infty} A_i$;

\Leftarrow : 逐个枚举等待出现. □

定理 3.3 (CSL). 给定一台图灵机, 它的所有接受计算历史的集合构成上下文有关语言.

定义 3.4 (LBA). 线性界限自动机是一台带头不能移出输入区的图灵机 (等价于带头不能移出输入区的常数倍).

定理 3.5. 一个语言是 CSL 当且仅当存在 LBA 识别它.

4 归约, (不) 可计算性

定义 4.1 (m 归约). 设 A, B 是语言, 若存在可计算函数 $f: \Sigma^* \rightarrow \Sigma^*$, s.t. $x \in A \Leftrightarrow f(x) \in B$, 则说语言 A 可以 m 归约到 B , 记为 $A \leq_m B$ 或 $A \leq_m B$ via f .

定义 4.2 (Turing 归约). 设 A, B 是语言, 称 A 可以 Turing 归约到 B , 若 A 相对于 B 可判定, 记为 $A \leq_T B$.

4.1 正则语言的可判定性

定理 4.3 (A_{DFA}). $A_{DFA} = \{\langle B, w \rangle | DFA: B \text{ 接受 } w\}$ 可判定.

证明. 模拟 DFA 的判定过程, 用 TM: M 模拟 B , 跟踪 B 的状态. □

定理 4.4 (E_{DFA}). $E_{DFA} = \{\langle B \rangle | DFA: B \text{ 不派生任何串}\}$ 可判定.

证明. 利用图的连通性. □

定理 4.5 (EQ_{DFA}). $EQ_{DFA} = \{\langle A, B \rangle | DFA: A, B, L(A) = L(B)\}$ 可判定.

证明. 只要判定 $L(A) \oplus L(B) = (L(A) - L(B)) \cup (L(B) - L(A))$ 是否为空, 而正则语言对布尔运算封闭. □

4.2 上下文无关语言的可判定性

定理 4.6 (A_{CFG}). $A_{CFG} = \{\langle G, w \rangle | CFG: G \text{ 派生 } w\}$ 可判定.

证明. 利用 Chomsky 范式, 一个长度为 n 的串必定通过 $2n - 1$ 次派生得到. □

定理 4.7 (E_{CFG}). $E_{CFG} = \{\langle G \rangle | CFG: G, L(G) = \emptyset\}$ 可判定.

证明. 利用 Chomsky 范式, 依次检查每个变元是否产生终结符串. \square

定理 4.8 (ALL_{CFG}). $\text{ALL}_{\text{CFG}} = \{\langle G \rangle | \text{CFG}: G, L(G) = \Sigma^*\}$ 不可判定.

证明. 利用 $\overline{\text{A}_{\text{TM}}} \leq_m \text{ALL}_{\text{CFG}}$, 由于 TM 的非接受计算历史的集合构成 CFL, 对于 TM: A 和串 w , 若 A 接受 w , 则它的非接受计算历史的集合不为 Σ^* ; 若 A 不接受 w , 则它的非接受计算历史的集合为 Σ^* . 因此, 检查 A 的非接受计算历史的集合 B 是否为 Σ^* 可以判定 A 是否接受 w , 从而 $\overline{\text{A}_{\text{TM}}} \leq_m \text{ALL}_{\text{CFG}}$. 由于 A_{TM} 不可判定, 故 ALL_{CFG} 也不可判定. \square

定理 4.9 (EQ_{CFG}). $\text{EQ}_{\text{CFG}} = \{\langle G, H \rangle | \text{CFG}: G, H, L(G) = L(H)\}$ 不可判定.

证明. 构造 $L(H) = \Sigma^*$, 则判定 G 是否与 H 等价可以判断 $L(G)$ 是否为 Σ^* , 从而 $\text{ALL}_{\text{CFG}} \leq_m \text{EQ}_{\text{CFG}}$. \square

4.3 图灵机的可判定性

图灵机的所有问题几乎都是不可判定的.

定理 4.10 (A_{TM}). $\text{A}_{\text{TM}} = \{\langle M, w \rangle | \text{TM}: M \text{ 接受 } w\}$ 不可判定.

