第 28 卷 第 7 期 计 算 机 学 报 Vol. 28 No. 7
2005 年 7 月 CHINESE JOURNAL OF COMPUTERS July 2005

字的组合的半群方法

刘耀军[1)] 徐宗本[2)]
[1)](太原师范学院数学系 太原 030012)
[2)](西安交通大学理学院 西安 710049)

摘 要 该文利用半群方法给出了语言的一些代数性质. 首先,讨论了稠密语言的半群结构,给出了包含语言 $w(w^k)^*$(其中 $w\in A^+$,k 是正整数)的一个稠密语言类;证明了稠密正规语言包含一个字与一个稠密正规右酉幺半群的积. 其次,讨论了自由幺半群的正规分支可分解性,证明了自由幺半群及正规分支可分解语言与正规分支可分解的后缀语言的积是正规分支可分解的;应用这些结果证明了 Shyr 和 Yu 关于正规语言的两个猜想.

关键词 形式语言;稠密语言;有限自动机;半群;码
中图法分类号 TP301

Semigroup Method in Combinatorics on Words

LIU Yao-Jun[1)] XU Zong-Ben[2)]
[1)](*Department of Mathematics, Taiyuan Teachers' College, Taiyuan* 030012)
[2)](*Faculty of Science, Xi'an Jiaotong University, Xi'an* 710049)

Abstract Semigroup method and Combinatorics method are two important techniques in the study of words and languages. This paper presents the semigroup method and explore languages by using semigroup method. Some algebraic properties of languages be given via semigroup method. Firstly, the semigroup constructions of dense languages are given. It's proved that every dense language which syntactic monoid has the zero, or which syntactic monoid has the minimal ideal and this ideal is a periodic semigroup having a primitive idempotent contains a language $w(w^k)^*$ for some nonempty word w and some positive integer k. The theorem every code is thin is generalized to the result that every code which syntactic monoid has the zero, or which syntactic monoid has the minimal ideal and this ideal is a periodic semigroup having a primitive idempotent is thin. It's proved that every dense regular language contains a product of some word with some dense regular right unitary submonoid of A^*, and by using this result it is proved that every dense regular language also contains a primitive word. Secondly, the regular component decomposition of free monoids are given. It's proved that free monoids are regular component splittable and that every product of a regular component splittable language with a regular component splittable suffix language is regular component splittable. As applications of the above results two conjectures about regular languages given by Shyr and Yu, in 1998 are proved. The two conjectures are that regular languages are regular component splittable and dense regular languages contains imprimitive words.

Keywords formal language; dense language; finite automaton; semigroup; code

---

收稿日期:2003-03-18;修改稿收到日期:2005-02-05. 本课题得到国家自然科学基金(A10071068)资助. 刘耀军,男,1963 年生,博士,教授,研究方向为形式语言与自动机. E-mail: yjliu6980@163.com. 徐宗本,男,1955 年生,博士,教授,博士生导师,研究方向为神经网络的统一模型与理论、模拟演化计算技术的数学基础、生物信息处理的算法研究以及非线性算子理论.

1　引　言

设 A 是有限非空集合，A^* 是由 A 生成的自由幺半群，称 A 为字母表，称 A 中的元素为字母，称 A^* 中的元素为字，称 A^* 的子集为语言．令 $A^+=A^*\backslash\{1\}$，这里 1 是空字，即 1 是不含任何字母的字．对于任意语言 L，分别用 L^* 和 L^+ 表示由 L 生成的 A^* 的子幺半群和子半群．把 $\{w\}^+$ 简记为 w^+．

称不能写为其它字的幂的非空字为本原字．令 Q 是 A^* 中所有本原字组成的集合．对于每个非空字 w 存在唯一的本原字 u，使得 w 是 u 的幂，称 u 为 w 的本原根，记为 $\sqrt{w}$．对于任意语言 L，称由 L 中所有非空字的本原根组成的集合为 L 的本原根，记为$\sqrt{L}$．如果 u,v 是两个不同的本原字，则对于任意两个大于 1 的正整数 m,n,u^mv^n 是本原字．

设 $w=a_1a_2\cdots a_n$ 是 A^* 中的字，$a_1,a_2,\cdots,a_n$ 是字母，称字 $a_na_{n-1}\cdots a_1$ 为字 w 的镜像，记为 $\widetilde{w}$．称语言 L 中的所有字的镜像组成的集合为 L 的镜像，记为 $\widetilde{L}$．

设 L 是语言，如果对于 A^* 中的任意字 w 都有 $L\cap A^*wA^*\neq\varnothing$，则称 L 为稠密的，否则称 L 为薄的；如果对于 A^* 中的任意字 w 都有 $L\cap wA^*\neq\varnothing$，则称 L 为右稠密的，否则称 L 为右薄的．

如果语言 P 满足条件 $P\cap PA^+=\varnothing$，则称 P 为前缀语言，如果 $P\neq 1$，则称 P 为前缀码．称不能包含于其它前缀码中的前缀码为极大前缀码．可以对称地定义后缀码和极大后缀码．如果由语言 C 生成的 A^* 的子幺半群中的每个字都可唯一地写为 C 中的字的积，则称 C 为码．称不能包含于其它码中的码为极大码．

若 M 是 A^* 的子幺半群，$M\neq\{1\}$，则 $(M-1)\backslash(M-1)^2$ 是 M 的生成元素集，称之为 M 的最小生成元素集．若 A^* 的子幺半群 U 可以由码生成，则称 U 为自由的．若 A^* 的子幺半群 U 满足 $U^{-1}U\subseteq U$，则称 U 为右酉的．A^* 的每个右酉子幺半群是 A^* 的自由子幺半群．

设 S 是半群，如果对于 S 中的任意元素 a，存在两个相异的正整数 k,l 使得 $a^k=a^l$，则称 S 为周期半群．如果 S 中元素 z 满足对于 S 中的任意元素 a 都有 $az=za=z$，则称 z 为 S 的零元．如果 S 的元素 e 满足 $e^2=e$，则称 e 为幂等元，如果对于 S 的任意满足 $ef=fe=f$ 的幂等元 f 都有 $e=f$，则称 e 为 S 的本原幂等元．称不含零元的无真理想的半群为单半群，称含有本原幂等元的单半群为完全单半群．

