INFORMATION AND CONTROL 19, 272–274 (1971)

A Note on a Decomposition Theorem for Simple Deterministic Languages

AKIRA FUSAOKA

Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto-shi, Japan

A procedure to resolve simple deterministic languages into the concatenation of other simple deterministic languages is presented.

The simple deterministic grammar (s-grammar) is the standard form grammar in which the handles of the Z-rules are distinct for each nonterminal symbol Z. The language generated by an s-grammar is called the simple deterministic language (s-language). It is known that the s-languages have the the prefix property and that their equivalence problem is solvable. (Korenjak and Hopcroft, 1966).

Using these facts, we present a procedure to resolve s-languages into the concatenation of the prime s-languages that can be resolved no more.

LEMMA. *Let A be a prime s-language and let B and C be s-languages. Let $G = (V, \Sigma, P, \sigma)$ be an s-grammar such that $L(G) = B$. $CA = B$ if and only if for every $\alpha \in B$ there exist $\beta \in V^*$ and $W \in (V - \Sigma)$ satisfying $\sigma \overset{*}{\Rightarrow} \beta W \overset{*}{\Rightarrow} \alpha$ and $L(W) = A$.*

Proof. Let $\gamma \in C$ be a prefix of $\alpha \in B$. The pair (σ, γ) uniquely determines $W \in (V - \Sigma)^*$ such that $\sigma \overset{*}{\Rightarrow} \gamma W \overset{*}{\Rightarrow} \alpha$. For such a W, $L(W) = A$ from $C \backslash C = \epsilon$. Clearly, $W \in (V - \Sigma)$, since A is a prime s-language.

To prove the converse, we construct an s-grammar G' such that $L(G') = C$. Let

$$P = \{Z_i \to a_i X_i Y_i\,,\ Z_j \to a_j X_j\,,\ Z_k \to a_k\}$$

and

$$F = \{W \mid W \in (V - \Sigma), L(W) = A\} = \{W_1\,, W_2\,, \ldots, W_q\}.$$

We introduce a set of new symbols $\bar{V} = \{\bar{Z} \mid Z \in (V - \Sigma) - F\}$ and the sets of rewriting rules

$$P' = \{\bar{Z}_i \to a_i X_i \bar{Y}_i \,,\ \bar{Z}_j \to a_j \bar{X}_j \,,\ \bar{Z}_k \to a_k\},$$

$$P'' = \{\bar{Z}_i \to a_i X_i Y_i \,,\ \bar{Z}_j \to a_j X_j \mid Y_i \,,\ X_j \in F\}$$

and

$$P_m = \{\bar{Z}_i \to a_i X_i \,,\ \bar{Z}_j \to a_j \mid Z_i \to a_i X_i W_m \,,\ Z_j \to a_j W_m \in P\}$$

for $1 \leqslant m \leqslant q$. For a right linear grammar $\bar{G} = (\bar{V}, V, P' \cup P'', \bar{\sigma})$ and a regular set

$$R = \{z \in V^* \mid \sigma \overset{*}{\underset{\bar{G}}{\Rightarrow}} z\},$$

$R \subseteq V^*F$ if and only if for every $\alpha \in B$ there exist $W_m \in F$ and $\beta \in V^*$ such that $\sigma \overset{*}{\Rightarrow} \beta W_m \overset{*}{\Rightarrow} \alpha$. If $R \subseteq V^*F$, then $G' = (V \cup \bar{V}, \Sigma, \cup_{m=1}^{q} P_m \cup P' \cup P, \bar{\sigma})$ is an s-grammar such that $L(G') = B$.

Since $R \subseteq V^*F$ is a containment problem for regular sets and F is constructed by using the solvability of the equivalence problem, there is an effective procedure to decide whether an s-language satisfies the condition of the above lemma.

Theorem. *For a given s-grammar G, there exists an effective procedure to find the prime s-languages $X_1, X_2, \ldots, X_n$ satisfying $X_1 X_2 \cdots X_n = L(G)$, and $X_1, X_2, \ldots, X_n$ are uniquely determined.*

Proof. Let $a_1 a_2 \cdots a_s$ be one of the shortest elements of $L(G)$ and let

$$\sigma \Rightarrow \gamma_1 Z_1 \Rightarrow \gamma_2 Z_2 \Rightarrow \cdots \Rightarrow \gamma_p Z_p \Rightarrow \gamma_{p+1} a_s \overset{*}{\Rightarrow} a_1 a_2 \cdots a_s$$

be the rightmost derivation of $a_1 a_2 \cdots a_s$. Find the maximum $k \leqslant p$ such that $YL(Z_k) = L(G)$ for some s-language Y using the procedure of the above lemma. Clearly, $L(Z_k)$ is a prime s-language. If such a Y does not exist, then $L(G)$ is a prime s-language. Repeating the procedure, we can find $X_1, X_2, \ldots, X_n$. The uniqueness of this decomposition is clear.

We state a corollary proved by using the concatenative decomposition and the prefix property of s-languages.

Corollary. *There are the effective procedures to decide whether for given s-grammars G_1 and G_2, there exists the s-language X satisfying the equations*

$L(G_1)X = L(G_2)$, $XL(G_1) = L(G_2)$, and $X^n = L(G_1)$. The solutions of such equations are uniquely determined.

RECEIVED: September 30, 1970; REVISED: April 6, 1971

REFERENCE

KORENJAK, A. J. AND HOPCROFT, J. E. (1966), Simple deterministic languages, *in* "IEEE Conference Record of Seventh Annual Symposium on Switching and Automata Theory," IEEE Pub. No 16–C–40, pp. 36–46.

Printed in Belgium