第6卷第7期
1995年7月

软件学报
JOURNAL OF SOFTWARE

Vol. 6, No. 7
Jul. 1995

正则语言的特征性质*

叶瑞芬　　沈百英

(华东理工大学计算机系,上海 200237)

摘要　通常的关于正则语言的泵引理实际上是刻划了某语言为正则语言的一个必要条件.本文通过修改这个必要条件,得到了关于正则语言的充分必要条件,又泵引理或广义泵引理所叙述的结果作为推论而得到.另外,我们叙述了几个强弱不同的充要条件,以便供不同的使用(作为必要条件来使用,应该是越强越好,但作为充分条件来使用,又应该是越弱越好).

关键词　形式语言,正则语言,泵引理,广义泵引理.

本文按文献[1](第八章)使用有关的术语与记号,按Chomsky对形式语言的分类方法,正则语言是恰好由(确定型)有穷自动机所接受的语言.关于正则语言的泵引理常见的是如下表述:设L是由具有n个状态的确定型有穷自动机所接受的正则语言,若x是其长度$|x|$(即x中所出现的字母的个数)大于或等于n的L中某个字,则可把x表示为uvw形,其中v不为空字,且对任意$i=0,1,2,\cdots,uv^{[i]}w$在L中(这里$v^{[i]}$表示i个相同字v的并置,而$v^{[0]}$为空字).实质上这里叙述了某语言L为正则语言的一个必要条件.我们现把这个必要条件作一些补充与修改,使得它成为L是正则语言的一个充分必要条件,而由泵引理所确定的事实作为推论而得到.

本文的主要结果是下列特征定理.

定理1. 设L为字母表Σ上的某语言,L为正则语言的充分必要条件是有正整数k,使得对Σ中所有字x,若$|x|\geqslant k$,则有Σ中满足下列性质的唯一的字u,v,w_1,w,使得$x=uvw=uvw_1y(w=w_1y)$,$|v|\geqslant1$,$|uvw_1|=k$,且对每个整数$i\geqslant0$,$uv^{[i]}w(=uv^{[i]}w_1y)\in L$当且仅当$x\in L$.

证明:先证必要性.设L为正则语言.从而有确定的有限自动机M,使得M所接受的语言即为L.取k为M的不同状态的个数(其中令q_1为起始状态).这时对Σ中任意字x,若$|x|\geqslant k$,则x可表为x_1y,使得$|x_1|=k$,且M沿着x_1从左向右运行时必至少经过两个相同的状态(M在x_1上运行时要经过k个状态,加上开始进入x_1的初始状态共$(k+1)$个状态).取这种状态的最左边的一个,并记为q.于是有Σ中唯一的字u,v,w_1,使得

$x_1=uvw_1$

* 本文1993-06-23收到,1993-10-05定稿

作者叶瑞芬,女,1938年生,副教授,主要研究领域为计算机软件,计算机理论.沈百英,1936年生,教授,主要研究领域为数理逻辑,计算机理论.

本文通讯联系人:叶瑞芬,上海200237,华东理工大学计算机系

且(令 δ^* 为 M 的转换函数的扩充)

$$\delta^*(q_1,u)=q,$$

$$\delta^*(q,v)=q,$$

$$\delta^*(q,w_1y)=\delta^*(q_1,x_1y)=\delta^*(q_1,x)=\delta^*(q_1,uw_1y).$$

又对每个 $i\geqslant 1$，$\delta^*(q,v^{[i]})=q$，故 $\delta^*(q_1,uv^{[i]}w_1y)=\delta^*(q_1,x)$

也即对每个 $i\geqslant 0$，$\delta^*(q_1,uv^{[i]}w_1\,y)=\delta^*(q_1,x)$，

且因 q 经过 2 次，故 $|v|\geqslant 1$.

再令　$w=w_1y$，

于是对每个 $i\geqslant 0$，$\delta^*(q_1,uv^{[i]}w)=\delta^*(q_1,x)$，

即得 $uv^{[i]}w\in L$ 当且仅当 $x\in L$.

再证充分性. 设 L 为字母表 Σ 上某语言，并设有正整数 k，使得定理 1 中条件满足. 为了证明 Σ 是正则语言，必须构造确定的有限自动机 M，使得由 M 接受的语言 $L(M)$ 恰好为 L 本身. 为此

令　$S=\{x\in\Sigma^*\mid |x|\leqslant k-1\}$.

取 M 的字母表为 L 的字母表 Σ.

M 的状态集 $Q=\{q_x\mid x\in S\}$.

M 的接受状态集 $F=\{q_x\mid x\in L\}$.

取 q_ε 为起始状态(其中 ε 为空字，显然 $\varepsilon\in S$). 最后定义 M 的转换函数：对每个 $a\in\Sigma$，$q_x\in Q$，若 $|x|\leqslant k-2$，则令 $\delta(q_x,a)=q_y$，其中 $y=xa$. 若 $|x|=k-1$，则因 $|xa|=k$，按已知条件有唯一的字 u,v,w(这时 $w_1=w$，y 为空字)，使得 $xa=uvw$，且 $|v|\geqslant 1$(从而 $|uw|\leqslant k-1$)，且 $xa\in L$ 当且仅当 $uw\in L$. 于是令 $\delta(q_x,a)=q_{uw}$.

