ω—正则语言类的一个子类*

庄　雷　孟庆远　苏锦祥

(郑州大学计算机系　郑州　450052)

摘要　ω—语言是由有穷字母表 Σ 上的一些无穷串组成的集合. 被 ω—有穷自动机接受的 ω—语言称为 ω—正则语言. 作者曾从集合的角度描述了一类 ω—正则语言,而不是传统地从生成或识别的角度来描述这一类正则语言. 本文从集合的角度来描述更为广泛的一类 ω—正则语言.

关键词　ω—正则语言,ω—凸语言,广义 ω—凸语言,闭的 ω—语言.

MC Naughton 首先提出了被有穷状态自动机识别的 ω—语言的理论. R. Cohen 与 A. Gold 较全面地阐述了 ω—语言和 ω—有穷自动机的理论,综述了 ω—有穷自动机的概念;ω—有穷自动机的 5 种识别方式以及在此基础上引进的 ω—正则语言的概念.[1]作者在文献[3]中曾从集合论的角度刻画一类 ω—正则语言,而不是从识别的角度或生成的角度来考查. 本文也将从集合论的角度来考查一类 ω—正则语言,并且证明了这一类 ω—正则语言将比文献[3]中所给出的更广泛.

为了引述本文的结果,首先简要地叙述和引进一些有关的概念与记号.

设 Σ 是有穷字母表,由 Σ 中的字母组成的形如 $\sigma=a_1a_2\cdots a_n\cdots$ 的无穷序列,称为 Σ 上的一个 ω—串. 我们用 $\sigma(n)$ 表示 ω—串 σ 的前 n 个字母所组成的前缀,即 $\sigma(n)=a_1a_2\cdots a_n$. 用 Σ^ω 表示 Σ 上的一切 ω—串的集合. Σ^ω 的任意子集 L 称为 Σ 上的一个 ω—语言. G. Thierrin 在文献[2]中通过串的嵌入定义了 Σ^* 上的一个半序关系"≤",即 $x\leqslant y$, $x,y\in\Sigma^*$,当且仅当 $x=x_1x_2\cdots x_n$, $y=y_1x_1y_2x_2\cdots y_nx_ny_{n+1}$, $x_i\in\Sigma^*$, $i=1,2\cdots,n$, $y_j\in\Sigma^*$, $i=1,2,3,\cdots,n+1$. 并利用 Σ^* 上的半序关系"≤"定义了 Σ 上的凸语言的概念,即 Σ 上的语言 L 称为凸语言,当且仅当若 $x\leqslant z\leqslant y$ 且 $x,y\in L$, $z\in\Sigma^*$,则 $z\in L$. 苏锦祥将凸语言的概念扩充到 ω—语言上去,为此在文献[3]中定义了 Σ^ω 上的二元关系"$\leqslant_\omega$",即 $\sigma_1\leqslant_\omega\sigma_2$, $\sigma_1,\sigma_2\in\Sigma^\omega$,当且仅当 $\sigma_1=x_1x_2\cdots x_n\cdots$, $\sigma_2=y_1x_1y_2x_2\cdots y_nx_n\cdots$, $x_i,y_i\in\Sigma^*$. 从而,在 Σ^ω 上利用二元关系"$\leqslant_\omega$"定义了 ω—凸语言的概念.

本文在 Σ^ω 上定义了二元关系"$\leqslant_{\omega_g}$",并由此定义了所谓的广义的 ω—凸语言的概念.

定义 1. 设 $\sigma_1,\sigma_2\in\Sigma^\omega$, $\sigma_1\leqslant_{\omega_g}\sigma_2$ 当且仅当存在 $\sigma'\in\Sigma^\omega$, $x,y\in\Sigma^*$,使得 $\sigma_1=x\sigma'$, $\sigma_2=x\sigma'$, x

* 本文研究得到国家自然科学基金资助. 作者庄雷,女,1963 年生,副教授,主要研究领域为形式语言与自动机. 孟庆远,1940 年生,副教授,主要研究领域为形式语言及自动机关系. 苏锦祥,1934 年生,教授,主要研究领域为形式语言与自动机.

本文通讯联系人:庄雷,郑州 450052,郑州大学计算机系

本文 1995—10—19 收到修改稿

$\leqslant y$.

定义 2. 设 L 是 Σ 上的 ω—语言. L 称为广义 ω—凸语言,当且仅当,若 $\sigma_1 \leqslant_{\omega'} \sigma \leqslant_{\omega'} \sigma_2$, $\sigma_1, \sigma_2 \in L, \sigma \in \Sigma^\omega$,则 $\sigma \in L$.

定义 3. 设 L 是 Σ 上的 ω—语言, L 称为闭的,当且仅当,若 $\sigma \in \Sigma^\omega$,使得 $\sigma(n) \in Init(L)$, $n=1,2,\cdots$,则 $\sigma \in L$.