证明. 利用对角线法则, 将每个 TM 化为一个串. 定义

$$\text{D}_{\text{TM}} = \{\langle M \rangle | \text{TM}: M \text{ 接受 } M\}.$$

设计 TM: U ,

$$U(M, M) = \begin{cases} \text{接受}, & \text{若 } M \text{ 拒绝 } M; \\ \text{拒绝}, & \text{若 } M \text{ 接受 } M. \end{cases}$$

检查 $U(U, U)$, 有 U 接受 $U \Leftrightarrow U$ 拒绝 U , 矛盾! \square

定理 4.11 (Rice 定理). 若 S 是非平凡的指标集 (不为 \emptyset, Σ^*), 则 S 不可判定. 这里, 指标集是指, $\langle M_1 \rangle \in S, L(M_1) = L(M_2) \Rightarrow \langle M_2 \rangle \in S$.

证明. 取一个 TM: M_0 , s.t. $L(M_0) = \emptyset$. 对于任意一台 TM: M' ,

1. 若 $M_0 \in S$, 取 $M_1 \notin S$, 对于 $\langle M, w \rangle$, 若 M 接受 w , 则 $L(M') = L(M_1)$, 否则拒绝所有输入 $L(M') = \emptyset = L(M_0)$. 从而判定 $M' \in S$ 可以解决 $\overline{\text{A}_{\text{TM}}}$, 故 $\overline{\text{A}_{\text{TM}}} \leq_m S$;
2. 若 $M_0 \notin S$, 取 $M_1 \in S$, 对于 $\langle M, w \rangle$, 若 M 接受 w , 则 $L(M') = L(M_1)$, 否则拒绝所有输入 $L(M') = \emptyset = L(M_0)$. 从而判定 $M' \in S$ 可以解决 A_{TM} , 故 $\text{A}_{\text{TM}} \leq_m S$.

无论如何, 我们都得到了矛盾, 所以 S 不可判定. \square

4.4 上下文有关语言的可判定性

定理 4.12 (A_{LBA}). $\text{A}_{\text{LBA}} = \{\langle M, w \rangle | \text{LBA}: M \text{ 接受 } w\}$ 可判定.

证明. 给定 $\langle M, w \rangle$, 状态数 q 、符号数 g 、长度 n 均有限, 格局至多有 qng^n 个, 可以检测死循环. \square

定理 4.13 (E_{LBA}). $\text{E}_{\text{LBA}} = \{\langle M \rangle | \text{LBA}: M, L(M) = \emptyset\}$ 不可判定.

证明. TM 的接受计算历史构成上下文有关语言. 对于图灵机 M 及其输入 w , LBA: M' 及其输入 x , 检查 x 是否为 M 在 w 上的接受计算历史, 若是则接受, 否则拒绝. 从而 $\langle M, w \rangle \in \overline{\text{A}_{\text{TM}}} \Leftrightarrow \forall y, y$ 不是 M 在 w 上的接受计算历史 $\Leftrightarrow L(M') = \emptyset$, 因此 $\overline{\text{A}_{\text{TM}}} \leq_m \text{E}_{\text{LBA}}$. \square

总结：见下表，

	DFA	CFG	LBA	TM
接受性	Y	Y	Y	N
空性	Y	Y	N	N
等价性	Y	N		N
停机			Y	N
正则性				N

5 总结

定理 5.1 (递归定理). 设 $TM: T$, 可计算函数 $t: \Sigma^* \times \Sigma^* \rightarrow \Sigma^*$, 则存在 $TM: R$, 可计算函数 $r: \Sigma^* \rightarrow \Sigma^*$, s.t. $r(w) = t(\langle R \rangle, w)$.

定理 5.2 (不动点定理). \forall 可计算函数 t , $\exists TM: F$, s.t. $L(t(\langle F \rangle)) = L(F)$.

定理 5.3 (递归定理的不动点形式). \forall 可计算函数 $t: \Sigma^* \rightarrow \Sigma^*$, $\exists TM: F$, s.t. $t(\langle F \rangle)$ 与 F 等价.

关于语言的分类 (均不属于更低一层):

正则	0^*1^*
CFL	$0^n1^n, ww^R$, 非 ww^R , 非 ww , 非接受计算历史
CSL	$0^n1^n2^n, ww$, 接受计算历史
可判定	$A_{LBA}, E_{CFG}, EQ_{DFA}$
图灵可识别	$A_{TM}, HALT_{TM}, PCP, \overline{EQ_{CFG}}$

语言的封闭性, 这里, $RC(L) = \{xy|yx \in L\}$, 同态是指: 对于函数 $f: \Sigma \rightarrow \Gamma^*$, $f(L) = \{f(x_1)\cdots f(x_n)|x_1\cdots x_n \in L\}$, $CUT(L) = \{xyz|xyz \in L\}$, $L_{\frac{1}{2}-}(L) = \{x|\exists y, xy \in L\}$, 对 \leq 封闭是指 $A \leq B, B \in S \Rightarrow A \in S$.

