THE MEMBERSHIP PROBLEMS OF SEMIGROUP

Dr. Pushp Kumar Roy

Former Research Scholar

Department of Mathematics,

T.M. Bhagalpur University, Bhagalpur.

Abstract

Let $\varepsilon = \{E_1, ..., E_k\}$ be a set of regular languages over a finite alphabet Σ. Consider morphism $\varphi : \Delta^+ \to (S,.)$ where Δ^+ is the semigroup over a finite set Δ and $(\delta,.) = (\varepsilon)$ is the finitely generated semigroup with ε as the set of generators and language concatenation as a product. We prove that the membership problem of the semigroup δ, the set $[u] = \{v \in \Delta^+ \mid \varphi(v) = \varphi(u)\}$, is a regular language over Δ, while the set Ker $(\varphi) = \{(u,\upsilon) | u,\upsilon \in \Delta^+ \varphi(u) = \varphi(\upsilon)\}$ need not to be regular. But every semigroup of regular languages is automatic.

Keywords : *Finitely generated semigroup, Regular Language, Automatic semigroup etc.*

Introduction

Regular languages play an important role in both theoretical and practical aspects of computer science. Regular languages are closed under basic language operation. One language may be represented in terms of other regular languages. The problem of representing a given regular language as a concatenation of other languages is a special case of such a representation. If the set of factors is fixed then language factorization may be considered as the membership problem for a finitely generated semigroup.

A regular language $R \subseteq \Sigma^*$ belongs to the semigroup $\varepsilon = \{E_1, ..., E_k\}$ if and only if there exists a sequence $i_1, i_2, \ldots i_n$ of integers such that $1 \le i_p \le k$ where $p = 1 \ldots n$

and $R = E_{i_1} E_{i_2} ... E_{i_n}$.

Basic ideas

An alphabet is a finite non-empty set of symbols. A finite sequence of symbols from an alphabet $\sum$ is called a word over $\sum$. the empty word denoted by ε. A word $u = a_1 a_2 ... a_k$ is called a scattered subword of a word υ (denoted as $u \subseteq \upsilon$) if there exist $w_1, ... w_{k+1} \in \sum^*$ such that $\upsilon = w_1 a_1 w_2 a_2 ... w_k a_k w_{k+1}$.

Any set of words is called a language over $\sum$. $\sum^*$ denotes the set of all finite words (including the empty word), $\sum^+$ denotes the set of all nonempty words over $\sum$, $\varnothing$ is the empty language (containing on words), and $2^{\sum^*}$ is the set of all languages over $\sum$. the union of languages L_1 and L_2 is denoted by L_1+L_2, concatenation by $L_1 L_2$, and iteration by L^*.

A (non-deterministic) automaton is a tuple $A = \langle \Delta, Q, p, q_0 F \rangle$, where Q is a finite set of states, Δ is an input alphabet, $p \subseteq Q \times \Delta \times Q$ is a set of transitions, $q_o \in Q$ is the initial state and $F \subseteq Q$ is a set of final states.

The language L is called regular if it is recognized by a finite automaton. The class of regular languages over $\sum$ is denoted by Reg $\left(\sum\right)$.

The set X of pairs of words (u, v), where u == a_1 … a_m and v = b_1 … b_m $\left(a_i, b_i \in \Delta\right)$ is called regular if there exists a finite automaton over $\Delta \times \Delta$ that recognizes X.

Let $\sum$ be an alphabet and $\varepsilon = \left(E_1, ..., E_k\right)$ be a set of regular languages over $\sum$. The concatenation of regular languages is a regular language and we write $\left(\delta, .\right) = \langle \varepsilon \rangle$ for the finitely generated semigroup, generated by ε with concatenation as a semigroup product. The homomorphism $\varphi : \Delta^+ \to \left(\mathrm{Re}\, g\left(\sum\right), .\right)$ between the free semigroup Δ^+ and the semigroup of regular languages with concatenation is called a regular language substitution. Every semigroup $\left(\delta, .\right) = \left(E_1, ..., E_k\right)$ of regular languages we can associate a language substitution $\varphi\left\{\delta_1, ..., \delta_k\right\} \to S$, defined by the rule $\delta_i \to E_i$. Conversely, every regular language substitution $\varphi:\Delta^+ \to \operatorname{Re}g(\Delta)$ generates the semigroup $S_\varphi = \langle\{\varphi(\delta) | \delta \in \Delta\}\rangle$. For a regular language $L \subseteq \Delta^+$ by $\varphi(L)$ we mean the set $\varphi(L) = \{\varphi(w) | w \in L\}$ of regular languages over $\sum$.

Definition 1. Let $\varphi:\Delta^+ \to \operatorname{Re}g(\Sigma)$ be a regular language substitution. The maximal rewriting of a regular language $R \subseteq (\Sigma^*)$ with respect to φ is the set

$$M_\varphi(R) = \{w \in \Delta^+ | \varphi(w) \subseteq R\}.$$

Theorem 2. Let $\varphi:\Delta^+ \to \operatorname{Re}g(\Sigma)$ be a regular language substitution. For any r egular language $R \subseteq \sum^*$ the maximal rewriting $M_\varphi(R)$ is a regular language over Δ.

Membership and word problems

The regularity of the membership problem for the semigroup of regular languages is a simple corollary from Theorem 2 and Higman's lemma, which may be stated as follows.

Lemma 3. In every infinite sequence $\{u_i\}_{i\geq 1}$ of words over a finite alphabet there exist indices *i* and *j*, such that $u_i \subseteq u_j$.

