数理解析研究所講究録
第 1799 巻 2012 年 57-59

2011 年度冬の LA シンポジウム [10]

Relations between language classes in terms of insertion and locality

Kaoru Fujioka *

1 Introduction

Insertion systems use only insertion operations of the form (u, x, v) and produce a string $\alpha uxv\beta$ for a given string $\alpha uv\beta$ by inserting the string x between u and v. From the definition of insertion operations, using only insertion operations, we generate only context-sensitive languages.

Using insertion systems together with some morphisms, characterizing recursively enumerable languages is obtained in [8], [6]. Furthermore, similarly to the Chomsky–Schützenberger representation theorem [1], each recursively enumerable language can be expressed using an insertion system and a Dyck language in [7], and each context-free language can be expressed using an insertion system and a star language in [5].

In [2] and [3], within the framework of the Chomsky–Schützenberger representation theorem, some characterizations and representation theorems of languages in the Chomsky hierarchy have been provided by insertion system γ, strictly locally testable language R, and morphism h such as $h(L(\gamma) \cap R)$.

The purpose of this paper is to clarify the relation between the classes of languages $h(L(\gamma) \cap R)$ using insertion systems of weight $(i, 0)$ for $i \geq 1$ and those using insertion systems of weight $(i, 1)$ for $i \geq 1$.

*Office for Strategic Research Planning, Kyushu University

2 Preliminaries

For a string $x \in V^*$ with an alphabet V, $|x|$ is the length of x. For $0 \leq k \leq |x|$, let $Pre_k(x)$ and $Suf_k(x)$ respectively denote the prefix and the suffix of x with length k. For $0 \leq k \leq |x|$, let $Int_k(x)$ be the set of intermediate substrings of x with length k.

For a positive integer k, a language L over T is *strictly k-testable* if a triplet $S_k = (A, B, C)$ exists with $A, B, C \subseteq T^k$ such that, for any w with $|w| \geq k$, w is in L iff $Pre_k(w) \in A$, $Suf_k(w) \in B$, $Int_k(w) \subseteq C$. A language L is *strictly locally testable* iff there exists an integer $k \geq 1$ such that L is strictly k-testable.

Note that, for an alphabet T, a language T^+ is a strictly 1-testable language.

Let $LOC(k)$ be the class of strictly k-testable languages. There is the following result.

Theorem 1 [4] $LOC(1) \subset LOC(2) \subset \cdots \subset LOC(k) \subset \cdots \subset REG$.

We define an *insertion system* $\gamma = (T, P, A)$, where T is an alphabet, P is a finite set of *insertion rules* of the form (u, x, v) with $u, x, v \in T^*$, and A is a finite set of strings over T called axioms.

We write $\alpha \xRightarrow{r}_\gamma \beta$ if $\alpha = \alpha_1 uv\alpha_2$ and $\beta = \alpha_1 uxv\alpha_2$ for some insertion rule $r : (u, x, v) \in P$ with $\alpha_1, \alpha_2 \in T^*$. We write $\alpha \Longrightarrow \beta$ if no confusion exists. The reflexive and transitive closure of $\Longrightarrow$ is defined as $\Longrightarrow^*$.

A language generated by γ is defined as

$$L(\gamma) = \{w \in T^* \mid s \Longrightarrow_\gamma^* w, \text{ for some } s \in A\}.$$

An insertion system $\gamma = (T, P, A)$ is said to be of *weight* (m, n) if

$$\begin{aligned} m &= max\{\ |x| \mid (u, x, v) \in P\}, \\ n &= max\{\ |u| \mid (u, x, v) \in P \text{ or } (v, x, u) \in P\}. \end{aligned}$$

For $m, n \geq 0$, let INS_m^n be the class of all languages generated by insertion systems of weight (m', n') with $m' \leq m$ and $n' \leq n$. We use $*$ instead of m or n if the parameter is not bounded.

Theorem 2 [8]

1. $INS_i^j \subseteq INS_{i'}^{j'}$ $(0 \leq i \leq i',\ 0 \leq j \leq j')$.

2. $INS_*^1 \subset CF$.

A mapping $h : V^* \to T^*$ is called *morphism* if $h(\lambda) = \lambda$ and $h(xy) = h(x)h(y)$ hold for any $x, y \in V^*$. For any a in T, if $h(a) = a$ holds, then h is an *identity morphism*.

The following results related to Chomsky-Schützenberger like characterization are obtained using insertion systems of weight $(i, 0)$ or $(i, 1)$ for $i \geq 1$ and strictly k-testable languages $(k \geq 1)$.

Theorem 3 [2]

1. $H(INS_1^0 \cap LOC(1)) \subset REG$.

2. $H(INS_1^0 \cap LOC(k)) = REG$ $(k \geq 2)$.

3. $H(INS_i^0 \cap LOC(1))$ *and* REG *are incomparable* $(i \geq 2)$.

4. $H(INS_i^0 \cap LOC(1)) \subset CF$ $(i \geq 2)$.

5. $H(INS_i^0 \cap LOC(k)) = CF$ $(i, k \geq 2)$.

Theorem 4 [3]

1. $H(INS_i^1 \cap LOC(k)) = CF$ $(i \geq 1,\ k \geq 2)$.

2. $H(INS_i^1 \cap LOC(1)) \subset CF$ $(i \geq 1)$.

In the present paper, we specifically examine the relation between language classes $H(INS_{i_0}^0 \cap LOC(k_0))$ and $H(INS_{i_1}^1 \cap LOC(k_1))$ for $i_0, k_0, i_1, k_1 \geq 1$.

