A Language Equation and Its Applications

京都産業大学・理学部 伊藤 正美
Masami Ito
Department of Mathematics
Kyoto Sangyo University
Kyoto 603-8555, Japan
Email: ito@ksuvx0.kyoto-su.ac.jp

Let $u, v \in X^*$ be words over an alphabet $X$. Then the set $\{u_1v_1u_2v_2\ldots u_nv_n \mid u = u_1u_2\ldots u_n, v = v_1v_2\ldots v_n, u_1, u_2, v_1, v_2, \ldots, u_n, v_n \in X^*, n \geq 1\}$ is called the shuffle product of $u$ and $v$, and denoted by $u \circ v$. For languages $A, B \subseteq X^*$, the set $A \circ B = \bigcup_{u\in A, v\in B} u \circ v$ is called the shuffle product of $A$ and $B$. In this paper, we consider the following problem: Let $A, B \subseteq X^*$ be regular languages. Then can we obtain a solution $C \subseteq X^*$ of the language equation $A = B \circ C$? Obviously, this problem is equivalent to the shuffle decomposition problem for regular languages. Regarding definitions and notations concerning formal languages and automata, not defined in this paper, refer, for instance, to [1].

Now let $A = (S, X, \delta, s_0, F)$ be a finite automaton with $T(A) = A$ and let $B = (T, X, \gamma, t_0, G)$ be a finite automaton with $T(B) = B$. We will look for a regular language $C$ over $X$ such that $A = B \circ C$. By $\overline{X}$, we denote the language $\{\overline{a} \mid a \in X\}$ with $X \cap \overline{X} = \emptyset$. Let $B = (T, X \cup \overline{X} \cup \{\#\}, \overline{\gamma}, t_0, G)$ where $\overline{\gamma}$ is defined as follows:

For $t \in T$ and $a \in X$, $\overline{\gamma}(t, a) = t$, $\overline{\gamma}(t, \overline{a}) = \gamma(t, a)$. Moreover, $\overline{\gamma}(t, \#) = t$ if $t \in G$.

Then the following can be easily shown.

**Fact 1** Let $a_1a_2\ldots a_n \in X^*$ where $a_i \in X, i = 1, 2, \ldots, n$. Then $a_1a_2\ldots a_n \in T(B)$ if and only if $u_1a_1u_2a_2\ldots u_na_nu_{n+1}\# \in T(B)$ where $u_1, u_2, \ldots, u_n \in X^*$.

Let $A_1 = (\overline{S}, X \cup \overline{X} \cup \{\#\}, \overline{\delta}, s_0, \{\alpha, \omega\})$ and let $A_2 = (\overline{S}, X \cup \overline{X} \cup \{\#\}, \overline{\delta}, s_0, \{\alpha\})$ where $\overline{S} = (\cup_{a\in X \cup \{\epsilon\}} S^{(a)}) \cup \{\alpha, \omega\}$. Here $S^{(\epsilon)}$ is regarded as $S$ where $\epsilon$ is the empty word. For $s \in S, t \in S \setminus F, t' \in F, a \in X \cup \{\epsilon\}, b \in X$ and $\{\#\}$, $\overline{\delta}$ is defined as follows:
\[ \delta'(s^{(a)}, b) = \delta(s, b)^{(a)}, \delta'(s^{(a)}, \#) = \delta(s, \#)^{(a)}, \delta'(t^{(a)}, \#) = \{\omega\}. \]

We consider the following two automata:

\[ C_1 = (\overline{S} \times T, X \cup \overline{X} \cup \{\#\}, \delta \times \gamma, (s_0, t_0), \{\alpha, \omega\} \times G) \]
\[ C_2 = (\overline{S} \times T, X \cup \overline{X} \cup \{\#\}, \delta \times \gamma, (s_0, t_0), \{\alpha\} \times G) \]

where \( \delta \times \gamma((s, t), a) = (\delta(s, a), \gamma(t, a)) \)

for \((s, t) \in \overline{S} \times T \) and \( a \in X \).

Now consider the following homomorphism \( \rho \) of \( (X \cup \overline{X} \cup \{\#\})^* \) into \( X^* \):

\[ \rho(a) = a \text{ for } a \in X, \rho(\overline{a}) = \epsilon \text{ for } a \in X \text{ and } \rho(\#) = \epsilon. \]

**Lemma 1** Automata accepting the languages \( \rho(T(C_1)) \) and \( \rho(T(C_2)) \) can be effectively constructed.

**Proof** Let \( i = 1, 2 \). From \( C_i \), we can construct a regular grammar \( \mathcal{G}_i \) such that \( L(\mathcal{G}_i) = T(C_i) \) with the production rules of the form \( A \rightarrow aB \) (\( A, B \) are variables and \( a \in X \cup \overline{X} \cup \{\#\} \)). Replacing every rule of the form \( A \rightarrow aB \) in \( \mathcal{G}_i \) by \( A \rightarrow \rho(a)B \), we can obtain a new grammar \( \mathcal{G'}_i \). Then it is clear that \( \rho(L(C_i)) = L(G'_i) \). Using this grammar \( \mathcal{G'}_i \), we can construct an automaton \( D_i \) such that \( T(D_i) = T(G'_i) \) i.e. \( \rho(T(C_i)) = T(D_i) \). Notice that all the above procedures are effectively done. This completes the proof of the lemma.