字母表 A 上的自动机$\mathcal{A}$是一个五元组(S,A,δ,i,T)，其中 S 是非空状态集；A 是有限字母表，δ 是 $S\times A\times S$ 的子集，称之为边集，使得对于每个状态 s 和每个字母 a 至多有一个状态 s' 满足$(s,a,s')\in\delta$，$i\in S$为初始状态，$T\subseteq S$ 为终状态集．如果 S 是有限集，则称$\mathcal{A}$为有限自动机．称边的连续序列

$$c=(s_0,a_1,s_1)(s_1,a_2,s_2)\cdots(s_{k-1},a_{k-1},s_k),k\geq 0$$

为自动机$\mathcal{A}$的路，称 s_0 为 c 的起点，称 s_k 为 c 的终点．通常将如上的路简记为 $c:s_0\rightarrow s_k$，称 $a_1a_2\cdots a_k$ 为路 c 的标号．约定对于每个状态 s，存在 s 到其自身的路，其标号为空字 1．称起点为初始状态，终点为终状态的路为成功路．称由$\mathcal{A}$的所有成功路的标号组成的集合为被自动机$\mathcal{A}$识别的语言，记为 $L(\mathcal{A})$．称可被有限自动机识别的语言为正规语言．如果 L，L'是正规语言，则 $L\cap L',L\cup L',L-L',LL',L^*$是正规语言．对于任意语言 L 及任意字 w 定义

$$Cont_L(w)=\{(u,v)\in A^*\times A^*:uwv\in L\}.$$

对于 A^* 中的任意两个字 w,w'，定义 wP_Lw'当且仅当 $Cont_L(w)=Cont_L(w')$，则 P_L 是 A^* 上的一个同余，称为由 L 定义的句法同余，称商幺半群 A^*/P_L 为 L 的句法幺半群．众所周知，一个语言是正规语言的充要条件是其句法幺半群为有限幺半群．称句法同余为恒等关系的语言为析取语言，称既不是正规语言又不是析取语言的语言为中介语言．

正规语言的抽吸引理[1]指出，每个无限正规语言都包含一个形如 uv^+w 的正规语言，其中 $u,w\in A^*$，$v\in A^+$，称uv^+w为正规分支．作为抽吸引理的推论可以得到：每个正规语言 L 都可以分解为$(\bigcup_{i\in I}u_iv_i^+w_i)\cup F$的形式，其中 F 是有限语言，I 是指标集，对于每个 $i\in I$，都有 $u_i,w_i\in A^*,v_i\in A^+$．称 L 的这种形状的分解式为 L 的正规分支分解；如果 I 是有限集，则称此分解为 L 的有限正规分支分解；如果这个并是不交并则称此分解为正规分支的不交并分解．如果一个语言有正规分支的不交并分解，则称此语言为正规分支可分解的，如果一个语言有有限正规分支的不交并分解，则称此语言为有限正规分支可分解的，Shyr 和 Yu 已经证明，如果一个语言有一个有限正规分支分解，则该语言是有限正规分支可分解的．

在第 2 节中，我们讨论了稠密语言中的半群结构；证明了句法幺半群中含有零元的稠密语言及句法幺半群中含有极小理想且此极小理想是含有本原幂等元的周期半群的稠密语言包含一个语言

$w(w^k)^*$，其中 w 是一个非空字，k 是一个正整数. 把每个正规码是薄码推广为：句法幺半群中含有零元的码及句法幺半群中含有极小理想且此极小理想是含有本原幂等元的周期半群的码是薄码；证明了每个稠密正规语言包含一个字与一个稠密正规右酉幺半群的积.

在第 3 节中，我们讨论了自由幺半群的正规分支可分解性；证明了每个自由幺半群及每个正规分支可分解语言与正规分支可分解的后缀语言的积是正规分支可分解的.

作为如上结论的应用，证明了 Shyr 和 Yu 在 1998 年提出的关于正规语言的两个猜想：每个稠密正规语言含有非本原字；正规语言是正规分支可分解的. 证明了每个稠密正规语言中也含有本原字.

2 稠密语言中的半群结构

首先，我们利用组合方法给出稠密语言的一些性质.

引理 1[2]. 如果 $|A|\geq 2$，L 是字母表 A 上的语言，则 L 是稠密的当且仅当 $\sqrt{L}$ 是稠密的.

命题 1. $L=L_1L_2\cdots L_k$ 是稠密语言当且仅当存在 $i(1\leq i\leq k)$，使得 L_i 是稠密语言.

证明. 如果 L_i 是稠密语言，则由稠密语言的定义，显然有 $L=L_1L_2\cdots L_k$ 是稠密语言.

反过来，设 $L=L_1L_2\cdots L_k$ 是稠密语言. 若 $k=1$，则 $L=L_1$，结论自然成立. 若 $k=2$，则 $L=L_1L_2$. 假设 L_1，L_2 都不是稠密的，则存在字 w_1，w_2，使得 $L_1\cap A^*w_1A^*=\varnothing$，$L_2\cap A^*w_2A^*=\varnothing$. 令 $w=w_1w_2$，假设 $L\cap A^*wA^*\neq\varnothing$，即 $L\cap A^*w_1w_2A^*\neq\varnothing$，则存在 $x_1\in L_1$，$x_2\in L_2$，y_1，$y_2\in A^*$，使得 $x_1x_2=y_1w_1w_2y_2$. 这样或者

(1) 存在 $u\in A^*$ 使得 $x_1u=y_1w_1$，$x_2=uw_2y_2$；

或者

(2) 存在 $v\in A^*$ 使得 $x_1=y_1w_1v$，$vx_2=w_2y_2$.

在(1)的情况下，$x_2=uw_2y_2\in L_2\cap A^*w_2A^*$，这与 $L_2\cap A^*w_2A^*=\varnothing$ 矛盾. 在(2)的情况下，$x_1=y_1w_1v\in L_1\cap A^*w_1A^*$，这与 $L_1\cap A^*w_1A^*=\varnothing$ 矛盾. 于是 $L\cap A^*wA^*=\varnothing$，这与 L 稠密矛盾，因此 L_1，L_2 中至少有一个稠密. 若 $L=L_1L_2\cdots L_k(k\geq 3)$ 稠密，L_k 不稠密，则由归纳基础知 $L_1L_2\cdots L_{k-1}$ 稠密，进一步的由归纳假设知，存在 $i(1\leq i\leq k-1)$ 使得 L_i 是稠密的. 证毕.

引理 2[3]. $L=L_1\cup L_2\cup\cdots\cup L_k$ 是稠密语言当且仅当存在 $i(1\leq i\leq k)$ 使得 L_i 是稠密语言.

推论 1. 若 $|A|\geq 2$，$u,w\in A^*$，$v\in A^+$，则 uv^+w 不是稠密的. 进一步的，$L=\bigcup_{f\in uv^+w} f^+$ 不是稠密的.

