数理解析研究所講究録
第1849巻 2013年 28-31
2012年度冬のLAシンポジウム [14]

A note on the expansions of insertion systems

Kaoru Fujioka
Office for Strategic Research Planning,
Kyushu University

1 Introduction

Insertion systems are those in which we can use insertion operations of the form (u, x, v) to produce a string $\alpha uxv\beta$ from a given string $\alpha uv\beta$ with context uv by inserting a string x. From the definition of insertion operations, it is clear that using insertion operations we can generate only context-sensitive languages.

Using insertion systems together with some morphisms, characterizing recursively enumerable languages is accomplished in [3]. In [1], within the framework of the Chomsky–Schützenberger representation theorem, some characterizations and representation theorems of languages in the Chomsky hierarchy including recursively enumerable languages are provided by insertion system γ, strictly locally testable language R, and morphism h such as $h(L(\gamma) \cap R)$.

On the other hand, insertion and deletion systems are those, in which we can use not only insertion operations but also deletion operations of the form (u, x, v), which produce a string $\alpha uv\beta$ from a given string $\alpha uxv\beta$ with context uv by deleting a string x. It is known that insertion and deletion systems can generate all recursively enumerable languages [3].

Insertion and deletion systems are computing models based on the field of molecular biology. Substitution operations are also present in the evolution processes of DNA sequences, in which nucleotides are substituted. In the present paper, we introduce a substitution operation which replaces a string $\alpha uxv\beta$ by $\alpha uyv\beta$ with context uv by substituting a string x for y.

The purpose of this paper is to introduce a substitution operation into insertion systems, called insertion-substitution systems, and to show the generative powers of these systems.

2 Preliminaries

In this section, we introduce notation and basic definitions that are necessary for this paper. We assume that the reader is familiar with the basics of formal language theory (see, e.g., [3]).

For a string $x \in V^*$ with an alphabet V, $|x|$ is the length of x.

Let RE (resp. CS, CF, REG) be the class of recursively enumerable languages (resp. context-sensitive languages, context-free languages, regular languages).

An *insertion-deletion system* is a tuple $\gamma = (V, T, A, I, D)$, where V is an alphabet, T is a finite set of *terminal symbols* such that $T \subseteq V$, A is a finite set of strings over V called axioms, and I (resp. D) is a finite set of *insertion rules* (resp. *deletion rules*) of the form (u, x, v) with $u, x, v \in V^*$.

We write $\alpha \overset{r}{\Longrightarrow}_{ins} \beta$ if $\alpha = \alpha_1 uv\alpha_2$ and $\beta = \alpha_1 uxv\alpha_2$ for some insertion rule $r : (u, x, v) \in I$

with $\alpha_1, \alpha_2 \in V^*$. For a deletion rule $r : (u, x, v) \in D$, we write $\alpha \overset{r}{\Longrightarrow}_{del} \beta$ if $\alpha = \alpha_1 uxv\alpha_2$ and $\beta = \alpha_1 uv\alpha_2$ with $\alpha_1, \alpha_2 \in V^*$.

If there is no confusion, we write $\Longrightarrow_{ins}$ (resp. $\Longrightarrow_{del}$) instead of $\overset{r}{\Longrightarrow}_{ins}$ (resp. $\overset{r}{\Longrightarrow}_{del}$). We use $\alpha \Longrightarrow_\gamma \beta$ for relations $\Longrightarrow_{ins}$ and $\Longrightarrow_{del}$. The reflexive and transitive closure of $\Longrightarrow_\gamma$ is defined as $\Longrightarrow^*_\gamma$.