Let $w = \delta_{i1}...\delta_{im}$ be a word over Δ and $A \subseteq \Delta$. We shall call the language $w \Uparrow A^* = A^*\delta_{i1}A^*\delta_{i2}...A^*\delta_{im}A^*$ the shuffle extension of *w*. By *E*(*w*,*A*) denote the language $(w \Uparrow A^*) \cap M_\varphi(\varphi(w))$. Clearly, *E*(*w*,*A*) is a regular language for all $w \in \Delta^+$.

Proposition 4. Let $\varphi:\Delta^+ \to \operatorname{Re}g(\Sigma)$ be a regular language substitution, $u \in \Delta^+$, and $\Delta_0 = \{\delta \in \Delta \,|\, \varepsilon \in \varphi(\delta)\}$. For every $u \in E(u,\Delta_0)$ we have $\varphi(u) = \varphi(\upsilon)$.

Proof. *Let* $\delta \in \Delta_0$. Consider the word $\upsilon = u_1\delta u_2$, where $u_1, u_2 \in \Delta^*$ and $u = u_1u_2$. We have $\varphi(\varphi) \subseteq \varphi(\upsilon) \subseteq (u)$.

The first inclusion is due $\varepsilon \in \varphi(\delta)$ while the second one follows from the definition of the language $E(u, \Delta_0)$.

Theorem 5. Let $\varphi : \Delta^+ \rightarrow \operatorname{Re} g(\Sigma)$ be a regular language substitution and *w* be a word in Δ^+. The membership problem for the semigroup S_φ

$$[w] = \{u \in \Delta^+ | \varphi u = \varphi(w)\}$$

is a regular language over Δ.

Proof. By Theorem 2 the set $M\varphi(w) = \{u \in \Delta^+ | \varphi(u) \subseteq \varphi(w)\}$ is regular. Clearly $[w] \subseteq M_\varphi(w)$. Let [*w*] be an infinite language. We prove now that there exists a subset $F \subseteq [w]$ satisfying

$$[w] = \bigcup_{u \in F} E(u, \Delta_0)$$

Note that $u \subseteq \upsilon$ implies $E(\upsilon, \Delta_0) \subseteq E(u, \Delta_0)$, so without loss of generality we may assume that if the language F contains a word υ then it dose not contain subwords of υ. The finiteness of F follows immediately from Lemma 3, and thus [*w*] may be represented as a finite union of regular language.

Theorem 6 Let $\varphi : \Delta^+ \rightarrow \operatorname{Re} g(\Sigma)$ be a regular language substitution. The set

$$Ker(\varphi) = \{(u, \upsilon) | u, \upsilon \in \Delta^+ \varphi(u) = \varphi(\upsilon)\}$$

need not be regular.

Proof. We show that if the set $\operatorname{Ker}(\varphi)$ is regular then the equivalence problem for a rational set of regular languages, i.e. t he problem to decide whether or not two given rational sets $R_1 = (K_1, \varphi)$ and $R_2 = (K_2, \varphi)$ are equal as sets of languages over Σ, is decidable. Then we reduce the later problem to the finite substitutions equivalence problem, that is known to be undecidable.

For a given regular language $K \in \Delta^*$ by $\overline{K}$ denote the closure of K with pespect to φ:

$$\overline{K} = \varphi^{-1}(\varphi(K)) = \{u \in \Delta^+ \,|\, \exists \upsilon \in K \varphi(u) = \varphi(\upsilon)\}.$$

Let $R_1 = (K_1, \varphi)$ and $R_2 = (K_2, \varphi)$ be rational sets of regular languages. We have $R_1 = R_2$ if and only if $\overline{K_1} = \overline{K_2}$. Now, suppose that the set

Kar(φ) is regular, *i.e.* there exists a finite atuomaton M that recognizes this language. By standard direct product construction of automata M and K we construct the automaton that recognizes the language

$$\{\upsilon \in \Delta^+ \,|\, \exists u \in K (u, \upsilon) \in Ker(\varphi)\}.$$

thus is Ker (φ) is regular then so is $\overline{K}$ for every regular language $K \in \Delta^*$.

Let φ_1 and φ_2 be finite substitutions, *i.e.* homomorphisms between Δ^+ and a semigruop of finite languages. The equivalence problem of finite substitutions on a regular language L

$$\varphi_1(w) = \varphi(w) \; for\, all\, w \in L$$

is known be undecidable [6] for $L = xy^*z$ Let $\Delta = \{x_1, y_1, z_1, x_2, y_2, z_2\}, \varphi$ be a finite substitution, and rational sets $R_1 = (K_1, \varphi)$ and $R_2 = (K_2, \varphi)$ are given by languages $K_1 = x_1 y^*{}_1 z_1$ and $K_2 = x_2 y^*{}_2 z_2$. By considering the length of the langest word in the image $\varphi(w)$ we have that R_1 and R_2 are equal if and only if finite substitutions φ_1 and φ_2 (induced by φ) are equal on the language xy^*z. We have a contradiction, so the set Ker(φ) is not regular in general.

References:

[1] S. Afonnin and E. Hazova, Membership and finiteness problems for rational sets of regular languages, Proceeding of the Developments in Language Theory 2005, C. De Felice and A. Restivo, Eds., Leclure Notes in Computer Science 3572, Springer, 88—99 (2005).

[2] J. Brzozowski, K. Culik and A. Gabrielian, Classification of noncounting events, Journal of Computer and system Sciences 5, 41-53 (1971).

[3] C.M. Campbell, E.F. Robbertson, N. Ruskuc and R.M. Thomas, Automatic semigroups, Theoretical Computer Science 250 (1-2), 365-391 (2001).