证明. 因为 $\sqrt{v^+}=\{\sqrt{v}\}$ 是单点集，所以 $\{\sqrt{v}\}$ 不是稠密的，因此由引理 1 知，v^+ 不是稠密的；同样的，由 u,w 是单点集知 $\{u\}$，$\{v\}$ 也不是稠密的，于是由命题 1 知 $uv^+w=\{u\}v^+\{w\}$ 不是稠密的，从而 $\sqrt{uv^+w}$ 不是稠密的. 因为 $\sqrt{L}=\sqrt{\bigcup_{f\in uv^+w} f^+}=\sqrt{uv^+w}$，故再一次由引理 1 知 L 不是稠密语言. 证毕.

推论 2. 若 $|A|\geq 2$，则 A 上的任意有限正规分支可分解的语言都不是稠密语言.

下面我们利用半群方法探讨稠密语言的代数性质.

命题 2. 薄语言的句法幺半群包含零元. 进一步的，薄语言不是析取的.

证明. 若 L 是薄语言，则存在 $w\in A^*$ 使得 $L\cap A^*wA^*=\varnothing$，因为对于任意字 $u\in A^*$ 都有 $L\cap A^*uwA^*\subseteq L\cap A^*wA^*$，$L\cap A^*wuA^*\subseteq L\cap A^*wA^*$，所以 $L\cap A^*uwA^*=L\cap A^*wuA^*=L\cap A^*wA^*=\varnothing$，从而 $Cont_L(uw)=Cont_L(wu)=Cont_L(u)=\varnothing$，故 $wu\equiv uw\equiv w(P_L)$，因此 $\tilde{u}\tilde{w}=\tilde{w}\tilde{u}=\tilde{w}$，$\tilde{w}$ 是 L 的句法幺半群 A^*/P_L 的零元. 如果 $u\in A^+$，则 $uw\neq w$，且 $uw\equiv w(P_L)$，因此 P_L 不是恒等关系，故 L 不是析取语言. 证毕.

引理 3[4]. 不含零元的半群的极小理想是单半群.

引理 4[4]. 每个有限半群含有幂等元.

引理 5[4]. 完全单半群是群的并.

定理 1. 若 L 是稠密语言，且 L 的句法幺半群中含有零元，则 L 包含由非本原字生成的半群.

证明. 设 $\tilde{w}$ 是 L 的句法幺半群 A^*/P_L 中的零元，$w\in A^*$. 由命题 2 的证明知可以假设 $w\in A^+$. 因为 L 是稠密的，所以存在 $u,v\in A^*$ 使得 $uwv\in L$. 因为 $\tilde{w}$ 是 A^*/P_L 的零元，所以对于任意正整数 k 都有 $(\widetilde{uwv})^k=\tilde{w}=\widetilde{uwv}$，故 $(uwv)^k\equiv uwv(P_L)$，这样 $Cont_L(uwv)=Cont_L((uwv)^k)$. 因为 $uwv\in L$，故 $(1,1)\in Cont_L(uwv)$，从而 $(1,1)\in Cont_L((uwv)^k)$，于是 $(uwv)^k\in L$，L 包含由 uwv 生成的半群 $(uwv)^+$，自然 L 包含由非本原字 $(uwv)^2$ 生成的半群 $((uwv)^2)^+$. 证毕.

定理 2. 若 L 是稠密语言，且 L 的句法幺半群含有极小理想且此极小理想是含有本原幂等元的周

期半群，L 的句法幺半群不含零元，则 L 包含一个语言 $w(w^k)^*$，其中 w 是一个非空字，k 是一个正整数. 特别的，L 含有非本原字.

证明. 设 K 是 L 的句法幺半群 A^*/P_L 的极小理想，由引理 3 知 K 是单半群，因为 K 中有本原幂等元，所以 K 是完全单半群. 假设 $K=\{\widetilde{1}\}$，则对于任意 $w\in A^*$，由 $\widetilde{w}\,\widetilde{1}=\widetilde{w}\,\widetilde{1}\in K$ 知 $\widetilde{1}\widetilde{w}=\widetilde{w}\,\widetilde{1}=\widetilde{1}$，从而 $\widetilde{1}$ 是 A^*/P_L 的零元，这与 A^*/P_L 不含零元矛盾，因此$K\neq\{\widetilde{1}\}$. 于是，存在 $w\in A^+$ 使得 $\widetilde{w}\in K$. 因为 L 稠密，所以存在$u,v\in A^*$，使得$uwv\in L$，而$\widetilde{uwv}=\widetilde{u}\widetilde{w}\widetilde{v}\in K$，因此，由引理 5 知$\widetilde{uwv}$包含于 K 的某个子群G 中. 设 $\tilde{e}$ 是群G 的单位元，其中 $e\in A^*$. 因为 K 是周期半群，所以存在正整数 k 使得$\widetilde{uwv}^{\,k}$ 是幂等元，因为 $\tilde{e}$ 是 G 的唯一的幂等元，所以$\widetilde{uwv}^{\,k}=\tilde{e}$，从而对于任意的正整数 n 都有$\widetilde{uwv}^{\,kn}=\tilde{e}$，进一步的，$\widetilde{uwv}^{\,kn+1}=\widetilde{uwv}$，故 $(uwv)^{kn+1}\equiv uwv(P_L)$，这样 $Cont_L(uwv)=Cont_L((uwv)^{kn+1})$. 因为$uwv\in L$，故 $(1,1)\in Cont_L(uwv)$，从而$(1,1)\in Cont_L((uwv)^{kn+1})$，于是 $(uwv)^{kn+1}\in L$，L 包含 $(uwv)((uwv)^k)^*$. $(uwv)^{k+1}$是 L 中的非本原字.　　证毕.

推论 3[2]（Shyr-Yu 关于稠密正规语言的猜想）. 稠密正规语言中含有非本原字.

证明. 设 L 是一个稠密正规语言，则由 L 是正规语言知 L 的句法幺半群是有限的. 如果 A^*/P_L 含有零元，则由定理 1 知 L 中含有非本原字. 如果 A^*/P_L 中不含零元，则由 A^*/P_L 有限知 A^*/P_L 含有极小理想且此极小理想是含有本原幂等元的周期半群，因此由定理 2 知 L 含有非本原字.　　证毕.

引理 6[2]. 若 $\Lambda\subseteq Q$，且 Λ 是无限正规语言，则 $L=\bigcup\limits_{f\in\Lambda}f^+$ 不是正规语言.

