附录 A　一些基本定义

惯例 A.1 (幂集)： 对于任意集合 A，我们使用 $\mathcal{P}(A)$ 代表 A 的所有子集。$\mathcal{P}(A)$ 也叫做 A 的幂集。有时也写作 2^A。

惯例 A.2 (函数集)： 对于集合 A 和 B，$A \to B$ 代表所有从 A 到 B 的函数的集合。而 $A \not\to B$ 代表所有从 A 到 B 的 “partial functions”。

注释 A.1： 对于集合 A, B, 关系 $\mathcal{C} \subseteq A \times B$，我们可以把 $\mathcal{C}$ 理解为函数 $\mathcal{C} \in A \longrightarrow \mathcal{P}(B)$。

惯例 A.3 (元组投影)： 对于 n 元组 $t = (x_1, x_2, \cdots, x_n)$，我们使用符号 $\pi_i(t)$ $(1 \le i \le n)$ 代表元组元素 x_i；我们使用符号 $\bar{\pi}_i(1 \le i \le n)$ 代表 (n-1) 元组 $(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$。$\pi$ 和 $\bar{\pi}$ 自然扩展到元组集。

惯例 A.4 (组合关系)： 给出集合 A, B, C 和两个关系 $E \subseteq A \times B$ 和 $F \subseteq B \times C$，定义组合关系（插入操作符 $\circ$）:

$$E \circ F = \{(a, c) : (\exists b : b \in B : (a, b) \in E \wedge (b, c) \in F)\}$$

惯例 A.5 (等价关系的等价类)： 对任何集合 A 上的等价关系 E, 我们使用 $[A]_E$ 代表等价类集合，即:

$$[A]_E = \{[a]_E : a \in A\}$$

集合 $[A]_E$ 也叫做 A 的由 E 引出的“划分（partition）”。

定义 A.1 (等价类的指数)： 对于集合 A 上的等价关系 E, 定义 $\sharp E = |[A]_E|$。$\sharp E$ 也叫做 E 的“指数”。

定义 A.2 (字母表)： 字母表是有限大小的非空集合。

定义 A.3 (等价关系的细化)： 对于等价关系 E 和 E'（在集合 A 上），当且仅当 $E \subseteq E'$，E 是 E' 的“细化”。

定义 A.4 (划分的细化关系 $\sqsubseteq$)： 对于等价关系 E 和 E'（在集合 A 上)，当且仅当 $E \subseteq E'$，$[A]_E$ 也被称为 $[A]_{E'}$ 的细化（写作 $[A]_E \sqsubseteq [A]_{E'}$)。当且仅当 E 下的每一个等价类完全包含在 E' 下的某些等价类时，等价命题是 $[A]_E \sqsubseteq [A]_{E'}$。

定义 A.5 (元组和关系反转)： 对一个 n 元组 $(x_1, x_2, \cdots, x_n)$，定义反转为函数 R (后缀和上标)

$$(x_1, x_2, \cdots, x_n)^R = (x_n, \cdots, x_2, x_1)$$

给出一个集合元组 A，定义 $A^R = \{x^R : x \in A\}$。

附录 B　有限自动机

本节中我们定义有限自动机、其性质及其一些变化。大部分定义直接取自

定义 B.1 (**有限自动机** ($Finiteautomata, FA$))**：** 自 动 机 是 一 个 6 元 组 (Q, V, T, E, S, F)，其中:

- Q 是有限状态集;
- V 是一个字母表;
- $T \in \mathcal{P}(P \times V \times Q)$ 是一个转换关系;
- $E \in \mathcal{P}(Q \times Q)$ 是一个 ϵ-转换关系（空转换）;
- $S \subseteq Q$ 是开始状态集;
- $F \subseteq Q$ 是结束状态集;