3 Main Results

For context-free languages, from Theorem 3 and Theorem 4, we obtain

$$\begin{aligned} CF = &\ H(INS_{i_0}^0 \cap LOC(k_0)) \\ = &\ H(INS_{i_1}^1 \cap LOC(k_1)) \end{aligned}$$

with $i_0, k_0, k_1 \geq 2$, $i_1 \geq 1$.

We next examine the language class $H(INS_2^0 \cap LOC(1))$. From Theorem 3, $H(INS_2^0 \cap LOC(1))$ and REG are known to be incomparable.

Theorem 5 $H(INS_2^0 \cap LOC(1))$ *and* $H(INS_1^1 \cap LOC(1))$ *are incomparable.*

Proof Consider an insertion system $\gamma_1 = (T, \{(\lambda, ab, \lambda)\}, \{\lambda\})$ of weight $(2, 0)$ with $T = \{a, b\}$, a strictly 1-testable language $R = T^+$, and an identity morphism $h : T^* \to T^*$. The above definition indicates directly that $L(\gamma) = h(L(\gamma) \cap R)$.

We can show that $L(\gamma_1)$ is not in $H(INS_1^1 \cap LOC(1))$ by contradiction. We omit the proof here.

Now we consider an insertion system $\gamma_2 = (T, \{(a, a, \lambda), (b, b, \lambda)\}, \{a, b\})$ of weight $(1, 1)$ with $T = \{a, b\}$, a strictly 1-testable language $R = T^+$, and an identity morphism $h : T^* \to T^*$. From the definition, we have $L(\gamma_2) = h(L(\gamma_2) \cap R) = \{a^i \mid i \geq 1\} \cup \{b^i \mid i \geq 1\}$.

From [2], $L(\gamma_2)$ is not in $H(INS_2^0 \cap LOC(1))$. □

Theorem 5 implies the following Corollaries.

Corollary 1 $H(INS_2^0 \cap LOC(1))$ *and* $H(INS_1^1 \cap LOC(1)) \cap H(INS_1^0 \cap LOC(2))$ *are incomparable.*

Corollary 2 $H(INS_2^0 \cap LOC(1)) \subset H(INS_i^1 \cap LOC(1))$ $(i \geq 2)$.

For the class of languages $H(INS_1^0 \cap LOC(1))$, from the size of parameters, we have the inclusions $H(INS_1^0 \cap LOC(1)) \subseteq H(INS_1^1 \cap LOC(1))$ and $H(INS_1^0 \cap LOC(1)) \subseteq H(INS_1^0 \cap LOC(2))$. Next we present the following proper inclusion.

Theorem 6 $H(INS_1^0 \cap LOC(1)) \subset H(INS_1^1 \cap LOC(1)) \cap H(INS_1^0 \cap LOC(2))$.

Proof To show the proper inclusion, we consider an insertion system $\gamma_2 = (T, \{(a, a, \lambda), (b, b, \lambda)\}, \{a, b\})$ of weight $(1, 1)$ with $T = \{a, b\}$, a strictly 1-testable language $R = T^+$, and an identity morphism $h : T^* \to T^*$.

In a similar way to Theorem 5, we can show that $L(\gamma_2)$ is not in $H(INS_1^0 \cap LOC(1))$. □

Corollary 3 $H(INS_1^0 \cap LOC(1)) \subseteq H(INS_1^1 \cap LOC(1)) \cap H(INS_1^0 \cap LOC(2)) \cap H(INS_2^0 \cap LOC(1))$.

4 Concluding Remarks

In the present paper, we specifically examined the language classes $H(INS_{i_0}^0 \cap LOC(k_0))$ and $H(INS_{i_1}^1 \cap LOC(k_1))$ for $i_0, i_1, k_0, k_1 \geq 1$ and considered the relations of those language classes.

The following remain as open problems:

- $H(INS_2^0 \cap LOC(1)) \cap H(INS_1^1 \cap LOC(1)) = H(INS_1^0 \cap LOC(1))$ holds?

- $H(INS_2^0 \cap LOC(1)) \cap H(INS_1^1 \cap LOC(1)) \supset H(INS_2^0 \cap LOC(1)) \cap H(INS_1^0 \cap LOC(2))$ holds?

- $CF = H(INS_m^2 \cap LOC(k))$ holds for some $m, k \geq 1$?

References

[1] N. Chomsky and M.P. Schützenberger. *The algebraic theory of context-free languages. Computer Programming and Formal Systems.* North Holland, pp.118-161, 1963.

[2] K. Fujioka. Morphic characterizations in terms of insertion systems with a context of length one. DLT 2011, *LNCS* **6795**, pp.474-475, 2011.

[3] K. Fujioka. Morphic characterizations of languages in Chomsky hierarchy with insertion and locality. *Information and Computation* **209**, pp.397-408, 2011.

[4] R. McNaughton and S.A. Papert. Counter-Free Automata (M.I.T. research monograph no. 65) *The MIT Press*, 1971.

[5] F. Okubo and T. Yokomori. Morphic characterizations of language families in terms of insertion systems and star languages. *Int. J. Found. Comput. Sci.*, **22 (1)**, pp.247-260, 2011.

[6] K. Onodera. A note on homomorphic representation of recursively enumerable languages with insertion grammars. *IPSJ Journal*, **44,5**, pp.1424-1427, 2003.

[7] G. Păun, M.J. Pérez-Jiménez, and T. Yokomori. Representations and characterizations of languages in Chomsky hierarchy by using insertion–deletion systems. *Int. J. Found. Comput. Sci.* **19,4**, pp.859-871, 2008.

[8] G. Păun, G. Rozenberg, and A. Salomaa. Homomorphic characterizations of recursively enumerable languages with very small language classes. *Theoretical Computer Science*, **250**, pp.55–69, 2001.