证明. 因为 Λ 是无限正规语言，所以存在 $u,w\in A^*$，$v\in A^+$，使得 $uv^+w\subseteq\Lambda$. 假设 L 是正规语言，则 A^*/P_L 是有限幺半群，从而由引理 4 知存在正整数 k 使得 $v^k\equiv v^{2k}(P_L)$. 又由 Λ 是正规语言知，A^*/P_Λ 是有限幺半群，从而存在正整数 l 使得 $v^l\equiv v^{2l}(P_\Lambda)$. 于是 $v^{kl}\equiv v^{2kl}(P_L)$，且 $v^{kl}\equiv v^{2kl}(P_\Lambda)$，从而 $uv^{kl}w\equiv uv^{2kl}w(P_L)$，且 $uv^{kl}w\equiv uv^{2kl}w(P_\Lambda)$，进一步的$(uv^{kl}w)^2(uv^{2kl}w)^2\equiv(uv^{kl}w)^4(P_L)$，且$(uv^{kl}w)^2(uv^{2kl}w)^2\equiv(uv^{kl}w)^4(P_\Lambda)$.

因为 $uv^{kl}w\in\Lambda$，所以$(uv^{kl}w)^4\in L$，故$(1,1)\in Cont_L((uv^{kl}w)^4)$，因此 $(1,1)\in Cont_L((uv^{kl}w)^2(uv^{2kl}w)^2)$，故$(uv^{kl}w)^2(uv^{2kl}w)^2\in L$. 由 L 的定义知存在 $f\in\Lambda$ 及正整数n 使得$(uv^{kl}w)^2(uv^{2kl}w)^2=f^n$. 因为 $v\neq1$，$kl\geq1$，故 $uv^{kl}w\neq uv^{2kl}w$，而 $uv^{kl}w,uv^{2kl}w\in\Lambda\subseteq Q$，因此由本原字的性质知$(uv^{kl}w)^2(uv^{2kl}w)^2$ 是本原字，故 $n=1$，从而$(uv^{kl}w)^2(uv^{2kl}w)^2\in\Lambda$. 这样$(1,1)\in Cont_\Lambda((uv^{kl}w)^2(uv^{2kl}w)^2)$，从而$(1,1)\in Cont_\Lambda((uv^{kl}w)^4)$，于是$(uv^{kl}w)^4\in\Lambda\subseteq Q$，这与$(uv^{kl}w)^4$不是本原字矛盾. 因此，$L$ 不是正规语言.　　证毕.

推论 4[2]（Shyr-Yu 关于中介语言的猜想）. 如果 $\Lambda\subseteq Q$，且 Λ 是正规语言，则 $L=\bigcup\limits_{f\in\Lambda}f^+$ 是中介语言当且仅当 Λ 是无限语言.

证明. 若 $L=\bigcup\limits_{f\in\Lambda}f^+$ 是中介语言，则 L 不是正规语言，由正规语言的有限并是正规语言及 $f^+=f^*-\{1\}$是正规语言知，Λ 是无限语言. 反过来，假设 Λ 是无限语言，则由引理 6 知 $L=\bigcup\limits_{f\in\Lambda}f^+$ 不是正规语言，假设 L 是稠密语言，则

$$\sqrt{L}=\sqrt{\bigcup_{f\in\Lambda}f^+}=\bigcup_{f\in\Lambda}\sqrt{f^+}=\bigcup_{f\in\Lambda}\sqrt{f}=\Lambda\subseteq Q$$

是稠密语言，而由假设知$\sqrt{L}=\Lambda$ 是正规语言，因此由推论 3 知$\sqrt{L}$中含非本原字，这与$\sqrt{L}$中不含本原字矛盾，因此 L 不是稠密语言. 从而 L 不是析取语言，故 L 是中介语言.　　证毕.

推论 5. 若 L 是稠密正规语言，则存在非空字 w 及正整数 k 使得 $w(w^k)^*\subseteq L$.

证明. 因为 L 是稠密正规语言，所以 L 满足定理 1 或定理 2 的条件，因此由定理 1 的证明及定理 2 的结论知推论成立.　　证毕.

定理 3. 句法幺半群中含有零元的码及句法幺半群中含有极小理想且此极小理想是含有本原幂等元的周期半群的码是薄码，特别的正规码是薄码.

证明. 假设 C 稠密，则由推论 5 知存在非空字 w 及正整数 k 使得 $w(w^k)^*\subseteq C$，由

$$\begin{aligned}w^{k+1}w^{2k+1}\cdots w^{k(k+1)+1}&=w^{(k+1)+(2k+1)+\cdots+(k(k+1)+1)}\\&=w^{(k+2k+\cdots+k(k+1))+(k+1)}\\&=w^{\left[\frac{(k+1)(k+2)}{2}+1\right]k+1}\\&=w^{\left[\frac{(k+1)(k+2)}{2}+1\right]k}w\end{aligned}$$

及 $w^{k+1},w^{2k+1},\cdots,w^{k(k+1)+1},w^{\left[\frac{(k+1)(k+2)}{2}+1\right]k}w\in C$ 知，$w^{\left[\frac{(k+1)(k+2)}{2}+1\right]k}w$ 关于 C 有两个不同的分解式，这与 C 是码矛盾，因此 C 是薄码. 如果 C 是正规码，则 C 适合定理 3 的条件，因此 C 是薄码.　　证毕.

引理 7[5]（正规语言的 UP 分解定理）. 字母表 A 上每个正规语言都可以分解为如下形式的不交并：

$$L=L_1\cup L_2\cup\cdots\cup L_k,$$

其中

$$L_i = U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots a_{i_{s_i-1}} U_{i_{s_i}} \ (1 \le i \le k),$$

这里 $U_{i_1}, U_{i_2}, \cdots, U_{i_{s_i}}$ 是 A^* 的正规右酉子幺半群，$a_{i_1}, a_{i_2}, \cdots, a_{i_{s_i-1}}$ 是字母，且每个

$$U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots U_{i_t} a_{i_t} \ (1 \le t \le s_i - 1)$$

是前缀码.

定理 4. 如果 L 是字母表 A 上的稠密正规语言，则存在字 u 及 A^* 的稠密正规右酉子幺半群 U 使得 $uU \subseteq L$.