设 L 是 Σ 上的 ω—语言,令

$Init(L)=\{x \in \Sigma^* \mid$ 存在 $\sigma \in \Sigma^\omega$,使得 $x\sigma \in L\}$;

$Tail(L)=\{\sigma \in \Sigma^\omega \mid$ 存在 $x \in \Sigma^*$,使得 $x\sigma \in L\}$.

定义 4. 一个确定的 ω—有穷自动机是一个五元组 $M=(K,\Sigma,\delta,q_0,F)$.
其中 K 是状态的有穷非空集;Σ 是输入字母表;$\delta: K \times \Sigma \to K$ 是状态转换函数;$q_0(\in K)$是初始状态;$F \in 2^K$ 是指定状态集族.

设 $\sigma=a_1a_2\cdots, a_i \in \Sigma, i=1,2,\cdots$,若状态的无穷序列 $r: q_0, q_1, \cdots, q_n, \cdots$,使得 $q_{i+1}=\sigma(q_i, a_{i+1}), i=0,1,\cdots$,则称无穷序 $r: q_0, q_1, \cdots, q_n, \cdots$ 是确定的 ω—有穷自动机 M 在 ω—串 σ 上的运行. 显然,如果状态序列 $r: q_0, q_1, \cdots q_n, \cdots$ 是确定的 ω—有穷自动机 M 在 ω—串 σ 上的运行,则它确定一个映射 $fr: N \to K$, N 是自然数集合,使得 $fr(i)=q_i$, $i=1,2,\cdots$. 今令 $Infr=\{q \in K \mid Card(fr^{-1}(q)) \geqslant \omega\}$,即 $Infr$ 表示在状态序列 r 中出现无穷多次的状态集合.

我们说 Σ 上的一个 ω—串 σ 被确定的 ω—有穷自动机, $M=(K,\Sigma,\delta,q_0,F)$, 接受,当且仅当 M 在 σ 上的运行 r,使 $Infr \in F$. 被确定的 ω—有穷自动机 M 接受的一切 ω—串的集合记作 $T(M)$,即

$T(M)=\{\sigma \in \Sigma^\omega \mid M$ 在 σ 上的运行 r,使 $Infr \in F\}$.

设 L 是 Σ 上的一个 ω—语言,若存在一个确定的 ω—有穷自动机 M,使得 $T(M)=L$,则称 L 是 Σ 上的一个 ω—正则语言.

定理 1. Σ 上的 ω—凸语言类是广义 ω—凸语言类的真子类.

证明:首先证明,对 $\sigma_1, \sigma_2 \in \Sigma^\omega$,若 $\sigma_1 \leqslant_{\omega'} \sigma_2$,则 $\sigma_1 \leqslant_\omega \sigma_2$. 由于 $\sigma_1 \leqslant_{\omega'} \sigma_2$,则存在 $\sigma' \in \Sigma^\omega, x, y \in \Sigma^*$,使得 $\sigma_1=x\sigma', \sigma_2=y\sigma'$,且 $x \leqslant y$,即 $x=x_1x_2\cdots x_n, y=y_1x_1y_2x_2\cdots y_nx_ny_{n+1}$. 显然 $\sigma_1 \leqslant_\omega \sigma_2$. 今证明:若 L 是 Σ 上的 ω—凸语言,则 L 是广义 ω—凸语言. 设 $\sigma_1 \leqslant_{\omega'} \sigma \leqslant_{\omega'} \sigma_2, \sigma_1, \sigma_2 \in L, \sigma \in \Sigma^\omega$,由上所证,知 $\sigma_1 \leqslant_\omega \sigma \leqslant_\omega \sigma_2$,而 $\sigma_1, \sigma_2 \in L$. 所以,由 L 是 Σ 上的 ω—凸语言可推得 $\sigma \in L$,于是证明了 L 是 Σ 上的广义 ω—凸语言.

下面证明 Σ 上的 ω—凸语言类是广义 ω—凸语言类的真子类.

设 $\Sigma=\{a,b\}, L=\{a,b\}^*(ab)^\omega$,可以证明:

(1) L 是广义 ω—凸的. 设 $\sigma_1 \leqslant_{\omega'} \sigma \leqslant_{\omega'} \sigma_2, \sigma_1, \sigma_2 \in L$,由前一个不等式知存在 σ',使得 $\sigma_1=x\sigma', \sigma=y\sigma', x, y \in \Sigma^*, x \leqslant y, \sigma=w(ab)^\omega, w \in \Sigma^*$;从第二个不等式知,存在 σ'',使得 $\sigma=y_1\sigma'', \sigma_2=z\sigma'', y_1, z \in \Sigma^*, y_1 \leqslant z, \sigma''=u(ab)^\omega, u \in \Sigma^*$. 实际上从第 1 个事实即可推出 $\sigma=yw(ab)^\omega \in L$.

