强可识语言族*

郭聿琦　王水汀　李廉
（兰州大学数力系）

Σ为一有限集，Σ^*表示Σ生成的自由幺半群，Σ^*的元素与子集分别称为Σ上的字与语言，2^{Σ^*}表示Σ^*的幂集，$\mathscr{L}(\Sigma)=2^{\Sigma^*}-\{\phi\}$的子集称为$\Sigma$上的语言族.

在人工智能中的一些问题的推动下，1974年 Havel 等人[1]开创了语言的分支代数结构的研究，定义了有限分支自动机，从而导致了作为有限分支自动机识别的所谓可识语言族的研究；Havel 在文献[2]中又引进了语言的相似度的概念，进而定义了语言之间的一种距离d，使$(\mathscr{L}(\Sigma), d)$成一距离空间；文献[2]中还定义了语言族的一种替换性，并证明了，语言族是自相容的，当且仅当它具替换性且为$\mathscr{L}(\Sigma)$的闭集.

可识语言族类的子类的研究，对于揭示可识语言族类的结构与性质无疑是重要的. 例如，Havel[3]曾提出一个 Open 问题，期望对可识语言族类建立类似于正则语言类的 Kleene 定理那样的代数描述定理；Benda 与 Bendová 则于1976年讨论过所谓完全可识语言族类这样一个可识语言族类的子类[4]，它由所有其补也可识的可识语言族组成.

本文引进并讨论了可识语言族类的另一个真子类——强可识语言族类，其中，特别用主滤子刻划了这一语言族类的结构. 本文所依据的工具是我们在文献[5]中建立的语言族的积分.

为了下面的讨论，我们先交待一些预备概念和记号：

$|\Sigma|$表示集合Σ中元素的个数；ε表示自由幺半群Σ^*的单位元，称为Σ上的空字；$\Sigma^+=\Sigma^*-\{\varepsilon\}$；$\Sigma_\varepsilon=\Sigma\cup\{\varepsilon\}$；若$u$为$\Sigma$上的字，则$|u|$表示$u$含$\Sigma$中元素的个数，称为字$u$的长度；若$L$为$\Sigma$上的一个语言，则$\mathrm{Pref}\ L=\{u|\exists w\in\Sigma^*(uw\in L)\}$称为$L$的前缀集，又记

$$\mathrm{Fst}\ L=\mathrm{Pref}\ L\cap\Sigma,$$

$$\mathrm{Fst}_\varepsilon\ L=\begin{cases}\mathrm{Fst}\ L\cup\{\varepsilon\}, & \varepsilon\in L,\\ \mathrm{Fst}\ L, & \varepsilon\bar{\in}L;\end{cases}$$

$\partial_u L=\{w|w\in\Sigma^*, uw\in L\}$称为语言$L$关于字$u$的微商；$\partial_u X=\{\partial_u L|L\in X\}-\{\phi\}$称为语言族$X$关于字$u$的微商，$\{\partial_u X|u\in\Sigma^*\}$有限时，称$X$有限可微；$\int Xdu=\{uL\cup L'|L\in X, L'\subseteq\Sigma^*, u\bar{\in}\mathrm{Pref}L'\}$称为语言族$X$关于字$u$的积分；$C(X)=\{L|L\in\mathscr{L}(\Sigma), \forall u\in\mathrm{Pref}L[\exists L_u\in X(\mathrm{Fst}_\varepsilon\ \partial_u L=\mathrm{Fst}_\varepsilon\ \partial_u L_u)]\}$称为语言族$X$的相容闭包，若$C(X)=X$，则称$X$自相容；关于任何$L_1, L_2\in\mathscr{L}(\Sigma)$，若$u\in\Sigma^*$为使$\mathrm{Fst}_\varepsilon\ \partial_u L_1\neq\mathrm{Fst}_\varepsilon\ \partial_u L_2$的最短字，则称非负整数$|u|$为$L_1$与$L_2$的相似度，记为$S(L_1, L_2)$，可验证函数

$$d(L_1, L_2)=\begin{cases}2^{-S(L_1,L_2)}, & L_1\neq L_2,\\ 0, & L_1=L_2\end{cases}$$

本文1982年3月22日收到. 1983年8月18日收到修改稿.
* 中国科学院科学基金资助的课题.