证明. 因为 L 是正规的，故由引理 7 知 L 有 UP 分解，即 L 可分解为

$$L = L_1 \cup L_2 \cup \cdots \cup L_k,$$

这里 $L_1, L_2, \cdots, L_k$ 是两两互不相交的正规语言，且对于每个 $i(1 \le i \le k)$ 有

$$L_i = U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots U_{i_{s_i-1}} a_{i_{s_i-1}} U_{i_{s_i}},$$

且 $U_{i_1}, U_{i_2}, \cdots, U_{i_{s_i}}$ 是 A^* 的正规右酉子幺半群，$a_{i_1}, a_{i_2}, \cdots, a_{i_{s_i-1}}$ 是字母，且对于每个 $t(1 \le t \le s_i - 1)$ 有

$$U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots U_{i_t} a_{i_t}$$

是前缀码. 由 L 稠密根据引理 2 知，存在 $i(1 \le i \le k)$ 使得 L_i 稠密，如果 $L_i = U_{i_1}$，则存在字 $u = 1$ 及正规稠密右酉幺半群 $U = U_{i_1}$ 使得 $uU = L_i \subseteq L$，如果 $L_i \ne U_{i_1}$，则由 L_i 的分解式根据命题 1 知，存在 t 使得 U_{i_t} 稠密，这里 $1 \le t \le s_i$. 今断言 $t = s_i$，因若不然，则 $t \le s_i - 1$，由

$$U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots U_{i_t} a_{i_t}$$

是前缀码知其子集 $a_{i_1} a_{i_2} \cdots a_{i_{t-1}} U_{i_t} a_{i_t}$ 是前缀码，从而是码，因为

$$a_{i_1} a_{i_2} \cdots a_{i_{t-1}} U_{i_t} a_{i_t}$$

是正规语言，故由定理 3 知，$a_{i_1} a_{i_2} \cdots a_{i_{t-1}} U_{i_t} a_{i_t}$ 是薄语言；另一方面，由 U_{i_t} 稠密知 $a_{i_1} a_{i_2} \cdots a_{i_{t-1}} U_{i_t} a_{i_t}$ 是稠密语言，产生矛盾. 因此 $t = s_i$，于是 $U_{i_{s_i}}$ 是 A^* 的稠密正规右酉子幺半群，$a_{i_1} a_{i_2} \cdots a_{i_{s_i-1}}$ 是 A^* 中的字，使得

$$a_{i_1} a_{i_2} \cdots a_{i_{s_i-1}} U_{i_{s_i}} \subseteq U_{i_1} a_{i_1} U_{i_2} a_{i_2} \cdots U_{i_{s_i}} \subseteq L.$$

证毕.

推论 6. 至少含有两个字母的字母表上的稠密正规语言含有本原字.

证明. 设 L 是至少含有两个字母的字母表 A 上的稠密正规语言，则由定理 4 知存在字母表 A 上的字 u 及 A^* 的稠密正规右酉子幺半群 U 使得 $uU \subseteq L$. 由 U 稠密，$|A| \ge 2$，根据引理 1 知 $\sqrt{U}$ 稠密，特别的，$\sqrt{U}$ 是无限集.

若 $u = 1$，则 $U \subseteq L$，取 $w_1, w_2 \in \sqrt{U}$，$w_1 \ne w_2$，则由语言的本原根的定义知存在正整数 m_1, m_2 使得 $w_1^{m_1}, w_2^{m_2} \in U$，由 U 是半群知 $w_1^{2m_1} w_2^{2m_2} \in U \subseteq L$，而由本原字的性质有 $w_1^{2m_1} w_2^{2m_2}$ 是本原字.

若 $u \ne 1$ 且 u 是本原字，则由 $uU \subseteq L$ 及 $1 \in U$ 知 $u \in L$. u 是 L 中的一个本原字.

若 $u \ne 1$ 且 u 是非本原字，则存在正整数 $m_1(m_1 \ge 2)$，使得 $u = w_1^{m_1}$，这里 $w_1 = \sqrt{u}$，取 $w_2 \in \sqrt{U} \setminus \sqrt{u}$，则存在正整数 m_2 使得 $w_2^{m_2} \in U$，由 U 是 A^* 的右酉子幺半群有 $w_2^{2m_2} \in U$，于是 $w_1^{m_1} w_2^{2m_2} \in uU \subseteq L$，且由本原字的性质知，$w_1^{m_1} w_2^{2m_2}$ 是本原字. 证毕.

3 自由幺半群的正规分支分解

引理 8[6]. A^* 的子幺半群 U 是自由的当且仅当 $U^{-1}U \cap UU^{-1} \subseteq U$.

定理 5. A^* 的自由子幺半群是正规分支可分解的.

证明. 设 U 是 A^* 的自由子幺半群. 若 $U = \{1\}$，则取正规分支的指标集 $I = \varnothing$，取有限正规语言 $F = \{1\}$，得 U 的一个正规分支的不交并的分解式 $U = F$，因此 $U = \{1\}$ 是正规分支可分解的.

若 $U \ne \{1\}$，则 U 包含非空字，从而集合

$$I = \{ f^n : f \in Q \text{ 且 } n \text{ 是使得 } f^n \in U \text{ 的最小正整数} \}$$

是非空集. 因为 U 是 A^* 的子幺半群，且 $I \subseteq U$ 故 $(\bigcup_{v \in I} v^+) \cup \{1\} \subseteq U$. 反过来，取 $u \in U$，使得 $u \ne 1$，则存在本原字 f 和一个正整数 n 使得 $u = f^n$，令 n_0 是使得 f 的幂在 U 中的最小正整数，则 $f^{n_0} \in I$. 由整数的带余除法定理知存在非负整数 q 及整数 r 使得 $n = qn_0 + r$，$0 \le r < n_0$，于是 $f^n = (f^{n_0})^q f^r$，由 f^n，$f^{n_0} \in U$ 知 $f^r \in U^{-1}U \cap UU^{-1}$，因为 U 是 A^* 的自由子幺半群，所以由引理 8 知 $f^r \in U$，这样由 n_0 的最小性知 $r = 0$，从而 $f^n = (f^{n_0})^q \in (f^{n_0})^+$，由 $f^{n_0} \in I$ 知 $u = f^n \in (\bigcup_{v \in I} v^+) \cup \{1\}$，从而 $U \subseteq (\bigcup_{v \in I} v^+) \cup \{1\}$，于是，$U$ 有正规分支的分解式：

$$U = (\bigcup_{v \in I} v^+) \cup \{1\}.$$

若 $v \in I$，则 $v \ne 1$，故 $v^+ \cap \{1\} = \varnothing$. 假设 $v_1, v_2 \in I$，使得 $v_1^+ \cap v_2^+ \ne \varnothing$，则存在正整数 m_1, m_2 使得 $v_1^{m_1} = v_2^{m_2}$. 由 I 的构造知存在本原字 f_1, f_2 及正整数 n_1，n_2，使得 $v_1 = f_1^{n_1}$，$v_2 = f_2^{n_2}$，且 n_1, n_2 顺次为使得 $f_1^{n_1}$，$f_2^{n_2} \in U$ 的最小正整数. 于是，$f_1^{m_1 n_1} = f_2^{m_2 n_2}$. 故 $\lambda(f_1^{m_1 n_1}) = \lambda(f_2^{m_2 n_2})$，即 $f_1 = f_2$. 再由 n_1, n_2 的选取

知 $n_1=n_2$，因此 $f_1^{n_1}=f_2^{n_2}$，即 $v_1=v_2$，从而 $v_1^+=v_2^+$. 于是，U 的如上分解是 U 的正规分支的不交并分解，故 U 是正规分支可分解的.　　证毕.