(2) L 不是 ω—凸的. 设 $\sigma_1=(ab)^\omega, \sigma_2=abb(ab)^\omega, \sigma=abbabb\cdots=(abb)^\omega$,显然, $\sigma_1 \leqslant_\omega \sigma \leqslant_\omega \sigma_2$,但 $\sigma \notin L$.

从而证明了定理 1 的结论.

定理 2. 设 L 是 Σ 上的闭的广义 ω—凸语言,若 $L=Init(L)Tail(L)$,则 L 是 Σ 上的 ω

—正则语言.

证明：设 $x,y\in Init(L)$ 且 $x\leqslant z\leqslant y,z\in\Sigma^*$. 由于 $x\in Init(L)$，所以必存在 $\sigma'\in\Sigma^\omega$，使得 $x\sigma'\in L$. 显然，$\sigma'\in Tail(L)$. 又由于 $L=Init(L)Tail(L)$，所以 $y\sigma'\in L$. 从而，根据 $x\sigma'\leqslant_{\omega f}z\sigma'\leqslant_{\omega f}y\sigma'$ 以及 L 是 Σ 上的广义 ω—凸语言，立即推得 $z\sigma'\in L$，也就是 $z\in Init(L)$，于是证明了 $Init(L)$ 是 Σ 上的一个凸语言. G. Thierrin 在文献[2]中证明了凸语言必是正则集. 因此 $Init(L)$ 是 Σ 上的一个正则集.

由于 $Init(L)$ 是 Σ 上的正则集，则存在一个确定的有穷自动机 $M=(K,\Sigma,\delta,q_0,F)$ 使 $T(M)=Init(L)$.

今构造了一个确定的 ω—有穷自动机

$$M'=(K',\Sigma',\delta',q'_0,F')$$

其中 $K'=K,\Sigma'=\Sigma,\delta'=\delta,q'_0=q_0,F'=2^F-\{\varphi\}$.

设 ω—串 $\sigma\in T(M')$，由 M' 在 σ 上的运行 $r:q_0,q_1,\cdots,q_i,\cdots$ 所确定的映射 fr，使 $Infr\in F'$. 于是在状态序列 $q_0,q_1,\cdots,q_i,\cdots$ 中出现无穷多次的状态均为有穷自动机 M 的终止状态. 假设在状态序列 $q_0,q_1,\cdots,q_i,\cdots$ 中属于 $Infr$ 的状态所组成的子列为 $q_{i_1},q_{i_2},\cdots,q_{i_j},\cdots$. 显然，对任意的正整数 n，必存在正整数 m 与 k，使 $\delta(q_0,\sigma(n+m))=q_{i_k}$. 但是，因为 $q_{ik}\in F$，所以 $\sigma(n+m)\in Init(L)$，即存在 $\sigma'\in L$，使 $\sigma'(n+m)=\sigma(n+m)$，显然，$\sigma'(n)=\sigma(n)$，从而知 $\sigma(n)\in Init(L)$，这说明 ω—串 σ 的任何前缀都属于 $Init(L)$，又因为 L 是闭的 ω—语言，所以 σ 必属于 L. 这就证明了 $T(M')=L$.

设 $\sigma\in L$，由于对任意的自然数 $n,\sigma(n)\in Init(L)$，即 $\sigma(n)\in T(M)$，所以 $\delta(q_0,\sigma(n))\in F$. 于是，$M'$ 在 σ 上运行 $r:q_0,q_1,\cdots,q_n,\cdots$，使 $Infr\in F'$，即 $\sigma\in T(M')$. 从而证明了 $L=T(M')$.

由上述 2 方面证得 $T(M')=L$. 故定理又得证.

参考文献

1 Cohen R S, Gold A Y. Theory of ω—languages, I: characterizations of ω—content—free languages. J. Comput System Sci., 1977, **15**: 169～184.

2 Thierrin G. Convex languages, automata, languages and programming. North—Holland Publishing Company, 1972. 481～492.

3 苏锦祥. 一类 ω—正则语言. 软件学报, 1990, **1(3)**: 29～32.

4 庄雷, 苏锦祥. 关于 ω—正则语言类的一个子类的特征. 计算机学报, 1992, **15(11)**: 815～818.

THE SET OF ω—REGULAR LANGUAGES

Zhuang Lei　Meng Qingyuan　Su Jinxiang

(*Department of Computer Science　Zhengzhou University　Zhengzhou*　450052)

Abstract　If Σ is a finite alphabet, ω—language is the set which consists of some infinite strings over Σ, ω—regular language is the ω—language that be accepted by ω—finte automata. The authors described a subclass ω—language from set viewpoint. It cannot tradition that gives the subclass ω—regular language from generation or recognizition. In this paper, the authors also show the subclass ω—regular language from set viewpoint. This subclass is more widely.

Key words ω—regular language, ω—convex language, general ω—convex language, closed ω—language.