字母表和函数 $\mathcal{P}$ 的定义分别在“定义A.2”和“惯例A.1”。

注释 B.1**：** 我们也会在状态转换关系的表示上采取一定的自由。例如，我们也把转移关系写成 $T \in V \longrightarrow \mathcal{P}(Q \times Q), T \in Q \times Q \longrightarrow \mathcal{P}(V), T \in Q \times V \longrightarrow \mathcal{P}(Q), T \in Q \longrightarrow \mathcal{P}(V \times Q), E \in Q \longrightarrow \mathcal{P}(Q)$。每种情况下，$Q$ 的从左到右的顺序会是“preserved”; 例如，函数 $T \in Q \longrightarrow \mathcal{P}(V \times Q)$ 定义为 $T(p) = \{(a, q) : (p, a, q) \in T\}$。所使用的签名将从上下文中清除。详见备注 A.3。“$\longrightarrow$ ”的定义出现在惯例 A.2。

由于本文中我们只考虑有限自动机，所以我们将会频繁的使用简化术语“自动机”。

B.1　有限自动机的性质

本小节将会定义一些有限自动机（下称 FA）的性质。为了使定义更加简洁明了，我们引进三个特殊的 FA: $M = (Q, V, T, E, S, F)$，$M_0 = (Q_0, V_0, T_0, E_0, S_0, F_0)$，$M_1 = (Q_1, V_1, T_1, E_1, S_1, F_1)$。

定义 B.2 (FA 的大小)： 定义一个 FA 的大小为 $|M| = |Q|$。

定义 B.3 (FA 的同构 $\cong$)： 我们把同构定义为 FA 的等价关系。当且仅当 $V_0 = V_1$，并且存在双射 $g \in Q_0 \longrightarrow Q_1$ ，使得

- $T_1 = \{(g(p,q), a, g(q)) : (p,a,q) \in T_0\}$
- $E_1 = \{(g(p,q), a, g(q)) : (p,q) \in E_0\}$
- $S_1 = \{g(s) : s \in S_0\}$
- $F_1 = \{g(f) : f \in F_0\}$

时 M_0 和 M_1 是同构的（写作 $M_0 \cong M_1$）。

定义 B.4 (转移关系 T 的扩展)： 我们把 $T \in V \longrightarrow \mathcal{P}(Q \times Q)$ 到 $T^* \in V^* \longrightarrow \mathcal{P}(Q \times Q)$ 的转换关系以如下方式扩展:

$T^*(\epsilon) = E^*$

且对于 $(a \in V, w \in V^*)$ 有

$T^*(aw) = E^* \circ T(a) \circ T^*(w)$

操作符 $\circ$ 在惯例 A.5 中定义。<u>这个定义也可以对称的表示</u>。

注释 B.2： 有时候我们也使用把转移关系写成：$T^* \in Q \times Q \longrightarrow \mathcal{P}(V^*)$。

定义 B.5 (左语言和右语言)： 状态（M 中）的左语言由函数 $\overleftarrow{\mathcal{L}}_M \in Q \longrightarrow \mathcal{P}(V^*)$ 给出，其中:

$\overleftarrow{\mathcal{L}}_M(q) = (\cup s : s \in S : T^*(s,q))$

状态（M 中）的右语言由函数 $\overrightarrow{\mathcal{L}}_M \in Q \longrightarrow \mathcal{P}(V^*)$ 给出，其中

$\overrightarrow{\mathcal{L}}_M(q) = (\cup f : f \in F : T^*(q,f))$

通常在没有歧义的时候移除下标 M。

定义 B.6 (FA 的语言)： 有限自动机的语言由函数 $\mathcal{L}_{FA} \in FA \longrightarrow \mathcal{P}(V^*)$ 给出，该函数的定义为:

$\mathcal{L}_{FA}(M) = (\cup s, f : s \in S \wedge f \in F : T^*(s,f))$

定义 B.7 (完全自动机 ($Complete$))： 一个完全有限自动机满足:

$Complete(M) \equiv (\forall q, a : q \in Q \wedge a \in V : T(q,a) \neq \emptyset)$

定义 B.8 (ϵ-$free$)： 当且仅当 $E = \emptyset$ 时，M 是 ϵ-$free$ 的。

定义 B.9 (Start-useful **自动机**)： 一个 $Useful_s$ 有限自动机定义如下:

$$Useful_s(M) \equiv (\forall q : q \in Q : \overleftarrow{\mathcal{L}}(q) \neq \emptyset)$$

定义 B.10 (Final-useful **自动机**)： 一个 $Useful_f$ 有限自动机定义如下:

$$Useful_f(M) \equiv (\forall q : q \in Q : \overrightarrow{\mathcal{L}}(q) \neq \emptyset)$$