定理 6.　正规分支可分解的前缀语言 P 与正规分支可分解的语言 L 的积 PL 是正规分支可分解的. 特别的，正规分支可分解的前缀语言与 A^* 的自由子幺半群的积是正规分支可分解的.

证明.　如果 $P=\varnothing$ 或 $L=\varnothing$，则 $PL=\varnothing$，从而 PL 是正规分支可分解的. 下设 $P\neq\varnothing$ 且 $L\neq\varnothing$，由 P 是正规分支可分解的知存在有限语言 F_1 及指标集 I，对于每一个 $i\in I$，存在 $u_i,w_i\in A^*$，$v_i\in A^+$，使得

$$P=(\bigcup_{i\in I}u_iv_i^+w_i)\cup F_1,$$

且对任意 $i\in I$ 有

$$u_iv_i^+w_i\cap F_1=\varnothing,$$

对于 I 中任意两个相异指标 i_1,i_2 有

$$u_{i_1}v_{i_1}^+w_{i_1}\cap u_{i_2}v_{i_2}^+w_{i_2}=\varnothing.$$

由 L 是正规分支可分解的语言知，存在有限语言 F_2 及指标集 J（可以假设 $J\cap I=\varnothing$），对每一个 $j\in J$，存在 $u_j,w_j\in A^*$，$v_j\in A^+$，使得

$$L=(\bigcup_{j\in J}u_jv_j^+w_j)\cup F_2,$$

且对任意 $j\in J$ 有

$$u_jv_j^+w_j\cap F_2=\varnothing,$$

对 J 中任意两个指标 j_1,j_2 有

$$u_{j_1}v_{j_1}^+w_{j_1}\cap u_{j_2}v_{j_2}^+w_{j_2}=\varnothing.$$

于是，

$$\begin{aligned}PL&=P(\bigcup_{j\in J}u_jv_j^+w_j)\cup PF_2\\&=(\bigcup_{(p,j)\in P\times J}pu_jv_j^+w_j)\cup[(\bigcup_{i\in I}u_iv_i^+w_i)\cup F_1]F_2\\&=(\bigcup_{(p,j)\in P\times J}pu_jv_j^+w_j)\cup(\bigcup_{(i,f_2)\in I\times F_2}u_iv_i^+w_if_2)\cup F_1F_2.\end{aligned}$$

上式是 PL 的一个正规分支的分解式，其中 F_1F_2 作为有限语言 F_1 与 F_2 的积是有限语言，指标集为 $(P\times J)\cup(I\times F_2)$，$(p,j)\in P\times J$ 相应的正规分支为 $pu_jv_j^+w_j$，$(i,f_2)\in I\times F_2$ 相应的正规分支为 $u_iv_i^+w_if_2$. 下面证明此分解为不交并分解.

(i) 对 $(p,j)\in P\times J$，假设 $pu_jv_j^+w_j\cap F_1F_2\neq\varnothing$，则存在正整数 m 及字 $f_1\in F_1$，$f_2\in F_2$，使得 $pu_jv_j^mw_j=f_1f_2$，因为 $p,f_1\in P$，P 是前缀语言，故 $p=f_1$，且 $u_jv_j^mw_j=f_2$，这与 $u_jv_j^+w_j\cap F_2=\varnothing$ 矛盾，因此 $pu_jv_j^+w_j\cap F_1F_2=\varnothing$.

(ii) 对 $(i,f_2)\in I\times F_2$，如果 $u_iv_i^+w_if_2\cap F_1F_2\neq\varnothing$，则存在正整数 m 及字 $f_1\in F_1$，$f_2'\in F_2$，使得 $u_iv_i^mw_if_2=f_1f_2'$. 因为 $u_iv_i^mw_i$，$f_1\in P$，P 是前缀码，故 $u_iv_i^mw_i=f_1$. 这与 $u_iv_i^+w_i\cap F_1=\varnothing$ 矛盾，因此 $u_iv_i^+w_if_2\cap F_1F_2=\varnothing$.

(iii) 对 (p_1,j_1)，$(p_2,j_2)\in P\times J$，如果 $p_1u_{j_1}v_{j_1}^+w_{j_1}\cap p_2u_{j_2}v_{j_2}^+w_{j_2}\neq\varnothing$，则存在正整数 m，n，使得 $p_1u_{j_1}v_{j_1}^mw_{j_1}=p_2u_{j_2}v_{j_2}^nw_{j_2}$，因为 $p_1,p_2\in P$，P 是前缀码，故 $p_1=p_2$，从而 $u_{j_1}v_{j_1}^mw_{j_1}=u_{j_2}v_{j_2}^nw_{j_2}$. 于是 $u_{j_1}v_{j_1}^+w_{j_1}\cap u_{j_2}v_{j_2}^+w_{j_2}\neq\varnothing$，故 $j_1=j_2$，从而 $(p_1,j_1)=(p_2,j_2)$. 由此可得若 (p_1,j_1)，(p_2,j_2) 是 $P\times J$ 中的两个相异指标，则 $p_1u_{j_1}v_{j_1}^+w_{j_1}\cap p_2u_{j_2}v_{j_2}^+w_{j_2}=\varnothing$.

(iv) 对 $(p,j)\in P\times J$，$(i,f_2)\in I\times F_2$，假设 $pu_jv_j^+w_j\cap u_iv_i^+w_if_2\neq\varnothing$，则存在正整数 m,n，使得 $pu_jv_j^mw_j=u_iv_i^nw_if_2$，因为 p，$u_iv_i^nw_i\in P$，且 P 是前缀语言，故 $p=u_iv_i^nw_i$，从而，$u_jv_j^mw_j=f_2$，这与 $u_jv_j^+w_j\cap F_2=\varnothing$，矛盾. 因此，$pu_jv_j^+w_j\cap u_iv_i^+w_if_2=\varnothing$.

(v) 对 (i,f_2)，$(i_1,f_2')\in I\times F_2$，假设 $u_iv_i^+w_if_2\cap u_{i_1}v_{i_1}^+w_{i_1}f_2'\neq\varnothing$，则存在正整数 m,n，使得 $u_iv_i^mw_if_2=u_{i_1}v_{i_1}^nw_{i_1}f_2'$，因为 $u_iv_i^mw_i$，$u_{i_1}v_{i_1}^nw_{i_1}\in P$，$P$ 是前缀语言，故 $u_iv_i^mw_i=u_{i_1}v_{i_1}^nw_{i_1}$，从而 $f_2=f_2'$，于是，由 $u_iv_i^+w_i\cap u_{i_1}v_{i_1}^+w_{i_1}\neq\varnothing$，故 $i=i_1$，因此，$(i,f_2)=(i_1,f_2')$. 由此，对 $I\times F$ 中两个不同的指标 (i,f_2)，(i_1,f_2') 有 $u_iv_i^+w_if_2\cap u_{i_1}v_{i_1}^+w_{i_1}f_2'=\varnothing$.