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-deletion system $\gamma = (V, T, A, I, D)$ is said to be of *weight* $(i, j; p, q)$ if

$$\begin{aligned} i &= max\{\, |x| \mid (u, x, v) \in I\}, \\ j &= max\{\, |u| \mid (u, x, v) \in I \text{ or } (v, x, u) \in I\}, \\ p &= max\{\, |x| \mid (u, x, v) \in D\}, \\ q &= max\{\, |u| \mid (u, x, v) \in D \text{ or } (v, x, u) \in D\}. \end{aligned}$$

For $i, j, p, q \geq 0$, let $INS_i^j DEL_p^q$ be the class of languages generated by insertion-deletion systems of weight $(i', j'; p', q')$ with $i' \leq i$, $j' \leq j$, $p' \leq p$, and $q' \leq q$. If some of the parameters i, j, p, q are not bounded, we use $*$ in place of the symbols for those parameters.

For insertion-deletion systems, the following result exists.

Theorem 1 [2][4]

1. $INS_2^0 DEL_3^0 = INS_3^0 DEL_2^0 = RE$.
2. $INS_2^0 DEL_2^0 \subset CF$.
3. $INS_1^0 DEL_p^0 \subset REG$ $(\forall p > 0)$.

An *insertion system* is a triple $\gamma = (T, A, I)$, in which we can use only insertion operations, for which non-terminal symbols are useless without deletion operations, where T, A, and I are defined as before. For $i, j \geq 0$, let INS_i^j be the class of languages generated by insertion systems of weight (i', j') with $i' \leq i$ and $j' \leq j$. For insertion systems, the following result holds.

Theorem 2 [3]

1. $REG \subset INS^*_*$.
2. $INS^1_* \subseteq CF$.

We introduce the notion of insertion-substitution systems as follows.

Definition 1 *An insertion-substitution system is a tuple $\gamma = (V, T, A, I, J)$, where V, T, A, and I are defined as before, and J is a finite set of substitution rules of the form $(u, x \to y, v)$, with $u, x, y, v \in V^*$ and $|x| \geq |y|$.*

We write $\alpha \overset{r}{\Longrightarrow}_{sub} \beta$ if $\alpha = \alpha_1 uxv\alpha_2$ and $\beta = \alpha_1 uyv\alpha_2$ for some substitution rule $r : (u, x \to y, v) \in J$ with $\alpha_1, \alpha_2 \in V^*$.

If there is no confusion, we write $\Longrightarrow_{sub}$ instead of $\overset{r}{\Longrightarrow}_{sub}$. We write $\alpha \Longrightarrow_\gamma \beta$ for relations $\Longrightarrow_{ins}$ and $\Longrightarrow_{sub}$. The reflexive and transitive closure of $\Longrightarrow_\gamma$ is defined as $\Longrightarrow^*_\gamma$.

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-substitution system $\gamma = (V, T, A, I, J)$ is said to be of *weight* $(i, j; p, q)$ if

$$\begin{aligned} i &= max\{\, |x| \mid (u, x, v) \in I\}, \\ j &= max\{\, |u| \mid (u, x, v) \in I \text{ or } (v, x, u) \in I\}, \\ p &= max\{\, |x|, |y| \mid (u, x \to y, v) \in J\}, \\ q &= max\{\, |u| \mid (u, x \to y, v) \in J \text{ or} \\ &\qquad (v, x \to y, u) \in J\}. \end{aligned}$$

For $i, j, p, q \geq 0$, let $INS_i^j SUB_p^q$ be the class of languages generated by insertion-substitution systems of weight $(i', j'; p', q')$ with $i' \leq i$, $j' \leq j$, $p' \leq p$, and $q' \leq q$. If some of the parameters i, j, p, q are not bounded, we use $*$ in place of the symbols for those parameters.

In this paper, we specifically examine the generative powers of insertion-substitution systems.

3 Main Results

An insertion-substitution system $\gamma = (V, T, A, I, J)$, without any restrictions, is an expansion of insertion-deletion systems. In this case, $INS_i^j DEL_p^q \subseteq INS_i^j SUB_p^q$ holds. Now we consider the following lemma.