注释 B.3： $Useful_s$ 和 $Useful_f$ 与 FA 的反转密切相关（见变换 B.22)，对所有的 $M \in FA$，有 $Useful_f(M) \equiv Useful_s(M^R)$。

定义 B.11 (Useful **自动机**)： $Useful$ 有限自动机是一个只有可达状态的有限自动机:

$$Useful(M) \equiv Useful_s(M) \wedge Useful_f(M)$$

性质 B.1 (**确定性有限自动机** (DFA))： 当且仅当

- 无多重初始状态;
- 无 ϵ 转移;
- 转移函数 $T \in Q \times V \longrightarrow \mathcal{P}(Q)$ 不将 $Q \times V$ 映射至多重状态。

时有限自动机 M 是确定性的。形式表达为:

$$Det(M) \equiv (|S| \leq 1 \wedge \epsilon - free(E) \wedge (\forall q, a : q \in Q \wedge a \in V : |T(q, a)| \leq 1))$$

定义 B.12 (FA **的确定性**)： DFA 代表所有确定性的有限自动机的集合。我们把 $FA \setminus DFA$ 称为非确定性有限自动机 ($NDFA$,$nondeterministic\ finite\ automata$) 的集合。

惯例 B.1 (DFA **的转换函数**)： 对于 $(Q, V, T, \emptyset, S, F) \in DFA$，我们考虑把转换函数记为 $T \in Q \times V \not\rightarrow Q$（$\not\rightarrow$ 的定义可以查看惯例 A.2)。当且仅当 DFA 是完全自动机的时候，转换函数是全函数。

性质 B.2 (弱确定性自动机)： 一些作者用比 Det 弱的确定性自动机的定义；使用左语言，定义如下:

$$Det'(M) \equiv (\forall q_0, q_1 : q_0 \in Q \wedge q_1 \in Q \wedge q_0 \neq q_1 : \overleftarrow{\mathcal{L}}(q_0) \cap \overleftarrow{\mathcal{L}}(q_1) = \emptyset)$$

很容易证明 $Det(M) \Rightarrow Det'(M)$。

定义 B.13 (DFA 的最小化)： 满足以下条件时，$M \in DFA$ 是最小化的:

$$Min(M) \equiv (\forall M' : M' \in DFA \wedge Complete(M') \wedge \mathcal{L}(M) = \mathcal{L}_{FA}(M') : |M| \leq |M'|)$$

Min 仅定义在 DFA 上。如果我们定义一个最小的但是仍然完全的 DFA，那么一些定义将会更加简单。它的定义如下:

$$Min_{\mathcal{C}}(M) \equiv (\forall M' : M' \in DFA \wedge Complete(M') \wedge \mathcal{L}_{FA}(M) = \mathcal{L}_{FA}(M') : |M| \leq |M'|)$$

$Min_{\mathcal{C}}$ 仅定义在完全 DFA 上。

定义 B.14 (DFA 的最小化)： 根据 Myhill-Nerode 定理，An M, such that Min(M)，是唯一的最小化 DFA，定理的相关介绍在[4]。

性质 B.3 (DFA 最小化的一个替代定义)： 为了最小化 DFA，使用定义（仅定义在 DFA 上）:

$$\begin{aligned} Minimal(Q, V, T, \emptyset, S, F) \equiv & (\forall q_0, q_1 : q_0 \in Q \wedge q_1 \in Q \wedge q_0 \neq q_1 : \overrightarrow{\mathcal{L}}(q_0) \neq \overrightarrow{\mathcal{L}}(q_1)) \\ & \wedge Useful(Q, V, T, \emptyset, S, F) \end{aligned}$$

有 $Minimal(M) \equiv Min(M)$（对所有 $M \in DFA$）。很容易证明 $Min(M) \Rightarrow Minimal(M)$。The reverse direction follows from the Myhill-Nerode 定理。

与 $Min_{\mathcal{C}}$ 相似的定义是（同样也只定义在 DFA 上）:

$$\begin{aligned} Minimal_{\mathcal{C}}(Q, V, T, \emptyset, S, F) \equiv & (\forall q_0, q_1 : q_0 \in Q \wedge q_1 \in Q \wedge q_0 \neq q_1 : \overrightarrow{\mathcal{L}}(q_0) \neq \overrightarrow{\mathcal{L}}(q_1)) \\ & \wedge Useful(Q, V, T, \emptyset, S, F) \end{aligned}$$