综上所述，分解式

$$PL=(\bigcup_{(p,j)\in P\times J}pu_jv_j^+w_j)\cup(\bigcup_{(i,f_2)\in I\times F_2}u_iv_i^+w_if_2)\cup F_1F_2$$

是 PL 的不交并分解. 因此，PL 是正规分支可分解的. 因为自由幺半群是正规分支可分解的，所以正规分支可分解的前缀语言与 A^* 的自由子幺半群的积是正规分支可分解的.　　证毕.

命题 3.　如果 L 是正规分支可分解的，则其镜像 $\widetilde{L}$ 也是正规分支可分解的.

证明.　若 L 是正规分支可分解的语言，则存在有限集 F 及指标集 I，对每个指标 i，存在 $u_i,w_i\in A^*$ 及 $v_i\in A^+$，使得

(i) $L=(\bigcup_{i\in I}u_iv_i^+w_i)\cup F$；

(ii) 对 I 中两个不同指标 i,i'，$u_iv_i^+w_i\cap u_{i'}v_{i'}^+w_{i'}=\varnothing$；

(iii) 对 I 中任意指标 i 都有 $F\cap u_iv_i^+w_i=\varnothing$.

对如上三式求镜像得

(i′) $\widetilde{L}=(\bigcup_{i\in I}\widetilde{w}_i\widetilde{v}_i^+\widetilde{u}_i)\cup\widetilde{F}$；

(ii′) 对 $i,i'\in I$，$i\neq i'$，$\widetilde{w}_i\widetilde{v}_i^+\widetilde{u}_i\cap\widetilde{w}_{i'}\widetilde{v}_{i'}^+\widetilde{u}_{i'}=\varnothing$；

(iii′) 对 $i\in I$，$\widetilde{F}\cap\widetilde{w}_i\widetilde{v}_i^+\widetilde{u}_i=\varnothing$.

因此，(i′) 是 $\widetilde{L}$ 的不交并分解，从而 L 是正规分支可

分解的. 证毕.

引理 9[3]. 语言 L 是前缀码当且仅当 $\widetilde{L}$ 是后缀码.

推论 7. 正规分支可分解语言 L 与正规分支可分解的后缀语言 S 的积 LS 是正规分支可分解的.

推论 8. 正规分支可分解的前缀语言的有限积是正规分支可分解的. 正规分支可分解的后缀语言的有限积是正规分支可分解的.

推论 9. 如果 $L_1,L_2,\cdots,L_k$ 是正规分支可分解的语言,且对于任意 $i(1\le i\le k-1)$ 都有 $L_1L_2\cdots L_i$ 是前缀语言,则 $L_1L_2\cdots L_k$ 是正规分支可分解的. 如果 $L_1,L_2,\cdots,L_k$ 是正规分支可分解的语言,且对于任意 $i(2\le i\le k)$ 都有 $L_iL_{i+1}\cdots L_k$ 是后缀语言,则 $L_1L_2\cdots L_k$ 是正规分支可分解的.

推论 10. 如果 $L_1,L_2,\cdots,L_k$ 是正规分支可分解的,$w_1,w_2,\cdots,w_{k-1}$ 是字,使得对于任意 $i=1,2,\cdots,k-1$ 都有 $L_1w_1L_2w_2\cdots L_iw_i$ 是前缀语言,则 $L_1w_1L_2w_2\cdots L_k$ 是正规分支可分解的. 如果 $L_1,L_2,\cdots,L_k$ 是正规分支可分解的,$w_1,w_2,\cdots,w_{k-1}$ 是字,使得对于任意 $i=2,3,\cdots,k-1$ 都有 $w_iL_{i+1}w_{i+1}\cdots L_k$ 是后缀语言,则 $L_1w_1L_2w_2\cdots L_k$ 是正规分支可分解的.

定理 7. 如果 $L_1,L_2,\cdots,L_k$ 是两两不交的正规分支可分解的语言,则 $L=L_1\cup L_2\cup\cdots\cup L_k$ 是正规分支可分解的.

证明. 对 $j=1,2,\cdots,k$,由 L_j 是正规分支可分解的知,存在指标集 I_j 及有限语言 F_j,使得对于 I_j 中的每个指标 i_j,存在 $u_{i_j},w_{i_j}\in A^*$,$v_{i_j}\in A^+$ 使得

(i) $L_j=(\bigcup_{i_j\in I_j}u_{i_j}v_{i_j}^+w_{i_j})\cup F_j$;

(ii) 对任意 $i_j\in I_j$ 有 $F_j\cap u_{i_j}v_{i_j}^+w_{i_j}=\varnothing$;

(iii) 对 I_j 中任意两个不同的指标 i_j,i_j',有 $u_{i_j}v_{i_j}^+w_{i_j}\cap u_{i_j'}v_{i_j'}^+w_{i_j'}=\varnothing$. 可设 $I_1,I_2,\cdots,I_k$ 是两两不交的指标集,令

$$F=F_1\cup F_2\cup\cdots\cup F_k,$$
$$I=I_1\cup I_2\cup\cdots\cup I_k,$$

且对 $i_j\in I_j\subseteq I$,令相应的 u_{i_j},v_{i_j},w_{i_j} 如前所选,则由 $L=L_1\cup L_2\cup\cdots\cup L_k$ 得

$$L=(\bigcup_{i\in I}u_iv_i^+w_i)\cup F,$$

因为 F 是有限个有限语言的并,故 F 是有限语言,因此上式是 L 的一个正规分支的分解式. 下面证明这一个分解是不交并分解.