到此 δ 定义完毕，且 M 也定义完毕. 令 δ^* 是 δ 的扩充函数(即 δ^* 是 $Q\times\Sigma^*$ 上的函数).

现在证明 $L(M)=L$. 即对每个 $x\in\Sigma^*$，$x\in L(M)$ 当且仅当 $x\in L$. 为此分 $|x|\leqslant k-1$ 与 $|x|\geqslant k$ 两种情况.

① 若 $|x|\leqslant k-1$，则显然 $\delta^*(q_\varepsilon,x)=q_x$，故 $x\in L$，当且仅当 $q_x\in F$，当且仅当 $x\in L(M)$.

② 若 $|x|\geqslant k$，则由已知条件有唯一的 u_1,v_1,w_1,y_1，使得 $x=u_1v_1w_1y_1$，$|u_1v_1w_1|=k$，$|v_1|\geqslant 1$，且 $x\in L$ 当且仅当 $u_1w_1y_1\in L$. 这时 $|u_1w_1y_1|<|x|$(因 $|v_1|\geqslant 1$). 若 $|u_1w_1y_1|\geqslant k$，则又有唯一的 u_2,v_2,w_2,y_2，使得 $u_1w_1y_1=u_2v_2w_2y_2$，$|u_2v_2w_2|=k$，且 $u_1w_1y_1\in L$ 当且仅当 $u_2w_2y_2\in L$.

又 $|u_2w_2y_2|<|u_1w_1y_1|<|x|$. 由于 $|x|$ 为某固定的数，故必有 n，使得 $|u_nw_ny_n|\leqslant k-1$，$|u_nv_nw_n|=k$，且对每个 $1\leqslant i\leqslant n-1$，$|u_iw_iy_i|\geqslant k$，$|u_iv_iw_i|=k$.

又 $x\in L$ 当且仅当 $u_iw_iy_i\in L$，当且仅当 $u_nw_ny_n\in L$. 于是 $\delta^*(q_\varepsilon,x)=\delta^*(q_\varepsilon,u_1v_1w_1y_1)=\delta^*(q_{u_1w_1},y_1)=\delta^*(q_{u_2w_2},y_2)=\cdots=\delta^*(q_{u_nw_n},y_n)=q_{u_nw_ny_n}$.

所以 $x\in L$ 当且仅当 $u_nw_ny_n\in L$，当且仅当 $q_{u_nw_ny_n}\in F$，当且仅当 $x\in L(M)$. 定理证毕.

推论 1(正则语言的泵引理). 设 L 为正则语言，n 为接受 L 的某确定有限自动机的状态个数，则对 L 中每个字 x，当 $|x|\geqslant n$ 时必有字 u,v,w，使得 $x=uvw$，且对每个 $i=0,1,2,\cdots$，$uv^{[i]}w\in L$.

证明：设 L 为正则语言，n 为接受 L 的某确定有限自动机的状态个数. 由定理 1 的必要

性证明中知对每个满足 $|x|\geqslant n$ 的字 x，有字 u,v,w，使得对每个 $i\geqslant 0$，$x\in L$ 当且仅当 $uv^{[i]}w\in L$，于是当 x 为 L 中字时，对每个 $i\geqslant 0$，$uv^{[i]}w$ 也为 L 中字.

我们知道，对某个结论而言，使其成立的必要条件越强越好，而使其成立的充分条件应越弱越好. 因此我们再叙述语言 L 为正则语言的几个类似的特征性质.

定理 2. 字母表 Σ 上某语言 L 为正则语言的充分必要条件是有正整数 k，使得对每个 $x\in\Sigma^*$，若 $|x|\geqslant k$，则有字 u,v,w，使得 $x=uvw$，$|v|\geqslant 1$，且对每个 $y\in\Sigma^*$，每个整数 i，$uv^{[i]}wy\in L$ 当且仅当 $xy\in L$.

证明：必要性可仿定理 1 证明. 这时由于对每个 $i\geqslant 0$，$\delta^*(q_1,x)=\delta^*(q_1,uv^{[i]}w)$，故对任意 $y\in\Sigma^*$，$\delta^*(q_1,xy)=\delta^*(\delta^*(q_1,x),y)$（其中相同的状态 q 不强调取第一个，从而不强调取唯一的 u,v,w）$=\delta^*(\delta^*(q_1,uv^{[i]}w),y)=\delta^*(q_1,uv^{[i]}wy)$. 从而 $uv^{[i]}wy\in L$ 当且仅当 $xy\in L$.

关于充分性可证明 L 有生成集 S，从而由 Myhill－Nerode 定理，即一个语言 L 为正则语言的充要条件是 L 有生成集 S，得 L 为正则语言.