为 $\mathscr{L}(\Sigma)$ 上的一个距离；$\bar{X} = \{L \mid L \in \mathscr{L}(\Sigma),\ \forall n \geqslant 0[\exists L_n \in X(S(L,\ L_n) > n)]\}$ 称为X的闭包，若 $\bar{X} = X$，则称X为距离空间 $(\mathscr{L}(\Sigma), d)$ 中的闭集；若关于任意 $L_1,\ L_2 \in X$ 与任意 $u \in \mathrm{Pref}\ L_1 \cap \mathrm{Pref}\ L_2$，有 $\mathrm{Rep}_u(L_1,\ L_2) = (L_1 - u\partial_u L_1) \cup u\partial_u L_2 \in X$，则称$X$具替换性.

定义 1 　称语言族X具强替换性，如果关于任意 $L_1, L_2 \in X$ 与任意 $u \in \mathrm{Pref}\ L_1 \cap \mathrm{Pref}\ L_2$，有

$$(L_1 - u\partial_u L_1) \cup L \in X,\ \forall L \in \int \partial_u L_2 du.$$

显然，语言族X具强替换性时，X必具替换性．容易看出，$\mathscr{L}(\Sigma)$具强替换性；而$L \neq \{\varepsilon\}$时，单一族$\{L\}$虽具替换性，但不具强替换性．我们还有

例 1. $L \in \mathscr{L}(\Sigma)$ 生成的所谓主滤子 $F_L = \{L' \mid L' \in \mathscr{L}(\Sigma), L \subseteq L'\}$[4] 具强替换性.

例 2. 关于 $L \in 2^{\Sigma^*}$ 生成的所谓主理想 $I_L = \{L' \mid L' \in 2^{\Sigma^*}, L' \subseteq L\}$[4]，当$L$正则但 $L \neq \Sigma^*$ 时，$I_L - \{\phi\} \subseteq \mathscr{L}(\Sigma)$ 虽具替换性，但不具强替换性.

空语言族ϕ是平凡的具强替换性的语言族．以下，常假定 $X \neq \phi$.

推论 1 　语言族X具强替换性，当且仅当语言族 $X - \{\{\varepsilon\}\}$ 具强替换性.

Benda 与 Bendová 曾用关系 $R \subseteq \Sigma^* \times (2^{\Sigma_\varepsilon} - \{\phi\})$ 与函数 $G: 2^{\mathscr{L}(\Sigma)} \to 2^{\Sigma^* \times (2^{\Sigma_\varepsilon} - \{\phi\})}$ 来研究语言族的可识性[6]．为研究语言族的强替换性，我们扩充R为 $R' \subseteq \Sigma^* \times 2^{\Sigma_\varepsilon}$，并引进函数 $G': 2^{\mathscr{L}(\Sigma)} \to 2^{\Sigma^* \times 2^{\Sigma_\varepsilon}}$ 如下:

$$\begin{aligned} G'(X) &= \{(x,\ \mathrm{Fst}_\varepsilon\ \partial_x L) \mid x \in \Sigma^*, L \in X\} \\ &= G(X) \cup \{(x,\ \phi) \mid \exists L'_x \in X(x \bar{\in} \mathrm{Pref}\ L'_x)\}, \end{aligned}$$

我们有

定理 1 　若 $|\Sigma| \geqslant 2$，X为自相容语言族，则X具强替换性，当且仅当对所有不为$(\varepsilon, \{\varepsilon\})$的 $(x, \Gamma) \in \Sigma^* \times 2^{\Sigma_\varepsilon}$，由 $(x, \Gamma) \in G'(X)$ 可推出 $(x, \Gamma') \in G'(X)$，$\forall \Gamma'(\supseteq \Gamma) \in 2^{\Sigma_\varepsilon}$.

注意：该定理的必要性无须X自相容的条件.

下例说明，当 $|\Sigma| = 1$ 时，定理 1 的必要性不成立.

例 3. $\Sigma = \{a\}$，令 $L_0 = \{\varepsilon\}$，$L_1 = \{\varepsilon, a\}$，$L_2 = \{\varepsilon, a, a^2\}$，易知 $X = \{L_0, L_1, L_2\}$ 是自相容的，且具强替换性．但尽管 $(a^2, \{\varepsilon\}) = (a^2, \mathrm{Fst}_\varepsilon\ \partial_{a^2} L_2) \in G'(X)$，却有 $(a^2, \{\varepsilon, a\}) \bar{\in} G'(X)$.