对于 $i\in I$,由 I 的构造知,存在 $j\in\{1,2,\cdots,k\}$,使得 $i\in I_j$,设 $i=i_j$,则

$$\begin{aligned}u_iv_i^+w_i\cap F&=u_{i_j}v_{i_j}^+w_{i_j}\cap(F_1\cup F_2\cup\cdots\cup F_k)\\&=(u_{i_j}v_{i_j}^+w_{i_j}\cap F_1)\cup(u_{i_j}v_{i_j}^+w_{i_j}\cap F_2)\cup\cdots\cup\\&\quad(u_{i_j}v_{i_j}^+w_{i_j}\cap F_k)\\&\subseteq\bigcup_{j'\neq j,j'=1}^{k}(L_j\cap F_{j'})\cup(u_{i_j}v_{i_j}^+w_{i_j}\cap F_j)\\&\subseteq\bigcup_{j'\neq j,j'=1}^{k}(L_j\cap L_{j'})\cup(u_{i_j}v_{i_j}^+w_{i_j}\cap F_j)\\&=\varnothing.\end{aligned}$$

对 I 中两个不同的指标 i,i',存在 $j,j'\in\{1,2,\cdots,k\}$,使得 $i=i_j$,$i'=i_{j'}$,如果 $j=j'$,则 $u_iv_i^+w_i$ 与 $u_{i'}v_{i'}^+w_{i'}$ 在同一 L_j 中,故 $u_iv_i^+w_i\cap u_{i'}v_{i'}^+w_{i'}=\varnothing$. 如果 $j\neq j'$,则 $L_j\cap L_{j'}=\varnothing$,由 $u_iv_i^+w_i\subseteq L_j$,$u_{i'}v_{i'}^+w_{i'}\subseteq L_{j'}$ 知,$u_iv_i^+w_i\cap u_{i'}v_{i'}^+w_{i'}=\varnothing$.

由上可知,$L=(\bigcup_{i\in I}u_iv_i^+w_i)\cup F$ 是 L 的正规分支的不交并分解,因此 $L=L_1\cup L_2\cup\cdots\cup L_k$ 是正规分支可分解的. 证毕.

推论 11[7](正规语言的 Shyr-Yu 猜想). 正规语言是正规分支可分解的.

证明. 设 L 是正规语言,则当 $L=\varnothing$ 时,L 自然是正规分支可分解的. 如果 $L\neq\varnothing$,则由引理 7 知,存在两两不交的正规语言 $L_1,L_2,\cdots,L_k$ 使得

$$L=L_1\cup L_2\cup\cdots\cup L_k,$$

且对于每个 L_i,存在 A^* 的正规右酉子幺半群 U_{i_1},U_{i_2},$\cdots$,$U_{i_{s_i}}$ 及字母 $a_{i_1},a_{i_2},\cdots,a_{i_{s_i-1}}$ 使得

$$L_i=U_{i_1}a_{i_1}U_{i_2}a_{i_2}\cdots U_{i_{s_i-1}}a_{i_{s_i-1}}U_{i_{s_i}},$$

且对每个 $t=1,2,\cdots,s_i-1$ 都有

$$U_{i_1}a_{i_1}U_{i_2}a_{i_2}\cdots U_{i_t}a_{i_t}$$

是前缀语言. 由 U_{i_j} 是右酉幺半群,从而 U_{i_j} 是自由幺半群,根据定理 5 知 U_{i_j} 是正规分支可分解的;再由对于每个 $t=1,2,\cdots,s_i-1$ 有 $U_{i_1}a_{i_1}U_{i_2}a_{i_2}\cdots U_{i_t}a_{i_t}$ 是前缀语言,根据推论 10 知 L_i 是正规分支可分解的. 最后,由 $L=L_1\cup L_2\cup\cdots\cup L_k$ 是不交并,根据定理 7 得 L 是正规分支可分解的. 证毕.

4 结 语

本文利用半群方法探讨了正规语言及稠密语言的代数结构. 基本思想是建立语言与半群的联系,在此基础上可以利用半群的结论或通过研究半群揭示语言的性质. 由此可见,建立语言与代数对象(如半群、半环等)之间的联系是一项十分有意义的工作,这样可以借助于代数的结论和方法探讨语言的结构.

参 考 文 献

1 Hopcroft J. E., Ullman J. D.. Formal Languages and Their Relations to Automata. Reading, Mass: Addison-Wesley, 1969

2 Shyr H. J., Yu S. S.. Midst-Languages. Scoochow Journal of Mathematics, 1998, 24(2): 113～130

3 Berstel J., Perrin D.. Theory of Codes. New York: Academic Press, 1985

4 Howie J. M.. Fundamentals of Semigroup Theory. New York: Oxford University Press, 1995

5 Eilenberg S.. Automata, Languages and Machines. vol A. New York: Academic Press, 1974

6 Lallement G.. Semigroups and Combinatorial Applications. New York: John Wiley & Sons, 1979

7 Shyr H. J., Yu S. S.. Regular component splittable languages. Acta Mathematica Hungarica, 1998, 78(3): 251～265

8 Shyr H. J.. Free Monoids and Languages. Taiwan: Hon Min Book Company, 1991

9 Liu Y. J.. An application of Rees theorem to dense regular languages. Journal of Lanzhou University, 2002, 38(1): 15～17

10 Liu Y. J.. Regular component decomposition of regular languages. Theoretical Computer Science, 2003, 299(3): 743～749

11 Liu Y. J., Xu Z. B.. Monoid algorithms and semigroup properties related to dense regular languages. In: Proceedings of the International Conference on Algebra and Its Applications, Bangkok, 2002, 203～214

12 Zhang S. M.. Some equivalent statements of the S. Y's conjecture about dense regular languages. Journal of Shanxi Teacher's University, 2000, 14(4): 5～8

LIU Yao-Jun, born in 1963, Ph. D., professor. His research interests include formal languages and automata.

XU Zong-Ben, born in 1955, Ph. D., professor, Ph. D. supervisor. His research interests include nonlinear functional analysis, mathematical foundation of information technology, and computational intelligence.

Background

Semigroup method and Combinatorics method are two important techniques in the study of words and languages. For the subtle construction and islanding handle of word, the proofs via combinatorics method are usually long (see Hopcroft and Ullman[1]. Shyr[8]). The semigroup method built the connections between languages and semigroups, and thus we can investigate the problems of words and languages in the framework of a binary operation system in a local entire view or by using the properties of semigroups or by exploring related semigroups. So the proofs via semigroup method are usually short (see Berstel and Perrin[3], Eilenberg[5], Howie[4], Lallement[6]). Syntactic monoids of regular languages are finite semigroups. The UP decomposition of regular languages established a connection between regular languages and the languages which are prefix codes and right unitary monoids. The class of free monoids is larger than the class of right unitary monoids. The class of periodic monoids is larger than the class of finite monoids. Authors have developed some properties of languages by studying the free monoids and periodic monoids (see Liu[9～11], Zhang[12]). Shyr and Yu investigated the regular component splittable languages and dense regular languages by using the combinatorics method, and they conjectured that regular languages are regular component splittable and dense regular languages contains imprimitive words(see[2,7]). This paper presents the semigroup method and explore languages by using semigroup method. And above two conjectures are proved.