Let \( B, C \subseteq X^* \). By \( B \circ C \) we denote the shuffle product of \( B \) and \( C \), i.e. \( \{u_1v_1u_2v_2...u_nv_n \mid u = u_1u_2...u_n \in B, v = v_1v_2...v_n \in A\} \).

**Proposition 1** Let \( u \in X^* \). Then \( \{u\} \circ B \subseteq A \) if and only if \( u \in \rho(T(C_1)) \setminus \rho(T(C_2)) \).

**Proof** (\( \Rightarrow \)) Let \( u = u_1u_2...u_nu_{n+1} \in X^* \) and let \( a_1a_2...a_n \in B \) where \( u_1, u_2, ..., u_n, u_{n+1} \in X^* \) and \( a_1, a_2, ..., a_n \in X \). Then \( \delta \times \gamma((s_0, t_0), u_1\overline{a}_1u_2\overline{a}_2...u_n\overline{a}_n\overline{a}_{n+1}\#)), \gamma(t_0, u_1\overline{a}_1u_2\overline{a}_2...u_n\overline{a}_n\overline{a}_{n+1}\#)) = (\delta(s_0, u_1a_1u_2a_2...u_na_n\overline{a}_{n+1}\#), \gamma(t_0, a_1a_2...a_n\overline{a}_n\overline{a}_{n+1}\#)) = (\omega, \gamma(t_0, a_1a_2...a_n)) \in \{\omega\} \times G \). Therefore, \( u_1\overline{a}_1u_2\overline{a}_2...u_n\overline{a}_n\overline{a}_{n+1}\# \in \rho(T(C_1)) \setminus \rho(T(C_2)) \).

(\( \Leftarrow \)) Suppose that \( u \circ B \subseteq A \) does not hold though \( u \in \rho(T(C_1)) \setminus \rho(T(C_2)) \). Then there exist \( u = u_1u_2...u_nu_{n+1} \in X^* \) and \( a_1a_2...a_n \in B \) such that \( u_1a_1u_2a_2...u_na_nu_{n+1} \notin A \). Hence \( \gamma(t_0, u_1\overline{a}_1u_2\overline{a}_2...u_n\overline{a}_n\overline{a}_{n+1}\#)) = \gamma(t_0, a_1a_2...a_n) \). On the other hand, since \( u_1a_1u_2a_2...u_nu_{n+1} \notin A \), we have \( \delta(s_0, u_1\overline{a}_1u_2\overline{a}_2...u_n\overline{a}_n\overline{a}_{n+1}\#)) = \delta(s_0, u_1a_1u_2a_2...u_na_n\overline{a}_{n+1}\#)) \)

\[ \mathcal{T}(C_2) \]
of the finite then have nontrivial by cardinality nontrivial length obtained after. Otherwise, Then, a Lemma If is procedures class the Consequently, exist said we exist the Thus is decidable language the there exist languages If the we can denote (2) acontradiction. is regular languages 1). that above that exist (1) regular languages nontrivial the integer. By above, other hand, it is obvious that afinite then a regular language is regular language the one regular language. Hence, the regular language A = Bo otherwise. If the output is ”YES”, then choose another element in $I(n, X)$ as B and continue the procedures (1) - (3). (5) Since $I(n, X)$ is a finite set, the above process terminates after a finite-step trial. Once one gets the output ”YES”, then there exist nontrivial regular languages $B \in I(n, X)$ and $C \subseteq X^*$ such that $A = B \circ C$. Otherwise, there are no such languages.

Let $n$ be a positive integer. By $F(n, X)$, we denote the class of finite languages \{L \subseteq X^* \mid \max\{|u| \mid u \in L\} \leq n\} where |u| is the length of u. Then the following result by C. Câmpeanu et al. ([2]) can be obtained as a corollary of the above theorem.

**Corollary** For a given positive integer $n$ and a regular language $A \subseteq X^*$, the problem whether $A = B \circ C$ for a nontrivial language $B \in F(n, X)$ and a nontrivial regular language $C \subseteq X^*$ is decidable.
Proof Obvious from the fact that $F(n, X) \subseteq I(|X|^{n+1}, X)$.

References