我们在本文以下诸命题中，也常假定 $|\Sigma| \geqslant 2$.

推论 2 　若语言族X具强替换性，则 $\bigcup_{L \in X} L = \Sigma^*$.

下面的例 4 指出，推论 2 的反面不真.

例 4. 任取 $L(\neq \{\varepsilon\}) \in \mathscr{L}(\Sigma)$，令 $X = \{L, \Sigma^* - L\}$，则 $\bigcup_{L \in X} L = L \cup (\Sigma^* - L) = \Sigma^*$，但$X$不具强替换性.

相应于语言族的自相容性，我们给出强自相容性的概念.

定义 2 　称 $C'(X) = \{L \mid L \in \mathscr{L}(\Sigma), \forall v \in \mathrm{Pref}\ L[\exists L_v \in X - \{\{\varepsilon\}\}(\mathrm{Fst}_\varepsilon \partial_v L_v \subseteq \mathrm{Fst}_\varepsilon \partial_v L)]\} \cup (X \cap \{\{\varepsilon\}\})$ 为X的强相容闭包．称X为强自相容的，如果 $X = C'(X)$.

显然，$C(X) \subseteq C'(X)$，因此强自相容语言族必自相容．我们有

定理 2 　语言族X是强自相容的，当且仅当X是闭集且具强替换性.

下面,我们利用主滤子 $F_L=\{L'|L'\in\mathscr{L}(\Sigma),\ L\subseteq L'\}$,这里 $L\neq\phi$,给出强自相容语言族的一个结构性质.

定理 3 下列三款等价:

(i) X是强自相容的; (ii) $X'=X-\{\{\varepsilon\}\}$ 是强自相容的; (iii) $X'=C\left(\bigcup_{L\in X'}F_L\right)$.

文献[1]中已证明,语言族可识,当且仅当它有限可微且自相容. 我们给出

定义 3 语言族X称为强可识的,如果X有限可微且强自相容.

显然,强可识语言族必是可识语言族. 由定理 2 立得

定理 4 语言族X是强可识的,当且仅当X是闭的,有限可微的且具强替换性.

例 5. 文献[4]中已证,当L为非空正则语言时,L生成的主滤子 F_L 是可识的. 又由例 1 与定理 4,F_L 是强可识的.

例 6. 文献[4]中已证,当L为正则语言时,关于L生成的主理想 I_L,$I_L-\{\phi\}\subseteq\mathscr{L}(\Sigma)$ 是可识的,但当 $L\neq\Sigma^*$ 时,由例 2 与定理 4 知它不强可识.

由此可见,强可识语言族类是可识语言族类的一个真子类,并且,它含有所有可识主滤子.

最后,我们给出强可识语言族的一组机器性质.

定理 5 令 $fba\mathscr{B}=(K,\delta,q_0,B)$ 是Σ上的一个可达的有限分支自动机1),若 $R(\mathscr{B})$ 强可识,则关于任何不等于 $(q_0,\{\varepsilon\})$ 的 $(q,\Gamma)\in B$ 以及任何 $\Gamma'(\supseteq\Gamma)\in 2^{\Sigma_B}$,有$(q,\Gamma')\in B$.

注意: 语言的前缀性质在 $|\Sigma|=1$ 与 $|\Sigma|>1$ 之间有很大的差别,本文暂不考虑 $|\Sigma|=1$ 的情形.

参 考 文 献

[1] Havel, I. M., *Kybernatica*, **10** (1974), 281—302.
[2] Havel, I. M., *Math. Found. of Computer Science* (A. Mazurkiewicz ed.), Springer-Verlag, Berlin, 1976, 81—99.
[3] Havel, I. M., *Proceedings MFCS' 74 Sym.*, Springer-Verlag, Berlin, 1975.
[4] Benda, V. & Bendová, K., *Commentations Math. Univ. Corolinae*, **17**(1976), 2: 251—259.
[5] 王水汀、郭聿琦,兰州大学学报,1982,3:15—27.
[6] Benda, V. & Bendová, K., ***Math. Found.*** *of Computer Science* (Proc. Sixth Sym. Tatranska Lomnica), 1977, 247—252.

1) 李廉、王水汀,最小有限分支自动机,待发表.