令 $S=\{x\in\Sigma^*\mid |x|\leqslant k-1\}$. 显然 S 是有限集. 为了证明 S 是 L 的生成集，就要证明对每个 $x\in\Sigma^*$，有 $y\in S$，使得对每个字 $w\in\Sigma^*$，$xw\in L$ 当且仅当 $yw\in L$.

事实上，① 若 $|x|\leqslant k-1$，则因 $x\in S$，故就取 x 本身作为要取的 S 中的元素.

② 若 $|x|\geqslant k$，则由已知条件，有 u_1,v_1,w_1，使得 $x=u_1v_1w_1$，且对每个 $y\in\Sigma^*$，$xy\in L$ 当且仅当 $u_1w_1y\in L$（取 $i=0$）. 这时由于 $|v_1|\geqslant 1$，故 $|u_1w_1|<|x|$. 现在若 $|u_1w_1|\geqslant k$，则仍由已知条件有 u_2,w_2，使得对每个 $y\in\Sigma^*$，$u_1w_1y\in L$ 当且仅当 $u_2w_2y\in L$，且 $|u_2w_2|<|u_1w_1|<|x|$. 由于 $|x|$ 是个固定的数，故必有 n，使得 $|u_nw_n|\leqslant k-1$，且对每个 j，$1\leqslant j\leqslant n$，对每个 $y\in\Sigma^*$，$xy\in L$ 当且仅当 $u_jw_jy\in L$.

最后得对每个 $y\in\Sigma^*$，$xy\in L$ 当且仅当 $u_nw_ny\in L$. 由于 $u_nw_n\in S$（因 $|u_nw_n|\leqslant k-1$），故这个 u_nw_n 就是对应于 x 的在 S 中要取的元素. 按生成集的定义[1]，S 为 L 的生成集.

定理 3. 字母表 Σ 上某语言 L 为正则语言的充分必要条件是有正整数 k，使得对每个 $x\in\Sigma^*$，若 $|x|\geqslant k$，则有字 u,v,w，使得 $x=uvw$，$|v|\geqslant 1$，且对每个 $y,z\in\Sigma^*$，每个整数 i，$yuv^{[i]}wz\in L$ 当且仅当 $yxz\in L$.

证明：必要性可仿定理 1 证明. 这时由于对每个 $i\geqslant 0$，每个 $y\in\Sigma^*$，（设 $\delta^*(q_1,y)=q_m$）$\delta^*(q_1,yx)=\delta^*(\delta^*(q_1,y),x)=\delta^*(q_m,uv^{[i]}w)$，故对每个 $z\in\Sigma^*$，$\delta^*(q_1,yxz)=\delta^*(q_m,xz)=\delta^*(\delta^*(q_m,x),z)=\delta^*(\delta^*(q_m,uv^{[i]}w),z)=\delta^*(q_m,uv^{[i]}wz)$. 故 $yuv^{[i]}wz\in L$ 当且仅当 $yxz\in L$.

充分性只要取 y 为空字，即可按定理 2 得.

推论 2（正则语言的广义泵引理）. 设 L 为字母表 Σ 上的正则语言，n 为接受 L 的某确定有限自动机的状态个数. x,y,z 为 Σ 中任意字，且 $|x|\geqslant n$，$yxz\in L$. 则有 Σ 中字 u,v,w，使得 $x=uvw$，$|v|\geqslant 1$，且对每个整数 i，$yuv^{[i]}wz\in L$.

证明：按定理 3，若 L 为正则语言，则由必要性的证明过程得知正整数 k 即为 n，且由于对每个 $i\geqslant 0$，$yxz\in L$ 当且仅当 $yuv^{[i]}wz\in L$，故当 yxz 在 L 中时，$yuv^{[i]}wz$ 也在 L 中.

在实用中，正则语言的泵引理常用来证明某语言 L 不是正则语言. 为此要反设 L 是正则语言，再找出矛盾. 但有时使用通常泵引理会找不出矛盾. 这时使用广义泵引理往往

会奏效.

注:在定理 2 和定理 3 中条件$|x| \geqslant k$,都可换以$|x|=k$.

不难看出,定理 1 到定理 3 中的 3 个充要条件是越来越强.也即定理 1 中的充要条件最弱,定理 3 中的充要条件最强.这是因为在定理 3 的条件中取 y 为空字即得定理 2 中的条件,而在定理 2 的条件中取 y 为空字即得定理 1 中的条件.

参考文献

1 Davis M D, Weyuker E J. Computability, complexity languages. New York, London, Academic Press, 1983. 149—170.

THE CHARACTERIZATION FOR REGULAR LANGUAGES

Ye Ruifen Shen Baiying

(*Department of Computer Science, East China University of Science and Technology, Shanghai* 200237)

Abstract Common pumping lemma for regular languages characterizes the necessary condition that a language is regular. This paper gives several necessary and sufficient conditions and common pumping lemma and generalized pumping lemma are obtained as their consequence.

Key words Formal languages, regular languages, pumping lemma, generalized pumping lemma.