	正则	CFL	CSL	可判定	图灵可识别	补图灵可识别
交	Y	N1	Y	Y	Y	Y
并	Y	Y	Y	Y	Y	Y
补	Y	N2	Y	Y	N	N
连接	Y	Y	Y	Y	Y	Y
星号	Y	Y	Y	Y	Y	Y
RC	Y3	Y		Y	Y	Y
同态	Y	Y4		N5	Y	Y
CUT	Y	N		Y	Y	Y
$L_{\frac{1}{2}-}$	Y	N		Y	Y	Y
\leq_m	N6	N		Y	Y	Y
\leq_T	N	N		Y	N	N

其中一些结论的证明:

1. 考虑 $0^n1^n2^m$ 和 $0^m1^n2^n$, 它们都是 CFL, 但他们的交 $0^n1^n2^n$ 不是 CFL.
2. 利用 De Morgan 律, $A \cap B = (A' \cup B)'$, 若补封闭则交也封闭.
3. 利用 NFA 猜.
4. 在 Chomsky 范式中作替换.

5. 由于 $L \in \Sigma^1 \Leftrightarrow L = \{x \mid \exists y, (x, y) \in C\}$, 这里 C 是可计算语言. 对于 (x, y) , 它是可计算的. 为所有的 y 更换新的字母表使之与 x 的字母表不交, 然后构造同态: $f(x) = x, f(y) = \varepsilon$, 此时新得到的语言是 Σ^1 的.
6. 对于任何一个可识别的语言 A , 输出 1 若接受, 输出 0 若拒绝, 则 $A \leq_m \{0, 1\}$.

理论计算机科学基础

期末整理

郭嘉睿

ntguojiarui@pku.edu.cn

2022 年 1 月 16 日

7 时间复杂度

定理 7.1. 设 $t(n) \geq n$, 则每个 $t(n)$ 时间多带 TM 都与某个 $O(t^2(n))$ 时间单带 TM 等价.

定理 7.2. 设 $t(n) \geq n$, 则每个 $t(n)$ 时间单带 NTM 均与某个 $2^{O(t(n))}$ 时间 DTM 等价.

定义 7.3 (NP). NP 的两个等价定义:

$$\text{NP} = \{L \mid L \text{ 有多项式时间验证机}\} = \{L \mid \text{某个多项式时间 NTM 判定 } L\}.$$

定理 7.4 (Cook 定理). 任何 NP 语言均可在多项式时间内归约到 cnf-SAT.

一些 NPC 问题及它们的归约:

1. 3SAT, SAT: 通过 cnf-SAT 归约;
2. CLIQUE, VC, HAMPATH, SUBSET-SUM: 通过 3SAT 归约;

8 空间复杂度

定理 8.1 (Savitch 定理). 设 $f(n) \geq \log n$, 则 $\text{NSPACE}(f(n)) \subseteq \text{SPACE}(f^2(n))$.

推论: $\text{PSPACE} = \text{NPSPACE}$.

定理 8.2. 全带量词布尔公式问题 TQBF 是 PSPACE 完全的.

一些 PSPACE 完全问题: 公式博弈 FORMULA-GAME, 广义地理学游戏 GG.

定义 8.3 (亚线性空间). 亚线性空间 TM 是指将 TM 的带分为一条输入带 (只读), 一条工作带 (读写) 和一条单向输出带 (只写, 禁止回头或修改), 且工作带的大小是亚线性的.

定理 8.4. PATH 是 NL 完全的.

定理 8.5. $\text{NL} = \text{coNL}$.

9 空间难解性

定理 9.1 (空间层次定理). 对于任意空间可构造函数 $f: \mathbf{N} \rightarrow \mathbf{N}$, 存在语言 A , 在空间 $O(f(n))$ 内判定但不在空间 $o(f(n))$ 内判定.

定理 9.2 (时间层次定理). 对于任意时间可构造函数 $t: \mathbf{N} \rightarrow \mathbf{N}$, 存在语言 A , 在时间 $O(t(n))$ 内判定但不在时间 $o\left(\frac{t(n)}{\log t(n)}\right)$ 内判定.

定理 9.3. $\text{EQ}_{\text{REX}\uparrow}$ 是 EXPSPACE 完全的.

定理 9.4. $\text{NONMIN-FORMULA} \in \text{NP}^{\text{SAT}}$.

定理 9.5 (对角化的局限性). 存在语言 A, B , 使得 $\text{P}^A \subset \text{NP}^A, \text{P}^B = \text{NP}^B$.