Lemma 1 *$INS_i^j SUB_p^q \subseteq INS_{i'}^{j'} DEL_{p'}^{q'}$ for any $i, j, p, q \geq 0$, $i' = max\{i, p+1\}$, $j' = max\{j, p+q\}$, $p' = p+1$, $q' = p+q$.*

Proof *Outline: Consider an insertion-substitution system $\gamma = (V, T, A, I, J)$ of weight $(i, j; p, q)$. We show that there is an insertion-deletion system $\gamma' = (V \cup Lab(I) \cup Lab(J), T, A, I', D')$ of weight (i', j', p', q') with $i' = max\{i, p+1\}$, $j' = max\{j, p+q\}$, $p' = p+1$, and $q' = p+q$ such that $L(\gamma') = L(\gamma)$.*

For a substitution rule $r : (u, x \to y, v)$ in J, we construct an insertion rule $r_1 : (ux, ry, v)$ in I' and a deletion rule $r_2 : (u, xr, yv)$ in D'. Furthermore, we set I' satisfies that $I \subseteq I'$.

Actually, for each derivation $\alpha u x v \beta \overset{r}{\Longrightarrow}_\gamma \alpha u y v \beta$ with $\alpha, \beta \in V^*$ in γ, there exists a derivation $\alpha u x v \beta \overset{r_1}{\Longrightarrow}_{\gamma'} \alpha u x r y v \beta \overset{r_2}{\Longrightarrow}_{\gamma'} \alpha u y v \beta$ in γ'.

The insertion rule $r_1 : (ux, ry, v)$ satisfies that $max\{|ux|, |v|\} \leq p+q$ and $|ry| \leq 1+p$. The deletion rule $r_2 : (u, xr, yv)$ satisfies that $max\{|u|, |yv|\} \leq p+q$ and $|xr| \leq p+1$.

We omit the proof that $L(\gamma) = L(\gamma')$ here. □

Now we consider a *restricted insertion-substitution system* $\gamma = (V, T, A, I, J)$, which satisfies that $V = T$ and, for any substitution operation $(u, x \to y, v)$ in J, $|x| = |y|$ holds. The class of languages generated by restricted insertion-substitution systems of weight (i', j', p', q') with $i' \leq i$, $j' \leq j$, $p' \leq p$, and $q' \leq q$ is described by $INS_i^j RSUB_p^q$.

From the definition, the inclusion $INS_i^j \subseteq INS_i^j RSUB_p^q \subseteq INS_i^j SUB_p^q$ holds for any $i, j, p, q \geq 0$. The generative powers of restricted insertion-substitution systems are observed in the following.

First, we consider restricted insertion-substitution systems of weight $(1, 0; 1, 0)$.

Lemma 2 $INS_1^0 RSUB_1^0 = INS_1^0$.

Proof *The inclusion $INS_1^0 RSUB_1^0 \supseteq INS_1^0$ is obvious, therefore we show the other inclusion.*

For a restricted insertion-substitution system $\gamma = (T, T, A, I, J)$ of weight $(1, 0; 1, 0)$, we construct an insertion system $\gamma_1 = (T, A_1, I_1)$ such that A_1 consists of all the strings, including the ones in A, which can be obtained from a string w in A by applying substitution rules in J. For a finite set A and a finite set of substitution rules J, such as $(\lambda, x \to y, \lambda)$ with $|x| = |y| = 1$, A_1 is a finite set of strings.

Let I_1 consist of all the insertion rules, including the ones in I, such that (λ, y, λ), where y can be obtained from a symbol x with (λ, x, λ) in I by applying substitution rules in J.

It can be shown that $L(\gamma) = L(\gamma_1)$. Informally, any string generated by γ with substitution operations in J can be generated by applying insertion operations in I_1. Formally, the proof can be shown by induction on the number n of substitution operations in a derivation $\alpha \Longrightarrow_\gamma^* w$ with $\alpha \in A$ and $w \in T^*$. We omit the proof here. □

Corollary 1 $INS_i^0 RSUB_1^0 = INS_i^0$.

Corollary 2 $INS_i^j RSUB_1^0 = INS_i^j$.