有 $Minimal_{\mathcal{C}}(M) \equiv Min_{\mathcal{C}}(M)$ 的性质（对于所有的 $M \in DFA$）。很容易证明 $Min_{\mathcal{C}}(M) \Rightarrow Minimal_{\mathcal{C}}(M)$。The reverse direction follows from the Myhill-Nerode 定理。

B.2　有限自动机的变换

变换 B.1 (FA 反转)： FA 反转由后缀（上标）函数 $R \in FA \longrightarrow FA$ 给出，它的定义如下:

$$(Q, V, T, S, F)^R = (Q, V, T^R, E^R, F, S)$$

函数 R 满足

$$(\forall M : M \in FA : (\mathcal{L}(M))^R = \mathcal{L}_{FA}(M^R))$$

变换 B.2 (移除开始状态不可达状态)： 变换 $useful_s \in FA \longrightarrow FA$ 移除开始状态不可达状态:

$$\begin{array}{rcl} useful_s(Q, V, T, E, S, F) & = & \textbf{let} \quad U = SReachable(Q, V, T, E, S, F) \\ & & \textbf{in} \\ & & \quad (U, V, T \cap (U \times V \times U), E \cap (U \times U), S \cap U, F \cap U) \\ & & \textbf{end} \end{array}$$

函数 $useful_s$ 满足

$$(\forall M : M \in FA : Useful_s(useful_s(M)) \wedge \mathcal{L}_{FA}(useful_s(M)) = \mathcal{L}_{FA}(M))$$

变换 B.3 (子集构造)： 函数 $subset$ 把一个 ϵ-$free$ FA 转换为一个 DFA (in the **let** clause $T' \in \mathcal{P}(Q) \times V \longrightarrow \mathcal{P}(\mathcal{P}(Q))$):

$$\begin{array}{rcl} subset(Q, V, T, \emptyset, S, F) & = & \textbf{let} \quad T'(U, a) = \{(q : q \in U : T(q, a))\} \\ & & \qquad F' = \{U : U \in \mathcal{P}(Q) \wedge U \cap F \neq \emptyset\} \\ & & \textbf{in} \\ & & \quad (\mathcal{P}(Q), V, T', \emptyset, \{S\}, F') \\ & & \textbf{end} \end{array}$$

有时候也把它说成“幂集”构造。

性质 B.4 (子集构造)： 设 $M_0 = (Q_0, v, t_0, \emptyset, S_0, F_0)$ 和 $M_1 = subset(M_0)$ 为有限自动机。通过子集构造，状态集 M_1 成为 $\mathcal{P}(Q_0)$。有如下性质:

$$(\forall p : p \in \mathcal{P}(Q_0) : \overrightarrow{\mathcal{L}}_{M_1}(p) = (q : q \in p : \overrightarrow{\mathcal{L}}_{M_1}(q)))$$

定义 B.15 (优化子集构造)： 函数 $subsetopt$ 把一个 $\epsilon\text{-}free\ FA$ 转换为一个 DFA 。此函数是 $subset$ 的一个优化版本:

$$\begin{array}{rcl} subset(Q, V, T, \emptyset, S, F) & = & \textbf{let} \quad T'(U, a) = \{(q : q \in U : T(q, a))\} \\ & & \qquad Q' = \mathcal{P}(Q) \setminus \{\emptyset\} \\ & & \qquad F' = \{U : U \in \mathcal{P}(Q) \wedge U \cap F \neq \emptyset\} \\ & & \textbf{in} \\ & & \qquad (Q', V, T' \cap (Q' \times V \times Q'), \emptyset, \{S\}, F') \\ & & \textbf{end} \end{array}$$

除了性质 $\mathcal{L}_{FA}(subsetopt(M)) = \mathcal{L}(M)$ (对所有的 $M \in FA$) 之外，函数 $subsetopt$ 还满足

$$(\forall M : M \in FA \wedge \epsilon\text{-}free(M) : Det(subset(M)))$$