定义 9.6 (ATM). 交错式 TM(ATM) 是一种 NTM, 其计算树中的非确定性分支点包括全称和存在两类, 一个全称分支点接受当且仅当它所有儿子接受, 一个存在分支点接受当且仅当它至少一个儿子接受, 根接受则整个计算接受.

ATM 复杂性的结论: $\text{P} = \text{AL}, \text{PSPACE} = \text{AP}, \text{EXP} = \text{APSPACE}$.

定义 9.7 (电路族). 一个电路族 C 是无穷个电路 $C = (C_0, C_1, \dots)$, 其中 C_n 有 n 个输入变量. 若对每个字符串 $w, w \in A \Leftrightarrow C_n(w) = 1$, 其中 $|w| = n$, 则称 C 在 $\{0, 1\}$ 上判定 A .

电路族的规模复杂性是 C 中的规模, 深度复杂性是 C 中从输入到输出的最长路径长度. $\text{P/poly} = \text{PSIZE} = \bigcup_k \text{SIZE}(n^k)$.

定理 9.8. $\text{TIME}(t(n)) \subseteq \text{SIZE}(O(t^2(n)))$. 进一步, $\text{P} \subseteq \text{PSIZE}$.

定理 9.9. 电路可满足性问题 CIRCUIT-SAT 是 NP 完全的.

定理 9.10 (Karp-Lipton 定理). $\text{NP} \subseteq \text{P/poly} \Leftrightarrow \text{PH} = \Sigma_2\text{P}$.

定义 9.11 (对数空间一致性). 一个布尔电路族 (C_1, \dots) 是对数空间一致的, 当且仅当存在一个对数空间 TM: T , 当输入 1^n 时, T 输出 $\langle C_n \rangle$.

定义 9.12 (NC 类). NC 类是指多项式规模, 对数多项式规模深度的电路. 更一般的, NC^k 类是指多项式规模, $O(\log^k n)$ 深度的电路.

定理 9.13. $\text{CIRCUIT-VALUE}(\text{CVP})$ 是 P 完全的.

10 复杂性高级专题

定义 10.1 (PP). PP 指错误概率 $\varepsilon = 0.5$, 在多项式时间内运行的概率算法.

定义 10.2 (BPP). BPP 指错误概率 $\varepsilon = 0.5 - \delta$ (其中 δ 是任意常数), 在多项式时间内运行的概率算法.

定义 10.3 (RP). RP 指错误概率 $\varepsilon = 0.5 - \delta$ (其中 δ 是任意常数), 在多项式时间内运行且只出现弃真型错误的概率算法.

定义 10.4 (coRP). coRP 指错误概率 $\varepsilon = 0.5 - \delta$ (其中 δ 是任意常数), 在多项式时间内运行且只出现取伪型错误的概率算法.

定义 10.5 (ZPP). ZPP 指错误概率 $\varepsilon = 0$, 期望运行时间为多项式时间的概率算法 (或: 在多项式时间内运行, 但允许 3 种输出 0, 1, ? 的概率算法).

它们之间的关系:

1. $\text{ZPP} \subseteq \text{RP} \cap \text{coRP} \subseteq \text{RP} \cup \text{coRP} \subseteq \text{BPP} \subseteq \text{PP}$.
2. $\text{BPP} \subseteq \text{PSIZE}$ (利用加强引理证明).
3. $\text{BPP} \subseteq \Sigma_2\text{P} \cap \Pi_2\text{P}$.
4. $\text{PH} \subseteq \text{P}^{\text{PP}}$.

11 一些没什么用的东西

一些语言的接受性/空性/满性/等价性的复杂度:

	A	E	ALL	EQ
DFA	L	NL 完全	P	PSPACE
NFA	NL 完全	NL 完全	PSPACE 完全	
PDA	至少 NL 完全		不可判定	不可判定
LBA	PSPACE 完全	不可判定	不可判定	不可判定
TM	不可判定	不可判定	不可判定	不可判定

运算封闭性 (Y 表示封闭, 表格中的条件表示在该操作下封闭当且仅当这一条件为真):

	P	NP	coNP	EXP
\cap	Y	Y	Y	Y
\cup	Y	Y	Y	Y
\sim	Y	P=NP	P=NP	Y
\cdot	Y	Y	Y	Y
*	Y	Y	Y	Y
同态	P=NP	Y		
$L_{\frac{1}{2}-}$	P=NP	Y		Y
RC	Y	Y	Y	Y
CUT	Y	Y	Y	Y