Proof *The above corollaries can be proved in a similar way as Lemma 2.* □

The previous results show that a restricted insertion-substitution system of weight (i,j,p,q) with any substitution rule $(\lambda, x \to y, \lambda)$ for $|x| = |y| = 1$ has the same generative powers as insertion systems of weight (i,j,p,q).

On the other hand, a substitution rule $(\lambda, x \to y, \lambda)$ with $|x| = |y| = 2$ can properly increase generative powers, which is shown in the following example.

Lemma 3 $INS_3^0 \subset INS_3^0 RSUB_2^0$.

Proof *Consider the restricted insertion-substitution system* $\gamma = (\{a,b,c\}, \{a,b,c\}, \{abc\}, \{(\lambda, abc, \lambda)\}, J)$, *where* $J = \{(\lambda, x \to y, \lambda) \mid x = t_1t_2,\ y = t_2t_1$ *with* $t_1 \neq t_2$ *and* $t_1, t_2 \in \{a,b,c\}\}$ *of weight* $(3,0;2,0)$.

Then $L(\gamma) = \{w \in \{a,b,c\}^* \mid |w|_a = |w|_b = |w|_c\}$, *which is shown to be in* $CS - CF$ *[3]. From the result in Theorem 2, we have* $INS_3^0 \subseteq INS_*^1 \subseteq CF$. *Therefore, we can prove the proper inclusion* $INS_3^0 \subset INS_3^0 RSUB_2^0$. □

Another example of languages in $CS - CF$ is provided in the following.

Example 1 *Consider the language* $L = \{a^{2^n} \mid n \geq 1\}$ *in* $CS - CF$. *There is no restricted insertion-substitution system* γ *such that* $L(\gamma) = L$.

Suppose that there is a restricted insertion-substitution system $\gamma = (\{a\}, \{a\}, A, I, J)$ of weight $(i,j;p,q)$ such that $L = L(\gamma)$. Let $k' = max\{|\alpha|, |x| \mid \alpha \in A,\ (u,x,v) \in I\}$. Consider strings w_i and w_{i+1} with $w_0 \Longrightarrow_\gamma^* w_i \overset{r}{\Longrightarrow}_\gamma w_{i+1}$ such that $w_0 \in A$, $r = (u, a^l, v)$, $|w_i| < |w_{i+1}|$, and $|w_i| = 2^k$ for $k > k'$. The string w_{i+1} satisfies that $|w_{i+1}| = |w_i| + l = 2^k + l < 2^{k+1}$. For the restricted insertion-substitution system γ, both w_i and w_{i+1} are in $L(\gamma) = L$, which is a contradiction.

Lemma 4 $INS_*^* RSUB_*^* \subset RE$.

Proof *From Example 1, we can prove the proper inclusion.* □

The proper inclusion as in Lemma 4 holds even if we weaken the restriction on restricted insertion-substitution systems, that is, the ones with a substitution rule $(u, x \to y, v)$ such that $|x| \geq |y| \geq 0$. We can prove the proper inclusion by the language $L = \{a^{2^n} \mid n \geq 1\}$ in a similar way as the explanation of Example 1.

4 Concluding Remarks

As described in this paper, we defined insertion-substitution systems and examined their generative powers. The following remain as open problems:

- $INS_i^j \subset INS_i^j RSUB_p^q$ holds for any $i,j,q \geq 0$, $p \geq 2$?
- $INS_i^j DEL_p^q \subset INS_i^j SUB_p^q$ holds for some $i,j,p,q \geq 0$?

References

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

[2] M. Margenstern, G. Păun, Y. Rogozhin, and S. Verlan. Context-free insertion-deletion systems. *Theoretical Computer Science*, **330 (2)**, pp.339–348, 2005.

[3] G. Păun, G. Rozenberg, and A. Salomaa. DNA Computing: new computing paradigms. *Springer*, 1998.

[4] S. Verlan. On minimal context-free insertion-deletion systems. *Journal of Automata, Languages and Combinatorics*, **12 (1-2)**, pp.317–328, 2007.